The microwave radiometer as a remote sensing device: design and application

de Maagt, P.J.I.

DOI:
10.6100/IR386537

Published: 01/01/1992

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
THE MICROWAVE RADIOMETER AS A REMOTE SENSING DEVICE:
DESIGN AND APPLICATION.

Proefschrift

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op
gezag van de Rector Magnificus, prof. dr. J.H. van Lint, voor een commissie aangewezen
door het College van Dekanen in het openbaar te verdedigen op dinsdag 1 december 1992
om 16.00 uur.

door

Peter Jacobus Irene de Maagt

goingen te Pauluspolder

© 1992 P.J.J. dc Maagt

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic, mechanical, including photocopy, recording, or any information storage and retrieval system, without the prior written permission of the copyright owner.
CONTENTS

Abstract 1
Korte Samenvatting 1

1. Introduction 3
   1.1. General Introduction 3
   1.2. Framework of the Research 6
   1.3. Scope of the Study and Survey of its Contents 6

2. Radiometer Antennas 11
   2.1. Introduction 11
   2.2. A General Optimization Method for Reflector Antenna Synthesis 12
   2.3. An Optimization Method for Radiometer Antennas 24
   2.4. A Review and Comparison of Some Asymptotic Techniques for Calculating the Wide-angle Radiation Pattern of Paraboloid Reflector Antennas 36
   2.5. Calculating the Wide-angle Radiation Pattern of an Offset Paraboloid Reflector Antenna 57
      2.5.1. Introduction 57
      2.5.2. Reflector Configuration 57
      2.5.3. Diffraction Point Location for an Offset Configuration 59
      2.5.4. Diffracted Field in the Symmetry Plane 63
      2.5.5. Diffracted Field in an Arbitrary \( \phi \)-Plane 64
      2.5.6. EEC Method for the Far-Field Caustic in the Symmetry Plane 66
      2.5.7. Numerical Results 67
   2.6. On the Design of Radiometer Antennas 71

3. Radiometer Receivers 103
   3.1. Introduction 103
   3.2. General Operation of Radiometers 104
      3.2.1. The Total Power Radiometer 105
      3.2.2. The Dicke Radiometer 106
3.2.3. The Noise Injection Radiometer 108
3.2.4. The Gain Balancing Radiometer 109
3.2.5. The Graham Radiometer 110
3.2.6. The Correlation Radiometer 110
3.2.7. Some Other Radiometer Types 112
3.3. A Novel Temperature Stabilization Technique 113
3.3.1. Introduction 113
3.3.2. The Test Set-up 113
3.3.3. Temperature Dependence of the Complete Radiometer 118
3.3.4. Temperature Dependence Measurement of the Radiometer on a Component Level 121

References Sections 3.1 to 3.3 124
3.4. The Temperature Dependence of Schottky versus Backward Diodes for Radiometry Applications 126
3.5. A Novel Radiometer Receiver Stabilization Method 137

4. Remote Sensing by Passive Imaging 161
4.1. Introduction 161
4.2. Basic Problems Related to Imaging Techniques 162
4.3. Some Inversion Techniques 165
4.3.1. Inversion by Multiple Convolution 165
4.3.2. Inversion by Maximum Likelihood 166
4.3.3. Inversion by Matrix Inversion 167
4.3.4. Other Inversion Techniques 167
4.3.5. Discussion 168
References Sections 4.1 to 4.3 169
4.4. Image Reconstruction using a Passive Microwave Radiometer 171
4.5. The Influence of Measurement Errors in the Object Velocity on the Reconstruction 191
4.6. The Transfer Function of the Algorithm 192
4.7. Influence of Shaping the Main Beam on the Trade-off between Resolution and Accuracy of Reconstruction. 194
4.7.1. The Influence of Measurement Errors in $T_a$ 197
4.7.2. The Influence of Measurement Errors in $G$ 197
4.7.3. Comparing the Results obtained with a Focused and Defocused Antenna System

4.8. Conclusions

5. Summary and Conclusions

Verantwoording
Acknowledgements
Curriculum Vitae
Abstract

This thesis deals with the development of novel approaches to the design of microwave radiometer systems. To offer a greater insight into the design procedures, the complete radiometer system is divided into two subsystems, namely the antenna and the receiver. These subdivisions are reflected in the structure of this thesis. Chapter 2 addresses the radiometer antenna. A design procedure for radiometer antennas is proposed and their theoretical performance is considered in relation to their expected performance in practice. Chapter 3 deals with the radiometer receiver. A novel radiometer temperature stabilization method is presented along with the results of testing it with a bread-board model.

The relatively new field of imaging in which microwave radiometry has shown its potential is studied in the remainder of this thesis. A fundamental analysis is included of the spatial and temporal filtering process of the observation instrument and the corresponding deconvolution procedures.

Korte Samenvatting

Dit proefschrift behandelt de ontwikkeling van nieuwe ontwerpbenaderingen van microgolf radiometersystemen. Om meer inzicht te bieden in de ontwerpprocedures is het complete radiometersysteem gesplitst in twee subsystemen, namelijk antenne en ontvanger. Deze onderverdeling is ook terug te vinden in de opbouw van dit proefschrift. In hoofdstuk 2 wordt de radiometerantenne behandeld. Een ontwerp procedure voor radiometer antennes wordt voorgesteld en de theoretisch haalbare performance wordt vergeleken met de praktisch realiseerbare performance.

In hoofdstuk 3 wordt de radiometerontvanger behandeld. In dit hoofdstuk wordt een nieuwe methode voor temperatuurstabilisatie van radiometerontvangers gepresenteerd, waarvan de mogelijkheden verkend zijn met een proefmodel.

Een relatief nieuw gebied waarin microgolfradiometrie zijn potentieel getoond heeft is imaging, hetgeen bestudeerd is in de rest van het proefschrift. Een fundamentele analyse van de ruimtelijke- en tijdsfilteringprocessen van het observatie instrument en de bijbehorende deconvolutieprocedures is uitgevoerd.
1. Introduction

1.1. General Introduction

Every object in the universe emits and absorbs radiation; in addition, it reflects natural and artificial radiation. The radiation emitted and reflected can be detected by means of remote sensing techniques after it has been passed through and modified by the intermediary medium, viz. the atmosphere. After radiation from the object has been received, it is possible to define the characteristics of the object in different domains: temporal (variation of power radiated with time), spatial (position, size, and shape), and spectral (distribution of power radiated with frequency).

It is difficult to establish a specific time or event that marks the beginning of remote sensing [1,2]. Some cite Aristotle's camera obscura (fourth century BC); some cite the experiments of Heinrich Hertz in 1886; others cite the experiments of Karl Jansky in 1928, or the two World Wars. A similar twilight zone can be found when trying to point out which part of the frequency spectrum should be used for remote sensing. The part that is used presently ranges from the radiospectrum to the infrared and visible regions. More consensus can be found in the classification of remote sensing systems; the most useful distinction that can be drawn is between the active and the passive remote sensing systems, according to whether the radiation emanating from the remote object originates from self-emission or via reflection from an artificial source. However, both sensing techniques have in common that the systems used for characterizing the object have to be designed to utilize whatever information the radiation contains in the domain of interest. The device developed for that purpose has to limit its response to the radiation that is needed in that domain.

The characteristics obtained with remote sensing techniques can play a major role in different disciplines. Although it is impossible to list all the applications, a few are mentioned below and a detailed summary can be found in [2]:

Meteorology
Oceanography
Glaciology
Geology, geomorphology and geodesy
Topography and cartography
Agriculture, forestry and botany
Hydrology
Disaster control
Planning applications
Military applications
Satellite communications.

With microwave satellite communications, where the growth in communications has prompted the use of higher frequency bands, exploration of higher bands by means of remote sensing is of prime importance. The reason for this is the fact that the required availability of a satellite link can be essentially limited by atmospheric attenuation. Passive remote sensing removes the need for expensive transmission devices at the frequency of interest. A powerful passive remote sensing tool for characterizing the atmospheric attenuation in the temporal domain is microwave radiometry.

The beginning of microwave radiometry can be dated back to the early Forties. In those days, astronomers made their first observations of the sun and moon [3,4]. The detection equipment used by Southworth [3] may be seen as one of the first radiometer receivers; however, this equipment suffered from "gain variation noise" [4]. In 1946, R.H. Dicke wrote his pioneering article, "The measurement of thermal radiation at microwave frequencies" [5], and introduced the use of "comparison radiometry". That technique became widely used in radiometry and some other engineers studied the issue of gain variations and stability [6,7]. Although nearly every development or device for improving those characteristics has found its way into radiometry design, the basic receiver proposed by Dicke is still the one that is most commonly used.

The use of microwaves for remote sensing applications was suggested by their ability to penetrate through clouds and rain, as well as through vegetation, plus the fact that an object's characteristics obtained with microwaves can compliment those obtained from other frequency regions. A serious drawback to microwave wavelengths is their inherent limited spatial resolution. Therefore, applications for microwave radiometry were mainly limited to oceanographical, meteorological, military, and satellite communications applications. However, in recent years it has been shown that it is possible to obtain object characteristics with spatial resolution comparable to the physical scale, and the applications of microwave radiometry are considered to be as diverse as mentioned before.

In contrast to the statement that "remote sensing systems have to be designed to limit their response to the domain of interest", most of the present-day radiometer system designs (especially ground-based ones) originated from communication-systems-based design procedures. It is not always possible to make the specific design criteria for remote sensing compatible with those used for Earth station communications. So, it is very important to develop new design procedures which are based on a radiometer as a remote...
sensing device consisting of a microwave antenna and highly sensitive microwave receiver.

Most radiometer systems use a communication antenna (often a simple front-fed reflector antenna). Optimization of the antenna from the point of view of remote sensing is necessary, and this is one valuable aspect of the work presented in this thesis. The optimization process produces theoretical limits for the best system which is optimal because it is impossible to construct a better device, subject to the given constraints. In other words, knowledge of the theoretically optimal system makes it unnecessary to consider a large number of possible modifications to physical antenna systems. Optimization is a process in which the various parameters are adjusted to obtain a desired result. However, it is necessary to evaluate the various design criteria by value judgments within given constraints. So the optimum can vary greatly depending upon the application. Furthermore, many desirable attributes of a system are mutually exclusive and can only be obtained at the expense of one another. Given the parameters, a design procedure can be developed. To be able to make a quantitative comparison between theoretical and practical antenna systems, full knowledge of the complete antenna's radiation pattern is needed. To compute these patterns time-efficiently, asymptotic approximations of the radiation integrals should be used.

The classical design of a radiometer receiver, which is basically still the same as that proposed by Dicke in 1946, involves continuous temperature stabilization and verification with a reference load to give a reliable output. The development of alternative designs will be the second aspect addressed by this thesis. It would appear that with modern technology (e.g. integrated microwave circuits, sensing devices, and on-line data acquisition), a better, more compact, and less expensive design is feasible by using different approaches to the stabilization problem. Developing of such a receiver could represent a real advance in the state of the art and, in this case, the specifications of classical receivers may be used as design goals. If, for example the temperature behaviour of the critical parts of a system can be characterized, temperature might be stabilized by compensating for any observed temperature excursions with PC-based software.

Most classical radiometer systems are employed to obtain characteristics in the temporal domain. The output of a radiometer used for such a purpose could be processed with inverse-transform procedures to unscramble a blurred object and to enhance objects of interest in the spatial domain. The most important objective is to obtain characteristics with spatial resolution comparable to the physical scale. This fairly new application of radiometer synthesis needs an investigation into the interaction between the spatial filtering of the antenna and the time filtering of the receiver. That is the third most important aspect of the thesis.
1.2. Framework of the Research

In the past few decades, research into radiowave propagation has contributed much to the exploration of the 11–14 GHz band for telecommunications applications. With the successful launch and commissioning of the Olympus satellite in 1989, a unique opportunity was offered to the scientific community for an expansion of propagation research into the 20 and 30 GHz bands. These frequency bands are being explored for new applications in satellite telecommunications and will probably be extensively exploited in the late Nineties.

At Eindhoven University of Technology (EUT), an extensive measurement campaign is performed with the Olympus propagation payload. This campaign includes the use of radiometers for the comparative analysis of attenuation and noise radiation. Much of the radiometer equipment had to be developed in–house, since commercial equipment for this application was not available. Analyzing the data from a multifrequency radiometer concerns both the telecommunications and the remote sensing aspects. Within that framework, research into optimal design and optimal data processing was started that is the subject of this thesis.

To enhance the exchange of ideas and experience, the Netherlands Coordinating Committee for the Olympus Propagation–experiments (NCOP) was established; whilst coordination of Olympus propagation experiments in Europe is in the hands of the Olympus Propagation Experimenters Group (OPEX).

1.3. Scope of the Study and Survey of its Contents

This thesis is intended to contribute to the development of novel design procedures for radiometer systems and to provide a system–oriented approach to radiometer design. A better insight into these complex design procedures involved will be obtained if the complete radiometer system is divided into two subsystems, namely the antenna and the receiver. This is reflected in the structure of the thesis. Chapter 2 looks at radiometer antennas. A general optimization method is discussed for optimizing several parameters simultaneously, with and without (integrated) pattern constraints. The performance of the resulting optimal system is then compared with that of different promising antenna configurations for radiometry. The performance parameters had to be computed in order to make a relevant comparison. For one parameter, the complete antenna pattern was needed and some asymptotic techniques were investigated in order to compute this time efficiently in a correct and transparent way. With the complete antenna pattern, it was possible to determine the relative importance of different parts of the antenna's pattern and
the different parameters. Using them as the basis, a design procedure for radiometer antennas has been proposed.

Chapter 3 deals with the radiometer receiver. The fundamental problems that can be encountered in radiometer receivers are discussed and the commonest radiometer receivers are surveyed and their advantages and shortcomings listed. The chapter also includes a novel radiometer stabilization method, the validity of which was tested with a bread-board model.

A relatively new field in which radiometry has shown its potential is imaging, this is studied extensively in the remainder of this thesis. A fundamental analysis was conducted on the spatial and temporal filtering processes of the observation instrument and the corresponding deconvolution procedures.

Chapters 2 and 3 include papers or letters previously published by the author in scientific literature during the research. The papers have been incorporated in this thesis and annotated in order to improve its coherence; furthermore, parts of chapters 2 and 4 were presented at the Fourth and Fifth International Symposia on Antennas (JINA) held in Nice in 1990 and 1992, and the Third Specialist Meeting on Microwave Radiometry and Remote Sensing held in Boulder (1992).
References

[1] Fischer, W.A.
HISTORY OF REMOTE SENSING.
In manual of remote sensing, ed. R.G. Reeves,
American Society of Photogrammetry, Falls Church, VA, 1975.

INTRODUCTION TO ENVIRONMENTAL REMOTE SENSING SENSING.

MICROWAVE RADIATION FROM THE SUN.

MICROWAVE RADIATION FROM THE SUN AND THE MOON

[5] Dicke, R.H.
THE MEASUREMENT OF THERMAL RADIATION AT MICROWAVE FREQUENCIES.

A BROAD-BAND MICROWAVE SOURCE COMPARISON RADIOMETER FOR ADVANCED RESEARCH IN RADIO ASTRONOMY.

A SWITCHED LOAD RADIOMETER.
2. Radiometer Antennas

2.1. Introduction

Antennas form an important part of a radiometric remote-sensing system. In the first instance, such antennas may resemble those used in communication systems. However, the differences are profound and the design approaches differ significantly. Communication antennas generally are optimized for the spatial filtering of coherent signals, which results in the optimization of the antenna gain and the realization of a prescribed sidelobe level. These criteria are not of paramount importance for a remote sensing device such as the radiometer and it is not always possible to make the specific design criteria for remote sensing compatible to those used for communications-system design. Radiometer antenna design requires a much more precise examination of different parameters than most antenna engineers are accustomed to and antenna optimization from the view of remote sensing is necessary. The various parameters have to be adjusted to approach a desired result as near as possible and it is necessary to establish the importance of the various criteria by value judgment. The number of relevant parameters is large and the requirements can often be mutually exclusive. At the same time, quite a lot of optimization procedures exist that focus on only one parameter. Therefore, it is of great interest to develop an optimization procedure that is able to deal with one or more radiometer antenna parameters.

This offers an opportunity to describe a theoretical optimal radiometer antenna. The advantage of knowing an optimal system is that it assures that it is impossible to construct a better device subject to the given constraints. In other words, knowledge of a theoretical optimal system will make it unnecessary to consider a large number of possible modifications to real antenna systems. As the optimum must not go beyond the possible or contain requirements that unnecessarily drive up costs, the choice of a radiometer antenna must take into consideration the optimal antenna's performance versus the achievable performance with a practical antenna.

To be able to compute the parameters of practical antennas, the complete far-field radiation pattern of those antennas have to be calculated. Firstly, this chapter addresses a general optimization method (in section 2.2); then, it is specifically applied to a radiometer antenna in section 2.3. The calculation of complete radiation patterns for different antenna configurations is discussed in sections 2.4 and 2.5. Section 2.6 combines all the tools presented in the preceding sections in order to determine an optimal practical radiometer antenna.
2.2. A General Optimization Method for Reflector Antenna Synthesis

P.J.I. de Maagt
Eindhoven University of Technology
Telecommunications Division, P.O.Box 513, 5600 MB Eindhoven, the Netherlands.

Abstract
This paper deals with an analytical approach to antenna synthesis. It presents an optimization method which is based on writing the design criteria as a ratio of two quadratic Hermitian forms, so that more than one antenna parameter (such as antenna and beam efficiency) can be optimized simultaneously, with and without pattern-structure constraints. The paper starts with the mathematical formulation; then the optimization method with and without the constraints to the far-field pattern is discussed. Finally, a comparison is made with the results obtained by others and examples are given, which clearly show the capability of the optimization procedure.

Summary
Generally speaking, the objective of antenna synthesis is to approach the best design realizable under the condition that the requirements with respect to radiation properties are met. Synthesis techniques can be divided in two categories. In one category, the solution is found via numerical manipulations; while in the other, the solution is found analytically. The latter has the advantage that it offers more insight into the effects and interactions between different design parameters. Furthermore, that method gives a closed form to both the aperture-field distribution and the far-field pattern. After optimization, the far-field will be known across the full angle-region of interest, which removes the need to compute the time consuming far-field integrals repeatedly.

The analytical method often uses the concept of partial radiation-patterns, which approximates the desired far-field pattern and the corresponding aperture-field distribution by means of a series of special source functions.
For the application of this method, two aspects require attention. Firstly, the selection of special functions can be governed by certain considerations, including: the simplicity of approximating the desired pattern with a minimum number of terms in a series, the property of orthogonality, the ease with which functions can be Fourier transformed or by the possibility of working with a series of functions which is familiar (some degree of arbitrariness cannot be denied). Secondly and more demanding, is the aspect that "the requirements with respect to radiation properties" can vary widely and can often be mutually exclusive.

Due to these two aspects, a whole range of synthesis procedures exists; most of them focus on the optimization of a specific antenna parameter such as aperture efficiency, beam efficiency, etc.([1]–[8]). If the (sidelobe-) structure of the antenna pattern is the design objective ([9]–[11]), there is no possibility of optimizing two (or more) antenna parameters, simultaneously.

This paper deals with the analytical approach and it presents an optimization method based on writing the design criteria as a ratio of two quadratic Hermitian forms, where more than one design criteria can be optimized simultaneously, with or without pattern (structure-) constraints. For optimization problems with constraints, most engineers apply Lagrange multipliers; however, the formulation used here, makes it possible to simplify the problem with the help of the Householder transformation. In the proposed optimization method, constraints are treated as an "advantage", because they reduce the number of variables to be adjusted.

The paper starts with the mathematical formulation; then, the optimization method without the constraints to the far-field pattern is discussed; finally, the process is repeated including the constraints. The paper ends by comparing the results obtained with this method and those found in literature, and examples which clearly show the capability of the optimization procedure.

Mathematical Formulation
In this section, the design objectives with respect to the antenna parameter(s) and/or far-field pattern structure are written in a form which is particularly suitable for the optimization. The mathematical formulation involves an antenna with a circular aperture. The aperture points are given by normalized aperture polar coordinates \((r, \phi')\) and the far-field observation point by spherical coordinates \((R, \theta, \phi)\).

The integral part \(g(u, \phi)\) of the far-field pattern \(E(R, \theta, \phi)\) is related to the aperture distribution \(f(r, \phi')\) by the integral [12]:
\[ g(u,\phi) = (\frac{D}{2})^2 \int_0^{2\pi} \int_0^{\frac{\pi}{2}} f(r,\phi) e^{jru \cos(\phi - \phi')} r \, dr \, d\phi' \quad (1) \]

with \( u = \frac{\pi D}{\lambda} \sin \theta \).

In the case of a rotationally symmetric equiphasic aperture—field—distribution, \( f(r,\phi') \) can be written as:

\[ f(r) = \sum_{n=0}^{N} a_n e_n(r) \quad 0 \leq r \leq 1 = a^T e \quad (2) \]

with \( a^T = (a_0, a_1, \ldots, a_N) \) and \( e^T = (e_0, e_1, \ldots, e_N) \).

Consequently, the rotationally symmetric far-field pattern \( g(u) \) can be written as the first order Hankel transform of \( f(r) \):

\[ \frac{g(u)}{2\pi (D/2)^2} = \int_0^1 f(r) J_0(u r) \, rdr = \sum_{n=0}^{N} a_n I_n(u; e_n) = a^T I \quad (3) \]

with \( I_n(u; e_n) = \int_0^1 e_n(r) J_0(u r) \, rdr \), \( a_n \) the excitation coefficients of the elementary real functions \( e_n(r) \), and \( J_0 \) the Bessel function of the first kind and zeroth order. Using the equations (2) and (3), it is possible to write most of the antenna parameters as a ratio of two quadratic Hermitian forms.

As an example this will be demonstrated for the aperture efficiency \( \eta_a \), the beam efficiency \( \eta_b \), and the normalized second moment \( \sigma^2 \). Similar expressions can be found for other antenna parameters ([9],[14]). These derivations need the equations for the power radiated by the aperture \( p_r \), the power radiated within a prescribed solid angle \( p_{\text{angle}} \) and the second moment \( \mu_2 \). The first two are given by:

\[ p_r = 2\pi (D/2)^2 \int_0^1 f^2(r) \, rdr, \quad p_{\text{angle}} = 2\pi \int_0^c p(u) \, du \quad (4) \]

with \( p(u) = g^2(u) \) and \( c = (\pi D/\lambda) \sin \theta_{\text{pre}} \), where \( \theta_{\text{pre}} \) is the prescribed angle.

The second moment of the far field radiated power with respect to the axis \( u=0 \) is found by integrating \( u^2 p(u) \). This leads to:
Using the equations (2)-(5), it is easy to derive the following formulas where the antenna parameters are written in the desired form:

\[
\begin{align*}
\eta_a &= \frac{2}{a} \mathbf{T}V(0) a \\
\eta_b &= \frac{a}{\mathbf{T}Aa} \\
\sigma^2 &= \frac{a}{\mathbf{T}Wa} \\
\end{align*}
\]

where \( V, A, X, W \) are \( N+1 \) square matrices with elements:

\[
\begin{align*}
A_{ij} &= \int_0^1 e_i e_j r dr, \quad V_{ij} = I(0; e_i)I(0; e_j) \\
X_{ij} &= \int_0^1 u V_{ij} du, \quad W_{ij} = \int_0^1 u^2 V_{ij} du. \\
\end{align*}
\]

The Optimization Procedure without Constraints

Consider the problem of optimizing a function which can be written as:

\[
h(a) = \frac{a}{\mathbf{T}A a} \\
\]

where \( a \) is an \( N+1 \)-element vector and \( A \) and \( B \) are \( N+1 \times N+1 \) real matrices. A basic theorem from linear algebra [14] is used to optimize the function. The theorem states that if \( A \) and \( B \) are Hermitian and if \( B \) is positive definite, the maximum (or minimum) of the quantity will be given by the largest (or smallest) eigenvalue determined by:

\[
A_a = \lambda B_a. \\
\]

So, the original optimization problem can be treated as a general eigenvalue problem. The matrices \( V, A, X \) and \( W \) satisfy these requirements. They are all Hermitian, as can be seen from (7), and positive definite, because they represent the power in the forward direction, the total radiated power, the power radiated in a prescribed solid angle and the spread of radiated power, respectively. Since all the matrices used are positive definite, the theorem is also valid for \( 1/h(a) \) so, minimizing \( h(a) \) can be treated in the same way as maximizing \( 1/h(a) \).

An interesting case appears if more than one antenna parameter has to be optimized
simultaneously. For a function which can be written as:

\[ h_1(\mathbf{a})h_2(\mathbf{a}) = \frac{\mathbf{a}^T \mathbf{A} \mathbf{a} - \mathbf{a}^T \mathbf{C} \mathbf{a}}{\mathbf{a}^T \mathbf{B} \mathbf{a} - \mathbf{a}^T \mathbf{D} \mathbf{a}} \] (10)

its optimization can be solved with ([15]):

\[
\left[\frac{1}{\mathbf{a}^T \mathbf{B} \mathbf{a}} \frac{\mathbf{a}^T \mathbf{C} \mathbf{a}}{\mathbf{a}^T \mathbf{D} \mathbf{a}} \right] \mathbf{a} = \frac{1}{\mathbf{a}^T \mathbf{B} \mathbf{a}} \left( \frac{\mathbf{a}^T \mathbf{A} \mathbf{a}}{\mathbf{a}^T \mathbf{D} \mathbf{a}} \mathbf{a} - \mathbf{a}^T \mathbf{B} \mathbf{a} \mathbf{C} \right) \mathbf{a} = \frac{1}{\mathbf{a}^T \mathbf{B} \mathbf{a}} \left( \mathbf{a}^T \mathbf{B} \mathbf{a} \mathbf{C} \mathbf{a} \right) \] (11)

\[ = \lambda \left( \frac{1}{\mathbf{a}^T \mathbf{B} \mathbf{a}} \mathbf{B} + \frac{1}{\mathbf{a}^T \mathbf{D} \mathbf{a}} \mathbf{D} \right) \mathbf{a} \text{ or } \mathbf{E} \mathbf{a} = \lambda \mathbf{F} \mathbf{a}. \]

The proof of (11) can be deduced from the derivative of a vector-valued function in an \(N+1\)-dimensional space. The optimization is now done iteratively. A suitable vector \(\mathbf{a}\) to start with is the eigenvector that corresponds to the largest eigenvalue of the two quadratic forms. After calculating the matrices \(\mathbf{E}\) and \(\mathbf{F}\), a generalized eigenvalue problem of exactly the same form as that in (8) is obtained. The eigenvector corresponding to the optimum solution of \(\mathbf{E} \mathbf{a} = \lambda \mathbf{F} \mathbf{a}\) is used in the next iteration. The computation can be continued until a maximum is reached with the desired degree of accuracy.

Optimization with Constraints

The problem of optimizing a function subject to \(M\) constraints is usually solved with the aid of Lagrange multipliers ([9]–[11]). However, due to the elegant mathematical formulation adopted here, it is possible to convert this \(N+1+M\) problem into a \(N+1-M\) problem. In this way, the manipulations needed to come to a solution are reduced when the number of constraints is increased. An explanation for this is that constraints reduce the number of variables which can be adjusted.

Antenna pattern constraints can be represented in the following way:

\[
\mathbf{l}_m^T \mathbf{u} = \mathbf{v}, \quad \mathbf{l}_m^T(\mathbf{u}) = \mathbf{v} \quad \text{or} \quad \mathbf{l}_m^T(0) = \mathbf{v} \Rightarrow \mathbf{q}_m = 0 \] (12)

where \(\mathbf{v}\) is the prescribed value in the direction \(\mathbf{u}_m\) relative to the value at \(u = 0\) and \(\mathbf{q}_m\) are constraint vectors \((m=1,\ldots,M<N+1)\). The function to be optimized now becomes:

\[
\mathbf{h}(\mathbf{a}) = \frac{\mathbf{a}^T \mathbf{A} \mathbf{a}}{\mathbf{a}^T \mathbf{B} \mathbf{a}}, \quad \mathbf{a}^T \mathbf{q}_m = 0 \quad (m = 1,\ldots,M \text{ and } M < N+1) \] (13)
This problem can be solved by supposing that $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M$ span an $M$ dimensional space $\mathcal{Y}$, and the $N+1-M$ dimensional space $\mathcal{X}$ is spanned by $\mathbf{w}_j$ ($j=M+1, \ldots, N+1$). Since $\mathbf{a}^T \mathbf{q}_n = 0$, the vector $\mathbf{a}$ must lie in the space $\mathcal{Y}$ and it can be written as:

$$\mathbf{a} = \sum_{j=M+1}^{N+1} \mathbf{w}_j \mathbf{c}_j = \mathbf{W}_c \mathbf{c}$$  \hspace{1cm} (14)$$

where $\mathbf{W} = [\mathbf{w}_{M+1}, \ldots, \mathbf{w}_{N+1}]$ is an $N+1 \times N+1-M$ matrix (the columns are formed by the vectors $\mathbf{w}_j$) and $\mathbf{c}$ is an $N+1-M$ vector. The problem has now been reduced to the determination of the vectors $\mathbf{c}$ and $\mathbf{w}_j$, and optimizing:

$$h(\mathbf{q}) = \frac{\mathbf{c}^T \mathbf{W}^T \mathbf{A} \mathbf{W} \mathbf{c}}{\mathbf{c}^T \mathbf{W}^T \mathbf{B} \mathbf{W} \mathbf{c}}$$  \hspace{1cm} (15)$$

with $\mathbf{W}^T \mathbf{A} \mathbf{W}$ and $\mathbf{W}^T \mathbf{B} \mathbf{W}$ being $N+1-M$ real square matrices (see fig. 1.a.).

Finding a basis for $\mathcal{Y}$ can be achieved in different ways. One way is via the Gram–Schmidt transformation, but the Householder transform with partial pivoting guarantees a better numerical stability. A property of the Householder transformation is that it reduces a $N+1 \times M$ matrix $\mathcal{V}$ ($\mathcal{V} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M]$) to an upper tridiagonal form; with $\mathbf{Q}$ an $N+1 \times N+1$ orthogonal matrix (Householder matrix) and $\mathbf{R}$ an $M \times M$ upper tridiagonal matrix (see fig. 1.b.). Defining $\mathbf{Q}$ as shown above, indicates that the last $N+1-M$ rows of $\mathbf{Q}$ ($=\mathbf{Q}_2$) form a basis for $\mathcal{X}$ (because $\mathbf{Q}_2 \mathcal{V} = \mathbf{0}$). Substituting $\mathbf{W} = \mathbf{Q}_2$ in equation (15) gives:
The advantage of using the Householder transform now becomes clear, because the matrix's product $Q_2^T A Q_2$ (or $Q_2^T B Q_2$) does not have to be evaluated by matrix multiplication, because it can be evaluated with two Householder transforms.

\[ h(x) = \frac{C Q_2^T A Q_2 C}{C Q_2^T B Q_2 C} \quad (16) \]

Comparison with Literature and Examples

The results obtained from the optimization method presented here can be checked and compared with others quoted in literature. Since it is not possible to include all the published results here, a selection was made. This selection intends to reach a large variety of source functions adopted and pattern requirements stated. The results obtained from the new procedure are identical to those of Kritskiy and Novosartov [6] (source function: Bessel $J_{\nu}(r)$ with $J_{\nu}(r)=0$; optimization of $\eta_0$); Mironenko [5] (source function: Zernike $R_{2n}^{m}(r)$; optimization of $\eta_0$); as well as, Kouznetsov [7] (source function $(1-r^2)^2$; optimization of $\eta_0$). Comparisons with Ling et al. [16] (source function: Bessel $J_{\nu}(r)$ with $J_{\nu}(r)$; minimization of \( \sigma^2 \) and Borgiotti [4] (source function: Bessel $J_{\nu}(r)$ with $J_{\nu}(r)=0$; optimization of $\eta_0$) show very small differences. The small difference with the results of Ling et al. is caused by the fact that they took $\alpha$ for the upper limit of the integral in equation (4). This is only allowed if the aperture distribution smoothly approaches zero at the edge. The difference in the case of Borgiotti is due to the poor accuracy with which his pattern was described (as quoted in Borgiotti [4], page 655).

Figure 2 shows some of the results obtained from the described optimization procedure with constraints. If there are requirements with respect to the sidelobe–peak levels within a specific angle region (fig.2.a.), the optimization procedure has to be done iteratively. Because the positions ($u_m$) of the peak levels are not known in advance, some starting positions have to be chosen. Suitable starting points will be those lying midway between the two nulls of the pattern in the unconstrained case. The starting points for any next iteration step will be midway between the old points and the position of the new maxima. This procedure is repeated till all sidelobe–peak levels have reached the desired level with a prescribed accuracy. If the problem requires the sidelobe envelope to be kept below a certain level (see fig.2.b.), it is better to start the procedure with only one constraint with respect to that sidelobe which is closest to boresight that exceeds the prescribed sidelobe envelope (the first sidelobe). If the level of the next sidelobe away from boresight exceeds the prescribed level, the procedure has to be repeated with two.
This is done for all sidelobes within the angle region of interest. In this way, it is possible to end with the highest number of variables which can be used for optimization purposes. Figure 2.c. shows an example of constraints to the main beam, in this case, a flat topped beam. The convergence rate for the procedure when antenna parameters are optimized using different source functions is shown in figures 3 and 4, for the unconstrained and constrained case, respectively. Figure 3.a. and 4.a. shows the optimization of $\eta_0$. The value $c$ for the upper boundary of the integral of (4) has been taken as 3.5. This value assures a narrow beam with a low first sidelobe, because it assures that even with an uniform illuminated aperture no sidelobe is in this region. Figures 3.b. and 4.b. show the optimization of $\eta_a \eta_b$ and figures 3.c. and 4.c. of $\eta_a \eta_b/\sigma^2$. The optimization of the latter product is interesting for ground-based radiometry purposes. A maximal $\eta_b$ will assure a high amount of power in a prescribed region and a minimal $\sigma^2$ will assure a small spread around the axis $u=0$, thus making the far-out sidelobes low. Including the maximization of $\eta_a$ will prevent the antenna from becoming too large, thereby, reducing the costs of the antenna.

Conclusions

The method presented in this paper makes it possible to optimize two (or more) antenna parameters simultaneously, with or without constraints. The results obtained with this method agree well with those found in literature. The examples given represent a small set of the variety of pattern requirements and optimization parameters that can be handled. From the resulting figures, it is possible to deduce the most suitable source function for optimization.

Acknowledgements

The author wishes to thank dr.M.H.A.J. Herben for his valuable help and discussions, and prof.G. Brussaard for his continued interest.
References

Design of circular apertures for narrow beamwidth and low sidelobes.

Analysis and synthesis of radiation patterns from circular aperture.

[3] Ruze, J.
Circular aperture synthesis.

Design of circular apertures for high beam efficiency and low sidelobes.

Synthesis of a finite-aperture antenna maximizing the fraction of power radiated in a
prescribed solid angle.

Deviation of the optimum field distributions for antennas with a circular aperture.

[7] Konosnov V.D.
Side lobe reduction in circular aperture antennas.

Optimum aperture monopulse excitations.

Constrained optimization of the performance indices of arbitrary array antennas.

Optimization of directivity and signal-to-noise ratio of an arbitrary antenna array.

Optimization of array performance subject to multiple power pattern constraints.

Microwave antenna theory and design.
McGraw-Hill Book Company, Inc. New York, 1949, MIT Radiation Laboratory Series,
Vol. 12.

The theory of matrices.

Optimisation of circular—aperture distributions for interference reduction.

A general optimization method for reflector antenna synthesis.
EUT report (to be published 1990), Eindhoven.

Radiation pattern of planar antennas with optimum and arbitrary illumination.
Radiometer Antennas

Figure 2: Different combinations of pattern (sidelobe-)structure constraints and antenna parameter optimization. The far-field patterns are shown, after optimization of:

a) $\eta_0$ (source function: Bessel, aperture-field distribution: $\sum_{n=1}^{15} a_n J_0(\nu_n r)$ with $J_0(\nu_n) = 0$, linearly decreasing sidelobe level)

b) $\eta_a \eta_b$ (source function: Zernike, aperture-field distribution: $\sum_{n=1}^{10} a_n R_0^0(r)$, sidelobe level $\leq -30$ dB)

c) $\eta_a \eta_b / \sigma^2$ (source function: $(1-r^2)^n$, aperture-field distribution: $\sum_{n=1}^{6} a_n (1-r^2)^n$, flat topped beam).
Figure 3: The convergence rate for the procedure, after unconstrained optimization of different antenna parameters, using different source functions. The value of the parameter is given against \( N \), where \( N \) is the number of elementary functions in the series. The corresponding optimal far-field patterns are also shown (\( N=15 \)).

a) \( \eta_0 \)  

b) \( \eta_0 \eta_0 \)  

c) \( \eta_0 \eta_0 / \sigma^2 \)
Figure 4: The convergence rate for the procedure, after constrained optimization (sidelobe peak level ≤ -30 dB) of different antenna parameters, using different source functions. The value of the parameter is given against $N$, where $N$ is the number of elementary functions in the series. The corresponding optimal far-field patterns are also shown ($N = 15$).

a) $\eta_b$  

b) $\eta_a \eta_b$  

c) $\eta_a \eta_b / \sigma^2$
2.3. An Optimization Method for Radiometer Antennas

In the previous section a general optimization method was presented and the most suitable optimization functions were selected. This method will be specifically applied to a radiometer antenna in this section.

Note: This section will be published in the proceedings of the JINA symposium, Proc.JINA'92 Int.Symp.Antennas, 12–14 November, 1992. Therefore, the numbering of Equations and references, as in the previous sections, does not follow that for the rest of the thesis.

An Optimization Method for Radiometer Antennas
by
P.J.I. de Maagt and J.W. Wittekamp
Eindhoven University of Technology
Faculty of Electrical Engineering
Telecommunications Division
P.O.Box 513, 5600 MB Eindhoven, The Netherlands.

Abstract

This paper examines an analytical method for optimizing radiometer antennas which is based on writing their design criteria as a ratio of two quadratic Hermitian forms, so that more than one antenna parameter (such as antenna efficiency or beam efficiency) can be optimized simultaneously, with or without pattern-structure constraints. This method was described by de Maagt [1] but is has been extended and applied to a radiometer antenna. Firstly, the mathematical equations for the parameters available for a radiometer antenna that has to be optimized need to be calculated. Further, the use of Zernike polynomials for the optimization will be elucidated and it is shown that combining the parameters as a sum or a product will lead to the same results. Finally, some values of optimal radiometer antenna parameters are presented.

Introduction

The growth of satellite communications throughout the world has prompted the use of higher frequency bands. However, those above 14 GHz are not so simple to use as the lower frequency bands. Atmospheric attenuation, particularly during heavy rain, can be so severe that it essentially limits the required availability of the satellite link; so, there is a need to explore the higher frequency bands with propagation experiments.
One way of performing propagation experiments is to use satellite beacon signals; however, when they are not available, an alternative is to use radiometry. Radiometry is a remote sensing technique which makes use of the fact that there is a relationship between radiowave absorption by the propagation medium and its emission of thermal radiation. A radiometer, which is a highly sensitive microwave receiver and antenna, is capable of measuring thermal noise; therefore it is capable of providing relevant information about propagation losses. A misapprehension concerning some of the present-day radiometer designs (especially, if ground-based) is that antennas based on communication systems are used, because a radiometer is basically a remote sensing instrument, for which one has to take into account other design criteria. It is not always possible to make such criteria specific for the criteria used in designing ground-station communication systems. The prime objective of "communication" antennas is to optimize antenna gain and the rest of the far-field pattern is of less important as long as the CCIR—requirements are satisfied. However, this does not apply to radiometry; since it is based on an integral relationship between antenna pattern and brightness temperature, the entire pattern must be considered as a whole. The main beam and the nearest sidelobes define the resolution of the remote sensing device and the far-out sidelobes define the sensitivity to noise from other objects (e.g. from the ground). In contrast to optimizing the gain factor (which is proportional to the aperture efficiency \( \eta_a \)), maximizing the fraction of power in a certain angle region is relevant (represented by the beam efficiency \( \eta_b \)), leading to a well defined resolution. However, optimizing \( \eta_b \) neglects the structure of the antenna pattern outside the angle region for which \( \eta_b \) was calculated. This is illustrated in figure 1.

From the viewpoint of optimizing \( \eta_b \), all the examples will be the same, but the resolution and sensitivity to noise from other objects realized with these patterns will not be the same. It should be clear that the first pattern will yield the best resolution and the lowest sensitivity to extraneous noise because most of the power is concentrated near to the main beam. So, the sidelobe structure cannot be ignored and must be taken into consideration. If a radiometer is designed to provide relevant information concerning microwave attenuation along a satellite-to-earth link, it will be desirable to receive as much power as possible in a well-defined region around the antenna's axis. In that way, the radiometer will be less sensitive to noise from outside this region. A figure of merit that represents this is the integrated pattern function \( h \). As will be shown, constrained optimization of the integrated pattern function \( h \) is analogous to optimizing the moments \( \mu_n \) of the pattern.

It follows that combining the beam efficiency \( \eta_b \) and the moments \( \mu_n \) should be maximized for a radiometer antenna, but a drawback of such a combined parameter is that optimization is inclined to result in an antenna with a very large beamwidth. In that case,
The resolution will decrease, that is undesirable, however, as these requirements are mutually exclusive, a compromise should be reached. A parameter that is directly related to the beamwidth, is the aperture efficiency and it is well-known that when \( \eta_a = 1 \), the beamwidth is smallest, while decreasing \( \eta_a \) leads to broadening of the beam. This implies that a trade-off between \( \eta_a \) and \( \eta_b \) can lead to a compromise between \( \eta_b \) and beamwidth. As a result, each radiometer antenna should be designed to maximize the combination of \( \eta_a \) (or beamwidth), \( \eta_b \) and \( \mu_n \).

Starting with the mathematical formulation of those parameters it will be explained how the circle polynomials of Zernike can be used in the optimization and the optimization of the integrated pattern by means of the moments of the pattern is discussed. Then, it is shown that different combinations of the parameters (as product or sum) lead to the same optimal results. Finally, the values for optimal radiometer antenna parameters are presented.

**Mathematical Formulation**

In this section, the antenna parameters and far-field pattern structure which are useful for optimizing radiometer antennas are written in an appropriate form. The mathematical formulation relates to an antenna with a circular aperture. The aperture points are given by normalized aperture polar coordinates \((r, \phi')\) and the far-field observation point by spherical coordinates \((R, \theta, \phi)\).

The integral part \( g(u, \phi) \) of the far-field pattern \( E(R, \theta, \phi) \) is related to the aperture distribution \( f(r, \phi') \) by the integral [2]:

\[
g(u, \phi) = \frac{(D/2)^2}{4\pi} \int_0^1 \int_0^{2\pi} f(r, \phi') e^{jru \cos(\phi-\phi')} r dr d\phi' \tag{1}
\]

with \( u = \frac{\theta}{D} \sin \theta \).

In the case of a rotationally symmetric equiphase aperture-field-distribution, \( f(r, \phi') \) can be written as:

\[
f(r) = \sum_{n=0}^{N} a_n e_n(r) = \mathbf{a}^T \mathbf{e} \tag{2}
\]

with \( \mathbf{a}^T = (a_0, a_1, \ldots, a_N) \) an excitation vector and \( \mathbf{e}^T = (e_0, e_1, \ldots, e_N) \) a set of orthogonal real functions.

Consequently, the rotationally symmetrical far-field pattern \( g(u) \) can be written as the
first order Hankel transform of $f(r)$:

$$
\frac{g(u)}{2\pi(D/2)^2} = \int_0^1 f(r)J_0(ur) r dr = \sum_{n=0}^N a_n \mathbf{I}_n(u; e_n) = \mathbf{a}^T \mathbf{I}
$$

(3)

where $I_n(u; e_n) = \int_0^1 e_n(r)J_0(ur) r dr$ and $J_0$ the Bessel function of the first kind and zero order.

When writing the aperture efficiency $\eta_a$, the beam efficiency $\eta_b$, the integrated power pattern $h$, and the moments $\mu_n$ as a ratio of two quadratic Hermitian forms, some other relations must be used. These are the power radiated by the aperture $p_r$, the power radiated within a prescribed solid angle $p_{r,\text{angle}}$, and the moments $\mu_n$. The first two are being given by:

$$
p_r = \frac{1}{2\pi(D/2)^2} \int_0^1 f^2(r) r dr, \quad p_{r,\text{angle}} = \frac{c}{2\pi} \int_0^\infty u p(u) du
$$

(4)

with $p(u) = g^2(u)$ and $c = (\pi D/\lambda) \sin \theta_{\text{pre}}$, where $\theta_{\text{pre}}$ is the prescribed angle.

The moments of the far-field power pattern with respect to the axis $u=0$ are found by integrating $u^n p(u)$ which will lead to:

$$
\mu_n = \int_0^{\pi D/\lambda} u^n p(u) u du = \int_0^{\pi D/\lambda} u^{n+1} p(u) du
$$

(5)

Using Eqs. (2)–(5), it is simple to write the antenna parameters as a ratio of two quadratic forms:

$$
\eta_a = \frac{2 \mathbf{a}^T \mathbf{V}(0) \mathbf{a}}{\mathbf{a}^T \mathbf{A} \mathbf{a}}, \quad \eta_b = \frac{\mathbf{a}^T \mathbf{X} \mathbf{a}}{\mathbf{a}^T \mathbf{A} \mathbf{a}}, \quad h(u) = \frac{\mathbf{a}^T \mathbf{Y} \mathbf{a}}{\mathbf{a}^T \mathbf{A} \mathbf{a}}, \quad \mu_n = \frac{\mathbf{a}^T \mathbf{W} \mathbf{a}}{\mathbf{a}^T \mathbf{A} \mathbf{a}}
$$

(6)

where $V,A,X,Y,W$ are $N+1$ square matrices with the following elements:

$$
A_{ij} = \int_0^1 e_i e_j r dr, \quad V_{ij} = \mathbf{I}(u; e_i) \mathbf{I}(u; e_j) \quad c u^{\pi D/\lambda}
$$

$$
X_{ij} = \int_0^\infty u V_{ij} du, \quad Y_{ij} = \int_0^\infty u V_{ij} du, \quad W_{ij} = \int_0^\infty u^{n+1} V_{ij} du.
$$

(7)
Eq. (7) shows that the integrated pattern function $h$ is a corollary of the beam efficiency; the beam efficiency has a fixed upper limit for the integration, while this value is variable in case of the integrated pattern function.

Selecting the elementary functions to be used in the optimization is governed by certain considerations, including: the ease of approximating the desired pattern with a minimum number of terms in a series, the property of orthogonality, and the ease with which the functions can be Hankel transformed.

Slepian [3] showed that the maximum value of $\eta_b$ could be attained when the illumination was a radial function, that is a solution of the Fredholm integral equation with a largest eigenvalue $\alpha(p)$ of:

$$ a(p)S_{00}(p,r) = \int_0^1 S_{00}(p,s) J_0(prs)sds $$

(8)

where the functions $S_{00}(p,r)$ are hyperspheroidal functions.

The Zernike circle polynomials are a limiting case of hyperspheroidal functions ($p \to 0$) and they can be shown to give good results for performing the optimization of $\eta_b$ or a combination which includes $\eta_b$ [4]. Moreover, the Zernike polynomials have the previously mentioned desirable properties.

When using the Zernike polynomials $R^{0}_{2n}(r)$ and the following relationship [5]:

$$ \int_0^1 R^{0}_m(r) J_\alpha(u)rdr = (-1)^{(2n-m)} \frac{J_{2n+1}(u)}{u} $$

(9)

Eq. (3) can be written as:

$$ \frac{g(\alpha)}{2\pi(D/2)^2} = \sum_{n=0}^{N} a_n \frac{J_{2n+1}(u)}{u} $$

(10)

So, $g = (R^{0}_0(r), R^{0}_2(r), \ldots, R^{0}_N(r))$

(11)

and $I(e) = (\frac{J_1(u)}{u}, \frac{J_3(u)}{u}, \ldots, \frac{J_{2n+1}(u)}{u})$.

(12)

The elements of the matrix $A$ are found from [5]:

$$ \int_0^1 R^{0}_{2i}(r)R^{0}_{2j}(r) rdr = \begin{cases} \frac{1}{4i+2} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} $$

(13)
and the elements of the matrix $X$ are found from [6, p.135]:

$$
\int_0^c J_{2i+1}^2(u) J_{2j+1}^2(u) \frac{du}{u} = c \left( \frac{J_{2i+1}^2(c) J_{2j+1}^2(c) - J_{2i+1}(c) J_{2j+1}(c)}{(2i+1)^2 - (2j+1)^2} \right) + \frac{J_{2i+1}^2(c) J_{2j+1}(c)}{2i+2j+2}
$$

if $i \neq j$, and modifying Hansen's equation [6, p.152]:

$$
\int_0^c J_{2i+1}^2(u) \frac{du}{u} = \frac{1}{2(2i+1)} \sum_{n=0}^{\infty} \epsilon_n J_{2i+1}^2(c) \quad \text{if } i=j
$$

where $\epsilon_n$ is Neumann's factor which is defined as:

$$
\epsilon_n = 1 \quad \text{for } n = 0
$$

$$
\epsilon_n = 2 \quad \text{elsewhere}.
$$

Finally, the elements of the matrix $W$ and $Y$ are given by:

$$
W_{ij} = \int_0^{\pi D/\lambda} u^{n-1} J_{2i+1}^2(u) J_{2j+1}^2(u) \, du, \quad Y_{ij} = \int_0^{\pi D/\lambda} u J_{2i+1}^2(u) J_{2j+1}^2(u) \, du
$$

which have to be calculated numerically.

The Optimization Procedure

The antenna parameters can be optimized by solving a general eigenvalue problem. Consider the problem of optimizing a function which can be written as a ratio of two Hermitian forms:

$$
h(a) = \frac{\mathbf{a}^T \mathbf{A} \mathbf{a}}{\mathbf{a}^T \mathbf{B} \mathbf{a}}
$$

where $\mathbf{a}$ is an $N+1$-element vector and $\mathbf{A}$ and $\mathbf{B}$ are $N+1 \times N+1$ real matrices. If $\mathbf{A}$ and $\mathbf{B}$ are Hermitian and if $\mathbf{B}$ is positive definite, the maximum (or minimum) of the ratio will be given by the largest (or smallest) eigenvalue determined from:

$$
\mathbf{A} \mathbf{a} = \lambda \mathbf{B} \mathbf{a}.
$$

The matrices $V,A,X,Y$ and $W$ satisfy these requirements. They are all Hermitian, as can be seen from Eq.(7), and positive definite because they represent: the power in the forward direction; the total radiated power; the power radiated within a prescribed solid angle or within a certain solid angle; and the moments of the far-field pattern, respectively.

Optimizing a product of ratios of Hermitian forms (see [1]) or a sum of ratios of Hermitian forms can be solved analogously to Eq.(18).

The problem of optimizing a function subject to $M$ constraints is usually solved with the aid of Lagrange multipliers. However, due to the elegant mathematical formulation adopted
here, it is possible to convert the $N+1+M$ problem into a $N+1-M$ problem, so that the steps needed to reach a solution are reduced as the number of constraints increases. In this case, optimization uses a Householder transform with partial pivoting in order to reduce the size of the problem. A detailed description can be found in [1].

**Optimization of the Integrated Pattern**

The integrated pattern function $h$ can be written as:

$$h(u) = \frac{2\pi}{P_r} \int_0^u p(u) u du$$  \hspace{1cm} (19)

The integrated pattern can be optimized by making use of the moments of the power pattern. Since the normalized power pattern $p(u)$ integrated over the half sphere is unity and because $p(u)$ is always larger than or equal to zero, $p(u)u$ can be considered to be a probability density function (pdf).

So:

$$\frac{2\pi}{P_r} \int_0^{\pi D/\lambda} p(u) u du = 1 = \int_0^{\pi D/\lambda} pdf(u) du$$  \hspace{1cm} (20)

where $pdf(u) = \frac{2\pi}{P_r} p(u) u$ the probability density function whose moments are defined as:

$$\mu_n = \int_0^{\pi D/\lambda} u^n pdf(u) du$$  \hspace{1cm} (21)

Using the Bienayme–Chebyshev inequality for the probability $P$ [7] gives:

$$P(|u| \geq \varepsilon) \leq \frac{\mu_n}{\varepsilon^n}$$  \hspace{1cm} (22)

Then, it is possible to write:

$$(1-h(\varepsilon)) \leq \frac{\mu_n}{\varepsilon^n}$$  \hspace{1cm} (23)

In this way the moments determine the behaviour of the tail of the integrated pattern $h(u)$. Since it is most important to maximize $h(u)$ for relatively small values of $u$, it is better to have as little power as possible in the tail of the integrated pattern and, consequently, it is important to minimize the moments of the pattern.

The constrained optimization of the integrated pattern is based on Eq. (23), the value $n$ of the moment being increased until the constraint is satisfied. If there are more constraints, the constrained optimization will begin with the constraint that is closest to boresight.
When it is satisfied, a check is performed if the other constraints are also satisfied. If not, the value of n can be increased until all the constraints are satisfied.

The validity of the paraxial approximation could be a problem, when determining the integrated pattern and the moments; however, a numerical evaluation of $p_x$ (Eq.(4)) and determination of $h(xD/\lambda)$ proved that the discrepancies can be neglected. This meant that the angle region, where the paraxial assumption ceased to hold, made no significant contribution to the integrated pattern. The relative difference between the total integrated pattern and total radiated power from the aperture for optimal values of $\eta_a$, $\eta_b$ and $0.5\eta_a + 0.5\eta_b$ is shown in figure 2. It became clear that even for optimizing $\eta_a$, which leads to a relatively high edge illumination, produced errors less than 3% for systems with dimensions larger than 10\lambda. Optimization of the moments gave an even smaller error due to relatively low edge illumination.

**Combinations of Parameters**

As stated before, combining different antenna parameters is important when considering the optimization of radiometer antennas. There are two ways of combining parameters: as a sum or as a product. It is uncertain whether these two ways will give the same optimization results or not, but it can be answered either intuitively or mathematically. As the matrices are positive definite, the solution of the eigenvalue problem has to be unique. It is easy to show mathematically that both ways have the same optimal value.

This has been shown in [8] for N parameters, but for the sake of clarity, here, only two parameters are combined.

Considering the weighted optimization of the parameter $s$:

\[
s = w \eta_x(a) + (1-w) \eta_y(a)
\]

which is optimal if $\frac{ds}{da} = 0$, which yields

\[
\frac{d\eta_x}{da} = \frac{w-1}{w} \frac{d\eta_y}{da}
\]

The weighted product $p$ is:

\[
p = \eta_x^a(a) \eta_y^{1-a}(a)
\]

and it yields an optimal value if:

\[
\frac{d\eta_x}{da} = \frac{a-1}{a} \frac{\eta_x}{\eta_y} \frac{d\eta_y}{da}
\]

The optimal values are equal if:

\[
a = \frac{\eta_x w}{\eta_y + (\eta_x - \eta_y)w}
\]
It is easy to show that these optimal values are identical when \( w = 1 \) or \( w = 0 \). When \( w = \frac{1}{2} \), a situation occurs which is representative of all \( w \in (0,1) \). Starting with optimization of the sum combination, two optimal values for \( \eta_x \) and \( \eta_y \) are obtained and inserting these values in Eq.(28) gives a value \( a \). Optimization of the product, with this value will give the same optimal values for \( \eta_x \) and \( \eta_y \). So, all the optimal values found when optimizing the sum combination are found when optimizing with the product combination. For reason of computing efficiency a weighted sum is preferable in the optimization.

**Results**

Figure 3.a shows the results for \( \eta_a \) and \( \eta_b \) after optimizing \( w \eta_a + (1-w) \eta_b \) and figure 3.b shows the results for \( \eta_b \) and the beamwidth. Each point on the curve represents an antenna system and to reach the corresponding aperture distribution, a shaped reflector antenna system is needed. The optimization process gives an optimal system in the sense that it is impossible to construct a better antenna, given the same constraints. In other words, knowing the theoretically optimal system makes it unnecessary to construct and test a large number of practical antenna systems. The result for \( \eta_a \) and \( \eta_b \) offers the possibility to determine the best multipurpose antenna for both radiometry (\( \eta_b \)) and communication purposes (\( \eta_a \)). The curve of \( \eta_b \) versus beamwidth is most suitable if the antenna is to be used for radiometry only [8].

An example of optimization subject to constraints with respect to the integrated pattern \( h \), a system which has to satisfy \( h(5)=98\% \) and \( h(7)=99\% \) is taken. The optimal values for \( \eta_a \) versus \( \eta_b \) and beamwidth versus \( \eta_b \) are shown in figure 4; while, figure 5 shows the corresponding integrated pattern for \( w = 0.5 \) (both constrained and unconstrained).

**Conclusions**

The method of optimization presented in this paper makes it possible to optimize radiometer antenna parameters with or without constraints. Optimization with constraints to the integrated pattern \( h \) can be done with the moments of the far-field pattern; that way, it is possible to optimize the integrated pattern analytically, a method which was not found in open literature. It is shown that different ways of combining the parameters (as a product or sum) lead to the same optimal results. A theoretically optimal system is described; it has the advantage of making sure that a better antenna can not be constructed, using the same constraints. In other words, with the knowledge of the theoretical optimum, it is unnecessary to consider constructing a large number of different practical antenna systems, in order to find the best.
Acknowledgements
The authors wish to thank prof. G. Brussaard and dr. M. H. A. J. Herben for their valuable help during interesting discussions.

Fig. 1. Different patterns with equal $\eta_b$.

Fig. 2. The relative error between the power radiated from the aperture and the total integrated power.
The optimal value of a) weighted combinations of $\eta_a$ and $\eta_b$ and b) combinations of beamwidth and $\eta_b$.

The optimal value of a) weighted combinations of $\eta_a$ and $\eta_b$ and b) combinations of $\eta_b$ and beamwidth, subject to the constraints $h(5)=0.98$ and $h(7)=0.99$. 
Fig 5. a) The integrated pattern for an optimal value of $0.5\eta_a+0.5\eta_b$.
b) The integrated pattern with the constraints $h(5)=0.98$ and $h(7)=0.99$.

References

[1] De Maagt, P.J.I.
A GENERAL OPTIMIZATION METHOD FOR REFLECTOR ANTENNA SYNTHESIS.

MICROWAVE ANTENNA THEORY AND DESIGN.

[3] Slepian, D.
PROLATE SPHERICAL WAVE FUNCTIONS

A SYNTHESIS METHOD FOR COMBINED OPTIMIZATION OF MULTIPLE ANTENNA PARAMETERS AND ANTENNA PATTERN STRUCTURE.

[5] Erdélyi, A.
TABLES OF INTEGRAL TRANSFORMS

A TREATISE ON THE THEORY OF BESSEL FUNCTIONS.

[7] Papoulis, A.
PROBABILITY & STATISTICS.

AN OPTIMIZED RADIOMETER ANTENNA; Theory and Design.
2.4. A Review and Comparison of Some Asymptotic Techniques for Calculating the Wide-angle Radiation Pattern of Paraboloid Reflector Antennas.

In the two preceding sections a framework was developed for specifying an optimal radiometer antenna pattern. As the optimum must not go beyond the possible or contain requirements that unnecessarily increase the costs, the choice of a radiometer antenna must take into account optimal antenna performance versus expected performance in practice. Setting the parameters for practical antennas requires knowledge of the complete radiation pattern of the antenna in question. Asymptotic techniques are used in order to compute the field time efficiently and there are two ways that the field can be determined asymptotically.

When the antenna is large in terms of wavelength, the process for determining the far-field pattern lends itself to a simple geometrical interpretation in terms of reflected and diffracted rays, satisfying Keller's extended version of Fermat's principle.

An alternative way is to evaluate the physical optics integral asymptotically by means of the stationary phase method.

Due to the fact that no references could be found in open literature giving a direct comparison between the results obtained in these two ways, the most common asymptotic techniques were applied to the same front-fed reflector antenna configuration. This comparison was published in a paper and it is reproduced here as section 2.4 under the same title.

Note: This section was published as a paper in Electromagnetics, vol.12 (1992), p.57-75; therefore, the numbering of Equations and references does not follow the rest of the thesis.
A Review and Comparison of Some Asymptotic Techniques for Calculating the Wide-angle Radiation Pattern of Paraboloid Reflector Antennas.

by
P.J.I. de Maagt
J. Chen*
M.H.A.J. Herben

Eindhoven University of Technology,
Faculty of Electrical Engineering,
Telecommunications Division,
PO Box 513, 5600 MB Eindhoven, the Netherlands.

* seconded to Eindhoven University of Technology from Chengdu Aircraft Company, PO Box 609, Chengdu, Sichuan, People's Republic of China.

Abstract

The paper starts with a review of calculation methods for determining the complete far-field radiation patterns of reflector antennas. Subsequently, a direct comparison is made between PO, APO, CAPO, GTD, UTD and EEC methods by applying all of these techniques to one and the same reflector antenna configuration. It is shown that both UTD, including surface-diffracted rays, and CAPO are satisfactory for calculating the scattered field from reflector antennas in wide-angle regions (away from the antenna axis). Correct results in both forward- and rear-axial region can be obtained with PO and EEC, respectively. Finally, special attention is paid to the transition angle between EEC and UTD.

1. Introduction

Reflector antennas have been used for about fifty years in radio astronomy, microwave communication and remote sensing. The demand for highly sensitive antenna systems requires accurate calculations of the wide-angle and the rear-direction antenna patterns in order to examine the effects of possible interference within this angle region. Several methods ([1]–[4]) have been introduced for calculating the far-field radiation patterns of reflector antennas. To find the angle region where the results obtained with these methods will be valid, some potential methods are reviewed and compared in this
paper. Those methods are PO, APO [5], CAPO [6], GTD [7], UTD [8] and EEC ([9]–[11]).

Physical Optics (PO) simply uses the currents derived from the application of the Geometrical Optics (GO) theory in order to approximate the currents on the reflector. By evaluating the contributions from all parts of the reflector to the field at an observation point, PO gives the total field. PO is generally used for calculating only the main-lobe and the first few sidelobes, because the wide-angle far-field calculation with this method is usually complicated and includes time consuming numerical integrals. But, if there are stationary points in the integral so that the stationary phase method can be used, the field integrals can be evaluated asymptotically. The method using stationary phase integration is called Asymptotic Physical Optics (APOL, and it was first applied by Rusch ([5],[12]) but later developed further by Knop [13]. It is considered to be a powerful method for calculating the reflector antenna radiation pattern; however, measurements [6] have revealed that APO can give errors up to 6 dB for some angle regions. Such errors can be explained by the incorrect GO approximation of currents at the edge of the reflector; therefore Knop and Ostertag [6] derived the corrected APO (CAPO) diffraction coefficients.

Another way to avoid using the time consuming PO integrals for wide-angle far-field calculations is to switch to the Geometrical Theory of Diffraction (GTD) ([14],[15]) introduced by Keller, which is an extension of the GO theory by adding diffracted rays to the usual GO ray. The corresponding diffracted waves are assumed to follow the laws of diffraction and to diverge according to GO laws. Consequently, the diffraction points and the paths of the rays can be found with the laws of diffraction, while the amplitude of the fields along the rays can be found from the principle of energy conservation. So, this theory will not only provide a qualitative explanation of diffraction in terms of the diffracted rays, but also permit a quantitative determination of the diffracted field. The initial value of the diffracted field at the diffraction points can be obtained by multiplying the field vector of the incident wave by the dyadic diffraction coefficient, which was first obtained by Keller when comparing his hypothetical diffraction expressions with Sommerfeld's exact solutions of various canonical problems. Although GTD results are not the exact solutions of the field equations, they are the leading terms in an series expansion of such solutions for high frequencies. GTD has been widely used for calculating the scattered fields from objects with large dimensions when compared with the wavelength. However, its results are invalid for some specific far-field angle regions.

Kouyoumjian and Pathak [8] extended Keller's GTD to the uniform geometrical theory of diffraction (UTD). UTD includes diffraction coefficients which are also valid in the shadow boundary regions where Keller's theory fails. Moreover, UTD gives a compact
form of the dyadic diffraction coefficient for electromagnetic waves obliquely incident on a curved edge (of a perfectly conducting reflector surface).

The equivalent edge current method (EEC or ECM) can produce the rear—direction patterns of paraboloidal reflector antennas ([9],[10]) which cannot be calculated with APO/CAPO or GTD/UTD. It is a method based on fictitious electric and magnetic currents flowing along the edge of a reflector. They are obtained from the GTD edge diffraction coefficients for the rear direction; therefore, EEC can be seen as a corollary of GTD. In the case of a axial-symmetric reflector antenna, EEC becomes a suitable calculation tool because the line—integral along the edge can be evaluated analytically by virtue of its symmetry. EEC in its closed form is often used for a small angle region around the rear direction. However, that is only possible if the GTD diffraction coefficients remain constant. The validity of such an assumption depends largely on the size of the antenna system under consideration. Michaeli [16] criticized this approach and proposed to use the corresponding diffraction coefficients for each observation angle instead. A drawback to this modified EEC is that the integrals do not have closed forms and that it, as PO, leads to a time consuming numerical integral.

Although there are many papers dealing with some of the high—frequency asymptotic techniques mentioned, none of them gives a direct comparison by applying the techniques to the same reflector antenna configuration. This is done in the present paper.

The paper starts with a description of the reflector configuration used for comparing the techniques and the feed patterns. Then, APO, CAPO, GTD and UTD are reviewed briefly and their performance is illustrated with numerical examples. This is followed by a comparison between the different high—frequency asymptotic techniques. Finally, the EEC method is discussed in order to obtain the far—field radiation pattern in the rear direction of the antenna and special attention is paid to the transition angle between the EEC and the high—frequency asymptotic techniques.

It should be noted that the review of basic techniques is not exhaustive. Both PO and aperture field integration methods are used extensively in the calculation of the main beam and first few sidelobes. A comparison between these techniques was given by Rahmat—Samii [19]. An alternative to the combined use of GTD and EEC, for the calculation of the wide—angle far—field pattern, is PTD. Results of a comparative study between these techniques were given by Knott and Senior [20].

2. Reflector Configuration and Feed Patterns

The antenna system under consideration is an axially symmetrical paraboloidal reflector with focal length f, diameter D (=2a) and subtended angle of 2a in the z>0 region
of the rectangular coordinate system (see Fig. 1). The z-axis is the symmetry axis of the antenna and the feed is at the focus (F) which is at the origin of the rectangular (x,y,z) coordinate system. The points O' and O'' are the intersections of the z-axis with the aperture plane and reflector surface, respectively. In the following, \((\rho,\phi,\xi)\) denotes the spherical coordinates to describe the reflection and diffraction points on the paraboloidal reflector and \((r,\theta,\phi)\) are the spherical coordinates indicating the far-field observation point.

The feed that is used has a cosine shaped gain function given by \(G_f(\psi)\) = \(2(n+1)\cos^n(\psi)\) [1] with the polarization properties of an \(y\)-polarized Huygens source.

The incident fields on the edge of the paraboloid then become:

\[
E_f(\rho_0, \alpha, \xi) = A_0 \frac{G_f(\alpha)}{\rho_0} e^{-jk\rho_0[\sin\xi \psi + \cos\xi \xi]} \tag{1.a}
\]

\[
\Pi_f(\rho_0, \alpha, \xi) = A_0 \frac{G_f(\alpha)}{\eta \rho_0} e^{-jk\rho_0[-\cos\xi \psi + \sin\xi \xi]} \tag{1.b}
\]

where: \(A_0\) is a normalization constant, \(\eta\) the intrinsic impedance of free space and \(k\) the wave number \((2\pi/\lambda)\). If the integrand of the diffraction integral is a fast-oscillating function, contributions to the field at an observation point \(P(r,\theta,\phi)\) come mainly from two points. With \(APO\) they will be the stationary phase points which coincide with the GTD diffraction points. For the specific antenna system shown in Figure 1, they will be the intersection points \(Q_i\) \((i=1,2)\), of the plane containing the lines \(O'P\) and \(O'F\) with the edge of the reflector.

In the following sections most of the calculations have been carried out for a paraboloid of \(15\lambda\) diameter, having a subtended angle of \(2\alpha = 120^\circ\) and illuminated by a feed with \(n=2\). It should be noted that possible blockage effects of struts and feed are not considered because they are unimportant for this comparative study. For the same reason, only \(E\)-plane far-field patterns will be shown.
3. APO/CAPO Analysis of the Radiation from Paraboloid Structures

According to GO, the currents induced on the reflector are:

\[
\mathbf{J}_s = \begin{cases} 
2(n_{refl} \times \mathbf{E}_r) & \text{illuminated surface} \\
0 & \text{shadowed surface}
\end{cases} 
\]  \hspace{1cm} (2.a)

with:

\[
\mathbf{H}_r(\rho, \varphi, \zeta) = \frac{1}{\eta} [\rho \times \mathbf{E}_r] 
\]  \hspace{1cm} (3)

and: \(n_{refl}\) the unit vector normal to the reflector surface.

The induced current distribution results in the PO formulation for the far-field radiation pattern, which is given by the following equation [1]:

\[
\mathbf{E}(r, \vartheta, \phi) = -j \frac{k \eta}{4 \pi r} e^{-jk \rho} \int_{S_{refl}} \left[ \mathbf{J}_s - (\mathbf{J}_s \cdot \hat{r}) \hat{r} \right] e^{jk(\rho \cdot \hat{r})} dS 
\]  \hspace{1cm} (4)

For the integral form of Eq. (4), it has been proven that the phase function of the integrand becomes fast oscillating with an increasing angle of observation \(\vartheta\), while the amplitude term remains a smoothly varying function. Then, the major contributions to the integral must come from points \(Q_i(x_{s1}, y_{s1})\) where the phase function is stationary. The method for calculating the contributions from these points is the stationary phase integration method. Rusch [5] obtained the APO diffracted field in that way and the contribution of the two edge points to the far-field were found to be given by [17]:

\[\text{Fig. 1. The Geometry of a Paraboloid Reflector}\]
\[
\begin{align*}
\left[ E_{\theta_1}(r, \theta, \phi) / E_{\phi_1}(r, \theta, \phi) \right] & \approx \left[ A_0 \left( \frac{G(r)}{\rho_0} e^{-jk\rho_0} \right) \right] e^{-jk r e^{j k \left( \text{asin} \theta + \rho_0 \text{cos} \cos \theta \right)}} \left[ \frac{1}{r} \frac{a}{\text{sin} \theta} \right]. \\
& \cdot \left[ -\frac{e^{-j(\pi/4)}}{2\sqrt{2}k} \right] \left[ \frac{2}{\text{sin} \theta - (\text{sin} \varphi / \text{cos} \theta)(1 + \text{cos} \theta)} \right] \left[ \text{sin} \varphi \left( \text{cos} \theta \text{sin} \theta + \text{sin} \theta \text{sin} \theta / \text{cos} \theta \right) / \text{cos} \varphi \right] \\
& \left[ \text{cos} (\varphi) \right] \\
\left[ \text{E}_{\theta_2}(r, \theta, \phi) / E_{\phi_2}(r, \theta, \phi) \right] & \approx \left[ A_0 \left( \frac{G(r)}{\rho_0} e^{-jk\rho_0} \right) \right] e^{-jk r e^{j k \left( \text{asin} \theta - \rho_0 \text{cos} \cos \theta \right)}} \left[ e^{j(\pi/2)} \frac{a}{\text{sin} \theta} \right]. \\
& \cdot \left[ -\frac{e^{-j(\pi/4)}}{2\sqrt{2}k} \right] \left[ \frac{2}{\text{sin} \theta + (\text{sin} \varphi / \text{cos} \theta)(1 + \text{cos} \theta)} \right] \left[ \text{sin} \varphi \left( \text{cos} \theta \text{sin} \theta - \text{sin} \theta \text{sin} \theta / \text{cos} \theta \right) / \text{cos} \varphi \right] \\
& \left[ \text{cos} (\varphi) \right]
\end{align*}
\] (5.a) (5.b)

However, Rusch's formulation results in a singularity at the shadow boundary \( \theta = \alpha \). Later, Knop [13] extended Rusch's formulas to include the shadow boundary region. Those formulas can be expressed concisely as follows [17]:

\[
\begin{align*}
\left[ E_{\theta_1}(r, \theta, \phi) / E_{\phi_1}(r, \theta, \phi) \right] & \approx \left[ A_0 \left( \frac{G(r)}{\rho_0} e^{-jk\rho_0} \right) \right] e^{-jk r e^{j k \left( \text{asin} \theta - \rho_0 \text{cos} \cos \theta \right)}} \left[ \frac{j a e^{j(\pi/4)}}{\sqrt{2} \text{sin} \theta} \right]. \\
& \cdot \exp \left\{ \frac{j a \text{sin} \theta}{2} \left[ \frac{\text{ctg} (\theta/2)}{\text{ctg} (a/2)} \right]^2 \right\} \left[ \left( C_+ - \frac{1}{2} \right) e^{j (S_+ - \frac{1}{2})} \right] U^+(\theta) \left[ \frac{\text{sin} \varphi (\text{cos} \theta \text{sin} \theta - \text{sin} \theta \text{sin} \theta / \text{cos} \theta)} {\text{ctg} (a/2)} \right] \\
& \left[ \text{cos} (\varphi) \right]
\end{align*}
\] (6)

where:

\[
\begin{align*}
C_+ = C(W^+), \quad S_+ = S(W^+) \quad \text{with} \quad W^+ = \left[ \frac{\text{cos} \theta}{\text{sin} a} \right] \left[ \frac{1}{\sqrt{2} \text{ctg} (a/2)} \right] \left[ \frac{\text{ctg} (\theta/2)}{\text{ctg} (a/2)} \right] \\
C(x) = \int_0^x \text{cos} \left( \frac{\pi x}{2} \right) dx, \quad S(x) = \int_0^x \text{sin} \left( \frac{\pi x}{2} \right) dx \\
\tan (a/2) = a/(2f), \quad U^+(\theta) = \begin{cases} 
1 & (\theta \geq a) \\
\frac{L}{\text{tan} a} & (\theta < a)
\end{cases}
\end{align*}
\] (6.a) (6.b) (6.c)

Measurements presented by Knop and Ostertag [6] revealed that APO gave errors up to 6dB in the shadow region (about 130°<\theta<175° in their case). They suggested [6] that the incorrect nature of the GO edge currents used in APO was the cause of errors in the far-field calculation. Further, they used Sommerfeld's exact results and the PO values.
in order to produce the equivalent diffraction problem for a half plane, so that they could give the ratio of Sommerfeld's exact results to their PO values in order to correct the APO results.

Those ratios can be expressed by [6]:

\[
\begin{align*}
K_s &= \frac{E_{\text{exact}}}{E_{\text{po}}} = \left| \frac{\sin(-\gamma^d_d/2)}{\cos(-\gamma^i_d/2)} \right| \\
K_h &= \frac{H_{\text{exact}}}{H_{\text{po}}} = \left| \frac{\cos(-\gamma^i_d/2)}{\sin(-\gamma^d_d/2)} \right|
\end{align*}
\]

(7)

where: \( E_{\text{exact}}/H_{\text{exact}} \) denotes Sommerfeld's exact results for the electric/magnetic field parallel to the edge of a half plane, \( E_{\text{po}}/H_{\text{po}} \) are the PO values for the electric/magnetic field parallel to the edge of a half plane, and \( \gamma^i_d/\gamma^d_d \) is the angle between the incident/diffracted ray and the paraboloid surface tangent to the plane of incidence/diffraction.

By multiplying the APO diffraction coefficients with these ratios, the corrected APO (CAPO) diffraction coefficients can be obtained.

Figure 2 shows the far-field patterns calculated with the APO used by Rusch, the APO modified by Knop and the CAPO of Knop and Ostertag.

![E-plane Far-field Radiation Pattern](image)

Fig. 2. The far-field patterns calculated with APO, modified APO and CAPO

Although the expressions for the far-field do not include the direct radiation from the feed it has been included in the figures. Figure 2 clearly illustrates that Rusch's APO has a singularity at the shadow boundary, while Knop's APO successfully prevents that and
results in a continuous pattern throughout the shadow boundary region. Furthermore, it is obvious that APO and CAPO differ in the shadow region, as pointed out by Knop and Ostertag [6], as well as outside that region.

4. GTD/UTD Analysis of the Radiation from Paraboloidal Structures

According to Keller's GTD ([14],[15]), contributions to the field at an observation point \( P(r,\theta,\phi) \), in the case of the symmetrical reflector antenna (Fig. 1), come mainly from two points \( Q_i \) \( (i=1,2) \). The expression produced by Keller [14] for the diffraction of a plane scalar wave at a straight edge of a half plane is:

\[
E_e = D E_i \frac{1}{4\pi} e^{-jkr}
\]

(8)

where: \( E_e \) is the scalar diffracted field; \( E_i \) is the scalar incident field at the diffraction point; \( r \) is the distance from the diffraction point to the observation point, and \( D \) is the diffraction coefficient given by:

\[
D_s = -\frac{e^{-j(\tau/4)}}{2\sqrt{2\pi}ks\sin\beta_0} \left[ \frac{1}{\cos^2\frac{\tau}{2}} \frac{1}{\cos^2\frac{d+\tau}{2}} \right]
\]

(9)

where: \( D_s \) is the diffraction coefficient under the soft boundary condition, \( D_h \) the diffraction coefficient under the hard boundary condition, and \( \beta_0 \) the angle between the unit vector in the direction of the incident ray and the tangent to the edge at the point of diffraction.

As mentioned before, Kouyoumjian and Pathak [8] extended Keller's GTD [14] but their UTD is based on the same principles. Their expression [8] for the diffracted field is:

\[
\mathbf{E}^d(P) = \mathbf{D} \mathbf{E}^i(Q_i) A(s_i^l,s_i^q)e^{-jk\mathbf{s}^q}
\]

(10)

where: \( \mathbf{D} \) is the dyadic diffraction coefficient; \( A(s_i^l,s_i^q) \) is the caustic divergence factor; \( s_i^l \) is the distance from the feed to the diffraction point; \( s_i^q \) is the distance from the diffraction point to the observation point; and \( \mathbf{E}^i(Q_i) \) is the incident field at point \( Q_i \).

It is obvious that Eq. (8) is a scalar formula and Eq. (10) is a vector formula which can be used to calculate field vectors. Due to the vector property of Eq. (10), its diffraction coefficient \( \mathbf{D} \) takes the form of a matrix. In the ray-fixed coordinate system [6], the matrix \( \mathbf{D} \) reduces to a diagonal dyadic matrix, having two non-zero elements \( (D_s \text{ and } D_h) \) given by:
\[
D_\phi = \frac{e^{-j(\pi/4)}}{2\sqrt{2\pi}k\sin\beta_0} \left[ \frac{F[kL^i\alpha(\gamma^d-\gamma^i)]}{\cos\frac{\gamma^d-\gamma^i}{2}} + \frac{F[kL^i\alpha(\gamma^d+\gamma^i)]}{\cos\frac{\gamma^d+\gamma^i}{2}} \right] (11)
\]

with:

\[
F(z) = 2\sqrt{2} \exp(jz) \int \exp(-jr^2)dr, \quad (11.a)
\]

\[
a(\gamma^d,\gamma^i) = 2\cos^2(\frac{\gamma^d-\gamma^i}{2}), \quad (11.b)
\]

\[L^i \text{ and } L^r \text{ are distance parameters.} \quad (11.c)
\]

The only difference between \(D_\phi\) in Eq. (11) and in Eq. (9) is the modified Fresnel integral which was introduced by Kouyoumjian and Pathak to remove the discontinuity in the far-field pattern at the shadow boundary. This discontinuity was produced by the edge diffraction coefficients used by Keller. It was shown in [8] that the integral \(F(z)\to 1\) when \(z\to \infty\) so that Eq. (11) will become Eq. (9) when the observation point is away from the forward-direction (\(\theta=\pi\)) and shadow boundary (\(\theta=\alpha\)).

The total field, away from the forward direction is the sum of the fields that originate from the diffraction points \(Q_{1,2}\) and it can be written as [17]:

\[
\begin{bmatrix}
E_{\theta,1}^d \\
E_{\phi,1}^d
\end{bmatrix} = \begin{bmatrix}
E_{\theta,1}^d \\
E_{\phi,1}^d
\end{bmatrix} + \begin{bmatrix}
E_{\theta,2}^d \\
E_{\phi,2}^d
\end{bmatrix}
\]

\[
A_0 \sqrt{G_f(a)} e^{-jk\rho_0} \left[ \frac{\rho_0 \sin\theta}{2\pi \sin\theta} \left[ \frac{F[2k\rho_0\sin^2(\frac{\alpha-\theta}{2})]}{\sin(\frac{\alpha-\theta}{2})} + \frac{1}{\cos(\frac{\alpha-\theta}{2})} \right] e^{-j[2k\rho_0\sin^2(\frac{\alpha-\theta}{2})+\frac{\pi}{4}]} + + \epsilon_\theta \left[ \frac{F[2k\rho_0\sin^2(\frac{\alpha+\theta}{2})]}{\sin(\frac{\alpha+\theta}{2})} + \frac{1}{\cos(\frac{\alpha+\theta}{2})} \right] e^{-j[2k\rho_0\sin^2(\frac{\alpha+\theta}{2})-\frac{\pi}{4}] \sin\phi} \right] \frac{e^{-jk\rho_0}}{\rho_0} \right] (12)
\]

with: \(\epsilon_\theta = \begin{cases} 1 & (\theta < \frac{\pi-\alpha}{2}) \\ 0 & (\frac{\pi-\alpha}{2} \leq \theta \leq \frac{\pi}{2}) \\ -1 & (\theta > \frac{\pi}{2}) \end{cases} \) (12.a)

\(\epsilon_\theta\) has been introduced, because at the diffraction point \(Q_2\), \(D\) changes its sign when the observation point moves from the region \(\theta < \frac{\pi-\alpha}{2}\) to the region \(\theta > \frac{\pi}{2}\). Furthermore, the single-diffracted ray from edge point \(Q_2\) does not contribute to the scattered field in
the region \(\left(\frac{\pi-\alpha}{2} \leq \theta \leq \frac{\pi}{2}\right)\), because the ray from \(Q_2\) to the observation point \(P\) will be blocked by the reflector.

Referring to Eq. (12), the pattern has discontinuities at \(\theta = \frac{\pi}{2}\) and \(\theta = \frac{\pi-\alpha}{2}\), due to \(\epsilon_0\). Since the single edge-diffracted field from \(Q_2\) is blocked by the reflector, the surface-diffracted field from the back surface of the reflector gives an important contribution to the far-field in that region. Therefore, the surface-diffracted field must be calculated for that region in order to obtain continuous radiation patterns.

The contributions from the surface-diffracted field can be calculated by replacing the edge diffraction coefficients for \(Q_2\) by the surface diffraction coefficients for the blocked angle region. It was stated in [14] that the field diffracted around a curved surface decreased exponentially with \(\lambda\) and was weaker than the field diffracted by an edge where the diffraction coefficient is proportional to \(\sqrt{\lambda}\).

Therefore, instead of using the UTD edge diffraction coefficients, the UTD edge diffraction coefficients used are multiplied by \(R_{se}\):

\[
R_{se} = (\sqrt{\lambda} - (\sqrt{\lambda} - e^{-s_s/\lambda}) \sin((\theta - \frac{\pi-\alpha}{2})/\frac{2 \pi}{a}))/\sqrt{\lambda}
\]

in order to obtain the surface-diffracted field. Here \(s_s\) denotes the length of the path over which the surface-diffracted wave propagates along the back of the reflector. Eq. (13) shows that the surface-diffracted field will indeed be weaker than the edge-diffracted field, because:

\[
\lim_{\lambda \to 0} e^{-s_s/\lambda} = 0
\]

The sin function has been taken as being analogous to the correction factor in the APO of Knop and Ostertag [6]. Furthermore, \(\sin((\theta - \frac{\pi-\alpha}{2})/\frac{2 \pi}{a})\) will be continuous even at the angles \(\theta = \frac{\pi}{2}\) and \(\theta = \frac{\pi-\alpha}{2}\). So, \(R_{se}\) is equal to unity outside the blocked region, and will gradually change from an edge diffraction coefficient (proportional to \(\sqrt{\lambda}\)) to a surface diffraction coefficient (proportional to \(e^{-s_s/\lambda}\)) in the region \(\frac{\pi-\alpha}{2} < \theta < \frac{\pi}{2}\).

Figure 3 shows the patterns calculated with the GTD, UTD and modified UTD (with surface diffraction) techniques. As can be seen in Figure 3, Keller’s GTD is discontinuous at the shadow boundary and at \(\theta = \frac{\pi}{2}\). The singularity at the shadow boundary will be prevented by using UTD; however for a completely continuous pattern the modified UTD is required.
5. Comparison of the Different High-Frequency Asymptotic Techniques
Comparing GTD and APO

Firstly, for ease and clarity, Keller's GTD and Rusch's APO diffraction coefficients are compared (because Kouyoumjian & Pathak's UTD and Knop's APO affect only the fields at or near the shadow boundary).

The GTD diffraction coefficients have been derived from Eq. (12) and can be written as:

\[
\begin{align}
[D_{sgtd,1}] &= \epsilon_0 e^{-j(\pi/4)} \frac{1}{2\sqrt{2\pi k}} \frac{1}{\sin(a-\theta/2) + \cos(\theta/2)} \\
[D_{hgtd,1}] &= \frac{1}{2\sqrt{2\pi k}} \frac{1}{\sin(a-\theta/2) + \cos(\theta/2)} 
\end{align}
\]

The APO diffraction coefficients can be obtained from Eq. (5):
The diffraction coefficients (the terms in square-brackets) in Eqs. (11) and (15) are illustrated in Figure 4. The comparison is not only performed for the $D_0$, but also for the $D_1$, whose graphic presentation was not included in [5]. The diffraction coefficients for the lower point have not been computed in the angle region $\frac{\pi - \alpha}{2} < \theta < \frac{\pi}{2}$, because the diffracted ray from the lower diffraction point to the observation point is blocked by the reflector there. Figure 4 shows that the diffraction coefficients for GTD and APO have similar trends and are discontinuous at $\theta = 60^\circ$ (shadow boundary). Furthermore, it appears that the differences between the diffraction coefficients could become quite large.

Secondly, the radiation patterns were calculated with the modified UTD and Knop's APO. The results that were obtained with those methods are plotted in Figure 5, where it can be seen that the difference between the resulting radiation patterns appears to be substantial despite the fact that differences between the diffraction coefficients shown in Figure 4 are "second order" according to Rusch [7].

Comparing UTD and CAPO

In the study by Knop and Ostertag [6], it was concluded that APO had to be replaced by CAPO. So, it is interesting to compare the results obtained from the modified UTD with those from CAPO. As shown in [6], the diffraction coefficients of both CAPO and UTD are numerically indistinguishable and consequently, the radiation patterns (see Figure 6) are almost identical too. Since the results obtained with CAPO agreed well with measurements [6], it is felt that UTD should give valid results too.
Fig. 4. GTD and APO diffraction coefficients

Fig. 5. The far-field patterns calculated with UTD (modified) and APO (modified)
The far-field patterns calculated with UTD (modified) and CAPO.

6. Rear-axial Region Pattern Analysis by EEC

The rear-axial direction of the symmetrical parabolic reflector has to be one of the caustics, because the diffracted rays from the whole edge of the reflector effectively contribute to the field on the axis of symmetry. Therefore, the UTD/CAPO method is not suitable as it only takes into account two edge points. A technique named the Equivalent Edge Current Method (EEC or ECM [9]–[11]) is usually used to calculate the field in the rear-axial direction. The EEC method requires both the diffraction theory (for the equivalent edge currents) and radiation field integrals (for the total contribution from the edge) in order to obtain the total field in the rear-axial region.

Using the GTD diffraction coefficients that were presented previously, the diffracted fields can be expressed by:

\[
E^d_{\phi, \theta} = E^i_{\psi} D_{sh} \frac{1}{\sqrt{s}} e^{-jkS}
\]

with \(s\) the distance from the diffraction point to the observation point.

Using the relationships between the diffracted and incident field components:

\[
H^d_{\phi} = \frac{1}{\eta} E^d_{\theta} \quad \text{and} \quad H^i_{\psi} = \frac{1}{\eta} E^i_{\psi}
\]

it is easy to derive:

\[
H^d_{\phi} = H^i_{\psi} D_{h} \frac{1}{\sqrt{s}} e^{-jkS}
\]
where: \( E_\phi^d, E_\phi^i / H_\phi^d, H_\phi^i \) are the electric/magnetic fields in the \( \phi \)-direction parallel to the edge of the reflector, \( E_\theta^d, E_\theta^i / H_\theta^d, H_\theta^i \) are the electric/magnetic fields in the \( \theta \)-direction perpendicular to the edge of the reflector, and \( \frac{1}{\sqrt{s}} \) is the caustic divergence factor.

If there is an electric current \( I_\phi^e \) and a magnetic current \( I_\phi^m \) at the edge and if each of these currents can locally be taken as infinite straight line currents by considering them as a local diffraction phenomenon at high frequencies, the fields due to these currents will be:

\[
E_\phi^d = -\eta k \frac{1}{\xi} e^{-j ks} \quad H_\phi^d = -\frac{k I_\phi^m}{\eta \xi} e^{-j ks}
\]

Comparing Eq. (19) with Eq. (18) and using the equivalence concept, it is possible to visualize the diffracted fields as being induced by an electric current \( I_\phi^e \) and a magnetic current \( I_\phi^m \) at the edge. These currents, called equivalent edge currents, are denoted by:

\[
I_\phi^e = -\frac{2 \sqrt{2 \pi k}}{\eta} e^{-j (\pi / 4)} E_r D_s \quad I_\phi^m = -\eta \frac{2 \sqrt{2 \pi k}}{k} e^{-j (\pi / 4)} H_r D_h
\]

The field in the rear-axial direction can now be obtained by integrating the fields produced by the equivalent currents on the complete edge of the reflector:

\[
E_r^e(r, \theta, \phi) = -\frac{k \eta}{\sqrt{A \pi} r} e^{-j k r} \int \left[ \langle \xi \rangle e^{j k \rho \phi} \dd \right] dl \quad (21. a)
\]

\[
H_r^m(r, \theta, \phi) = -\frac{k \eta}{\sqrt{A \pi} r} e^{-j k r} \int \left[ \langle \xi \rangle e^{j k \rho \phi} \dd \right] dl \quad (21. b)
\]

The integrals above can be expressed in the form of Bessel functions so that the field in the rear-axial direction becomes [17]:
\[
\begin{align*}
\begin{bmatrix}
E_\theta \\
E_\phi
\end{bmatrix}
&= \begin{bmatrix}
E_\theta^0 \\
E_\phi^0
\end{bmatrix} + \begin{bmatrix}
E_\theta^m \\
E_\phi^m
\end{bmatrix} = \left( A_0 \left[ \mathbf{G}(a) e^{-jk_0} \right] \frac{ae i \delta}{4r} e^{-jk_0 r} \right) \\
\end{align*}
\]

where:

\[
\begin{align*}
\nu_0 &= D_n \frac{2 \sqrt{2} k e^{i\pi/4}}{\sin\left(\frac{\alpha}{2}\right)} - 1, \quad \nu_h = D_h \frac{2 \sqrt{2} k e^{i\pi/4}}{\sin\left(\frac{\alpha}{2}\right)} + 1 \quad (22.a) \\
k_0(\vec{r} \cdot \vec{\tau}) &= k_0 \cos \theta \sin \theta \cos (\phi - \xi) + k_0 \cos \theta \cos \theta = \tau \cos (\phi - \xi) + \delta \quad (22.b)
\end{align*}
\]

It should be noted that the angle \( \theta \) still appears in Eq.(22). Ratnasiri et al. [18] justified this by stating that \( D_n \) and \( D_h \) can be kept constant in the rear—axial region. In this way, the formula can also be used to calculate the field in the angle region close to the rear—axial direction. Such an approach was criticized by Michaeli [16] who proposed to use the corresponding diffraction coefficient for each observation angle. A drawback to using the modified EEC is that the integrals no longer have closed form expressions and, like PO, it leads to a time consuming numerical integration.

An explicit connection between Mitzner's incremental length diffraction method and Michaeli's EEC was established by Knott [21]. Figure 7 shows the rear—direction patterns that were calculated with both EEC and UTD for two different systems: viz. for the reference system where \( D/\lambda = 15 \) and for a much larger system where \( D/\lambda = 150 \). It can be seen for the system with a large \( D/\lambda \) that the closed form of the EEC gives good results. For the system with a small \( D/\lambda \), the EEC and the modified EEC differ in the region where normally the computation is switched to UTD; so even in this case, it is valid to use the standard EEC.

**Transition Region Between EEC and UTD**

As said earlier, the EEC currents are derived from the GTD diffraction coefficients. Although GTD fails in the rear—axial direction, EEC is able to provide the correct fields in that direction due to the fact that EEC is not based on the contributions from two edge points, but on the integration of the contributions from the complete edge. The close relationship between EEC and GTD makes it very likely that a smooth transition between the two methods will be obtained and Figure 7 confirms it for antenna systems with \( D/\lambda \geq 15 \). To see how close to the rear—axial direction the UTD results are still valid, the far—field at and near the rear—axial direction was calculated using both EEC and UTD. The influence of \( D/\lambda \) and \( n \) of the feed pattern function on the transition angle \( \theta_{GE} \) is shown in
Figure 8. The transition angle $\theta_{GE}$ is determined as the angle where the difference between the maxima of the pattern lobes, as calculated with the two methods, becomes smaller than 0.1 dB. From the figure, it appears that $\theta_{GE}$ is almost inversely proportional to $D/\lambda$ and does not depend on the value of $n$ of the feed function. The latter is just as expected because the UTD field (Eq.(12)) and the EEC field(Eq.(24)) have a common term, $\sqrt{G_f(a)}$.

Fig.7. The far-field radiation patterns in the rear-axial direction, calculated with UTD and EEC for a small and a large reflector antenna system

Fig.8. The transition angle $\theta_{GE}$ as a function of $D/\lambda$ and $n$
7. Conclusions

Starting from Rusch's APO, it is shown that, if Knop's modified APO is used, continuous radiation patterns for the wide-angle region are obtained. However, measurements have revealed that, due to the incorrect nature of the edge currents used, APO will give unreliable results; therefore, Knop and Ostertag's CAPO should be used.

Following in the same way as with Keller's GTD, it is shown that when Kouyoumdjian and Pathak's UTD diffraction coefficients are used, supplemented with the contribution of the surface-diffracted rays, continuous radiation patterns for the wide-angle region are obtained.

The patterns calculated with both CAPO and the modified UTD (with surface diffraction included) appear to be almost identical; however, CAPO requires more computing time than the UTD method, due to the fact that different mathematical equations are involved.

Since the high-frequency asymptotic techniques fail in the two far-field caustic directions of the axially symmetrical parabolic reflector antenna under consideration, the pattern in the forward- and rear-axial region must be calculated with the PO and EEC, respectively. There is a smooth transition between UTD and EEC; the transition angle between them being inversely proportional to $D/\lambda$ and independent of the reflector's edge illumination. Up to that angle the standard EEC and the modified EEC proposed by Michaili give almost identical results for antenna systems with $D/\lambda \geq 15$. So, for those systems, it is recommended that the EEC be used in its closed form rather than the modified EEC which contains a time consuming numerical integration.

8. Acknowledgements

The authors would like to thank the Chinese Government and Chengdu Aircraft Company for giving Mr. Chen the opportunity and funds to work at Eindhoven University of Technology for one year.

Also, they want to thank Professor G. Brussaard for his continued interest in their work and Professor J. Boersma for his valuable advice when drafting this paper.

Finally, thanks go to Dr. P.R. Attwood for reading this paper and correcting the grammar.
References

1. Silver, S.
MICROWAVE ANTENNA THEORY AND DESIGN.

2. Collin, R.E. and F.J. Zucker
ANTENNA THEORY.

3. Balanis, C.A.
ANTENNA THEORY — ANALYSIS AND DESIGN.

4. Rusch, W.V.T. and F.D. Potter
ANALYSIS OF REFLECTOR ANTENNAS.

5. Rusch, W.V.T.
PHYSICAL—OPTICS DIFFRACTION COEFFICIENTS FOR A PARABOLOID.

A NOTE ON THE ASYMPTOTIC PHYSICAL OPTIC SOLUTION TO SCATTERED FIELDS FROM A PARABOLOIDAL ANTENNA.

7. James, G.L.
GEOMETRICAL THEORY OF DIFFRACTION FOR ELECTROMAGNETIC WAVES.

8. Kouyoumjian, R.G. and P.H. Pathak
AN UNIFORM GEOMETRICAL THEORY OF DIFFRACTION FOR AN EDGE IN A PERFECTLY CONDUCTING SURFACE.

9. Ryan, C.E. and L. Peters
EVALUATION OF EDGE—DIFFRACTED FIELDS INCLUDING EQUIVALENT CURRENTS FOR THE CAUSTIC REGIONS.

10. James, G.L. and V. Kerdemelidis
REFLECTOR ANTENNA RADIATION PATTERN ANALYSIS BY EQUIVALENT EDGE CURRENTS.

11. Knott, E.F. and T.B.A. Senior
EQUIVALENT CURRENTS FOR A RING DISCONTINUITY.

12. Rusch, W.V.T.
ANTENNA NOTES.
Electromagnetics Institute, Technical University of Denmark, Lyngby, NB84, 1974, vol. II.

13. Knop, C.P.
AN EXTENSION OF RUSCH'S ASYMPTOTIC PHYSICAL OPTICS DIFFRACTION THEORY OF A PARABOLOID ANTENNA.

14. Keller, J.B.
GEOMETRICAL THEORY OF DIFFRACTION.
15. Keller, J. B.
DIFFRACTION BY AN APERTURE.

16. Michaeli, A.
EQUIVALENT EDGE CURRENTS FOR ARBITRARY ASPECTS OF OBSERVATION.

WIDE-ANGLE RADIATION PATTERN CALCULATION OF PARABOLOIDAL REFLECTOR ANTENNAS: A COMPARATIVE STUDY.

THE WIDE ANGLE SIDE LOBES OF REFLECTOR ANTENNAS.

19. Rahmat-Samii, Y.
A COMPARISON BETWEEN GO/APERURE-FIELD AND PHYSICAL OPTICS METHODS FOR OFFSET REFLECTORS.

20. Knott, E.F. and Senior, T.B.A.
COMPARISON OF THREE HIGH-FREQUENCY DIFFRACTION TECHNIQUES.

THE RELATIONSHIP BETWEEN MITZNER'S ILDC AND MICHAELI'S EQUIVALENT CURRENT.
2.5. Calculating the Wide-angle Radiation Pattern of an Offset Paraboloid Reflector Antenna.

2.5.1. Introduction

In the previous section it was shown that UTD is a satisfactory method for calculating the scattered field of symmetrical front-fed reflector antennas in the wide-angle regions; however, for proper results in the forward and rear-axial regions, PO and EEC must be used, respectively. This section deals with the offset paraboloidal reflector antenna [1]. Compared with a front-fed symmetrical paraboloidal reflector antenna, the offset configuration is free from aperture-blockage by the feed system and, consequently, it has better VSWR and radiation properties [2]. Its main disadvantage is an inherent crosspolarization. Due to its advantages, the offset parabolic reflector antenna would appear to be a more promising configuration.

However, designing an offset parabolic reflector antenna requires a complex theoretical analysis, because its asymmetric geometry results in numerical calculation difficulties. Most of the analyses for offset reflector antennas are based on the work done by Cook et al.[3], either by making use of the same geometry, or by following a similar approach but with a different geometry. All the methods can be considered as being related to the Physical Optics (PO) theory, by using the current-distribution method or the aperture-field integration method. Since these methods can only be used in a limited angle region, another method is needed to calculate the wide-angle radiation pattern.

In this section, the UTD is applied to calculating the wide-angle radiation pattern of an offset reflector antenna. The papers published on this subject are few [4]. For the analysis of an offset reflector configuration, a two-dimensional diffraction model, like that used for the symmetrical configuration, is generally inadequate and a three-dimensional model is needed. Another difference from analyzing a symmetrical antenna, occurs because the caustic, which is in the rear-axial direction for a symmetrical antenna, now appears in the symmetry plane at an angle with the rear-axial direction, due to the asymmetric geometry.

2.5.2. Reflector Configuration.

An offset parabolic reflector configuration is either composed of a single-reflector and a feed at its focus, or two reflectors for which the main offset reflector is illuminated by a combination of feed and sub-reflector. In this section, an offset parabolic single reflector antenna is considered.
The geometry of the antenna is shown in figure 2.1. The offset parabolic reflector is a portion of a paraboloid of revolution around the \(z\)-axis with a focal length of \(f\).

![Fig. 2.1. The Geometry of an Offset Parabolic Single Reflector Antenna](image)

The paraboloid is illuminated by a feed within a cone with a half-subtended angle \(\alpha\) measured from the \(z'\)-axis (the axis of the feed).

Since the feed axis differs from the axis of symmetry for the paraboloid, the prime \((x',y',z')\) coordinates system has been introduced. The corresponding \((\rho';\phi';\zeta')\) spherical coordinates system is used to describe the vectorial radiation pattern of the feed. The same feed pattern will be used as with the symmetrical antenna configuration discussed in the previous section.

Due to its asymmetric offset geometry, the radiation pattern of the complete antenna system for \(x'\) and \(y'\) feed polarization will be different. Therefore, the kind of polarization used will be clearly stated.

It is worth noting that the projection of the edge of the reflector on the \(x-y\) plane (figure 2.2(a)) is a circle with diameter:
and that the edge curve lies on a plane parallel to the y-axis, making an angle $\psi_{po}$ with the x-axis (figure 2.2(b)), where $\psi_{po}$ is given by:

\[
\tan \psi_{po} = \frac{\sin \varphi_0}{\cos \varphi_0 + \cos \psi_0}
\]  

(a) The Projection on the x-y Plane  
(b) The Projection on the x-z Plane  

Fig. 2.2. The Projection of the Edge Curve for an Offset Paraboloidal Reflector Aperture

2.5.3. Diffraction Point Location for an Offset Configuration

The GTD expressions produced by Keller ([5],[6]) for the diffraction of a plane scalar wave at a straight edge of a half plane were extended by Kouyoumjian and Pathak [7]. Their expression for the diffracted field is:

\[
E^d(P) = D E^i(Q_1) A(s_i,s_q^d) e^{-jks_q^d}
\]  

where: $D$ is the dyadic diffraction coefficient; $A(s_i,s_q^d)$ is the caustic divergence factor; $s_i$ is
the distance from the feed to the diffraction point; \( s_i^f \) is the distance from the diffraction point to the observation point; and \( \mathbf{E}^i(Q_i) \) is the incident field at point \( Q_i \).

For the diffraction of an incident spherical wave at a curved edge, the caustic divergence factor takes the following form [7]:

\[ A(r_c, s_i^f) = \frac{\rho_c}{s_i^f(\rho_c + s_i^f)} \]  \hfill (2.4)

where

\[ \frac{1}{\rho_c} = \frac{1}{\rho_{n_i}} - \frac{n \cdot (s_i^f - \bar{s}_i^f)}{\rho_g \sin^2 \beta_0} \]  \hfill (2.5)

with

- \( \rho_c \) the distance between the caustic at the edge and the second caustic of the diffracted ray,
- \( \rho_g \) the radius of curvature of the edge at the diffraction point,
- \( \bar{n} \) the unit vector normal to the edge at \( Q_i \) and directed away from the center of the curvature,
- \( \rho_{n_i} \) the radius of the curvature of the incident wavefront at the edge—fixed plane of incidence which contains the unit vectors \( \bar{s}_i^f \) and the unit vector \( \mathbf{T} \) tangent to the edge at \( Q_i \),
- \( \bar{s}_i^f \) the unit vector in the direction of the incident ray,
- \( \bar{s}_i^d \) the unit vector in the direction of the diffracted ray,
- \( \beta_0 \) the angle between \( \bar{s}_i^f \) and the tangent \( \mathbf{T} \) to the edge at the point of diffraction.

It is shown in [7] that, using a ray—fixed coordinate system, the diffraction coefficient can be represented by a soft and a hard scalar diffraction coefficient, given by:

\[ D_h = -c^{-j(\pi/4)} \left[ \frac{F[kL^i 2 \cos^2(\frac{\gamma_i^d - \gamma_i^f}{2})] F[kL^i 2 \cos^2(\frac{\gamma_i^d - \gamma_i^f}{2})]}{2\sqrt{2\pi} \sin \beta_0} \frac{\cos(\frac{\gamma_i^d - \gamma_i^f}{2})}{\cos(\frac{\gamma_i^d + \gamma_i^f}{2})} \right] \]  \hfill (2.6)

with
the scalar diffraction coefficient for the soft boundary condition,
\( D_s \)
the scalar diffraction coefficient for the hard boundary condition,
\( D_h \)
the angle between the incident ray and the paraboloidal surface
tangent, which is perpendicular to the plane of incidence,
\( \gamma_i \)
the angle between the diffraction ray and the paraboloidal surface tangent,
which is perpendicular to the plane of diffraction,
\( \gamma_d \)
the distance parameters as defined in [7].
\[
F(z) = 2j \sqrt{z} \exp(jz) \int_{-\infty}^{\infty} \exp(-j\tau^2) \, d\tau
\]
involving a Fresnel integral.

According to the GTD and UTD theories, contributions to the far field at an observation
point \( P(r, \theta, \phi) \) come mainly from a finite number of points \( Q_i \). If the diffraction points are
known, the total field can be obtained by simply adding the individual contributions
together. Therefore, the locations of the diffraction points have to be found. In contrast to
the simplicity of finding the diffraction points for the case of a symmetrical antenna
configuration, determining the diffraction points for an offset configuration is rather
complex.

Such a determination can be performed in two ways; the first is based on a UTD ray
tracing technique. In that case, a diffraction point coordinate transformation [4] is most
convenient, for efficiently describing the phenomena related to the diffraction. The second
way uses the stationary phase method [8].

As shown in [4], both ways will reduce to the same result and the relationship between the
observation point and the reflector coordinates that have to be satisfied by the diffraction
points are given by the following set:

\[
\begin{align}
\sin \xi' &= 0 \text{ for } \phi = 0 \\
\text{or} \\
\tan \theta &= -2 \left[ \left( \cos \psi \cos \alpha \xi' + \cos \xi' - \sin \psi \sin \alpha \right) \sin \phi - \left( \sin \xi' \left( \cos \psi + \cos \alpha \right) \cos \phi \right) \sin \xi' \sin \psi \right] \\
&\quad \left( \left( \cos \psi \cos \alpha \xi' + \cos \xi' - \sin \psi \sin \alpha \right) \sin \phi - \left( \cos \psi \cos \alpha \right) \sin \xi' \cos \phi \right)^2 - \sin^2 \phi \sin^2 \xi'
\end{align}
\]

An interesting case appears if \( \phi = 0 \) (the symmetry plane of the reflector). In that case, the
diffraction points found are the upper point \( Q_u (\xi' = 0) \) and the lower point \( Q_l (\xi' = \pi) \) as
given by Eq. (2.7.a), which is analogous to the symmetrical configuration (normal incidence,
\( \beta_0 = \pi/2 \)).
For $\phi = 0$, Eq. (2.7.b) becomes:

$$\tan \theta = \frac{2(\cos \alpha + \cos \psi_0)\sin \psi_0}{(\cos \psi_0 + \cos \alpha)^2 \sin^2 \psi_0}$$  \hspace{1cm} (2.8)$$

which can be rewritten as:

$$\tan \theta = \frac{2\sin \psi_0}{\cos \alpha - \cos \psi_0} \frac{1}{1 - \sin^2 \psi_0 - \cos \alpha} = \frac{2\tan \psi_{\phi_0}}{1 - \tan^2 \psi_{\phi_0}}$$  \hspace{1cm} (2.9)$$

with the solution:

$$\theta = 2\psi_{\phi_0}$$  \hspace{1cm} (2.10)$$

which is $\xi_L$-independent. This means that, for this specific direction, the complete edge will contribute and that there is a second caustic direction in which the diffracted rays from the complete edge give an in-phase contribution to the far-field. The caustic is fully determined by the offset geometry ($\psi_0$ and $\alpha$).

The mathematical proof for the existence of a caustic at this particular angle is shown in [4]. However, to prevent from getting mired in algebraic details, a more conceptual proof will be given. As stated before, the projection of the edge-curve on the $x-z$ plane is a line given by Eq. (2.2). This means that the complete edge is in a plane perpendicular to the $x-z$-plane. If all the rays from the edge to the far-field observation point in the forward-direction are mirrored at this plane, a second caustic is obtained (see Fig. 2-3); this is due to the fact that the mirroring has no effect on the length of the rays.

![Fig. 2.3. The Second Far-field Caustic: a) Side View; b) Back-side View](image-url)
2.5.4. Diffracted Field in the Symmetry Plane.

As the location of the diffraction points are known, the diffracted field can be calculated, but it is complicated, because the diffraction rays for most $\phi$-planes do not hit the reflector edge perpendicular to its tangent. The oblique incidence makes an analysis much more difficult and a 3-D diffraction model is needed. However, in the case of $\phi = 0$, the angle of incidence $\beta_0$ is $\pi/2$ and a transparent two-dimensional model is adequate. Then the total field away from the two caustic directions $\theta = \pi$ and $\theta = 2\psi_{p0}$, can be given by:

\[
\begin{bmatrix}
E_{\theta}^d \\
E_{\phi}^d
\end{bmatrix} = \begin{bmatrix}
E_{\theta}^d \\
E_{\phi}^d
\end{bmatrix} + \begin{bmatrix}
E_{\theta 1}^d \\
E_{\phi 1}^d
\end{bmatrix} + \begin{bmatrix}
E_{\theta 2}^d \\
E_{\phi 2}^d
\end{bmatrix}
\]

\[
= \left(\frac{P}{2\pi}\right) e^{-jkr} \left(\sqrt{G_\theta(|\theta-\psi_0|)} + \frac{G_\phi(\theta,\phi,\psi_0)\sin \cos \psi_{p0}}{8\pi \sin \left(\frac{x-\theta}{2}\right) \sin \left(\frac{\theta - 2\psi_{p0}}{2}\right) (\cos \psi_{p0} + \cos \phi)}
\right)
\]

\[
\left\{ \begin{array}{l}
\frac{1}{\rho_0^2} \cdot \epsilon_{\theta 1} \left[ \frac{F(2k\rho_0 \sin^2 \left(\frac{\psi_0 + a - \theta}{2}\right))}{\sin \left(\frac{\psi_0 + a - \theta}{2}\right)} \right] + \frac{1}{\rho_0^2} \cdot \epsilon_{\phi 1} \left[ \frac{F(2k\rho_0 \sin^2 \left(\frac{\theta - \psi_0 + a}{2}\right))}{\sin \left(\frac{\theta - \psi_0 + a}{2}\right)} \right] \cdot e^{-j\left[k\rho_0(1 - \cos(\psi_0 + a - \theta)) + \frac{\pi}{4}\right]} + \\
\frac{1}{\rho_1^2} \cdot \epsilon_{\theta 2} \left[ \frac{F(2k\rho_1 \sin^2 \left(\frac{\theta - \psi_0 + a}{2}\right))}{\sin \left(\frac{\theta - \psi_0 + a}{2}\right)} \right] + \frac{1}{\rho_1^2} \cdot \epsilon_{\phi 2} \left[ \frac{F(2k\rho_1 \sin^2 \left(\frac{\theta - \psi_0 + a}{2}\right))}{\sin \left(\frac{\theta - \psi_0 + a}{2}\right)} \right] \cdot e^{-j\left[k\rho_1(1 - \cos(\psi_0 + a)) - \frac{\pi}{4}\right]} \end{array} \right\}
\]

where

\[
\epsilon_{\theta 1} = \begin{cases} 
1 & \text{for } (0 \leq \theta < \frac{3\pi}{2} + \psi_{p0}) \\
0 & \text{for } (\frac{3\pi}{2} + \psi_{p0} < \theta < \frac{3\pi}{2}) \\
-1 & \text{for } (\frac{3\pi}{2} + \psi_{p0} < \theta < 2\pi)
\end{cases}, \quad \epsilon_{\phi 2} = \begin{cases} 
1 & \text{for } (0 \leq \theta < \frac{3\pi}{2} + \psi_{p0}) \\
0 & \text{for } (\frac{3\pi}{2} + \psi_{p0} < \theta < \frac{3\pi}{2}) \\
-1 & \text{for } (\frac{3\pi}{2} + \psi_{p0} < \theta < 2\pi)
\end{cases}
\]

The terms $\epsilon_{\theta 1}$ and $\epsilon_{\phi 2}$ have been introduced in order to account for the blocking effect of the reflector on the diffracted rays from the edge of the reflector and to take care of the sign change when the observation point moves from one angle region to another.
2.5.5. The Diffracted Field in an Arbitrary \( \phi \)-Plane

As opposed to the diffracted field in the symmetry plane, it is difficult to write concisely the total diffracted field in an arbitrary \( \phi \)-plane; therefore, the caustic divergence factor will be given, followed by the dyadic diffraction coefficient. In that way, all the tools are present in order to determine the diffracted field as given in Eq. (2.3).

**The Caustic Divergence Factor**

As can be seen from Eqs. (2.4) and (2.5), calculating the caustic divergence factor needs expressions for \( \vec{n}, \beta_0, \vec{s}^1, \vec{s}^d \) and \( \rho_0 \). The first two are related in some way to the edge tangent unit vector \( \vec{T} \) which is given by:

\[
\vec{T} = T_x \hat{x} + T_y \hat{y} + T_z \hat{z}
\]

with

\[
T_x = \frac{-\sin \xi' (\cos \psi_0 + \cos \alpha)}{T_s}
\]

\[
T_y = \frac{\cos \psi_0 \cos \phi_0 \xi' \cos \xi' + \cos \xi' - \sin \psi_0 \sin \alpha} {T_s}
\]

\[
T_z = \frac{\sin \xi' \sin \psi_0} {T_s}
\]

\[
T_s = \sqrt{[\sin \xi' (\cos \psi_0 + \cos \alpha)]^2 + [\cos \psi_0 \cos \phi_0 \xi' + \cos \xi' - \sin \psi_0 \sin \alpha]^2 + [\sin \xi' \sin \psi_0]^2}
\]

The unit vector \( \vec{n} \) and the angle of incidence \( \beta_0 \) can now be deduced according to:

\[
\vec{n} = \text{det} \left[ \begin{array}{cc} T_y & T_z \\ 0 & \cos \psi_0 \end{array} \right] \vec{x} + \text{det} \left[ \begin{array}{cc} T_z & T_x \\ \cos \psi_0 & \sin \psi_0 \end{array} \right] \vec{y} + \text{det} \left[ \begin{array}{cc} T_x & T_y \\ \sin \psi_0 & 0 \end{array} \right] \vec{z}
\]

\[
\cos \beta_0 = \frac{\vec{T} \cdot \vec{\rho}} {T_s} = -\frac{\sin \psi_0 \cdot \sin \xi'}{T_s}
\]

The vectors \( \vec{s}^1 \) and \( \vec{s}^d \) become:
\[
\vec{s}_i = \vec{\rho} = \begin{bmatrix} \sin \psi_0 \cos \alpha + \cos \phi \cos \psi_0 \sin \alpha \vec{x} \\ \sin \alpha \sin \phi \vec{y} \\ \cos \psi_0 \cos \alpha - \cos \phi \sin \psi_0 \sin \alpha \vec{z} \end{bmatrix}, \quad \vec{s}_d \times \vec{\tau} = \begin{bmatrix} \sin \theta \cos \phi \vec{x} \\ \sin \theta \sin \phi \vec{y} \\ \cos \theta \vec{z} \end{bmatrix}
\]  
\tag{2.15},

and \( \rho_e^i \) and \( \rho_e^g \) are given by:

\[\rho_e^i = \vec{\rho} \quad \text{and} \quad \rho_e^g = \frac{[a^2 \sin^2 \gamma + b^2 \cos^2 \gamma]^\frac{3}{2}}{ab(\sin^2 \gamma - \cos^2 \gamma)} \]
\tag{2.16}

with

\[a = \frac{2f \sin \alpha}{(\cos \psi_0 + \cos \alpha) \cos \psi_0} \quad \text{and} \quad b = \frac{2f \sin \alpha}{\cos \psi_0 + \cos \alpha} \]

and \( \sin(\gamma) = \vec{n}_0 \cdot \vec{y} \)

where \( \vec{n}_0 \) is the unit vector pointing from the center \( Q_0 \) of the edge curve to the edge of the offset reflector.

**The Diffraction Coefficient**

The general expression for the dyadic diffraction coefficient given in Eq. (2.6) of section 2.5.3 shows that angles, \( \gamma^i \) and \( \gamma^d \) must be known. The angle \( \gamma^i \) can be calculated from:

\[\cos(\gamma^i) = \vec{N}_s \cdot \vec{n}_{\text{refl}}. \]
\tag{2.17}

where \( \vec{N}_s \) and \( \vec{n}_{\text{refl}} \) represent the normal vector of the incident and tangent planes respectively, and are given by:

\[\vec{N}_s = \det \begin{bmatrix} T_y & T_x \\ \rho_y & \rho_z \end{bmatrix} \vec{x} + \det \begin{bmatrix} T_x & T_y \\ \rho_z & \rho_x \end{bmatrix} \vec{y} + \det \begin{bmatrix} T_y & T_z \\ \rho_x & \rho_y \end{bmatrix} \vec{z} \]
\tag{2.18}

\[\vec{n}_{\text{refl}} = -\cos \xi \sin \frac{\psi}{2} \vec{x} - \sin \xi \sin \frac{\psi}{2} \vec{y} - \cos \frac{\psi}{2} \vec{z} \]
\tag{2.19}

Similarly, \( \gamma^d \) is obtained from:

\[\cos(\gamma^d) = \vec{N}_d \cdot \vec{n}_{\text{refl}}. \]
\tag{2.20}
where, \( \mathbf{N}_d \) is the unit normal vector of the diffraction plane, namely:

\[
\mathbf{N}_d = \text{det} \left[ \begin{array}{c} T_y \ T_z \\ r_y \ r_z \end{array} \right] \mathbf{x} + \text{det} \left[ \begin{array}{c} T_z \ T_x \\ r_z \ r_x \end{array} \right] \mathbf{y} + \text{det} \left[ \begin{array}{c} T_x \ T_y \\ r_x \ r_y \end{array} \right] \mathbf{z} \tag{2.21}
\]

Furthermore, it can be shown [4] that the distance parameters \( L_i \) and \( L_r \) for a paraboloidal reflector and a spherical incident wave, are respectively given by:

\[
L_i = \rho \quad \text{and} \quad L_r = \infty \tag{2.22}
\]

Inserting Eqs.(2.13) to (2.16) into Eq.(2.4) and Eqs.(2.17) to (2.22) into Eq.(2.6), followed by putting the outcome into Eq.(2.3) gives the opportunity to determine the total diffracted field for any arbitrary \( \phi \)-plane.

### 2.5.6. EEC Method for the Far-field Caustic in the Symmetry Plane

In sections 2.5.4 and 2.5.5, the diffracted field in any \( \phi \)-plane was given. However, UTD fails at caustics and the Equivalent Edge Current method (EEC [9],[10]) has to be employed to determine the field at and around this caustic.

As already stated, in contrast to a symmetrical configuration, there is no caustic in the backward direction but at a specific angle in the symmetry plane.

The EEC currents for that particular caustic direction are given by [4]:

\[
I_T^{e}=-\frac{\sqrt{2}}{\sqrt{\pi}}k \mathbf{e}^{-j\pi/4} \mathbf{E}_0^d \sin \beta_0 \tag{2.23.a}
\]

\[
I_T^{m}=-\frac{\sqrt{2}}{\sqrt{\pi}}k \mathbf{e}^{-j\pi/4} \mathbf{F}_0^d \sin \beta_0 \tag{2.23.b}
\]

By using [2]:

\[
E_\theta^0(\tau,\theta,\phi)=-\frac{j\eta}{4\pi\tau} \mathbf{e}^{-jkr} \int_1 \mathbf{[}T_{\mathbf{e}}^0-(\mathbf{T}^0 \cdot \mathbf{T})\mathbf{T}^\mathbf{e}] \mathbf{e}^{jk\rho_{\mathbf{T}} \cdot \mathbf{T}} \mathrm{d}l \tag{2.24.a}
\]

\[
H_\theta^0(\tau,\theta,\phi)=\frac{j\eta}{4\pi\tau} \mathbf{e}^{-jkr} \int_1 \mathbf{[}T_{\mathbf{m}}^0-(\mathbf{T}^0 \cdot \mathbf{T})\mathbf{T}^\mathbf{m}] \mathbf{e}^{jk\rho_{\mathbf{T}} \cdot \mathbf{T}} \mathrm{d}l \tag{2.24.b}
\]

and \( E_\theta^0 = \eta H_\phi^0, \ E_\phi^0 = -\eta H_\theta^0 \tag{2.25} \)

it is possible to derive the total diffracted field in the vicinity of the caustic:
2.5.7. Numerical Results

The calculation methods described in the previous sections have been applied to an offset configuration with the system parameters $\psi_0=45^\circ$ and $a=30^\circ$ which results in $\psi_p=24.2^\circ$ and $f/D=0.79$. The feed had a cosine shaped gain function with $n=2$, and the polarization properties of a Huygens source with $x'$- or $y'$-axis polarization. Using the UTD for the wide-angle region pattern calculation, and the EEC and PO in or near the caustic directions, the complete radiation pattern in the symmetry plane (the $x-z$ plane) and asymmetry plane (the $y-z$ plane) can be found, and these are shown in figures 2.4 and 2.5 for $x'$-axis feed polarization. Figures 2.6 and 2.7 shows similar patterns for $y'$-axis polarization.

For the plots shown, the horizontal axis is the far-field observation angle measured from the forward-direction of the reflector antenna. The notation $(180^\circ-\theta)<0$ indicates the half plane where $\phi=0$ for the symmetry plane or $\phi=\pi/2$ for the asymmetry plane; while, $(180^\circ-\theta)>0$ indicates the half plane where $\phi=\pi$ for the symmetry plane or $\phi=3\pi/2$ for the asymmetry plane. The vertical axis is the gain function in dB.
**Fig. 2.4.** Far-field Radiation Pattern of an Offset Paraboloidal Reflector Antenna in the Symmetry plane for $x'$-axis polarization.

**Fig. 2.5.** Far-field Radiation Pattern of an Offset Paraboloidal Reflector Antenna in the asymmetry plane for $x'$-axis polarization.
Fig. 2.6. Far-field Radiation Pattern of an Offset Paraboloidal Reflector Antenna in the Symmetry plane for y'-axis polarization.

Fig. 2.7. Far-field Radiation Pattern of an Offset Paraboloidal Reflector Antenna in the Symmetry plane for y'-axis polarization.
References

[1] Rudge, A. W. and N.A. Adatia
OFFSET—REFLECTOR ANTENNAS: A REVIEW.

[2] Silver, S.
MICROWAVE ANTENNA THEORY AND DESIGN.

THE OPEN CASSEGRAIN ANTENNA: PART I ELECTROMAGNETIC DESIGN AND ANALYSIS.

Eindhoven: Faculty of Electrical Engineering, Eindhoven University of Technology, 1991
HUT-report 91-F-252.

GEOMETRICAL THEORY OF DIFFRACTION.

DIFFRACTION BY AN APERTURE.

AN UNIFORM GEOMETRICAL THEORY OF DIFFRACTION FOR AN EDGE IN A PERFECTLY CONDUCTING SURFACE.

A STATIONARY PHASE METHOD FOR THE COMPUTATION OF THE FAR FIELD OF OPEN CASSEGRAIN ANTENNAS.

[9] Ryan, C. E. and L. Peter
EVALUATION OF EDGE—DIFFRACTED FIELDS INCLUDING EQUIVALENT CURRENTS FOR THE CAUSTIC REGIONS.

[10] James, C.L. and V. Kerdemelidis
REFLECTOR ANTENNA RADIATION PATTERN ANALYSIS BY EQUIVALENT EDGE CURRENTS.
2.6. On the Design of Radiometer Antennas
by
P.J.I de Maagt, J.W. Wittekamp, and M.H.A.J. Herben
Eindhoven University of Technology
Faculty of Electrical Engineering
Telecommunications Division
P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

Abstract
This paper studies the design of ground-based radiometer antennas. In order to produce general functional and operational specifications and to describe the interaction of an antenna pattern with its noisy surroundings, a model for the brightness temperature of its environment is discussed. This is followed by a description of a method for finding the best template for the integrated pattern function and the corresponding template for the antenna pattern. Different contributions to the uncertainty of an antenna's output are discussed, such as those due to the mainlobe, the sidelobes and the antenna positioner. A theoretically optimal antenna configuration is presented and the performance parameters for both a front-fed and an offset parabolic reflector configuration are compared with those of the optimal system. Finally, a radiometer antenna design procedure is presented.

1. Introduction
Antennas are an important part of any radiometric remote-sensing system. In the first instance, such antennas may resemble those used in communication systems; however, the approaches to their design differ significantly and, consequently, the differences can be considerable. Communication antennas are generally optimized for the spatial filtering of coherent signals, that results in an optimization of antenna gain and realization of a prescribed sidelobe envelope. Those criteria are not of paramount importance for a radiometer antenna, which should filter an incoherent noise signal coming from one specific direction from among the noise signals coming from all the other directions.

It is not always possible to make the specific design criteria for a radiometric remote sensing system compatible with those for a radio communications-system. Radiometer antenna design work requires more careful consideration of particular parameters than is
usually the case with antenna designs and antenna optimization from the viewpoint of remote sensing is necessary. The various parameters have to be adjusted to obtain an acceptable result and the importance of the various criteria is judged from practical experience. However, some of the criteria are difficult to quantify, for example, the technical risk and high accuracy performance during unattended operation. Furthermore, the weight that is placed on some of the parameters will depend on the use intended. For space applications, the physical weight and sensitivity to extreme temperature differentials may assume prime importance, while for ground-based operation, they are of secondary concern. Space systems are beyond the scope of this paper which deals with only ground-based radiometer antennas.

Although that would appear to narrow the scope, the number of relevant parameters is still diverse and the various ways of combining them makes it difficult to reach an uniform design procedure for radiometer antennas. However, there is consensus that the most important design parameters are: the beam efficiency \( \eta_i \) and the integrated pattern function \( h \). The quantitative values of those parameters differ from design to design and it is better to say that mostly the parameters are specified to be as "high as practical possible". Other common design criteria are manufacture, cost, and technical risk. Another important feature is that an antenna system has to interface with a radiometer receiver and both of these components of the radiometer system should be matched.

In order to produce general functional and operational specifications and to describe the interaction of an antenna pattern with its surroundings, a model of the environment is needed. The model developed should describe as many of the foreseeable geometrical and meteorological situations as possible. The uncertain meteorological state of the atmosphere and the seasonal fluctuations at the antenna site can cause an uncertain and variable environment. Due to the non-zero width of the main beam, the presence of sidelobes and the finite accuracy of the antenna pointing mechanism, the radiometric predictions are uncertain. This uncertainty is the basis for describing an optimal template for the integrated pattern function. While the constraints to the integrated pattern are the basis for calculating a theoretically optimal radiometer antenna. The advantage of knowing the optimal system is that it assures that it is impossible to construct a better system under the given conditions. In other words, such knowledge makes it unnecessary to consider an unnecessarily large number of possible modifications for a practical antenna system. As the optimum must not go beyond what is possible or contain requirements that unnecessarily raise the costs, the choice of a radiometer antenna must take into account the optimal antenna performance versus the expected performance in practice. Therefore, two existing antenna system candidates are compared in relation to the specified design goals.
and desired system performance. These candidates are front-fed and offset parabolic reflector antennas. The first is included because it can be manufactured easily without too much technical risk. Structural considerations suggest that a front-fed configuration is the best choice, due to its symmetrical geometry (duplicated parts). The offset antenna is included because an initial review suggested that an unblocked configuration would be particularly suitable for radiometer applications.

The comparative study involves a detailed analysis of various key performance parameters such as the beam efficiency, beamwidth, integrated pattern function and aperture efficiency. For computing the integrated pattern function, the complete far-field radiation patterns of the related antennas are calculated using the Uniform Theory of Diffraction (UTD) complemented by the Equivalent Edge Current method (EEC) and Physical Optics (PO) in order to obtain the pattern in and near to the caustic directions \[1,2\].

A description of the system's performance in terms of its parameters provides an overview of the important design criteria and that should offer an opportunity to direct and monitor the development of a system. Although it is not possible to include all the relevant parameters and although some of the equations are evaluated by using the characteristics of the Olympus satellite \[3\] and the Eindhoven University of Technology (EUT) ground station, the methodology and conclusions given here are generally applicable and it is believed that the framework produced is conceptually sound, so that it can be used as a guideline.

Firstly, this paper addresses the model used for generating the brightness temperature of the atmosphere and the Earth's surface. Then, a description follows of the method used to determine an optimal integrated pattern template and the corresponding template for the antenna pattern. Different contributions to the uncertainty of the antenna output are also discussed. A theoretically optimal antenna configuration is presented and the performance parameters for both a front-fed and an offset parabolic reflector configuration are compared with those of the optimal system. Finally, a radiometer antenna design procedure is described.

2. Simulating the Influence of the Atmosphere and the Earth's Surface.

The model that is used to simulate the antenna's surroundings can be divided into two parts: a model for the atmosphere and a model for the Earth's surface. Both parts will be discussed initially.

A representation of atmospheric absorption and emission is given by the radiative transfer equation:
The Microwave Radiometer as a Remote Sensing Device

\[ T_{\text{sky}}(s) = \int_{0}^{s} T(s) \alpha(s) \exp(-\int_{s'}^{s} \alpha(s') ds') ds + T_{\text{ext}} \exp(-\int_{0}^{s} \alpha(s) ds) \]  

(1)

where \( T_{\text{sky}} \) is the brightness temperature, \( T(s) \) is the physical temperature of the atmosphere at the distance coordinate \( s \), \( \alpha(s) \) is the absorption coefficient of the atmosphere, and \( T_{\text{ext}} \) is the brightness temperature contributed by extraterrestrial sources. Centimeter and millimeter wave absorption is dominated by the water vapor, oxygen and liquid water in the atmosphere. They are related to the distributions of various meteorological parameters. To be able to estimate the influence of the atmosphere, profiles of temperature, pressure, and water vapour are required. It is possible to use detailed profiles for a specific time and place, but climatic means are more useful for estimating typical propagation effects. The profiles used, are the ones provided by the CCIR [4].

Calculating attenuation due to water vapour, oxygen and liquid water is performed using the 1989 version of Liebe’s Millimeter-wave Propagation Model [5]. The atmospheric profiles are divided into 200m thick layers and pressure, water vapour, and temperature are assumed to be constant for each layer. Because the properties of the atmosphere change with height, the layers are assumed to be horizontally stratified with respect to the Earth’s spherical surface.

Using Eq.(1), it is easy to calculate the zenith brightness temperature; however, for non-zenithal angles the path length \( s \) in Eq.(1) has to be increased according to the secant law:

\[ s_{\text{zen}} = \frac{s_{\text{zen}}}{\cos(\psi)} \]  

(2)

where \( s_{\text{zen}} \) is the zenithal distance and \( \psi \) the zenithal angle.

For low elevation angles (from 0 to about 10°), the effects of both refraction and the Earth’s curvature have to be included and \( s \) has to be increased according to [6]:

\[ s = s_{\text{zen}} \left[ 1 - \left( \frac{10^6+N_i}{10^6+N_j} \right) \left( \frac{r_E+h}{r_E+h_0} \right)^2 \right]^{-\frac{1}{2}} \]  

(3)

where \( N_i, N_j \) are the real part of the complex refractivity at ground level \( h_0 \) and the height \( h \), respectively, and \( r_E = 6357 \cdot 10^3 \) is the Earth radius.

The brightness temperature calculated for the Olympus beacon frequencies of 12.5, 20, and 30 GHz according to the above procedure are shown in figure 1.
Radiometer Antennas

Fig. 1. The brightness temperature at 12.5, 20, and 30 GHz as a function of the zenithal angle.

An accurate estimate of the brightness temperature down to an elevation angle of $10^\circ$ is given by:

$$ T_{\text{sky}} = \frac{a}{\cos(\psi)} + b $$

(4)

and for low elevation angles by [7]:

$$ T_{\text{sky}} = c + d\psi $$

(5)

The corresponding coefficients for the three Olympus beacon frequencies are shown in table 1.

Table 1 The coefficients that describe $T_{\text{sky}}$ for the Olympus beacon frequencies.

<table>
<thead>
<tr>
<th>$f$ (GHz)</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.5</td>
<td>2.796</td>
<td>1.802</td>
<td>26.048</td>
<td>-38.366</td>
</tr>
<tr>
<td>20</td>
<td>11.298</td>
<td>4.768</td>
<td>26.734</td>
<td>-37.217</td>
</tr>
</tbody>
</table>

Next, the model for the Earth's surface will be discussed; its influence is simulated in figure 2. Radiation that is incident on the Earth's surface will be partly specularly reflected and partly diffuse reflected, while the remainder will be absorbed, causing emission of radiation [8].
Although the incident radiation is randomly polarized, this is not the case for reflected radiation, because reflections from the Earth's surface depend on polarization, which is represented by the Fresnel reflection coefficients. Those coefficients are only valid for "smooth" surfaces, and have to be corrected if the surface is "slightly" rough. Deciding whether the surface is rough or not requires the Rayleigh or Fraunhofer criteria [9]:

\[
s > \frac{\lambda}{\text{acos}(\phi)} \tag{6}
\]

where \( s \) is the standard deviation of the surface roughness, \( \alpha = 8 \) for the Rayleigh criterium and \( \alpha = 32 \) for the Fraunhofer criterium [8].

It is evident from Eq.(6) that for grazing incidence, the ground will always appear to be smooth. However, Brussaard [10] showed that most of the radiation from the ground entering the antenna comes from within a radius of 20 m around the antenna, assuming an antenna height of 2 m. Therefore, this "grazing angle effect" can be neglected. The effect of rough ground on reflected radiation is modeled by Beckmann and Spizzichino [11]. However, as shown in [11, p.246], the corrected Fresnel reflection coefficient drops rapidly to zero with increasing roughness. Furthermore, as the surface becomes rougher the radiation emitted becomes polarization independent [9, p.830].

It should be noted that the Fresnel reflection coefficient increases with increasing moisture content of the soil, but rough ground is less sensitive to moisture content than smooth ground [12]. Vegetation also affects the reflection coefficient; it is indicated by King [13] that even short grass will eliminate all specular reflection from the underlying ground at 35 GHz.
In view of the above considerations, the ground is assumed to be rough and the Earth's surface is taken to be a randomly polarized noise source, whose influence is given by:

\[
T_{\text{ground}} = \begin{cases} 
T_r(\psi) + [(1-\epsilon)T_{\text{sky}}(r-\psi) + \epsilon T_{\text{surf}}]e^{-\tau r} & \pi/2 \leq \psi \leq \beta \\
T_r(\psi) + \epsilon T_{\text{surf}} e^{-\tau r} & \beta \leq \psi \leq \pi 
\end{cases}
\]

with \( \epsilon \) the emissivity, \( T_{\text{sky}} \) the brightness temperature of the atmosphere, \( T_{\text{surf}} \) the physical temperature of the surface, and \( T_r \) the noise temperature generated between the antenna and a reflection point at a distance \( r \), which can be found by applying the radiative transfer equation which gives:

\[
T_r(\psi) = T_{\text{layer}} (1-e^{-\tau r})
\]

where \( T_{\text{layer}} \) is the physical temperature of the lowest air layer, and \( \tau r \) is the atmospheric attenuation between the reflection point and the antenna.

3. Uncertainty in the Antenna Output Related to the Antenna Pattern.

The antenna temperature, at a given frequency, is usually written as:

\[
T_A = \frac{1}{4\pi} \int G(\theta,\phi) T_b(\theta,\phi) \, d\Omega
\]

where \( G(\theta,\phi) \) is the antenna gain pattern and \( T_b \) is the brightness temperature of the surroundings.

Eq.(9) clearly shows that all the surroundings contribute to the antenna's noise temperature. If it is the objective to determine the brightness temperature from one specific direction or region, the contributions from other directions are unwanted. If these noise contributions could exactly be represented by the model discussed in section 2, their influence could be removed. As already stated, the atmosphere is dynamic and it is not possible to predict its state exactly. A realistic impression of the uncertainty of the atmospheric brightness temperature can be found with the following procedure.

Theoretically, the average clear-sky values are shown in figure 1, and they are taken as the starting point. Then, the probability of encountering other values can be found from measured cumulative distributions, so that the uncertainty becomes:

\[
\Delta T_{\text{sky}}(\psi) = |T(\psi) - T_{\text{sky}}(\psi)|
\]

where \( T(\psi) \) follows from the cumulative distributions and \( T_{\text{sky}}(\psi) \) is given in figure 1.

As such data was not readily available on the EUT ground station, the data presented by Slobin [14] was used instead. Slobin gave information for 15 distinct climatic regions and presented cumulative distributions of the brightness temperature for each region. Based on
The Microwave Radiometer as a Remote Sensing Device

The Nicrowave Radiometer

meteorological data [15], the region most similar to the Eindhoven area was selected. It should be noted that, as the data was presented for zenith brightness temperatures, a
secant and linear transformation is used, similar to Eq.(4) and Eq.(5) when calculating the corresponding non–zenith values, with the proviso that the brightness temperature does not exceed 290 K.

Besides the uncertainty in the sky noise, additional uncertainty is due to radiation from the ground, which can be estimated by taking the first order partial derivative of Eq.(7), so that:

$$
\Delta T_{\text{ground}} = [((\epsilon-1)T_{\text{sky}}(\tau-\psi) - \epsilon T_{\text{surf}} + T_{\text{layer}})e^{-\tau r}A_{r} + (1-\epsilon^{-\tau r})A_{\text{layer}} + 
(1-\epsilon)e^{-\tau r}A_{\text{sky}} + (T_{\text{surf}} - T_{\text{sky}}(\tau-\psi))e^{-\tau r}A_{\text{epsilon}} + \epsilon e^{-\tau r}A_{\text{surf}}]
$$

if $\pi/2 \leq \psi \leq \beta$, and

$$
\Delta T_{\text{ground}} = [-\epsilon T_{\text{surf}} + T_{\text{layer}}]e^{-\tau r}A_{r} + (1-\epsilon^{-\tau r})A_{\text{layer}} + T_{\text{surf}} e^{-\tau r}A_{\text{epsilon}} + \epsilon e^{-\tau r}A_{\text{surf}}
$$

if $\beta \leq \psi \leq \pi$.

The corresponding values are listed in Table 2.

Table 2 Values of the parameters and their uncertainties contributing to $\Delta T_{\text{ground}}$.

<table>
<thead>
<tr>
<th>symbol</th>
<th>value</th>
<th>$\delta$</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{\text{sky}}(\psi)$</td>
<td>Eq.(4),(5)</td>
<td>Eq.(10)</td>
<td>[K]</td>
</tr>
<tr>
<td>$T_{\text{surf}}$</td>
<td>290</td>
<td>10</td>
<td>[K]</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.9</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$a,\tau$</td>
<td>0.2$a,\tau$</td>
<td>[Np]</td>
</tr>
<tr>
<td>$T_{\text{layer}}$</td>
<td>290</td>
<td>10</td>
<td>[K]</td>
</tr>
</tbody>
</table>

The procedure for determining the uncertainty of an antenna's output signal is based on assuming that the main beam contribution is wanted and the rest is unwanted. In that way, it is possible to determine the influence of the non–zero width of the main beam, the near–in sidelobes, the spillover lobe, the far–out sidelobes, and the backlobe. For that purpose, the template shown in figure 3 is introduced.

It should be noted that the proper coordinate transforms have to be used in order to correctly calculate the integral relationship of Eq.(9), because the antenna pattern is given in antenna coordinates ($\theta,\phi$) and the brightness temperature in zenithal–coordinates.
Fig. 3. Template for the antenna pattern. Where I=main beam, II=near-in sidelobes, III=spillover lobe, IV=far-out sidelobes, and V=backlobe.

In figures 4 to 14, contributions to uncertainty by different parts of the pattern are illustrated for $f=30$ GHz, starting with the spillover lobe, which normally constitutes a major part of the power received from outside the main beam. Figure 4 shows the templates used. The main beam contains 90% of the power while the remaining 10% is located in the spillover lobe. The figures can easily be adapted to any arbitrary percentage of power received via the spillover lobe; for example if the main beam contains 99% of the power and the spillover lobe 1%, the results then have to multiplied by 0.1 in order to find the corresponding uncertainty in this case. Figures 5 and 6 show the influence of the spillover lobe for two different values of the elevation angle. In most cases, the uncertainty for 90% of the time should be considered, because the radiometer has to operate in all weather conditions. The remaining 10% represents rainy conditions when another model should be used for calculating the uncertainty of the brightness temperature, assuming that a homogeneous rain cell surrounds the antenna [10]. However, that is beyond the scope of this paper.

The plots in figures 5 and 6 clearly show that uncertainty at the 90% level increases with a decreasing elevation angle. The effect of the position of the spillover lobe is clearly visible,
and in general it is valid to say that when the spillover lobe is directed at the ground it has a rather high value for low percentages of the time but remains constant and results in a rather low value for high percentages of the time. Next, the influence of the far-out sidelobes (and backlobe) is considered and the templates used are shown in figure 7. In figures 8 and 9, the cumulative uncertainty is shown for elevation angles of 30° and 70°. The additional uncertainty due to the far-out sidelobes can easily be determined by combining those figures with figures 5 and 6. Scaling down the results shown in figure 5 and 6 to 6.7% and subtracting them from the results shown in figures 8 and 9, will give the additional uncertainty. As can be seen, the differences between the two templates are small. To complete this analysis, the near-in sidelobes have to be included and the templates used are shown in figure 10. From the results plotted in figures 11 and 12, it can be seen that the sky noise entering the antenna from close to the Earth's surface contributes most to the uncertainty. The additional influence of the near-in sidelobes can be determined easily by scaling down the previous figures accordingly. Furthermore, the influence of the elevation angle is clearly visible from figures 13 and 14 where the uncertainty for the pattern templates SP/F/N 1 and SP/F/N 4 have been plotted for four different values of the elevation angle.

As a concluding remark, it is valid to say that the region nearest the Earth's surface has a marked influence on the total uncertainty. Therefore, an antenna should always be designed in such a way that a negligible part of the spillover lobe and the sidelobes is directed towards the region nearest the Earth's surface.
Fig. 4. Pattern templates for the spillover lobe.

Fig. 5. The absolute uncertainty of the measured brightness temperature, due to the spillover lobe at an elevation angle of 30°.

Fig. 6. The absolute uncertainty of the measured brightness temperature, due to the spillover lobe at an elevation angle of 70°.
The Microwave Radiometer as a Remote Sensing Device

Fig. 7. Pattern template for the spillover lobe and far-out sidelobes.

Fig. 8. The absolute uncertainty of the measured brightness temperature, due to the spillover lobe and far-out sidelobes, at an elevation angle of 30°.

Fig. 9. The absolute uncertainty of the measured brightness temperature, due to the spillover lobe and far-out sidelobes, at an elevation angle of 70°.
Fig. 10. Pattern template for the spillover, far-out and near-in sidelobes.

Fig. 11. The absolute uncertainty of the measured brightness temperature, due to the spillover lobe, far-out and near-in sidelobes, at an elevation angle of 30°.

Fig. 12. The absolute uncertainty of the measured brightness temperature, due to the spillover lobe, far-out and near-in sidelobes, at an elevation angle of 70°.
The absolute uncertainty of the measured brightness temperature, for template SP/F/N 1 at different elevation angles.

As stated earlier, the brightness temperature observed in the main beam is considered to be an exact replica of the atmospheric brightness temperature in the boresight direction of the antenna. However, due to the non-zero width of the main beam it is better to say that it represents a spatially averaged brightness temperature. So, a different uncertainty exists.

The brightness temperature in antenna coordinates is given by:

\[ T_{\text{sky}}(\theta, \phi) = \frac{a}{\sin \theta \cos \theta - \cos \phi \sin \phi} + b \] (12)

where \( \theta \) is the elevation angle and \( a \) and \( b \) are as given by Eq.(4).

So, the main beam's contribution to the noise measured at the antenna terminals is given by:
\[ T_{A_{mb}} = \frac{1}{4\pi} \int_{\Omega_{mb}} G(\theta, \phi) T_b(\theta, \phi) \, d\Omega \]  
(13)

and the corresponding average brightness temperature is given by:

\[ T_{mb} = \frac{\int_{\Omega_{mb}} G(\theta, \phi) T_b(\theta, \phi) \, d\Omega}{\int_{\Omega_{mb}} G(\theta, \phi) \, d\Omega} \]  
(14)

where \( \Omega_{mb} \) is the solid angle of the main beam.

The uncertainty due to a finite beamwidth is estimated by using the 90\% value of the cumulative distribution for the upper part of the main beam and the clear sky value for the lower part (see figure 15), so that:

\[ \Delta T_{mb} = T_{mb}(\Omega_{mb}) - T(\Omega_{mb} \rightarrow 0) \]  
(15)

Fig. 15. Uncertainty due to the non-zero width of the main beam.

To estimate this uncertainty, the main beam is modelled as:

\[ G(\theta, \phi) \sim \left[ \frac{J_1(a\theta)}{2a\theta} \right]^2 \]  
(16)

where "a" controls the beamwidth. The resulting uncertainty is shown in figure 16; it shows that the effect decreases with increasing elevation angle.
The absolute uncertainty in brightness temperature due to the non-zero width of the main beam.

The previously discussed results show that, for a $\eta_b$ of 90%, which is often quoted as an objective in the present proposals, an uncertainty of at least 4 K exists at 30 GHz. If that is reflected in the calibration uncertainties of 0.5 to 1 K which are typical of radiometer receivers, there appears to be an interfacing problem. The implications of such results for the design of a practical antenna system will be discussed in section 6. Firstly, a theoretically optimum antenna system will be considered in section 4, which can be used as a reference for practical antennas.


The previous section offers the opportunity for determining the template that will assure the least uncertainty for a high percentage of the time. For the EUT Olympus antenna, the elevation angle is 30° and the corresponding templates could be used; however, as already stated, the integrated pattern function is not the only design objective. A certain value of $\eta_b$ within the main-beam width is also required. However, optimization of a high $\eta_b$ and a integrated pattern for a given template is liable to result in an antenna system with a very wide beam. In that case, the uncertainty due to the non-zero width of the beamwidth will increase, and that is unwanted. Since those two requirements are mutually exclusive, the one can only be obtained at the expense of the other. A parameter that is directly related to the beamwidth is the aperture efficiency. It is well known that when $\eta_a$ is unity the smallest beamwidth is assured, and decreasing $\eta_a$ produces a wider beam. This implies that a trade-off between $\eta_a$ and $\eta_b$ will lead to the same results as a trade-off between $\eta_a$ and
the beamwidth. An optimal set of combinations of \((\eta_a, \eta_b)\), obtained from maximizing of the weighted sum \(w \eta_a + (1-w) \eta_b\) and for the corresponding \((\eta_b, \text{beamwidth})\), are determined for the optimal integrated pattern template as defined in the previous section [27]. The optimal values are calculated using an algorithm developed by de Maagt [16,17], then it was extended to make it suitable for radiometry [18]. The results are shown in figure 17. The \(\eta_a\) versus \(\eta_b\) curve would allow a sort of multipurpose antenna to be developed for both radiometry (high \(\eta_b\)) and communication purposes (high \(\eta_a\)). It will be shown in the following sections, that the curve of \(\eta_b\) versus beamwidth is better for the design of radiometer antennas.

![Graphs showing \(\eta_a\) versus \(\eta_b\) and \(\eta_b\) versus beamwidth](image)

Fig.17. The optimal value of a) \(\eta_a\) and \(\eta_b\) and the corresponding values of b) \(\eta_b\) and beamwidth \((\theta_{mb} \text{ is the half beamwidth})\) while the integrated pattern satisfies the template SP/F/N 4.

5. Practical Antenna Systems.
In the previous section, a theoretical optimal radiometer antenna was presented. As an optimum must not be impractical or contain requirements that unnecessarily drive up the costs, the choice of a radiometer antenna has to take into account optimal antenna performance versus practical performance with existing antennas. Therefore, two main antenna system candidates are compared in relation to the specified design goals and desired system performance. Those candidates are a front-fed and an offset parabolic reflector antenna. The first is included because of its ease of manufacture and low technical risk. Structural considerations suggest that a front-fed configuration is the best choice,
because of its symmetric geometry (duplicated parts). The offset configuration is included because an initial review suggested that it was particularly promising for radiometer applications. For a fair comparison between the performance of symmetrical and offset antennas, it is decided that their (projected) apertures should be equal. Furthermore, the offset ($\psi_o$) and subtended angle ($\psi_a$) of the offset configuration are directly derived from the subtended angle ($\psi$) of the symmetrical antenna (see figure 18). The value of the feed clearance ($\psi_c$) is taken to be 5°.

![Relation between symmetrical and offset antenna configurations.](image)

The comparative study involves an analysis of the key performance parameters; namely, beam efficiency, beamwidth, and integrated pattern function. Furthermore, the aperture efficiency is included in order to facilitate a comparison with the ideal multipurpose antenna.

To be able to compute such parameters in a time efficient manner, the far-field radiation patterns of the antennas were calculated using UTD [19] complemented with EEC [20] and PO near the caustic directions [1,2]. The UTD is an extension of geometrical optics (GO) that adds diffraction rays to the usual GO ray.

The diffracted field $E^d(P)$ resulting from a diffraction point $Q_i$ at the edge of the reflector can be expressed by [19]:

$$E^d(P) = E^i(Q_i) \cdot D A(s^i_s^d) e^{-jks^d}$$  \hspace{1cm} (17)

where $D$ is the dyadic diffraction coefficient, $A(s^i_s^d)$ is the caustic divergence factor, $s^i$ is the distance from the feed to the point of diffraction, $s^d$ is the distance from the point of
diffraction to the observation point P, and \( \mathbf{E}_i(Q_1) \) is the incident field originating from the feed and incident at point Q_1.

The feed pattern is chosen to resemble that of a corrugated or scalar feed horn, which has been widely used in the design of many reflector–antenna configurations. Feed patterns of that type can accurately be fitted with the function [21,22]:

\[
G_r(\psi) = \begin{cases} 
2(n+1)\cos^n(\psi) & (\psi \leq \frac{\pi}{2}) \\
0 & (\psi > \frac{\pi}{2})
\end{cases}
\]  

(18)

which describes the shape of the main lobe of most common feed patterns for the correct choice of n (a positive real).

Knowing the complete far-field pattern, it is possible to calculate the parameters \( \eta_a \) and \( \eta_b \) as well as the function h. The parameter \( \eta_a \) needs only the antenna gain for its determination; while, calculating the beam efficiency requires an integration of the gain function over the main beam from \( \theta = 0 \) to \( \theta = \theta_{mb} \) (where \( \theta_{mb} \) is the half beamwidth). For each \( n \)-value of the feed, a value for \( \eta_b \), \( \eta_a \), the beamwidth and an integrated pattern is found. It has been determined that a grid of 50 radial and 10 \( \phi \)-cuts in combination with spline interpolation is adequate to obtain the beam efficiency and the integrated pattern function accurate to within a few hundreds of a percent for \( D/\lambda > 10 \). Figures 19 and 20 show \( \eta_b \) versus \( \eta_a \) for the symmetrical and offset configurations, respectively, for different values of \( F/D \) (the values of \( F_{off}/D \) are related to \( F_{sym}/D \) according to figure 18). The values of \( \eta_a \) and \( \eta_b \) are found to be independent of \( D/\lambda \), because they appear to depend only on the field distribution across the reflector aperture. For a given \( F/D \) and \( n \) such a distribution does not change if \( D/\lambda \) varies and, consequently, \( \eta_a \) and \( \eta_b \) remain unaltered. Of course, the beamwidth is dependent on \( D/\lambda \); this is illustrated in figures 21 and 22. In figures 19 to 22, the theoretically optimal values are also depicted. It is clear that, even with the simple configurations, a satisfactory performance can be achieved. Figures 23 and 24 show the integrated pattern for both systems, for different values of \( n \). The impact of the spillover lobe on the integrated pattern function is clearly visible from these figures. As stated previously, the region nearest the Earth has a great influence on uncertainty. It is clear from figures 23 and 24 that the spillover lobe constitutes a major part of the power outside the main beam. Therefore, the part of the spillover lobe that is directed towards the region nearest the Earth's surface should be negligible.
Fig. 19. The values of $\eta_a$ versus $\eta_b$ for different values of $n$ with the symmetrical antenna for different values of $F/D$.

Fig. 20. The values of $\eta_a$ versus $\eta_b$ for different values of $n$ with the offset antenna for different values of $F/D$. 
Fig. 21. The values of $\eta_b$ versus beamwidth for different values of $n$ with the symmetrical antenna.

Fig. 22. The values of $\eta_b$ versus beamwidth for different values of $n$ with the offset antenna.
Fig. 23. The integrated pattern function for different values of $n$ for the symmetrical antenna. $D/\lambda = 25$, $F_{sym}/D = 1.42$.

Fig. 24. The integrated pattern function for different values of $n$ for the offset antenna. $D/\lambda = 25$, $F_{off}/D = 1.10$. 
The values shown in figures 19 to 22 may be misleading. Although antenna theory does not explicitly dictate the fundamental limits of the efficiencies, there are practical limitations, namely:
- spillover
- blockage
- surface tolerance

These various items will be discussed briefly in order to qualify their influences.

- spillover
Spillover efficiency represents the energy radiated beyond the edge of the reflector and it is often combined with aperture efficiency in order to give the antenna efficiency $\eta$.

- blockage
Blockage will reduce the antenna's efficiency by decreasing the gain and affecting the sidelobe level. In [25], it is shown that blockage only slightly influences beam efficiency.

- surface tolerance
In practice, a reflector cannot be built without surface imperfections. The effect of surface tolerance $\varepsilon$ on antenna efficiency has been determined by Ruze [24] as:

$$\eta_a = e^{-\frac{\varepsilon^2}{2}} \left[ \eta_{a0} + \frac{2 \pi}{D} \left( \sum_{n=1}^{\infty} \frac{\varepsilon^{2n}}{n \cdot n!} \right) \right]$$

(19)

where $\varepsilon$ is the rms phase error related to the rms surface tolerance by $\varepsilon = 4\pi\varepsilon/\lambda$ and $c$ is a correlation distance.

Surface imperfections do not only reduce antenna efficiency, but also beam efficiency. The latter is given by [25]:

$$\eta_b = e^{-\frac{\delta^2}{2}} \left[ \eta_{b0} + \frac{2 \pi}{D} \left( \sum_{n=1}^{\infty} \frac{\varepsilon^{2n}}{n \cdot n!} (1 - e^{-\frac{\pi\varepsilon }{n}}) \right) \right]$$

(20)

where $\eta_{b0}$ is the beam efficiency for a perfect surface.

In the case of $\varepsilon << D$, it can be shown that the second term has a negligible effect so that Eq.(20) reduces to:

$$\eta_b = e^{-\frac{\delta^2}{2}} \eta_{b0} = e^{-\left[\frac{4\pi\varepsilon}{D} \frac{D}{\lambda}\right]^{2}} \eta_{b0}$$

(21)

Figure 25 shows the effect of surface irregularities on beam efficiency. For each $D/\lambda$ value the ratio of surface tolerance to diameter is taken as $1 \cdot 10^{-4}$ and $5 \cdot 10^{-4}$, respectively which is complied with current commercial technology [28].
Fig. 25. The influence of surface irregularities on beam efficiency; the upper curves for $\epsilon/D = 10^{-4}$, and the lower curves for $\epsilon/D = 5 \times 10^{-4}$.

a) a symmetrical configuration  
b) an offset configuration
Another influence that can be encountered in the trade-off process is the accuracy of the antenna positioner. Of course, it will not affect a comparison in favor of the one or the other antenna type, because both systems are troubled with the same phenomena. However, in the overall trade-off, that effect has to be taken into account. Also, that makes a difference between communication and radiometer antennas, because a communication antenna can be pointed to a certain signal source; however, that is not possible with a radiometer antenna. So, it is important to know the exact direction in which the main beam of the antenna points. The effect of the finite accuracy of the antenna positioner can be included as follows. It is convenient to rewrite Eq. (1) in terms of the brightness temperature, as follows:

\[ \text{\( T_b = T_b^{\text{ext}} e^{-\tau} + T_m(1-e^{-\tau}) \)} \]  (22)

where \( T_m \) is the mean radiating temperature and \( \tau \) is the opacity of the atmosphere. In the case of a stratified atmosphere, the opacity can be expressed by:

\[ \tau = \frac{\tau_0}{\cos(\psi)} \]  (23)

with \( \tau_0 \) the total zenith opacity. Sensitivity of the observed brightness temperature relative to the elevation angle can be given by:

\[ \frac{dT_b}{d\psi} = -\tau \tan(\psi) (T_b^{\text{ext}} - T_m) e^{-\tau} \]  (24)

It is plotted in figure 26. In the worst situation (low elevation angle and high opacity) an uncertainty of 5 K per degree of elevation is found. The resultant uncertainty depends on the antenna mechanism; e.g., 0.5 K in the case of a pointing accuracy of 0.1 degree.

![Fig. 26. Absolute uncertainty due to pointing errors, with the opacity as the parameter, and \( T_m = 275 \text{ K} \).](image-url)
6. Design Procedure

In the previous sections, some tools have been presented which could be used to derive an optimal radiometer antenna system. In this section, the same tools will be used to develop a design procedure and it will be explained how target values can be determined and achieved. The first step is to assure that the radiometer receiver and antenna are correctly balanced. A practical value of the radiometer receiver calibration uncertainty is 1 K. It is therefore desirable to make the uncertainty due to the antenna approximately equal to 1 K. The uncertainty consists of individual contributions from the main-beam (figure 16), from the rest of the pattern (figures 4 to 14) and from the pointing mechanism (figure 26). They should be rss-added and conform to the objective of achieving 1 K uncertainty.

From the elevation angle, the frequency and the tolerated uncertainty introduced by the complete pattern, excluding the main beam, two design parameters can be derived. Firstly, the desired position of the spillover lobe (leading to a desired subtended angle for the symmetrical antenna, and at a given feed clearance angle leading to a subtended and offset angle of the offset antenna and, consequently to an F/D ratio) and, secondly, the desired value of the beam efficiency \( \eta_b \). From the acceptable uncertainty due to the main beam, the desired beamwidth can be derived. Given a maximal beamwidth and a minimal \( \eta_b \), it is possible to determine a D/\( \lambda \) and surface tolerance \( \epsilon/D \) from the curves shown in figure 25. At this point, a trade-off is possible between D/\( \lambda \) and \( \epsilon/D \) that is governed by the costs.

After selecting one of those curves, the \( n \)-value of the feed is determined (leading to the edge taper required). In that way, the complete antenna system is determined for both configurations.

Of course the trade-off has not ended yet, because a cost is attached to the system. If it is too high, a trade-off between costs and performance has to be made. The complete procedure is shown schematically in figure 27. The values of A\( \Delta T \), B\( \Delta T \), and C\( \Delta T \) in that figure should be rss added in order to conform with the requirement of 1 K; however, most uncertainty comes from the pattern excluding the main beam (A\( \Delta T \)). As an example, a similar procedure is followed as for designing the EUT radiometer antenna that operates at 30 GHz and an elevation angle of 30°. Those results are presented in table 3 for the two antenna configurations. That table includes \( \eta_b = 0.9 \) which is the frequently quoted objective, and \( \eta_b = 0.975 \), because this value leads to an uncertainty in the order of approximately 1 K. The first three columns of table 3 contain the objectives and the last three contain the resulting antenna configuration parameters.
Table 3 Results with the optimized symmetrical and offset antenna configurations $F_{sym}/D = 1.42$ and $F_{off}/D = 1.10$.

<table>
<thead>
<tr>
<th>antenna</th>
<th>$\theta_{mb}$ [$^\circ$]</th>
<th>$\eta_b$ [%]</th>
<th>D/\lambda</th>
<th>$\epsilon/D$</th>
<th>taper [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sym</td>
<td>3</td>
<td>90</td>
<td>50</td>
<td>5E-4</td>
<td>-21.9</td>
</tr>
<tr>
<td>sym</td>
<td>3</td>
<td>97.5</td>
<td>50</td>
<td>1E-4</td>
<td>-18.1</td>
</tr>
<tr>
<td>sym</td>
<td>5</td>
<td>90</td>
<td>25</td>
<td>5E-4</td>
<td>-13.0</td>
</tr>
<tr>
<td>sym</td>
<td>5</td>
<td>97.5</td>
<td>25</td>
<td>3E-4</td>
<td>-20.5</td>
</tr>
<tr>
<td>off</td>
<td>3</td>
<td>90</td>
<td>50</td>
<td>1E-4</td>
<td>-11.8</td>
</tr>
<tr>
<td>off</td>
<td>3</td>
<td>97.5</td>
<td>50</td>
<td>1E-4</td>
<td>-18.5</td>
</tr>
<tr>
<td>off</td>
<td>5</td>
<td>90</td>
<td>25</td>
<td>5E-4</td>
<td>-12.6</td>
</tr>
<tr>
<td>off</td>
<td>5</td>
<td>97.5</td>
<td>25</td>
<td>3E-4</td>
<td>-18.9</td>
</tr>
</tbody>
</table>

It can be surmised from this table that a similar performance could have been achieved with either configuration. However, in the case of the symmetrical antenna, an F/D ratio would be needed that is difficult to obtain commercially. Another aspect is blockage. Although it does not effect the quantitative value of $\eta_b$ significantly, blockage by struts introduces "scatter cones" [26] that point in a certain region. If they are directed towards the region nearest the Earth's surface, uncertainty will increase and the performance will deteriorate. An offset antenna can be designed easily to prevent blockage. Furthermore, an offset configuration can be positioned in one of two different ways. Either with the feed above the reflector, or with the feed under the reflector. In the first case, it is easy to direct the spillover lobe towards the Earth in order to ensure low uncertainty. So, the offset configuration appears to be the most promising for radiometry purposes.

7. Conclusions
In this project, the interaction between a radiometer antenna and its surroundings was modelled. Due to the dynamic nature of the surroundings it is not possible to predict its state. If the objective is to measure the brightness temperature in one direction or from a small region, this leads to an uncertainty in the antenna output because the complete noisy surroundings contributes. It has been found that the spillover lobe is a major contributor to that uncertainty and that the sky noise entering the antenna from the region nearest the Earth's surface has a big influence on the total uncertainty. Based on that uncertainty, a appropriate template for the integrated pattern is deduced and the corresponding theoretically optimal antenna is determined to serve as a reference system. The symmetrical and offset configurations were studied in order to compare the ideal theoretical with practical antenna configurations. The description of a system's performance in terms of its parameters provides an insight into the most important design criteria. Combining
the parameters with the uncertainty of the antenna’s output signal enables to design a practical procedure and to determine target values for its parameters. Compared with an optimal system’s performance, it gives an indication of what could be gained from modifications to the simple antenna configurations, but it is clear that even with the simple configurations satisfactory performance can be achieved. Furthermore, it is shown that an offset parabolic reflector configuration is particularly suitable for radiometry. Although it is not possible to provide a full coverage of all the relevant parameters, the methodology and conclusions in the paper apply generally. Some of the equations have been evaluated, using the characteristics of the Olympus satellite [4] and the Eindhoven University of Technology (EUT) ground station, but it is believed that they form a framework that is conceptually sound and may be used as a guideline.

8. Acknowledgements
We are indebted to Drs. M. Mathiesen, W. Kummer, J.B. Snider, M. Pitsiladis, F. Davarian, R.S. Orton, G. Doro, and D. Thompson for their interesting correspondence, and to Professor G. Brussaard for his valuable help. Furthermore, thanks go to L.J.J.C. Franssen for implementing the "Liebe model". Finally, thanks go to Dr. P.R. Attwood for reading this paper and correcting its grammatical sense.
Fig. 27. Schematic representation of the design procedure.

\[ \sqrt{A^2 + B^2 + C^2} = 1 \]
References


CLIMATE ATLAS OF THE NETHERLANDS.  

de Maagt, P.J.I.  
A GENERAL OPTIMIZATION METHOD FOR REFLECTOR ANTENNA SYNTHESIS.  

de Maagt, P.J.I.  
A SYNTHESIS METHOD FOR COMBINED OPTIMIZATION OF MULTIPLE ANTENNA PARAMETERS AND ANTENNA PATTERN STRUCTURE.  

De Maagt, P.J.I. and J. Wittekamp.  
AN OPTIMIZATION METHOD FOR RADIOMETER ANTENNAS.  

Kouyoumdjian, R.G. and P.N. Pathak.  
AN UNIFORM GEOMETRICAL THEORY OF DIFFRACTION FOR AN EDGE IN A PERFECTLY CONDUCTING SURFACE.  

Ryan, C. E. and L. Peters.  
EVALUATION OF EDGE-DIFFRACTED FIELDS INCLUDING EQUIVALENT CURRENTS FOR THE CAUSTIC REGIONS.  

Kildal, P-S.  
FACTORIZATION OF THE FEED EFFICIENCY OF PARABOLOIDS AND CASSEGRAIN ANTENNAS.  

Rusch, W.V.T.  
THE CURRENT STATE OF THE REFLECTOR ANTENNA ART.  

Ruze, J.  
ANTENNA TOLERANCE THEORY—a review.  

Croswell, W.F. and M.C. Bailey.  
RADIOMETER ANTENNAS.  

GROUND RADIATION SCATTERED FROM FEED SUPPORT STRUTS: A SIGNIFICANT SOURCE OF NOISE IN PARABOLOIDAL ANTENNAS.  

FORWARD SCATTERING FROM SQUARE CYLINDERS IN THE RESONANCE REGION WITH APPLICATION TO APERTURE BLOCKAGE.  

Wittekamp, J.W.  
AN OPTIMIZED RADIOMETER ANTENNA; Theory and Design.  

Ha, T.T.  
DIGITAL SATELLITE COMMUNICATIONS.  
3. Radiometer Receivers

3.1. Introduction

In 1901, Max Planck established the relationship between the thermodynamic temperature and the noise emission spectrum of a black body. A radiometer is a device that measures thermal radiation. A microwave radiometer basically consists of an antenna (scanning or non-scanning) and a very sensitive receiver. The antenna collects thermal radiation and transfers it to the receiver where it is amplified, detected and recorded.

In a block-diagram form, the radiometer resembles a familiar superheterodyne receiver; however, due to the special function of a radiometer, that is the detection of noise, the differences are profound and additional stringent requirements, such as high gain and low system noise, are placed on the radiometer receiver. Furthermore, its performance characteristics will differ from those needed for communications receivers.

Sensitivity, for example, is not measured as a signal amplitude but as the minimal change in the noise input which the system is able to measure. Another difference is the bandwidth. A relatively narrow bandwidth is usual for superheterodyne receivers, while very wide bandwidths are a characteristic of radiometer receivers. Furthermore, radiometers place higher demands on stability of system noise and gain than superheterodyne receivers. As a result of such stringent requirements, the use of exclusive microwave components is often essential, and almost every significant attempt at increasing sensitivity or improving stability has found its way into radiometric system designs.

Therefore, the systems may appear to be unusual at first sight. In sections 3.1 and 3.2, the fundamental problems encountered in radiometry are discussed and the advantages and disadvantages of the most popular radiometer receivers are presented. A novel approach to the radiometer stabilization problem appears to be feasible. That idea is validated with a first bread-board model which is described in section 3.3. Special attention has been given to the detector diodes in section 3.4. A second test set-up and the results obtained with it are described in section 3.5. Finally, some conclusions are drawn in section 3.6.
3.2. General Operation of Radiometers

In essence, a radiometer is used to yield useful information from incoherent noise-like radiation. Whatever the source may be, the radiation is related to the brightness temperature of that source. If a radiometer measures the power in a specified bandwidth $B$ with a power amplification factor of $G$, this radiation can be expressed as a receiver output power given by:

$$P_{\text{out}} = kBG T_a$$  \hspace{1cm} (3.1)

where $T_a$ is a measure of the output power of an antenna which can be directly related to the thermal radiated power to be measured, and $k$ is Boltzmann's constant.

Unfortunately, this is not a realistic representation because the receiver itself will also generate noise. The statistical properties of the "input noise" will not differ from those of the noise originating in the receiver itself, and the receiver output power can be more realistically represented by:

$$P_{\text{out}} = kBG (T_a + T_{\text{rec}}) = kBG T_{\text{sys}}$$  \hspace{1cm} (3.2)

where $T_{\text{rec}}$ is the receiver noise temperature, and $T_{\text{sys}}$ is the system noise temperature.

The microwave power given by Eq.(3.2) has to be detected in order to find some measure for its mean. An attractive choice is the square law detector, because the resulting detector output voltage will be proportional to the input power and consequently to the input signal.

A more founded choice of the optimal detector law was given by Kelly et al.[1]. In terms of optimal sensitivity, Kelly showed that a square law detector should be considered as a maximum-likelihood estimator and that such a detector was superior to detectors using any other law. That conclusion was based on the assumption that receiver gain fluctuations were random and independent of frequency. Furthermore, filtering by the receiver was assumed to be fixed. A more general derivation was given by Olsen [2] who reached the same conclusion.

Irrespective of the detector used, the main problem in radiometry is to discriminate between the input noise and the noise added by the receiver. Eq.(3.2) linking the input to the output, gives the impression that there is an unique output for a given input. However, the receiver's characteristics such as gain and receiver noise temperature are not constant with time due to power-supply variations, ambient temperature changes, mechanical stresses, and aging. This gives rise to an uncertainty in the output.

The supply voltage can be kept nearly constant by using voltage-regulators. They give...
very accurate voltage control, have a very low temperature dependence, and are available for various fixed or adjustable voltages. The stress effect can be minimized by ensuring the mechanical rigidity of components and connections in the system; while the effect of aging can be minimized by careful and regular calibration. In that way, most of the problems are alleviated, but it still leaves the effect of temperature on the receiver's characteristics. Therefore, classical radiometer receivers need continuous temperature stabilization in order to give a reliable output.

Different radiometer types have been developed through the years, each of them trying to solve the problem of the uncertainty in the output in its own way. Therefore, a discussion of the advantages and disadvantages of the most popular radiometer types will be worthwhile.

3.2.1. The Total Power Radiometer

![Block diagram of a total power radiometer.](image)

The total power radiometer is shown diagrammatically in figure 3.1. Basically, it is a superheterodyne receiver followed by a square law detector, a video amplifier and an integrator. The RF-amplifier amplifies the noise signals within a certain bandwidth centered on an RF-frequency. The mixer translates this RF-band into the same bandwidth centered on an IF-frequency. The IF-amplifier provides further amplification. Usually the bandwidth of the IF-amplifier and the corresponding IF-filter is much smaller than that of the RF-amplifier. The predetection bandwidth $B$ is largely determined by the IF bandpass filter characteristics, so that $B = B_{IF}$. The detector and integrator are used to find some measure for the mean of the microwave power. The resulting output voltage is given by:

$$V_{out} = c B G (T_a + T_{rec})$$  \hspace{1cm} (3.3)

with $G$ the total gain of the receiver and $c$ a proportionality constant which includes the voltage sensitivity of the detector diode.
The total power radiometer is an attractive design thanks to its simple configuration. In the absence of any imperfections, it should be the radiometer with the highest sensitivity. In the field of radiometry, sensitivity [3] is defined as the minimum change in $T_{sys}$ that is necessary to produce a detectable change in the radiometer's output. Here a detectable change is defined as the change in the dc-level of the output voltage equal to the standard deviation of its ac-component. According to this definition, the ideal sensitivity will be:

$$\Delta T = \frac{T_{a} + T_{ref}}{\sqrt{B \tau}}$$  \hspace{1cm} (3.4)

where $\tau$ is the integration time of the integrator.

Eq.(3.4) gives the sensitivity which takes into account uncertainties due to system's noise fluctuations. However, as stated in the previous section, gain variations will degrade the radiometer's performance and the sensitivity becomes:

$$\Delta T = (T_{a} + T_{ref}) \left[ \frac{1}{B \tau} + \left( \frac{\Delta G}{G} \right)^2 \right]$$  \hspace{1cm} (3.5)

Eq.(3.4) shows that for the typical case of a one-second integration time and several MHz bandwidth, the term representing gain fluctuations may be the limiting factor as far as the ultimate sensitivity is concerned.

If absolute accuracy is considered, the total power radiometer cannot be regarded generally as sufficiently stable to satisfy reasonable requirements of absolute accuracy. This is due to the fact that no extra countermeasures have been incorporated in order to overcome the problem of gain and system noise-instabilities. Therefore, even though the total power radiometer has the potential for the highest sensitivity, the sophistication of the instrumentation needed to reach a design in terms of reasonable requirements of stability has precluded it from being the normal choice for radiometry.

3.2.2. The Dicke Radiometer

The stability problem of the previous radiometer was partly solved by Dicke [4]. The Dicke radiometer is basically a total power radiometer, with the addition of a microwave switch, a load at a well-defined temperature, and a synchronous detector between the square law detector and the integrator. The receiver monitors the input signal for only half the measuring time and during the other half, it monitors the reference noise temperature of the load. If the switch is operated at a rate higher than the highest frequency of gain
fluctuations, the gain can be considered to be constant during one switch period. In this switched system, additional problems are the difficulties of determining the best switching rate and the optimum waveform that controls the switch. The problem of optimum waveform is related to the fact that the switch has to provide a duty cycle of 50% in order to cancel out the effect of the instabilities. This was discussed by Colvin [5] who found that the square waveform was best. He also showed that the synchronous detector must also have a square waveform and the bandpass between the square law detector and synchronous detector must be sufficiently broad to allow passage of the higher harmonics of the switching frequencies [5]. The problem of the optimum switching rate is related to the gain instabilities and its analysis can be a complicated procedure. Although a limited number of studies have been conducted in order to evaluate the nature of the system's gain fluctuations, it is generally observed that:

1) the bulk of the gain fluctuation spectrum lies at frequencies below 1 Hz (Hersman and Poe [6]),
2) for frequencies below 1 Hz, the power spectral density of $G_S$ (fluctuation spectrum) decreases with increasing frequency as $f^{-a}$ ($a > 0$) or faster (Yerbury [7]),
3) above 1 Hz, a nearly uniform spectrum has been observed (Hersman and Poe [6]),
4) practically no fluctuations with frequencies above 1 kHz can exist (Ulaby et al. [8]).

Figure 3.2. The Dicke radiometer.

Figure 3.2 shows the layout of the essential elements of a Dicke radiometer receiver. The input of the radiometer is switched with a frequency $F_{\text{switch}}$ to the antenna noise temperature and the reference noise temperature. When $F_{\text{switch}}$ is so rapid that $G$, $T_a$ and $T_{\text{rec}}$ can be considered to be constants, $V_{\text{out,1}}$ and $V_{\text{out,2}}$ can be given as:

$V_{\text{out,1}} = c B (T_a + T_{\text{rec}}) G$ during the first half period of $F_{\text{switch}}$, and
$V_{\text{out,2}} = c B (T_r + T_{\text{rec}}) G$ during the second half period.
If $1/P_{\text{switch}} \ll \tau$ the output to the integrator is:

$$V_{\text{out}} = c B (T_a - T_r) G$$

Comparing Eq.(3.3) and Eq.(3.6) shows that $T_{\text{rec}}$ has been eliminated and as long as $T_a - T_r$ is significantly less than $T_a + T_{\text{rec}}$, the effect of any gain instabilities has been reduced. That results in a better immunity to instabilities. The penalty paid for the improvement of stability is an increase in the ideal sensitivity by a factor 2. A factor of $\sqrt{2}$ is introduced by the fact that the input signal is monitored only half the time. Another factor of $\sqrt{2}$ comes from the comparison of two noise-like signals. The sensitivity, including gain variations, then becomes:

$$\Delta T = \sqrt{\left[ \frac{2(T_a + T_{\text{rec}})^2}{B \tau} + \frac{2(T_r + T_{\text{rec}})^2}{B \tau} + \left( \frac{\Delta G}{G} \right)^2 (T_a - T_r)^2 \right]}$$

Additionally, there is an increase in the system's complexity. Nevertheless, the Dicke radiometer is not capable of eliminating the sensitivity degradation by the receiver gain fluctuations unless the reference noise temperature equals the input noise temperature. Such a balanced situation can be accomplished by adjusting the reference noise temperature or by adjusting the predetection gain.

3.2.3. The Noise-Injection Radiometer

Fig. 3.3. The noise-injection radiometer.
Figure 3.3 shows the block diagram of a noise-injection radiometer. This type of configuration is the most complex mentioned so far. It is capable of eliminating the uncertainties introduced by $T_{\text{rec}}$ and $G$ variations, since it includes the components that are subject to instabilities in a feedback loop [9] (again an increase in complexity), which continuously adjusts the reference noise temperature to the input noise temperature. This leads to:

$$V_{\text{out}} = c (T_a' - T_r)G = 0 \text{ if } T_a' = T_r$$

with $T_a' = T_a + T_{\text{inj}}$.

So, a knowledge of $T_{\text{inj}}$ and $T_r$ gives sufficient information for determining $T_a$. In that way, the accuracy of $T_a$ will depend only on the accuracy of $T_r$ and $T_{\text{inj}}$ and is independent of the radiometer gain and receiver noise fluctuations.

The sensitivity of a noise-injection radiometer is very easy to deduce from Eq. (3.7) by setting $T_a = T_r$:

$$\Delta T = 2(T_a + T_{\text{rec}}) \frac{1}{B \tau}.$$  \hspace{1cm} (3.9)

The absolute accuracy of the noise-injection radiometer seems to be about the same as for the Dicke radiometer (however, the noise injection radiometer is sometimes reported to have better accuracy [10]).

3.2.4. The Gain Balancing Radiometer

![Fig.3.4. The gain balancing radiometer.](image)
Another method for obtaining a balanced situation is to adjust the IF gain in front of the detector [11]. The block diagram for this type of radiometer is shown in figure 3.4. The function of the feedback loop is to adjust the gain so that the output voltage is zero. This gives:

\[ V_{\text{out}} = 0 \text{ if } G_f(T_a + T_{\text{rec}}) = G_a(T_{\text{ref}} + T_{\text{rec}}) \]  

(3.10)

with \( G_f \) the fixed gain and \( G_a \) the adjustable gain.

A drawback to this method is that variations in \( T_{\text{rec}} \) will result in a measurement error in \( T_a \). In contrast, the technique employing noise injection in order to balance the radiometer is insensitive to drifts in \( T_{\text{rec}} \).

### 3.2.5. The Graham Radiometer

In all radiometer configurations, sensitivity and high stability are the prime design objectives. So far, the designs that have been mentioned aimed mainly at providing good stability. Several radiometers have been designed to increase the sensitivity over that of the Dicke radiometer and to approach the ideal sensitivity. Graham [12] proposed the use of two predetection chains (see figure 3.5), each of which is switched between the antenna and the reference noise source in such a way that at least one chain is always monitoring the antenna input. The net result is an increase in sensitivity by a factor of \( \sqrt{2} \). However, the price that has to be paid is the necessity of two predetection chains as well as the extra input and output coupling hardware.

### 3.2.6. The Correlation Radiometer

An increase of \( \sqrt{2} \) in sensitivity can also be gained by using a correlation radiometer [13] (see figure 3.6). In that type, the antenna output is divided between two receivers, the outputs of which are brought together for correlation purposes. The receiver noise temperatures in the two receiver chains are uncorrelated and they cancel each other out. Care must be taken to provide maximum isolation between the two receivers, otherwise receiver noise from one receiver entering the other will produce an extra correlated signal which will introduce an error when determining \( T_a \).

As in the case of Graham's radiometer, the extra sensitivity might not warrant the increased cost and complexity of the system.
Fig. 3.5. The Graham radiometer.

Fig. 3.6. The correlation radiometer.
3.2.7. Some Other Radiometer Types

Other radiometer configurations do exist, but they are modifications of the types which have been discussed already. Therefore, they are not elaborated here. Some examples are the phase-switching radiometer [14], frequency-switching radiometer [15], the Selove radiometer [16] and the polarization radiometer [17].
3.3. A Novel Temperature Stabilization Technique

3.3.1. Introduction
The discussions in the previous sections show that achieving an increased accuracy implies increasing the system's complexity and costs; also, that a temperature stable environment is needed to give a reliable output when the simple conventional total power radiometer is used. The temperature stabilization, which most commonly consists of an enclosure in which the temperature is maintained at an extremely constant level, may account for a high percentage of the total receiver costs and system complexity. Despite all engineering efforts and costs to produce the best stabilization circuitry, temperature gradients inside the stabilized enclosure cannot be eliminated in practice because the active components generate heat locally. However, a temperature stabilized environment has become a generally adopted solution and it has been used ever since the first total power radiometer receiver was designed. It is to be expected that with modern technologies (e.g. integrated microwave circuits, temperature sensing devices, and on-line data preprocessing), a compact and less expensive design will be feasible when using a novel approach to the stabilization problem. If for example, the temperature behaviour of the system's components can be characterized, temperature stabilization could be replaced by compensating for observed local temperature excursions using PC-based software. Since a PC is already used for the data acquisition, the advantage of using PC-based software for such a stabilization is clear. Stabilization in this way requires a knowledge of the accuracy with which the receiver components can be characterized and the determination of the system's critical components. The above reasoning provided the motivation for starting the research into the design of a novel temperature stabilization technique. Firstly, the test set-up used for characterizing the temperature dependence of the complete receiver and its components will be described. The possibility of software-based temperature stabilization is discussed and special attention is paid to the determination of the system's critical components.

3.3.2. The Test Set-up
Validating the reasoning described in the previous section required a test set-up to be designed. The device to be tested was the IF-part of an EUT radiometer designed for use in the Olympus propagation experiments project. The radiometer was originally a Dicke radiometer but it could easily be reconfigured to operate in the total power mode. The reason for using that configuration was the fact that it was simple, small and
manageable during the experiments. The radiometer, which is shown in figure 3.7, had an operating frequency of 12.7 GHz, a bandwidth of 100 MHz and an integration time of one second. The receiver had a very compact design in which the parts were mounted in separate copper compartments. That construction offered adequate electrical shielding between the separate parts. The IF-part of the total power radiometer is found in the upper compartments of the copper block. In figure 3.7, the following parts can be distinguished:

- an IF-amplifier (one A1 Watkins-Johnson and one 201 Avantek amplifier)
- a triple stage Cauer bandpass 10–110 MHz filter
- an input matching circuit for the detector (passive elements and one BF494 transistor)
- a square law detector diode (JISCII3486 Hewlett-Packard Schottky diode)
- a video amplifier (OPA27 Burr-Brown amplifier)

The input noise source was a microwave load, so the input noise temperature is equal to the physical temperature of the load.

In order to determine the temperature characteristics of the radiometer system's parts, some sort of temperature sensor needs to be incorporated. The criteria for selecting the method of temperature measurement and the corresponding sensor were:

- the method had to be simple in order to prevent unnecessary and undesirable complexity;
- the method had to have the potential of adequate electrical shielding from the on-line computers and other electrical equipment nearby in order to prevent their interference with the receiver system;
- the sensors had to ensure flexibility for easy handling and locating;
- the sensors should not be expensive; and
- last but not least, the sensors had to be capable of recording accurately temperatures.

For classical receivers using hardware temperature stabilization, sensitivities of 0.1 K are routinely achieved; therefore, their specifications were adopted for the development of the software-stabilized radiometer. The temperature measurement system had to be capable of a resolution of 0.1 K.

Thermocouples were chosen, to cope with all these criteria. They were J-type Iron-Constantan (Fe-CuNi) thermocouples with a 1.2 mm diameter (coating included).

In an ideal situation, the temperature of each component would be measured; however, such a complex measurement set-up needed to be simplified and choices had to be made so that only the active components were monitored.

Monitoring the physical temperature of active devices can be very complicated, because every component is fitted inside a package, making it almost impossible to reach the active region of the device. So in practice, there will be a temperature difference between the device temperature and the package temperature. Measuring the package
temperature correctly required thermocouples in physical contact with the packing material.
The thermocouples were transited through small holes in the receiver's enclosure and were attached to the packages with heat conducting paste and shielded from ambient temperature with perspex screening. To make sure that no measurement ambiguities originated from shifts of the sensors, a construction was designed to assure a fixed placement of the sensors during the experiments (see figure 3.7.).
The complete set-up was put in a climate chamber manufactured by Statham, which ensured a very stable temperature, between -50 and +150 degrees Celsius. During the measurements, no interference with the climate chamber's circuitry was detected.

The signals from the temperature sensors were recorded with a 30-channel chart recorder from Siemens (Multireg C1730). It also allowed measurement data to be sent to a computer through an RS-232 interface. That data was recorded with an PC, using the program SIMESP which was supplied with the chart recorder. The complete set-up is shown in figure 3.8.

It is worth mentioning that the sensors were calibrated at different temperatures before the experiments. The thermocouples were calibrated at zero degree by putting them into a tube filled with silicon oil, standing in melting ice (see figure 3.9.a). The silicon oil prevented short circuiting and had the property of establishing a homogeneous temperature distribution. The accuracy of the calibration was about 0.005 K. Calibration at other temperatures were performed by putting the sensors into a massive block of copper (heat conducting paste was applied to ensure a good thermal connection); then it was held at a constant temperature (see figure 3.9.b). Small temperature changes had no effect because of the block's mass. The temperature of the block was monitored with a calibrated thermometer (resolution of 0.05K). In that way, it was secured that the accuracy of temperature measurements were better than 0.08K.
Fig. 3.7. The EUT total power radiometer receiver and the construction to ensure a fixed placement of the temperature sensors during the experiments.
Fig. 3.8. The complete test set-up.

Fig. 3.9. a) Calibration of a sensor using melting ice for 0°C.  
b) Calibration using a block of copper for other temperatures.
3.3.3. Temperature Dependence of the Complete Radiometer

The measurements that were performed can be divided into two categories:
1) temperature dependence measurements of the complete radiometer,
2) temperature dependence measurements of the radiometer on a component level.

The first category was intended to get an impression of the amount of degradation in sensitivity, while the second set of measurements were performed in order to determine the most critical components of the receiver.

In this section, the results of the first set of measurements will be discussed. The signals that were recorded during the experiments represented the receiver's output and seven local temperatures. The thermocouples monitored the temperature of:
1) the A1 Watkins–Johnson amplifier
2) the 201 Avantek amplifier
3) the BF494 transistor
4) the square law detector diode HSCH3486.
5) the video amplifier
6) the radiometer enclosure
7) the environment.

The measurements were performed at fixed environment temperatures ranging from -10 to +25 degrees Celsius at one degree intervals. After activating the measurement set-up, a period for stabilization was allowed before the data was recorded. The measurement records shown in figure 3.10 clearly shows that the stabilization period was typically 10 minutes.

Numerous measurements were taken in order to assess the problems of hysteresis and reproducibility. Figure 3.11 shows that the stabilized receiver appeared to suffer little or not at all from hysteresis. The differences between the downward (from +25° to -10° degree Celsius) and the upward (from -10° to +25° degree Celsius) sessions are less than the radiometer sensitivity. Figure 3.12 depicts the reproducibility of the measurements; again the difference between two sessions, separated in time, was less than the sensitivity which equals 0.1 K.

Other sessions were performed at fixed input temperatures: one where the input temperature was held at 0° C (273 K) and one where the temperature load was liquid nitrogen (77 K). The results are shown in figure 3.13 and give a first impression of the feasibility of using software-based temperature stabilization. The temperature measurement error of 0.08K resulted in an uncertainty in the output voltage of 6 mV in the worst case. For the system under consideration that implied a sensitivity degradation of 1.8 K. It was felt that this was not a competitive performance specification and, therefore, an attempt was made to find out which parts of the radiometer receiver were more sensitive to temperature variations.
Fig. 3.10. A record of the temperature dependence for the complete radiometer. The ambient temperature was 10 degrees Celsius.

Fig. 3.11. The output voltage versus ambient temperature, for stabilized measurements. $+T_{amb} = +15^\circ C \rightarrow -10^\circ C$ $-T_{amb} = -10^\circ C \rightarrow +25^\circ C$
Fig. 3.12. Reproducibility of the measurements.

Fig. 3.13. Output voltage versus ambient temperature of 0 degree load and liquid nitrogen load.
3.3.4. Temperature Dependence Measurements of the Radiometer on a Component Level

The effects of local temperature changes of the active devices were measured in order to find the temperature dependence of each individual active device. In that way, it was possible to decide the critical components of the radiometer. Local temperature changes are difficult to record with the set-up used; but heating or cooling devices such as Peltier elements could define temperature precisely. However, the set-up used required the physical dimensions of such devices to be smaller than 2 mm which was impractical. Another option was to change the temperature, by blowing hot or cold air over a component. This could be done by guiding air through a capillary tube, diameter 2 mm, with one end close to the device and the other connected to a compressor. If the capillary tube was surrounded by melting ice or boiling water it was possible to cool or heat the air inside the capillary tube. To improve heat exchange, a tube with a larger diameter (8 mm) was used inside a bucket which contained hot water or cold ice. To prevent loss of heat, insulation was provided by means of glass fiber thread. The problem of water condensation on the components was circumvented by filtering the air through silica gel which absorbs water vapour. With that set-up (see figure 3.14) it was possible to change the local temperature of the active devices. A typical cooling record is shown in figure 3.15, it shows that the temperature of the video-amplifier could be changed, while the temperature of the other components remained the same.

Many local cooling and heating measurement sessions showed that almost all the temperature dependence of the receiver could be attributed to the Schottky diode detector. This implies that another type of diode should be used which was less temperature dependent; yet, software stabilization would still be feasible. Therefore, it was decided to compare the temperature dependence of different types of diodes. One type of diode that seemed particularly promising for use in a software stabilized radiometer was the tunnel diode because it is said to be extremely independent of temperature. In the next section, the relative merits of the common Schottky diode and the less frequently used tunnel diode will be compared.
Fig. 3.15. Local cooling of the video amplifier.
References Sections 3.1 to 3.3

THE SENSITIVITY OF RADIOMETRIC MEASUREMENTS.

ON THE OPTIMUM RADIOMETER.

RADIO ASTRONOMY RECEIVERS.

[4] Dicke, R.H.
THE MEASUREMENT OF THERMAL RADIATION AT MICROWAVE
FREQUENCIES.

[5] Colvin, R.S.
A STUDY OF RADIO ASTRONOMY RECEIVERS.

SENSITIVITY OF THE TOTAL POWER RADIOMETER WITH PERIODIC
ABSOIUTE CALIBRATION.

[7] Yerbury, M.J.
A GAIN–STABILIZING DETECTOR FOR USE IN RADIO ASTRONOMY.

MICROWAVE REMOTE SENSING.

A MICROWAVE FEEDBACK RADIOMETER.

[10] Skou, N.
MICROWAVE RADIOMETER SYSTEMS: DESIGN AND ANALYSIS.

A SWITCHED LOAD RADIOMETER.

[12] Graham, M.H.
RADIOMETER CIRCUITS.

ELEVEN–CENTIMETER BROADBAND CORRELATION RADIOMETERS.

[14] Ryle, M.
A NEW RADIO INTERFEROMETER AND ITS APPLICATION TO THE
OBSERVATION OF WEAK RADIO STARS.

[15] Drake, F.D.
RADIO–ASTRONOMY RADIOMETERS AND THEIR CALIBRATION.
[16] Selove, W.
A DC COMPARISON RADIOMETER

Polarization of 20—cm wavelength radiation from radio sources.
3.4. The Temperature Dependence of Schottky versus Backward Diodes for Radiometry Applications

Peter J. de Maagt, Tom F. Buss, and Theo G. van de Roer.
Eindhoven University of Technology
PO Box 513
5600 MB Eindhoven, The Netherlands

Key Terms
Square law detection, radiometry, Schottky diode, tunnel diode, backward diode.

Abstract
In this paper the temperature behaviour of Schottky and backward diodes operating as square law detectors at microwave frequencies is presented. To make a quantitative comparison between both diodes, the most important diode parameters are examined. The criteria for comparing them are voltage sensitivity and video resistance.

1. Introduction
The performance of a detector diode depends on the following factors: the rectification efficiency, output impedance and noise properties of the diode which determine the response of the diode to incident microwave radiation. The input impedance, and noise properties of the video amplifier (needed to obtain a desired output voltage gain) will affect the overall detector performance. The most popular detector diodes for microwave applications are the Schottky and tunnel diodes, the latter in the form of a backward diode. The Schottky diode uses the Schottky barrier [1], a metal-to-semiconductor junction, as a rectifier. The current in a Schottky barrier junction depends on majority carrier conduction and the Schottky diode is therefore potentially capable of operating at high frequencies. The tunnel diode is a PN-junction device whose doping is made purposely high in order to produce a very narrow junction across which electrons can tunnel easily. The tunneling effect leads to an N-shaped I–V-characteristic in the forward direction. For detection purposes, reduced doping densities are used so that the negative resistance region, which is
unwanted for this application, is reduced to almost zero and a strong curvature of the I–V characteristic near zero-bias is obtained. This tunneling phenomenon is also a majority carrier effect which makes the tunnel diode also suitable for use at microwave frequencies. Often Schottky diodes (independent of video amplifiers) are preferred because of their higher sensitivity. However, it is possible that better overall performance is obtained using a tunnel diode, when other system parameters (such as temperature dependence, low noise operation, impedance matching) are important.

That situation arises in radiometry applications. In such an application the diode input levels are very low. Therefore, the noise of the video amplifier may limit the dynamic range of the radiometer. In practice, low-noise amplifiers often have low input impedances and impedance matching between detector diode and video amplifier may become a problem.

As all diode performance specifications presented in data-sheets are given for very high load resistances, the presumed higher sensitivity of the Schottky diode may turn out to be an illusion.

Another aspect of the classical radiometer receiver is the required temperature stabilization. Temperature variations will affect the performance characteristics of the radiometer receiver (gain, sensitivity, etc.) and will lead to fluctuations in the output signal. To obtain a better immunity to temperature effects, a price has to be paid in the form of temperature stabilization. A serious drawback of such stabilization is the engineering effort and costs. If it could be possible to characterize the temperature behaviour of each part of the system, temperature stabilization could be replaced by computer correction of raw measured data using PC-based software. This requires a knowledge of the temperature behaviour of the receiver parts, the detector diode being one of them. In literature it is often quoted that the tunnel diode has "small" sensitivity variations over a "large" temperature range and that those variations are "much smaller" compared to those of the Schottky diode, but quantitative data is never provided.

This paper discusses in detail the effect of temperature on the most important parameters of the tunnel and Schottky diodes in order to make a quantitative comparison possible. The parameters chosen for comparison are voltage sensitivity and video resistance. The temperature dependence of the voltage sensitivity is determined both theoretically and experimentally; the temperature dependence of the video resistance will be derived theoretically. The results presented here are generally applicable, although a radiometry application was the motivation for starting the research.
The current transport in a Schottky diode is very similar to a standard PN–junction, the main difference being that minority carriers play practically no part in the transport process. As with a standard PN–junction device the dominant current transport process for Schottky diodes is the thermionic current \([1,3]\). This leads to the Schottky diode equation, which is very similar to the standard PN–diode equation:

\[
I = I_0 \left( \exp\left(\frac{qV}{kT}\right) - 1 \right)
\]

with

\[
I_0 = A^* A T^2 \exp\left(-\frac{q\phi_B}{kT}\right),
\]

\[
A^* = \frac{4qA^* r m^* k^2}{h^3},
\]

\(A^*\): the effective Richardson constant,

\(k\): the Boltzmann constant,

\(q\): the electron charge,

\(m^*\): the effective mass,

\(\phi_B\): the Schottky barrier height.

This standard diode–equation holds also for PN–junctions made of relatively pure semiconductors (any impurity doping represents a small fraction of the total density of the material). If the doping concentrations are increased, a situation is reached where the material is called degenerate. In such a situation the thermionic current as given by Eq.(1) is not necessarily the dominant process \([2,3]\). The total current, \(I\), now consists of three components: the tunneling, the excess and the thermionic current and each of these gives the most important contribution to the total current in a particular region of the I–V characteristic (see figure 1.a).

![Fig. 1. Current processes for degenerate semiconductors.](image)

- **a)** different current processes
- **b)** backward and tunnel diode.
For square law operation the diode has to be used in the range where the tunneling current is dominant. To move the region of strongest curvature as close as possible to the origin and to reduce the negative resistance which in this case is unwanted, the peak of the tunneling current is lowered by reducing the doping densities on both sides (see figure 1.b).

The tunnel current can be given in a closed form by [4]:

\[ I = I_p \left( V/V_p \right) \exp \left( 1 - V/V_p \right) \]  

(2)

where \( I_p \) and \( V_p \) are the peak current and peak voltage, respectively (see figure 1.b).

For calculation of the peak voltage, the relation between the charge carrier densities and doping degeneration is needed [4]. The doping degeneration is defined as \( E_n = E_f - E_c \) on the n-side and \( E_p = E_v - E_i \) on the p-side. For, \( n_n \), the density of electrons in the conduction band and \( p_p \), the density of holes in the valence band it is possible to write:

\[ n_n = 4\pi \left[ \frac{2m_n^* kT}{h^2} \right]^{3/2} \Gamma(1.5) \mathcal{F}_{\nu^2}(\Delta_n, E_n/kT) \]  

(3)

\[ p_p = 4\pi \left[ \frac{2m_p^* kT}{h^2} \right]^{3/2} \Gamma(1.5) \mathcal{F}_{\nu^2}(\Delta_p, E_p/kT) \]  

(4)

with

\[ \Delta_n = (E - E_c)/kT \quad \Delta_p = (E_v - E)/kT \]

\( E_v \): the energy at the valence band maximum,
\( E_c \): the energy at the conduction band minimum,
\( E_f \): the Fermi energy level,
\( m_n^*, m_p^* \): the effective masses for electrons and holes, respectively.
\( \Gamma \): the gamma function

The function \( \mathcal{F}_{\nu^2}(\epsilon, \eta) \) is a Fermi–Dirac integrals which can be calculated efficiently by using the following approximation [5]:

\[ \mathcal{F}_{\nu^2}(\epsilon, \eta) \approx \frac{1}{\Gamma(1.5)} \int_1^{\infty} \frac{\epsilon}{1 + \exp(\epsilon - \eta)} d\epsilon \approx \frac{4}{3} \left[ \eta^2 + 1.7 \right]^{3/4} \]

From Eqs. (3) and (4), \( E_n \) and \( E_p \) can be solved assuming that the values of \( n_n \) and \( p_p \) are equal to the donor and acceptor concentrations \( N_d \) and \( N_a \), respectively.

If \( E_n \) and \( E_p \) are determined, it is easy to determine \( V_p \) with the following relation [4]:

\[ V_p = \frac{E_n + E_p}{3q} \]  

(5)
The equation for the peak current density is given by [6]:

\[ J_p = \frac{g m^*}{36 \hbar^3} F_\perp D \exp \left[ -\frac{\tau (m^*)^2 (E_g)^2}{24 \hbar F} \right] \]

with

\[ F_\perp = \frac{\sqrt{2 \hbar F}}{\tau \hbar (m^*)^2 V_e}, \quad F = \left[ \frac{q^2}{2 \varepsilon} \right] \frac{1}{2N^* V_d}, \]

\[ N^* = \frac{N_a N_d}{N_a + N_d}, \]

\( m^* \): the effective mass for tunneling,
\( V_d \approx \frac{V_e}{q} \): the diffusion potential,
\( \varepsilon \): the permittivity,
\( D \approx q V_p \): the overlap integral,
\( E_g \): the energy bandgap,
\( \hbar \): the reduced Planck constant.

The temperature dependence of the energy bandgap is given by:

\[ E_g = E_g(0) - \frac{a T^2}{(T + \delta)} \]

where \( E_g(0) \), \( a \), and \( \delta \) are constants dependent on the material:

\( E_g(0) = 0.7437/1.170 \), \( a = 4.77 \times 10^{-4} / 4.73 \times 10^{-4} \) and \( \delta = 235/636 \) for Ge and Si, respectively.

The diffusion potential is given by:

\[ V_d = \frac{kT}{q} \ln \left[ \frac{N_a N_d}{n_i} \right] \]

with \( n_i \) the intrinsic density.

Using Eqs. (1) and (2) as a basis, the Schottky and tunnel diode detector characteristics as a function of temperature will be derived in the next section.

3. Temperature Dependence of Detector Characteristics

In this section the current sensitivity \( \beta(T) \), the voltage sensitivity \( \gamma(T) \) and the video resistance \( R_v(T) \) will be derived. These parameters are related by \( \gamma = \beta R_v \), and they are defined as follows:

\[ \beta = \frac{I_{sc}}{P_{in}}, \quad \gamma = \frac{V_{oc}}{P_{in}} \quad \text{and} \quad R_v = \frac{V_{oc}}{I_{sc}} \]

where the subscripts \( sc \) and \( oc \) refer to short circuit and open circuit, respectively (see...
figure 2). In general, these are frequency dependent quantities.
In the low-frequency limits, $\beta$ and $R_v$ become (see figure 2):

$$\beta_0 = \frac{1}{2} \frac{d^2 I}{dV^2} R_j$$

(10)

$$R_v = R_j + R_s$$

(11)

with $R_j$ the active small signal junction resistance given by:

$$R_j = \frac{dV}{dI}$$

(12)

and $R_s$ the diode series resistance which is, assuming that it resides mostly in the epitaxial layer and semiconductor substrate, given by [7]:

$$R_s = \frac{w}{\pi \eta \mu_i \mu_2} + \frac{1}{4\pi \eta \mu_i \mu_2}$$

(13)

with $w$ the thickness of the epitaxial layer, $r$ the radius of the contact, $N_1$ and $N_2$ the epitaxial layer majority carrier density and the substrate majority carrier density ($N_d$ or $N_a$), respectively and $\mu_1$ and $\mu_2$ the corresponding carrier mobility.

The temperature dependence of $\mu$ can be approximated by [3]:

$$\mu = \mu(T_0) \left( \frac{T}{T_0} \right)^c$$

(14)

where $c$ is a constant dependent on material; $c = -1.66$ and $-2.33$ for n- and p-type Ge, $c = -2.42$ and $-2.20$ for n- and p-type Si, respectively.

Due to the presence of parasitic junction capacitance ($C_j$), package capacitance ($C_p$) and series resistance the diode has to be represented by an equivalent circuit as in figure 2.

Fig. 2. The diode equivalent circuit.

In this case, the sensitivity $\beta_0$ is degraded to:

$$\beta = \frac{\beta_0}{1 - \omega^2 C_j C_p R_j R_s + \omega^2 C_j \frac{R_s R_j}{R_s + R_j} (C_p (R_j + R_s) + R_j C_j)} = \beta_0 f(\omega)$$

(15)
where \( C_j \) is given by:

\[
C_j = \varepsilon A \left( \frac{q}{2\varepsilon V_d} \right) N^* \tag{16}
\]

\( \gamma \) is derived easily from combination of Eq.(11) and Eq.(15), because:

\[
\gamma = R_v \beta_0 \tag{17}
\]

For a Schottky diode the \( \beta \) and \( R_j \) are, using Eqs.(1), (10) and (12), given by:

\[
\beta = \frac{q}{2kT} f(\omega) \tag{18}
\]

\[
R_j = \frac{k}{qAA^*T} e^{\frac{q\psi}{kT}} \tag{19}
\]

For a tunnel diode, using Eqs.(2), (10) and (12), it is possible to obtain Eqs.(20) and (21) (to facilitate comparison with the Schottky diode, the forward and reverse bias directions are interchanged, so that the characteristics of figure 1 are turned upside down):

\[
\beta = \frac{1}{V_P} f(\omega) \tag{20}
\]

\[
R_j = \frac{V_P}{A\psi e^{\exp(1)}} \tag{21}
\]

Substitution of the values of \( V_P \) (Eq.(5)), \( J_P \) (Eq.(6)), \( R_s \) (Eq.(13)), and \( C_j \) (Eq.(16)), gives the relations between the temperature and \( \beta, \gamma, \) and \( R_v \) for Schottky and tunnel diodes.

**4. Comparison of the Schottky and the Tunnel Diodes.**

In the previous section, the temperature dependence of the most important detector parameters were derived. In this section, both theoretically and experimentally obtained results are shown. The diodes used in the experiments are a zero-bias Schottky diode (type 7908-Z1) and a tunnel diode (type 7908-T2) from EMC Technology at an operating frequency of 1.1 GHz. As detailed information about the diodes could not be obtained from the manufacturer, most of the material parameters are obtained from measurements (the \( I-V \)-curves were measured point by point with a pulsed input signal). The results are summarized in table 1. Here \( R_s, C_j \) and \( C_p \) are measured at a temperature of 298 K.

These parameters are substituted in the relationships between detector parameters and temperature. Figure 3 shows the theoretically predicted temperature behaviour of the voltage sensitivity for both Schottky and tunnel diode. To obtain the measured curves a Statham climate chamber was used. A constant power input signal from a signal generator
was supplied to the diode input and its output voltage was measured with a high impedance multimeter \((R_l = 10 \, \text{MO})\). As the output impedance of the diode itself is \(R_v\), the effect of this load resistance on the diode output voltage is given by:

\[
V_{out} = V_{\text{diode out}} \frac{R_l}{R_l + R_v}
\]

(22).

In figure 4, the measured voltage sensitivity is shown for both diodes.

**Table 1 Material parameters**

<table>
<thead>
<tr>
<th>parameters</th>
<th>Schottky</th>
<th>tunnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A , [\mu m^2])</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>(R_s , [\Omega])</td>
<td>17</td>
<td>30</td>
</tr>
<tr>
<td>(C_j , [pF])</td>
<td>1.9</td>
<td>-</td>
</tr>
<tr>
<td>(C_p , [pF])</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(N_a , [cm^{-3}])</td>
<td>-</td>
<td>4.5 E19</td>
</tr>
<tr>
<td>(N_d , [cm^{-3}])</td>
<td>-</td>
<td>4.5 E19</td>
</tr>
<tr>
<td>(\phi_b , [eV])</td>
<td>0.5</td>
<td>-</td>
</tr>
</tbody>
</table>

**Fig. 3.** The calculated voltage sensitivity \(\gamma\) as a function of temperature.
Fig. 4. The measured voltage sensitivity $\gamma$ as a function of temperature.

It is clear that there is an extremely good agreement between measured and predicted values, except in the region $T < 260$ K. Furthermore, it is clearly visible that the voltage sensitivity of the tunnel diode is nearly independent of temperature.

Figures 5 and 6 show the influence of the load resistance $R_1$ on the Schottky and tunnel diodes, respectively. These figures show that the influence of the load resistance is large for the Schottky diode and illustrate that the performance specifications presented in data-sheets have to be treated with care. The reason for the influence of $R_1$ is the great dependence of $R_v$ on temperature for the Schottky diode, which is shown in figure 7.

Fig. 5. The calculated voltage sensitivity $\gamma$ as a function of temperature for the Schottky diode at different values of $R_1$. 
Fig. 6. The calculated voltage sensitivity $\gamma$ as a function of temperature for the tunnel diode at different values of $R_1$.

Fig. 7. The calculated $R_v$ as a function of temperature

5. Conclusions
This paper presents a quantitative comparison of Schottky and tunnel diodes. The bases for comparison are voltage sensitivity ($\gamma$) as a function of temperature and video resistance ($R_v$) as a function of temperature. Furthermore, the influence of load resistance on voltage sensitivity is discussed.

The measured results of the voltage sensitivity are found to agree well with the predicted values. The supposition that tunnel diodes have a lower temperature dependence is confirmed. Furthermore, the sensitivity of Schottky diodes drops drastically when the load resistance is lowered.
References

[1] Schottky, W.
HALBLEITERTHEORIE DER SPÜRRSCHICHT
Naturwissenschaften, 26, 843 (1938).

NEW PHENOMENON IN NARROW GERMANIUM P–N JUNCTIONS

[3] Saw, S.M.
PHYSICS OF SEMICONDUCTOR DEVICES

THE PREDICTION OF TUNNEL DIODE–CURRENT CHARACTERISTICS

[5] Blakemore, J.S.
SEMICONDUCTOR STATISTICS.

THEORY OF TUNNELING

VARIABLE IMPEDANCE DEVICES
Note: This section was published as a paper in International Journal of Infrared and Millimeter Waves, Vol.13 (1992), p.1075–1097. Therefore, the numbering of its Equations and references do not agree with the previous section.

3.5. A Novel Radiometer Receiver Stabilization Method
by
P.J.I. de Maagt, R.A.E. Oerlemans, J.C.A.M. van Gestel, and M.H.A.J. Herben
Eindhoven University of Technology
Telecommunications Division
PO Box 513, 5600 MB Eindhoven, The Netherlands

Abstract
A novel approach to solving the temperature stabilization problem of radiometer receivers is discussed, in which the classical temperature stabilization is replaced by a software-based compensation for any local temperature deviations observed. It is shown that a tunnel diode has to be incorporated as a square law detector in order to validate this new approach. The resulting measurement uncertainty is calculated and an approximation of the effect of aging is given.

1. Introduction
The use of microwave radiometers for an exploration of new satellite communication channels is based on the fact that there is a relationship between absorption of the propagation medium and emission of thermal radiation. Radiometry is a technique which enables thermal radiation to be measured by means of a highly sensitive microwave receiving system. The thermal radiation can be represented by an equivalent noise temperature $T_{\text{in}}$ at the input of the receiver and it will result in a receiver output voltage of:

$$V_{\text{out}} = cBG T_{\text{in}}$$

(1)

where $c$ is a proportional constant, $B$ the bandwidth of the receiver, and $G$ the total gain of the receiver.

Unfortunately, the statistical properties of the "input" noise do not differ from the noise originating in the receiver itself and the receiver output voltage is more realistically represented by:

$$V_{\text{out}} = cBG(T_{\text{in}} + T_{\text{rec}}) = cBG T_{\text{sys}}$$

(2)
The Microwave Radiometer as a Remote Sensing Device

where $T_{\text{rec}}$ represents the equivalent receiver noise temperature and $T_{\text{sys}}$ the equivalent system noise temperature.

The main problem in radiometry is discriminating between the input noise and the noise added by the receiver. Eq. (2) links the input to the output and gives the impression that there is an unique output for a given input. However, the receiver's characteristics, such as gain and receiver noise temperature, are not constant with time due to power-supply variations, ambient temperature changes, mechanical stress, and aging. This gives rise to an uncertain output.

The supply voltage can be kept nearly constant by using voltage-regulators. These regulators have a high voltage accuracy, a very low temperature dependence, and are available with different fixed or adjustable voltages. Dependence on mechanical stress can be minimized by maintaining the mechanical rigidity of components and connections in the system, and the effect of aging can be minimized by careful and frequent calibrations. That will alleviate most of the problems but it ignores the fact that the receiver characteristics depend on temperature. Therefore, classical radiometer receivers need a continuous temperature stabilization in order to obtain a reliable output.

The total power radiometer, which is an attractive design due to its high sensitivity and simple configuration, still has a limited performance due to the residual gain and receiver noise instabilities which limit its wider use in radiometry.

Such stability problems were partly solved by Dicke in 1946 [1] by introducing "comparison radiometers" which use a reference noise source and a switch. The Dicke receiver monitors the input signal for just half of the measuring time, and during the other half it monitors the reference noise temperature. If the switch operates at a rate higher than the highest frequency of relevant gain fluctuations, the gain can be considered constant during each switch period. Based on this assumption, it is possible to eliminate the influence of receiver noise, but the penalty paid for that improvement in stability is raised sensitivity by a factor of 2 and an increase in the system's complexity. Nevertheless, the Dicke radiometer is not capable of completely eliminating the effect of the receiver gain fluctuations (unless the reference noise temperature equals the input noise temperature). They can be eliminated by enclosing the parts of the system subject to instabilities in a feedback loop [2] (thereby, increasing the system's complexity) which continuously adjusts the reference noise temperature to the input noise temperature. The accuracy of such a balancing radiometer will depend on the accuracy of the reference temperature and feedback circuit and will be independent of gain and receiver noise fluctuations.

Over the years, other radiometer configurations with different conformation and
accuracy have been proposed, but they are generally modifications or combinations of the three types mentioned.

The above discussion shows that increasing the accuracy implies increasing the system's complexity and costs. The temperature stabilization, which usually consists of an enclosure in which the temperature is extremely stable, may account for a high percentage of the receiver's total costs and system's complexity. Despite considerable technical effort and costs, temperature gradients inside the stabilized enclosure cannot be avoided entirely in practice, because the active components generate heat locally. However, a temperature stabilized environment has become the generally adopted solution and, in this way, stabilization has always been used since the first radiometer receiver designs.

It seems that with modern technology (e.g. integrated microwave circuits, small temperature sensing devices, and on-line data preprocessing), a compact and less expensive design is feasible, if a new approach to the stabilization problem is used. For example, if the temperature behaviour of the system's components can be characterized, temperature stabilization may be replaced by compensating for observed temperature deviations with PC-based software.

Stabilization in this way requires accurate knowledge of how the receiver's components can be characterized and determining the critical parts of the system. Such reasoning was the motivation for starting the research described in this paper. Firstly, the temperature measurement system is discussed, followed by a discussion of the test set-up used for characterization of the temperature dependence of the complete receiver and its components. A detailed statistical analysis of the measurements is done and the possibility of software-based temperature stabilization is discussed. Special attention is given to the determination of the final measurement uncertainty and to an estimation of the effect of aging.

2. The Temperature Measurement System
In order to determine the temperature characteristics of the radiometer system, a temperature sensor needs to be incorporated. The criteria for selecting a particular temperature measurement method and its corresponding sensor are: the method must be simple enough to prevent unnecessary and undesirable complexity; the method must have the potential for adequately shielding it electrically from on-line computers and other electrical equipment in order to prevent their interference with the receiving system. Also, the sensors must ensure flexibility so that they are easy to handle and easy to locate; the sensors should be inexpensive, while last but not least, the sensors must be able to measure
the temperature accurately.
In classical receivers, the ambient temperature of the radiometer is stabilized within 0.1 K. With the development of a software-stabilized radiometer, the specifications of classical radiometer receivers must be stated as design objectives; therefore, the temperature measurement system has to be able to measure with a 0.1 K accuracy.
To comply with all of these criteria, thermocouples were chosen. The ones used were J-type iron-Constantan thermocouples with a 1.2 mm diameter (coating included). They were connected to a computer interface board which was designed specifically for use with thermocouples and the measured thermocouple voltages were amplified and multiplexed. The board also offered on-line cold-junction compensation. When taking measurements, it became clear that the performance of the cold-junction compensation circuit depended on the ambient temperature and it had to be corrected for this relationship.
The thermocouples were calibrated at zero degrees, by putting them in a tube filled with silicon oil, surrounded by melting ice. The silicon oil prevented short circuiting and had the property of establishing a homogeneous temperature distribution. The accuracy of the calibration was about 0.005 K. Calibration at other temperatures was performed by putting the sensors in a massive block of copper (heat conducting paste was applied to ensure a good thermal connection) and it was kept at a constant temperature. Small rapid temperature changes would have no effect due to the mass of the block. The temperature of the block was monitored with a calibrated thermometer (resolution of 0.05 K). Consequently, the accuracy of temperature measurements was better than 0.1 K.

3. The Measurement Set-up
Initial tests were carried out with the IF-section of a classical total power radiometer of the Eindhoven University of Technology (EUT). The radiometer has an operating frequency of 12.7 GHz, a bandwidth of 100 MHz and an integration time of 1 sec. This receiver forms a very compact design in which the receiver's components are mounted in separate copper compartments (see Figure 1).
Measurements with the initial set-up revealed that most of the receiver's dependence on temperature could be attributed to the Schottky diode used as a square law detector [3]. It also became clear that the temperature dependence of the Schottky detector diode prevented a performance comparable with those of classical receivers; therefore, it was necessary to consider using another detector diode.
The most popular diodes for microwave detector applications are the Schottky and tunnel diodes, the latter in the form of a backward diode. For classical radiometer receivers, normally Schottky diodes are preferred because of their higher sensitivity. However, a better overall performance can be obtained with a tunnel diode; particularly when other system parameters are important, such as temperature dependence, low noise operation, and impedance matching. In [4] it can be seen that the higher sensitivity of the Schottky diode is an illusion when low impedance video amplifiers are connected to the output in order to ensure low-noise operation. The sensitivity of Schottky diodes appears to drop drastically when the load resistance is lowered. It is also shown in [4] that the tunnel diode has a much lower temperature dependence than the Schottky diode. So, it may be that a tunnel diode should be employed instead of a Schottky diode for the design of the novel radiometer described here.

As the existing EUT radiometer did not allow experimental flexibility, another test set-up with the same operating parameters was constructed. The main building block of this set-up was a low-cost TVRO-down converter which has an RF to IF gain of 52 dB. This down converter is followed by an IF-amplifier of 10 dB, a bandpass filter whose 100 MHz bandwidth is centered around an IF-frequency of 1.4 GHz, a detector diode and a video-amplifier plus integrator. All of the components of this set-up are connected to each other with SMA-connectors (see Figure 2).

With that set-up, a fair comparison was obtained between the results with the tunnel-diode configuration and those obtained with the Schottky diode configuration, because only the diodes needed to be exchanged while the predetection part of the set-up remained the same. Since a mismatch can lead to significant errors, great care had to be taken when connecting and disconnecting the components in order to ensure reproducibility. To prevent instabilities due to a bad contact, the connectors were cleaned every time before connection.
The ambient temperature of the radiometer and its components was maintained within a climate chamber that generated a stable and homogeneous temperature by keeping the air circulating. When taking measurements, no interference caused by the climate chambers circuit was detected. The ambient temperature and the local temperatures of individual radiometer components were continuously monitored (see Figure 2). Furthermore, a period of stabilization was allowed before each measurement.

4. Statistical Analysis of the Measurements

A description of the statistical methods used will be given first. A calibration session typically consists of obtaining a number of points near the curve which represents the relationship between the unknown variable (the dependent variable) and a selected variable (the controlled variable). As the dependent variable was subject to a certain amount of experimental variation, due to the various instabilities in the system, the location of these points was uncertain. If the points obtained are displayed, it is often possible to fit the data with a smooth curve by eye; however, for proper scientific modeling, an objective statistical approach is required.

In the experiments described below, the controlled variable mostly was a local temperature and the dependent variable was a noise power or voltage somewhere in the system. Furthermore, every point near the curve for a particular output parameter represents an arithmetic average of a series of measurements at that particular temperature.

One problem is that, the arithmetic average (\( \overline{y} \)) may differ from the true value (\( \mu \)) due to the experimental variation and the limited size of the series of measurements. The point estimate \( \overline{y} \) of \( \mu \) gives no idea of the precision of the estimate and it is often preferable to use an interval estimate. If a series consists of \( N \) measurements and each measurement is taken from a population with an unknown mean \( \mu \) and standard deviation \( \sigma \), the distribution of \( \overline{y} \) will be approximately normal with an unknown mean \( \mu \) and standard deviation \( \sigma/\sqrt{N} \) (the central limit theory [5]). The a\( \alpha \)-confidence interval for \( \mu \) is given by:

\[
P(-c < \overline{y} - \frac{\mu}{\sigma/\sqrt{N}} < c) = \alpha \Rightarrow P(+\overline{y} - c\sigma/\sqrt{N} < \mu < \overline{y} + c\sigma/\sqrt{N}) = \alpha
\]

where \( P \) denotes the probability.
Fig. 2. The test set-up used with SMA-connectors.
The two end points of the interval are called the confidence limits. The value of the confidence interval $a$ depends on the application, taking into account the risk of making a wrong decision; for the measurements, the confidence level was taken as 99%. The confidence interval assumes that there is a % confidence that the interval will contain $\mu$. Usually, the standard deviation $\sigma$ is unknown and it has to be replaced by the series standard deviation $s$. In that case, the confidence interval can be determined in a similar manner; if $\sigma$ is replaced by $s$, the distribution is no longer normal but becomes a student-t distribution [5].

After determining the estimates $\bar{y}$ of $\mu$ with satisfactory confidence, a polynomial regression curve can be fitted to these points. The regression curve can be shown to give a good estimate (maximum likelihood [5]) of the unknown parameter when the following assumptions hold: the distribution of $y$ follows a normal distribution with a mean of $a_0+a_1x+...+a_mx^m$ and the conditional variance ($\sigma^2_{y|x}$) of a variable $y$ for a fixed value of $x$ is constant. Thus the conditional variance of the dependent variable does not depend on the value of $x$. A normality check for $y$ has to be performed with every regression.

Another problem of regression is the need to determine the lowest degree of the polynomial which adequately describes the measured data. That is done by using the $F$-test [5]. Here $F$ is the ratio between the increase in residual sum of squares and the estimated residual variance, which theoretically follows an $F$-distribution and can be written as:

$$F = \frac{\sum \left( \bar{y}_i - \sum_{k=0}^{m} a_kx_i^k \right)^2 - \sum \left( \bar{y}_i - \sum_{k=0}^{m-1} a_kx_i^k \right)^2}{\sum \left( \bar{y}_i - \sum_{k=0}^{m} a_kx_i^k \right)^2/(n-m-1)}$$

Here $R_m$ and $R_{m-1}$ are the residual sum of squares for the polynomial fits of orders $m$ and $m-1$, respectively, and $n$ the number of $x$-values considered. If a polynomial fits the data adequately then $R_m/(n-m-1)$ will be a good estimate of the conditional variance $\sigma^2_{y|x}$. If a polynomial of degree $m$ does fit the data better than a polynomial of degree $m-1$, then $R_m$ will be substantially smaller than $R_{m-1}$, and the $F$-ratio will be large. When checking how significant that better fit is, it is necessary to determine a predicted value of $F$ from its distribution, by calculating the value of the point $F(a,\nu_1,\nu_2)$ where $a$ is the significance level, and $\nu_1$ and $\nu_2$ are the degrees of freedom of two different polynomial fits. If the calculated $F$-value is less than the predicted value, the increased accuracy is not significant and it is unnecessary to try a higher order polynomial fit.

It should be stressed that the $F$-test is only valid for normally distributed residuals. So, in
order to examine whether the test is permissible, the residuals should be subjected to a
normality test. Another important observation is the fact that statistical significance tests
are not the only criteria for choosing a suitable model. Measurement instruments have
limited accuracy, which results in a practical significance. If the regression model used,
yields a curve which fits the data points properly, the residual variance \( \sigma^2 \) will have to
be of the same order as the variance of the measurement errors.

5. Determining the Temperature Dependence of Radiometer Components
Firstly, the temperature dependence of each radiometer component was considered.

a) RF down converter plus load.

The first temperature characteristic that is presented here is that of the complete RF down
converter block plus microwave load. An outline of the set-up is shown in Figure 3.

![Fig.3. The RF down converter block plus load.](image)

Here, \( T_{\text{load}} \) represents the temperature of the load, and \( T_{\text{rf}} \) and \( T_{\text{if}} \) represent the
temperatures of the RF and IF compartments of the down converter, respectively.

To prevent the load from becoming heated by the converter block, a stainless steel
transition waveguide should be used. Furthermore, two cho-seal RF-gaskets were used to
block the heat transfer from the down converter to the load. A small hole was made in the
rear of the ferrite for a temperature sensor in order to measure the temperature of the
microwave load. In that way, the temperature of the actual noise producer was measured
and not that of the waveguide. The temperature behaviour of the front-end was measured
with a spectrum analyzer having a resolution of 100 kHz. The measured output power of
the converter block plus load versus the RF-compartment temperature and the frequency
is shown in Figure 4.a. The corresponding regression surface is shown in Figure 4.b. From
the measurements, it appeared that the difference between \( T_{\text{rf}}, T_{\text{if}} \) and \( T_{\text{load}} \) was constant.

To ensure that an accurate estimate of the output power of the pre-detection section is
obtained, regression analysis was performed on both variables. As the spectrum analyzer has an accuracy of 0.1 dB and because the power ranged from −68 dBm to −66 dBm, the practical significance is in the order of $4 \times 10^{-9}$ mW.

Fig 4. Measured RF-compartment temperature and frequency dependence of the down converter plus load. a) measured data b) regression surface.

Tables 1 and 2 illustrate the process for finding the best regression surface, based on the following relationship:

$$P = c_0 + a_1T + a_2T^2 + b_1f + b_2f^2 + c_1Tf$$

where $P$ is the power measured in mW, $T$ is the RF-compartment temperature in K and $f$ is the frequency in MHz.

Table 1 The polynomial coefficients after regression

<table>
<thead>
<tr>
<th></th>
<th>$c_0$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$c_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>2.010e-07</td>
<td>-1.242e-09</td>
<td>-9.342e-12</td>
<td>-2.253e-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>5.334e-07</td>
<td>-1.137e-09</td>
<td>-9.342e-12</td>
<td>-2.253e-10</td>
<td>9.609e-13</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>5.251e-07</td>
<td>-1.137e-09</td>
<td>-9.342e-12</td>
<td>-2.253e-10</td>
<td>9.609e-13</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>2.407e-06</td>
<td>-1.137e-09</td>
<td>-9.342e-12</td>
<td>-2.253e-10</td>
<td>9.609e-13</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Results of the F-test

<table>
<thead>
<tr>
<th></th>
<th>$s_Y^2 \mu$</th>
<th>F</th>
<th>$F(a_1, \nu_1, \nu_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-b</td>
<td>4.423e-09</td>
<td>2.087e+03</td>
<td>6.850</td>
</tr>
<tr>
<td>b-c</td>
<td>4.048e-09</td>
<td>2.078e+01</td>
<td>6.850</td>
</tr>
<tr>
<td>b-d</td>
<td>4.325e-09</td>
<td>5.669e+00</td>
<td>6.850</td>
</tr>
<tr>
<td>c-e</td>
<td>3.935e-09</td>
<td>6.843e+00</td>
<td>6.850</td>
</tr>
<tr>
<td>b-f</td>
<td>3.950e-09</td>
<td>2.892e+01</td>
<td>3.950</td>
</tr>
<tr>
<td>e-f</td>
<td>3.950e-09</td>
<td>3.206e+01</td>
<td>6.850</td>
</tr>
</tbody>
</table>

As can be seen from Table 2, regression model c shows a significant improvement over models a and b (see a-b and b-c). Model c shows the significance of a second order model in $T$ (see b-c), while model d does not show such significance for the frequency $f$ (see b-d). Additionally, model e just fails to reach the statistically significance level of 99% (see c-e). Including the cross term $Tf$ does not show a significant improvement for model e (see e-f). Furthermore, it can be seen that a higher order regression in $T$ leaves the coefficients $b_1$ and $b_2$ unchanged, while a higher order regression in $f$ leaves the coefficients $a_1$ and $a_2$ unchanged. This means that $T$ and $f$ are independently controlled variables, which explains why model f produces no improvement over model e (see e-f). Models c, e, and f have statistical standard deviations in the order of magnitude of practical measurement accuracy. From these considerations, model c was chosen as the best model to characterize the output of the down converter. The corresponding regression surface is shown in Figure 4b.

The effect of this functional relationship on the error in the dependent variable is given by the propagation of error [5]. The propagation of error for model c is given by:

$$s_{pc}^2 = \left( \frac{\partial P}{\partial T} \right)^2 s_T^2 + \left( \frac{\partial P}{\partial f} \right)^2 s_f^2 = (a_1 + 2a_2T)^2s_T^2 + b_1^2s_f^2$$

where $s_T$ is the accuracy of the temperature measurement (0.08K) and $s_f$ is the accuracy of the frequency measurement of the spectrum analyzer (<10^-10 MHz).

This error is negligible with respect to the error due to measurement accuracy of the spectrum analyzer ($s_{meas}$). Therefore, the total r.m.s. error of the determination of the regression curve is given by:

$$s_t = \sqrt{s_{meas}^2 + s_{pc}^2 + s_{pe}^2} \sqrt{s_{meas}^2} = 4 \times 10^{-9} \text{ mW.}$$
b) IF-amplifier.

An IF-amplifier was used to ensure that both the detector diodes operated in the upper part of their square law region. In that way, the output voltage of the detector was kept as high as possible, which in turn, reduced any problems due to the noise generated by the video amplifier. The IF-amplifier set-up is shown in Figure 5 and the amplifier gain as a function of the amplifier temperature and frequency appears in Figure 6.

Figure 5 shows that the thermocouple was placed within the metal package, just on top of the IF-amplifier. The curves shown in Figure 6 were obtained from supplying a constant input signal to the amplifier while measuring the output with the spectrum analyzer.

![IF-amplifier and location of thermocouple](image)

**Fig. 5.** The IF-amplifier and the location of the thermocouple.

![Amplifier gain as a function of local temperature and frequency](image)

**Fig. 6.** The amplifier gain as a function of the local temperature and frequency.

- a) measured data
- b) regression surface.

The regression analysis performed on the IF-amplifier gain was similar to the regression analysis on the down converter block. The resulting polynomial coefficients are presented in Table 3.
Table 3 The polynomial coefficients after regression

| diode | $c_0$     | $a_1$     | $a_2$ | $b_1$     | $b_2$     | $b_3$     | $s_{y|x}$ |
|-------|-----------|-----------|-------|-----------|-----------|-----------|-----------|
| Schottky | -7.237e+03 | -2.196e-01 | 3.799e-04 | 1.512e+01 | -1.045e-02 | 2.404e-06 | 2.187e-01 |
| tunnel  | 1.866e+00  | -7.716e-03 | 1.563e-05 |           |           |           | 4.634e-04 |

The variables $f$ and $T$ were found to be independently controlled variables and the error of propagation was again negligible with respect to the measurement accuracy of the spectrum analyzer.

c) bandpass filter.
Both the bandwidth, and the insertion loss of the bandpass filter appeared to be independent of the local temperature; furthermore, the ripple of the filter was less than the measurement accuracy.

d) diodes.
The Schottky and tunnel diodes were fitted in a special package with SMA connectors. As stated previously, the advantage of that configuration was its flexibility, but a disadvantage was the difficulty of measuring the real diode temperature. A construction was designed for fixing the thermocouple to the package (see Figure 7). A pincer-like construction and isolation material were used to screen the thermocouple from temperature fluctuations around it.

A constant signal power from a signal generator was supplied to the diode input and its output was measured with a high impedance multimeter ($R_1 >10M\Omega$). Figure 8 clearly shows that the voltage sensitivity of the Schottky diode is highly dependent on temperature, while the sensitivity of the tunnel diode is nearly temperature independent. Examination of those curves shows that the tunnel diode has a higher sensitivity than the Schottky diode when the diode temperature is less than 262 K.

Regression analysis produced the coefficients given in Table 4. Also, the propagation of error was negligible for this system component with respect to the measurement accuracy.

Table 4 The polynomial coefficients after regression

| diode     | $c_0$     | $a_1$     | $a_2$ | $a_3$     | $s_{y|x}$ |
|-----------|-----------|-----------|-------|-----------|-----------|
| Schottky  | 2.368e+02 | -2.850e+00 | 1.117e-02 | -1.423e-05 | 3.163e-02 |
| tunnel    | 1.866e+00 | -7.716e-03 | 1.563e-05 |           | 4.634e-04 |
The sensitivity of the Schottky and tunnel diode as a function of the diode temperature.

The video amplifier was not dependent on the temperature; however, an offset drift of a few $\mu$V was measured (see figure 9), which appeared to be not directly related to the local temperature. Although this drift can result in a poor post-detection performance of the total power radiometer, it was not important for our comparative study.
Fig. 9. The offset drift of the video amplifier.

6. Determination of Temperature Dependence of the Complete Set-up

In this section, measurements taken for the complete set-up are considered. The complete radiometer was placed in the climate chamber and the ambient temperature was registered. Firstly, the results of the radiometer with a Schottky detector diode are given below.

Table 5 Regression of the Schottky diode radiometer output on ambient temperature; the first session (1), the second session (2) and the combined session (1+2).

<table>
<thead>
<tr>
<th></th>
<th>c_0</th>
<th>a_1</th>
<th>a_2</th>
<th>a_3</th>
<th>a_4</th>
<th>s^y^ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.017e+03</td>
<td>-1.099e+01</td>
<td>3.925e−02</td>
<td>-4.622e−05</td>
<td></td>
<td>6.825e−02</td>
</tr>
<tr>
<td>2</td>
<td>-5.840e+03</td>
<td>8.662e+01</td>
<td>-4.811e−01</td>
<td>1.185e−03</td>
<td>-1.091e−06</td>
<td>4.514e−02</td>
</tr>
<tr>
<td>1+2</td>
<td>8.345e+02</td>
<td>-9.033e+00</td>
<td>3.226e−02</td>
<td>-3.789e−05</td>
<td></td>
<td>6.267e−02</td>
</tr>
</tbody>
</table>
Figure 10 shows the output voltage of this radiometer as a function of the ambient temperature during two measurement sessions, which offer an opportunity to test the short-term reproducibility of the measurements. Table 5 gives the results of the regression analysis for the first and the second sessions, and a combination of the two. To test the reproducibility, the confidence interval of the first session can be calculated and if verification shows that the data points of the second session lie within that interval it can be concluded that the measurement could be reproduced. It is also possible to calculate the confidence interval of the combined session and verify whether the data points of the first and second session both lie within this interval. The latter method is depicted in Figure 11, which takes into account the drift of the video amplifier. The value of the drift has been obtained from the worst-case value listed in the data-sheets for the video-amplifier and making that value equal to the $3\sigma$ value of the real drift that was assumed to be Gaussian distributed. Figure 11 clearly shows that the measurements can be reproduced at the 99% confidence level.

The same procedure can be used for the tunnel diode configuration and the results are given in Table 6 (regression) and Figures 12 (measurements) and 13 (confidence intervals).
Fig. 11. Verification of the reproducibility for the Schottky diode configuration.

Table 6 Regression of the tunnel diode radiometer output on ambient temperature; the first session (1), the second session (2) and the combined session (1+2).

<table>
<thead>
<tr>
<th></th>
<th>c_0</th>
<th>a_1</th>
<th>a_2</th>
<th>a_3</th>
<th>a_4</th>
<th>s_{Y^* x}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.143e+01</td>
<td>-2.650e-02</td>
<td></td>
<td></td>
<td></td>
<td>3.848e-02</td>
</tr>
<tr>
<td>1+2</td>
<td>1.939e+02</td>
<td>-1.970e+00</td>
<td>6.884e-03</td>
<td>-8.113e-06</td>
<td></td>
<td>5.001e-02</td>
</tr>
</tbody>
</table>

Fig. 12. The output voltage of the tunnel diode configuration as a function of the ambient temperature.
Fig. 13. Verification of the reproducibility for the tunnel diode configuration.

As with the Schottky diode set-up the measurements are reproducible at the 99% confidence level.

7. Relationship between Ambient and Local Temperatures

For each of the stabilized measurements previously discussed, the ambient and local temperatures were registered. That gives an opportunity to find the relationships between the ambient and local temperatures, which are summarized in Table 7.

Table 7 Local temperatures as a function of the ambient temperature.

|          | \( c_0 \)   | \( a_1 \)   | \( a_2 \)   | \( S_{Y|x} \)  |
|----------|-------------|-------------|-------------|-----------------|
| \( T_{load} \) | 4.707e+00  | 9.831e-01  |             | 0.760e-01      |
| \( T_{rf} \)    | 3.788e-01  | 9.837e-01  |             | 0.150e-01      |
| \( T_{if} \)    | 3.712e-01  | 9.834e-01  |             | 0.629e-01      |
| \( T_{amp} \)   | 7.599e+00  | 9.798e-01  |             | 0.336e-01      |
| \( T_{Schottky} \)| 4.458e+00  | 9.822e-01  |             | 0.486e-01      |
| \( T_{tunnel} \)| -2.563e+01 | 1.197e+00  | -3.794e-04 | 0.986e-02      |
| \( T_{video} \) | 6.950e+00  | 9.737e-01  |             | 0.590e-01      |

Each measurement session offered an opportunity to determine those relations; however, statistical analysis showed that all the curves were statistically identical. Therefore, the relationships given in Table 7 may be considered to represent each of the measurements. As a consequence, the need to measure all local temperatures continuously becomes
questionable, particularly for the situations in which the ambient temperature of the radiometer fluctuates only slowly, so that the temperature distribution within the radiometer can be considered to be stable.

8. The Feasibility of Software Compensation

Sections 5, 6 and 7 dealt with measuring the temperature characteristics of the radiometer system's components and the complete system, and the relationships between ambient and local temperatures. To discover the practical consequences of those characteristics on software-based compensation for radiometer temperature deviations, this section will concentrate on the additional error introduced by this method when compared with the classical method of temperature stabilization.

The output of the total power radiometer set-up can be expressed by:

\[ V_{\text{out}} = kBG(T_{in} + T_{rec})G_{\text{amp}}G_{\text{video}} \]  

(10)

with \( G \) the gain of the down converter, \( G_{\text{amp}} \) the gain of the IF-amplifier, \( \gamma \) the voltage sensitivity of the detector diode, and \( G_{\text{video}} \) the gain of the video amplifier.

In practice it is impossible to distinguish a change in one of the parameters from a change in the input signal \( T_{in} \) but the additional uncertainty of the radiometer can be found by assuming that changes in any of the parameters are caused by the input signal.

The changes represented by \( s_{\text{conv}}, s_{\text{amp}}, s_{\gamma}, \) and \( s_{\text{video}} \) are mainly caused by the fact that the regression curves are estimations of the real curves.

The relative quadratic error in the radiometer output is the sum of the relative quadratic errors introduced by the individual components. If software stabilization is based on the curves for the separate components, it will lead to:

\[ \frac{s_{\text{out}}^2}{V_{\text{out}}^2} = \frac{s_{\text{conv}}^2}{G_{\text{conv}}^2} + \frac{s_{\text{amp}}^2}{G_{\text{amp}}^2} + \frac{s_{\gamma}^2}{\gamma^2} + \frac{s_{\text{video}}^2}{G_{\text{video}}^2} \]  

(11)

In other words, if a software correction is made in that way, the quadratic errors will accumulate. Figure 14 shows the total additional uncertainty obtained by summing the individual contributions of the down converter, the diodes, and the IF-amplifier. As the video amplifier was temperature independent, the error \( s_{\text{video}}^2 \) equals zero. The values of the other errors can be found as \( s_{\gamma}^2 \) in Tables 2 to 4. It should be noted that the transfer function of the converter was obtained while the input was varying. This effect could be canceled because the temperature dependence of both the converter gain and the receiver noise was found to be linear, and given by:
\[ G_{\text{conv}} = 1.51 \times 10^6 - 4.03 \times 10^3 T_{\text{amb}} \]
\[ T_{\text{ref}} = -101.9 + 2.020 T_{\text{amb}} \]  

When the software stabilization is performed using the curve which characterizes the dependence of the complete set-up on the ambient temperature, the additional error will be as shown in Figure 15 (the effect of varying input has been canceled). Examining this figure shows that the typical additional error for the tunnel diode configuration is 2.5K. This uncertainty has to be considered as an additional "calibration" error introduced by the software stabilization. For comparison, the conventional hot and cold load calibration procedure introduces a typical uncertainty of 1 to 1.5 K.

![Figure 14](image-url)  
**Fig. 14.** The additional uncertainty in T_in obtained by summing the errors introduced by the individual components.

![Figure 15](image-url)  
**Fig. 15.** The additional uncertainty in T_in for the complete radiometer.
The figures 14 and 15 also reflect the expected inferiority of the Schottky diode. As a consequence, the method of software stabilization described is only viable if a tunnel diode is incorporated as a square law detector device.

Comparing these figures indicates that software stabilization should not be based on the curves of the individual components. However, it must be noted that figure 14 includes the measurement accuracy of different measurement instruments. More accurate instruments for determining the temperature dependence of the system's individual parts, could decrease the error. If the separate relative errors are considered it appears that in this case the most critical parts are the IF-amplifier and the diodes.

Another effect that can be estimated from the calibration curves is aging, which causes the performance of the radiometer to change with time. That could result in a constant change in output level (a variation of \(c_0\)), or it could result in a change in the slope of the calibration curves. Any variation in \(c_0\) will result in a shift of the uncertainty curves shown in Figures 14 and 15 which can be removed by the hot and cold load calibration. The effect of a change in the slope is illustrated in Figure 16; it shows the additional uncertainty for 1% and 2% slope variations.

![Graph showing additional uncertainty in T in for different values of the slope.](image)

Fig. 16. The additional uncertainty in \(T_\text{in}\) for different values of the slope.

To test the long-term reproducibility, measurement sessions were performed at different times separated by several months. Since statistically equal curves were obtained from these sessions, the variations of the slope considered in figure 16 would not occur over a period of several months.
8. Conclusions
The research described here has produced some functional relationships which describe radiometer performance as a function of ambient and local temperatures. Those relationships can serve as the input to a software-based temperature stabilization procedure. In that way, readily available computing power can be used to keep track of the system's performance and the output signal from the radiometer can be constantly corrected with the updated performance data. The advantages of such an approach to the temperature stabilization problem are quite striking: when a personal computer is already being used for data acquisition, no extra temperature stabilization hardware is needed and, therefore, the complete configuration will be simple and economic. The disadvantage of this approach is the difficulty of obtaining the temperature curves which requires some experience, judgment and time. Practice has shown the value of using a tunnel diode as a detection device in order to ensure accurate compensation. The additional uncertainty of this radiometer configuration was found to be 2.5 K which makes this approach a valid one. Since all the data for the temperature curves was obtained in a stabilized environment, the effect of rapid temperature fluctuations is not included. However, no significant loss in performance is expected because typical "off-the-shelf" radiometers are mounted in a metal enclosure. Such enclosures act as a temperature buffer and that results in a partially stabilized situation. Furthermore, it is very easy to use a low-cost temperature insulation buffer (e.g. polystyrene), so that temperature differentials across the radiometer parts can be neglected. In a stabilized situation, the relationships between ambient and local temperatures are fixed, and it is unnecessary to monitor the local temperatures. Furthermore, it is advisable to have as few sensors as possible, because every extra calibration curve introduces an additional uncertainty to the output.
A first order approximation of the effect of aging has been obtained in the form of additional uncertainty due to a change in the slope of the calibration curve. Different measurements separated by several months showed no aging effects.

9. Acknowledgements
The authors are indebted to Prof. G. Brussaard for initiating and motivating this research topic, and to dr. P.R. Attwood for reading this paper and correcting the grammar.
References

[1]  Dick, R.H.
THE MEASUREMENT OF THERMAL RADIATION AT MICROWAVE FREQUENCIES.

A SWITCHED LOAD RADIOMETER.

THE CORRECTION OF TEMPERATURE VARIATIONS IN A RADIOMETER.

THE TEMPERATURE DEPENDENCE OF SCHOTTKY VERSUS BACKWARD DIODES.
Accepted for publication in Microwave and Optical technology letters (June 1992 issue).

[5]  Chatfield, C.
STATISTICS FOR TECHNOLOGY.
4. Remote Sensing by Passive Imaging

4.1. Introduction

Remote sensing is, in a broad sense, collecting information about an object without coming into physical contact with it. There are two major forms of remote sensing. The first form can be referred to as being quantitatively oriented, because it stresses the quantitative aspects and the long-term statistical properties of the data. The second form can be referred to as image oriented, because it emphasizes the image aspects of the object and relies greatly on the generation of an image. The previous chapters were mainly concerned with the first form of remote sensing, while this chapter studies the second form. Imaging remote sensing techniques can be further subdivided, according to the distance between the sensor and the object resulting in long-range or near-field techniques. The near-field sensing techniques include many of the medical, industrial and scientific applications. Examples are the investigation of deep hyperthermia [1], the detection of thermal gradients in the brain [2], vision for robotics [3], and non-destructive testing [4]. A recent survey of the near-field imaging techniques was done by Bolomey [5] and, although important, that area is not within the scope of this chapter; actually, it emphasis on long-range remote sensing techniques.

Passive remote sensing of the atmosphere is based on the natural electromagnetic noise emission of the atmosphere, and therefore, is able to provide relevant information about atmospheric processes and parameters (such as the liquid water content, water vapour content, air temperature gradient, and so on). The use of passive remote sensing was originally restricted to IR-observations, but in recent years the use of microwave radiometry for this purpose has developed rapidly. It was motivated by the lesser extinction of microwaves in contrast to IR-waves when hydrometeors are present. As a consequence microwave frequencies allow nearly all-weather operation. Furthermore, a related characteristic of microwaves is that their interactions with media are primarily sensitive to the dielectric properties of the media in such a way that microwaves can be expected to provide indirect access to any physical or chemical factor which these dielectric properties are dependent upon, such as composition, water content and temperature. Another advantage of microwave radiometry rests in its use of the same frequencies as satellite communication. It is possible to predict the behaviour of communication links at those frequencies by using the knowledge that is obtained from sensing the atmosphere. A major disadvantage of microwave radiometry is its substantially lower spatial resolution when compared with infrared systems.
Active sensors can overcome that limitation by exploiting the phase coherency of the received signals [10]. It should be mentioned, that doing so would increase the system's complexity. The inherent nature of noise radiation prevents analogous techniques from being used in passive systems.

A partial enhancement of spatial resolution can be obtained by exploiting the relative motion between the actual situation and the microwave radiometer system. Several relevant techniques have been proposed [6]–[11]; however, some of them use approximations of questionable validity. The purpose of this chapter is to present the results of research concerning the practical possibilities of spatial reconstruction that use as few approximations as possible.

In section 4.2 the basic problems related to imaging are discussed. Section 4.2 is followed by a brief description of some of the existing imaging techniques; while, a new robust imaging technique is presented in section 4.4. The influence of the object velocity on the reconstruction is explained in section 4.5. Section 4.6 discusses the performance of a reconstruction technique for antennas with shaped main beam. Section 4.7 gives the transfer function of the algorithm. Finally, the conclusions of this research are given in section 4.8. It should be noted that in the figures of sections 4.4 to 4.7 the curves of the antenna temperature are scaled by a factor of \( r \) in order to prevent crowded plots.

4.2. Basic Problems Related to Imaging Techniques

Most atmospheric remote sensing problems can be reduced to a Fredholm integral of the first kind. Writing this form specifically for a radiometer system (antenna + receiver), the integral takes the following form:

\[
T_A(t) = \frac{1}{4\pi} \int_{4\pi} T(t, \theta, \phi) G(\theta, \phi) d\Omega \quad (4.1)
\]

where \( T_A \) is the antenna noise temperature, \( T(t, \theta, \phi) \) the brightness temperature of the scene and \( G(\theta, \phi) \) the radiometer antenna pattern.

This relationship represents the influence of the radiometer antenna pattern and because of the finite dimensions of the antenna, resulting in a non-zero beamwidth, the antenna temperature measured may differ significantly from the brightness temperature of the scene. It is better to say that the temperature measured represents a spatially filtered brightness temperature.

Furthermore, the radiometer receiver can integrate the signal \( T_A \) with time, so that \( T_A \) becomes:
Remote Sensing by Passive Imaging

\[ T_a(t) = \frac{1}{\tau} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} T_A(\xi) \, d\xi \]  

(4.2)

Where \( \tau \) is the integration time of the radiometer receiver.

If a more detailed image of the scene is needed, the influences of antenna and receiver have to be canceled. This process is called imaging, deconvolution or microwave inversion. However, a problem inherent to radiometric observation is that the inversion is a very complex matter. To understand more about inversion, it is convenient to enumerate the ways that the brightness temperature of the scene can change before it reaches the output of a radiometer receiver. The different changes are the following:

a) addition of cosmic radiation and radiation from the environment, including reflections.

b) alteration of the radiation energy by absorption and attenuation.

c) integration of the scene with the radiation pattern characteristics of the radiometer antenna and receiver.

d) addition of systematic and statistical changes in the radiation detected by the electronic components of the detection equipment.

If the changes in a) and b) are taken for granted, the problem becomes a case of eliminating, or at least reducing, the changes in c) and d).

To visualize the changes related to c), consider equation (4.1). For simplicity, the influence of the radiometer receiver can be disregarded. The problem of the smoothing action of the antenna pattern can be demonstrated by considering the following example of brightness temperature and antenna pattern:

\[ T_1(\theta, \phi) = T(\theta, \phi), \quad T_2(\theta, \phi) = T(\theta, \phi) + C \cos(\omega \theta) \]

(4.3)

The difference between \( T_{A_1} = G \ast T_1 \) and \( T_{A_2} = G \ast T_2 \), where * denotes the integral operator of Eq.(4.1), is given by:

\[ |T_{A_1} - T_{A_2}| = \frac{C}{4\pi} \frac{4\pi}{\Omega_a} \int_0^{\Omega_a} \cos(\omega \theta) \, d\Omega \times \frac{C}{\Omega_a \omega} \]  

(4.4)

So, by making \( \omega \) sufficiently large or \( C \) sufficiently small, that difference can be reduced considerably until it becomes impossible to distinguish between \( T_1 \) and \( T_2 \). This means that integrating with the antenna pattern is equivalent with low-pass filtering.
This result also forms an introduction to the changes related with d). Due to the smoothing action of the antenna-pattern, small-amplitude or spatially-rapid variations that occur in $T(\theta,\phi)$ will have no effect on $T_A$. Or in another way, a change detected in $T_A$ will have to come from a relatively large or slow change in $T(\theta,\phi)$. However, antenna temperature is being measured with a non-ideal radiometer receiver and that introduces additional noise to the already noisy signal $T_A$. So, a change in $T_A$ can also originate from the radiometer receiver. Performing the inversion in that case will lead to errors when reconstructing the original brightness temperature. This makes clear that the problem of inversion is very complex.

The related problems originate from the physical relationship as given in Eq. (4.1). These problems can not be solved by increasing the accuracy and efficiency of the numerical algorithms used in the inversion process. That is because the first kind Fredholm integral is a well-known example of an ill-posed problem; this means that Eq. (4.1) does not satisfy the following properties [12]:

1) for every $T_A$ there exists a solution $T$;
2) the solution $T$ is unique;
3) the problem is stable, such that a small change in $T$ leads to a small change in $T_A$

and vice versa.

The discussion above makes it obvious that the second and third properties are not satisfied, and that the problem is ill-posed.

The first step towards a solution to the problem is finding a problem definition suited for inversion. In general, the Fredholm integral of the first kind can be written in a matrix form as [13]:

$$T_a = G T$$

(4.5)

where the elements $T_{ai}$ and $T_i$ of the vectors $T_a$ and $T$, respectively, represent the values at a certain time $t=t_i$ and $G$ is a matrix representing the antenna pattern.

Two basic schemes exist to solve this matrix problem: (1) iteration and (2) matrix inversion. However, the two ways are not unrelated. Fleming [14] showed that for an iteration scheme, a corresponding inverse matrix method exists, and conversely. In the next section, some of the types of inversion that are most widely known will be presented briefly along with their advantages and disadvantages.
4.3. Some Inversion Techniques

In this section a brief description of some inversion techniques is given, as an introduction to the problems that most of the techniques have in common.

4.3.1. Inversion by Multiple Convolution

The first technique that can be used for the purpose of inversion is multiple convolution. In this method, the problem of inverse convolution is overcome by rewriting the problem \( F(T) = 0 \) in the form \( T = f(T) \) and then starting an iteration process.

The original problem is given by:

\[
F(T(\theta, \phi, t)) = T_a(t) - G(\theta, \phi) * T(\theta, \phi, t)
\]

or in a form suitable for iteration:

\[
F(T_n(\theta, \phi, t)) = T_a(t) - G(\theta, \phi) * T_n(\theta, \phi, t)
\]

The solution to this problem is found by successive substitutions:

\[
T_{n+1}(\theta, \phi, t) = T_n(\theta, \phi, t) - C_n F(T_n(\theta, \phi, t))
\]

(4.8)

where \( C_n \) is a constant that guarantees convergence (if \( C_n = 1/F'(T_n) \), where \( ' \) denotes the derivative, this method becomes the well known Newton's method). The iteration process is terminated if:

\[
|T_{n+1}(\theta, \phi, t) - T_n(\theta, \phi, t)| \leq \varepsilon
\]

(4.9)

Combining Eq.(4.8) with Eq.(4.9) shows that, when Eq.(4.9) is satisfied \( T_{n+1}(\theta, \phi, t) \) is also a good approximation for a zero of \( F(T_n(\theta, \phi, t)) \). Initially it is surmised that to start the iteration, \( T_n \) must be the noisy data \( T_a(t) \) itself.

\[
T^0(\theta, \phi, t) = T_a(t)
\]

\[
T^{n+1}(\theta, \phi, t) = T_n(\theta, \phi, t) - C_n (T_a(t) - G * T_n(\theta, \phi, t))
\]

A drawback to this method (known as van Cittert's method [15]) is that it does not include the presence of noise and does not guarantee convergence with noisy data. So, the constant \( C_n \) will have to be chosen carefully.
Dittel who applied this method for the deconvolution of radiometric data [7] took the value for $C_n$ as:

$$C_n = C \left[ 1 - e^{-\frac{t}{\tau}} \left\{ \frac{T_n(\theta, \phi, t)}{V(\theta, \phi, t)} - \frac{(A(\theta, \phi, t) + B(\theta, \phi, t))}{\sigma(\theta, \phi, t)} \right\}^2 \right] (4.10)$$

Where $C$ is a constant, $A(\theta, \phi, t)$ and $B(\theta, \phi, t)$ are the lower and upper boundaries for $T_n(\theta, \phi, t)$, while $V(\theta, \phi, t)$ and $\sigma(\theta, \phi, t)$ are parameters that control the effect of stochastic events [7]. However, the method only gives satisfactory results with a proper selection of the values for $C$, $V(\theta, \phi, t)$ and $\sigma(\theta, \phi, t)$. Those values have to be determined on a trial-and-error basis.

### 4.3.2. Inversion by Maximum Likelihood

Another method is inversion by maximum likelihood; it uses the vicinity of the data point $T_1$ in order to calculate the most likely estimate of the value of $T_i$. The iteration process is derived using Bayes–theorem [16].

If $T_a$ and $T$ are written as the vectors $T_a$ and $T$ with elements $T_{ai} = T_a(t_i)$ and $T_i = T(t_i)$, it is possible to write:

$$P(T_i | T_a) = \frac{P(T_a | T_i)P(T_i)}{\sum_j P(T_a | T_j)P(T_j)} (4.11)$$

with $P$ denoting the probability.

Furthermore,

$$P(T_i) = \sum_k P(T_i | T_a)P(T_a) (4.12)$$

Inserting Eq. (4.11) into Eq. (4.12) gives:

$$P(T_i) = \sum_k \frac{P(T_a | T_i)P(T_i)P(T_a)}{\sum_j P(T_a | T_j)P(T_j)} (4.13)$$

This equations can be solved iteratively as follows:

$$P_{n+1}(T_i) = P_n(T_i) \sum_k \frac{P(T_a | T_i)P(T_a)}{\sum_j P(T_a | T_j)P_n(T_j)} (4.14)$$
This equation can be reduced to a more workable form [16] by putting:

\[ P(T_i) = \frac{T_i}{\Sigma T_i} \quad P(T_{ai}) = \frac{T_{ai}}{\Sigma T_{ai}} = \frac{T_{ai}}{\Sigma T_i} \]

since the process is conservative, and

\[ P(T_{i|T_{ak}}) = P(G_{ik}) = \frac{G_{ik}}{\Sigma G_{ik}} \]

This leads to:

\[ T_{i,n+1} = T_{i,n} \Sigma \frac{G_{ik} T_{ak}}{\Sigma G_{j,k} T_{j,n}} \]  \hspace{1cm} (4.15)

It should be noted that this method appeared to suffer from instabilities when abrupt intensity peaks were present [7]; therefore a correction factor had to be added to guarantee convergence [8] like with the multiple convolution.

4.3.3. Inversion by Matrix Inversion

The most transparent way to solve the inversion is by matrix inversion. However, direct inversion is very unstable and some kind of regularization has to be performed. This can be accomplished by applying the Twomey—Philips technique [8].

Twomey extended a method of Philips so that it used a controlled amount of smoothing for solving the matrix system produced, but no systematic method for determining the amount of smoothing required was given.

The solution to the inversion becomes:

\[ T = T_0 + (G^T G + \gamma I)^{-1} (G^T T_a - G^T G T_0) \]  \hspace{1cm} (4.16)

where \( T_0 \) is the initial guess, \( I \) is the identity matrix, and \( \gamma \) is a "stabilization term".

The advantage of the matrix inversion is the short computing time that it requires, because the solution can be found without iteration.

4.3.4. Other Inversion Techniques

Some other mathematical techniques have been proposed to overcome the problems of inversion. They include singular value decomposition [9], singular function decomposition [10], and the Fourier transform [11]. The solution, given by most of those techniques could
The Microwave Radiometer as a Remote Sensing Device

best be described as a regularization of the ill-posed problem. The regularization can be accomplished in different ways. One way being to approximate the true antenna pattern with another more appropriate pattern (this often includes neglecting the sidelobes). In [10] and [11] this was done by truncating a series expansion of the original antenna pattern (in [10], a series of eigenfunctions was used and in [11], a Fourier series). Another way is singular value decomposition of the matrix.

However, those ways have in common that basically the function of the regularization, is to ignore the contributions from the antenna pattern outside a central region (where the antenna pattern falls below a certain level), and to make the contributions equal to 0. These corresponding numerical small contributions lead to relatively large rounding errors during the process of deconvolution, giving the impression that the noise is being amplified.

4.3.5. Discussion

Each of the method discussed, whether regularized or not involves critical parameters whose "optimal" value is crucial for the accuracy of the inversion, which make the process to converge and the computer time to decrease. As a consequence, the methods require trial-and-error constants, or need the proper truncation of a series. In the next section, an inversion technique will be presented which is easy to implement, as well as robust and reliable. In the technique proposed, no fuzzy constants are present and no series need to be truncated. Furthermore, the method makes optimum use of the measurement data that is available.
References Sections 4.1 to 4.3


170

The Microwave Radiometer as a Remote Sensing Device

INVERSION TECHNIQUES FOR REMOTE SENSING OF ATMOSPHERIC PROFILES

COMPARISON OF LINEAR INVERSION METHODS BY ESTIMATION OF THE DUALITY BETWEEN ITERATIVE AND INVERSE MATRIX METHODS.
in: INVERSION METHODS IN ATMOSPHERIC REMOTE SOUNCING.

[15] Cittert, P.H. van,
ZUM EINFLUß DER SPALTREITE AUF DIE INTENSITÄTVERTEILUNG IN SPEKTRALLINIEN.

[16] Richardson, W.H.
BAYESIAN-BASED ITERATIVE METHOD OF IMAGE RESTORATION
Note: This section is a revised version of a paper by the author and a colleague that was published in the proceedings of the Specialist Meeting on Microwave Radiometry and Remote Sensing, Proc. 9th Int. Conf, Boulder, 14-16 January, 1992, p.229-236, ed. by E. Westwater. Therefore, the sequence of numbering for the Equations and references does not follow that used in the rest of this thesis.

4.4. Image Reconstruction Using a Passive Microwave Radiometer
by
Peter J. de Maagt* and Hennie G. ter Morsche**
Eindhoven University of Technology
PO BOX 513, 5600 MB Eindhoven, the Netherlands.
* Faculty of Electrical Engineering, Telecommunications Division
** Department of Mathematics and Computing Science.

Abstract
This paper presents an image reconstruction technique which is based on passive microwave radiometry. The method has been applied to various (noisy) input signals and the results are presented in order to illustrate the technique's performance. Special attention is paid to the effect of statistical and systematic errors (in both the measured antenna noise temperature and the antenna pattern) on the reconstruction.

1. Introduction
Passive remote sensing of the atmosphere is based on the natural electromagnetic noise emission of the atmosphere and, therefore, it is capable of providing relevant information about atmospheric processes and parameters (such as liquid water content, water vapour content, the air temperature gradient and so on). The use of passive remote sensing was originally restricted to IR-observations, but in recent years the use of microwave radiometry for this purpose has developed rapidly. This has been justified by the lesser extinction of microwaves in clouds as opposed to IR-observations of the cloudy atmosphere which are only capable of giving information of its upper (or lower) boundaries. Another advantage of microwave radiometry lies in the use of these frequencies for satellite communication. It is possible to obtain a prediction of the behaviour of communication links at those frequencies from the knowledge obtained from sensing the atmosphere. A major disadvantage of microwave radiometry is its substantially lower spatial resolution.
The Microwave Radiometer as a Remote Sensing Device

when compared with infrared systems.
A partial enhancement of spatial resolution can be obtained by exploiting the relative motion between the observed scene and the microwave radiometer system. Several techniques have been proposed to this end [1]–[5]. However, some methods use approximations, the validity of which may be questionable. The purpose of this paper is to show the results of research concerning the practical possibilities of spatial reconstruction where as few approximations as possible are made.
The paper starts with the relationship between the radiometer input and output; then the reconstruction technique is discussed, followed by several examples which clearly show the capability of this method for restoring spatial resolution.

II. Relationship between the Observed Scene and the Observation Instrument
Consider a radiometer system (antenna + receiver) which monitors a distributed target, referred to as a "scene". Due to the relative motion between the scene and the radiometer, a sequence of radiometric measurements can be taken for (partially) overlapping positions of the antenna footprints. The influence of the radiometer antenna pattern can be given by the following relationship:

\[ T_A(t) = \frac{1}{4\pi} \int T(t, \theta, \phi) G(\theta, \phi) d\Omega \]  

where \( T_A \) is the antenna noise temperature, \( T(t, \theta, \phi) \) is the brightness temperature of the scene and \( G(\theta, \phi) \) is the radiometer antenna pattern.
Because of the finite dimensions of the antenna, which results in a non-zero beamwidth, the measured antenna temperature may differ significantly from the brightness temperature of the scene. It is better to say that the measured temperature represents a spatially filtered brightness temperature. Furthermore, the radiometer receiver integrates the signal \( T_A \) in time, and \( T_A \) becomes:

\[ T_a(t) = \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} T_A(\xi) d\xi \]  

where \( \tau \) is the integration time of the radiometer receiver.
Hence, the image of the scene is altered due to the integration with the radiometer pattern characteristics (spatial integration of the antenna and time integration of the receiver).
Furthermore, the signal \( T_a \) is obscured by the addition of both systematic and statistical changes by the electronic components of the detection equipment. If a more detailed image of the scene is needed, it is necessary to eliminate those influences (or at least to reduce them) leading to inverse methods in microwave imaging. However, the problem inherent to
Radiometric observation is that inversion is ill-posed and the solution cannot be improved by increasing the accuracy and efficiency of the numerical algorithms used in the inversion process. Luckily, inversion can sometimes be facilitated by assuming knowledge about the object and data properties.

III. Modelling the Relationship between the Observed Scene and the Antenna Pattern

In order to be able to perform inversion the following model can be used. The basis for the modelling is provided by Eq.(1). Firstly, the scene is projected on a plane given by the vectors \( \mathbf{x} \) and \( \mathbf{v} \), where \( \mathbf{x} \) represents the direction of the movement, and represented by a noise distribution \( T(x,y) \).

![Figure 1. The coordinate systems.](image)

Fig.1 shows the coordinates used. Due to the fact that measurements can be taken at different relative positions between the antenna contours and the scene, it is possible to estimate \( T(x,y) \) in the direction of the movement and to gain knowledge about the boundaries of the scene. At Eindhoven University of Technology (EUT) a non-scanning antenna is part of the radiometer system; therefore, the antenna "smoothing" effect cannot be canceled in the direction perpendicular to the direction of the movement and the problem is converted to a one dimensional one in order to simplify the calculations. If a scanning antenna had been used this simplification would not have occurred.

Assuming that the movement, with constant \( v \) along the \( x \)-axis is linear in the time
The Microwave Radiometer as a Remote Sensing Device

interval of measurements it is possible to write:

\[ T_a(t) = \frac{1}{\pi} \int_{-1}^{1} \left[ \frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{T(x+\xi,y)}{R^2(x,y)} G(x,y) \, dx \, dy \right] d\xi = \]

\[ = \frac{1}{\pi} \int_{-1}^{1} \left[ \frac{1}{4\pi} \int_{-\infty}^{\infty} T(x,y) \frac{G(x+\xi,y)}{R^2(x,y)} \, dx \, dy \right] d\xi \]

\[ = \frac{1}{\pi} \int_{-1}^{1} \left[ \frac{1}{4\pi} \int_{-\infty}^{\infty} T(x,Y) \frac{G(x+\xi,y)}{R^2(x,y)} \, dx \, dy \right] d\xi \]

\[ T_a(t) \approx \frac{1}{4\pi} \int_{-\infty}^{\infty} T'(x)G'(x,t,\tau) \, dx \]

with \( T'(x) = T(x,Y) \) and \( G'(x,t,\tau) = \frac{1}{\pi} \int_{-1}^{1} \left[ \frac{\omega G(x+\xi,y)}{R^2(x,y)} \right] d\xi \)

where \( T(x,Y) \) is an average value with respect to the \( y \)-coordinate and which is assumed to be frozen for the period \( \tau \), needed to obtain \( T_a(t) \).

IV. Reconstruction

The problem of determining \( T'(x) \) can now be dealt with in various ways [1],[2],[3],[4] but it can be visualized as in figure 2.

Figure 2. The projection of the set \( T'(x) \) to \( T_a \).
Due to the fact that the problem is ill-posed, one measured $T_a$ corresponds to a set of $T'(x)$ and it is necessary to develop an inversion which gives an unique $T'(x)$ as solution to the problem. Different inversion methods will result in different solutions $T'(x)$. However, as a radiometer system is involved, it is rational to treat the imaging problem from the radiometric point of view. It is possible to draw a parallel with the traditional use of the radiometer for accurate measurement of $T_a$ as a function of time. In that case, the objective will be to reconstruct the signal $T_a$ in time from a noisy measurement and to get the best possible estimate of the average of the antenna brightness temperature. If $T_a$ is written as follows:

$$T_a(t) = <T_a(t)> + \delta T_a(t)$$
(6)

where $<>$ indicates ensemble averaging, the objective is to minimize the time variations $\delta T_a(t)$. With an imaging radiometer, where the objective is to reconstruct the signal $T'(x) = <T'(x)> + \delta T'(x)$, it is therefore reasonable to try and get the best possible estimate for the average of $T'(x)$, while minimizing the spatial variations $\delta T'(x)$.

This can be accomplished by minimizing:

$$\int_{-\infty}^{\infty} T'^2(x) dx = <T'(x)>'^2 + \int_{-\infty}^{\infty} \delta T'(x)^2 dx$$
(7)

with the constraint that $T'$ has to satisfy the measurements given by Eq.(5).

Using the following formulation:

$$T_{ai} = \frac{1}{T} \int_{\frac{t_i}{2}}^{\frac{t_i + \tau}{2}} T_A(\xi) d\xi = \frac{1}{4\pi} \int_{-\infty}^{\infty} T'(x) G_i'(x) dx$$
(8)

with $G_i'(x) = \frac{1}{\tau} \int_{\frac{t_i}{2}}^{\frac{t_i + \tau}{2}} \left[ \int_{-\infty}^{\infty} \frac{G(x + v_x \xi, y) dy}{R^2(x, y)} \right] d\xi$

it is possible to write the problem as: reconstruct $T'(x)$ so that it satisfies

$$T_{ai} = \frac{1}{4\pi} \int_{-\infty}^{\infty} T'(x) G_i'(x) dx \; (1 \leq i \leq N) \; \text{and minimizes} \; \int_{-\infty}^{\infty} T'^2(x) dx$$
(9)

Eq.(9) can be seen as a dot product in a function space, so that the problem can be written as:
The Microwave Radiometer as a Remote Sensing Device

\[ T_{ai} = \frac{1}{4\pi} G_{i} \cdot T' \]

and

\[ \min\{|T' \cdot T'|\} = \min\{| |T'| | \} \]  \hspace{1cm} (10)

where \( |T'| | \) represents the "length" of \( T \) in the function space.

In Appendix A it is shown that the solution to this problem can be written as (for more detail see [6]):

\[ T' = \sum_{i=1}^{N} a_i G_i' \]  \hspace{1cm} (11)

So, \( T_{ai} = \frac{1}{4\pi} G_{i}' \cdot T' \Rightarrow T_a = (\sum_{j=1}^{N} a_j G_j') \Rightarrow T_{ai} = (\frac{1}{4\pi}) G_{i}' \cdot \mathbf{G} \mathbf{T}_a \Rightarrow \mathbf{T}_a = (\frac{1}{4\pi}) \mathbf{G} \cdot \mathbf{T}_a \]  \hspace{1cm} (12)

or,

\[ \mathbf{T}_a = \mathbf{G}_a \]  \hspace{1cm} (13)

with \( \mathbf{T}_a = \begin{bmatrix} \mathbf{T}_{a_1} & \mathbf{T}_{a_2} & \cdots & \mathbf{T}_{a_N} \end{bmatrix} \), \( \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} \), \( \mathbf{G} = \begin{bmatrix} G_{1}' \\ G_{2}' \\ \vdots \\ G_{N}' \end{bmatrix} \) and

\( \mathbf{G} \) a \( N \times N \) matrix with elements \( G_{ij} = \int_{-\infty}^{\infty} G_{i}'(x) G_{j}'(x) \, dx. \)

It is simple to determine the reconstruction vector \( \mathbf{a} \) from:

\[ \mathbf{a} = \mathbf{G}^{-1} \mathbf{T}_a \]  \hspace{1cm} (14)

V. Trade-off between Resolution and Noise

The above reconstruction technique enables \( T' \) to be written as a summation of shifted versions of the transformed antenna pattern \( G' \). As an example, the process of reconstruction is illustrated in figure 3.

The pattern of \( G' \) is given in figure 3.b and \( G_{i}'(x) \) in figure 3.c, if \( N=4 \), \( a_1=0.3 \), \( a_2=0.5 \), \( a_3=0.4 \), and \( a_4 = 0.2 \), respectively. \( T' \) will be as given in figure 3.d.

In terms of resolution, it is desirable to have as many slightly shifted \( G_i' \)'s as possible; then, it is possible to reconstruct \( T' \) with all its details. However, calculating \( \mathbf{a} \) in such a case can lead to unacceptable errors in the reconstruction. If \( G_{3}' \) is a slightly shifted version of \( G_{i}' \)
there will be a marked interdependence between the rows and columns of $G$, resulting in a large condition number of $G$. Since $T_a$ is contaminated with noise from the radiometer receiver, the process of determining $a$ can become unstable. For improving the condition number, it is necessary to reduce the interdependence, which means that there has to be as little overlap as possible. Therefore, the reconstruction of $T'$ needs a trade-off between the resolution and the accuracy of reconstruction.

VI. Measurements Contaminated with Noise

Statistically errors due to the radiometer receiver

The method explained in the previous sections selects a $T'$ that satisfies Eq.(10), but it neglects a-priori knowledge of the existence of measurement errors in the signal $T_a$. One contributor to the measurement error is the radiometer receiver which introduces a statistical error equal to the radiometer resolution $\epsilon$. In practice, random errors in the antenna pattern can be neglected, because their influence is minimal, due both to spatial and time integrations of the pattern. Therefore, the problem, as stated in Eq.(10), is more realistically represented by:

$$\left\| \left( \frac{1}{4\pi} \right) G \cdot T' - T_a \right\| \leq \epsilon \quad \text{and} \quad \min \left\{ \left\| T' \right\| \right\}$$

and figure 2 changes into figure 4.

As the set $T'(x)$ becomes larger, it becomes possible to find a smoother (a more "well-behaved") solution to the ill-posed problem. In a way similar to the method described in Appendix A, it is possible to show that the solution to this problem can again be written as a linear combination of antenna patterns, so using Eq.(11) the problem
becomes:
\[ | | G\alpha - T_a | | \leq \epsilon \text{ and } \min \{ 4\pi (\alpha^T G\alpha) \} \]  
(16)

In Appendix B, it is shown that this problem is equivalent to:
\[ | | G\alpha - T_a | | = \epsilon \text{ and } \min \{ 4\pi (\alpha^T G\alpha) \} \]  
(17)

With the help of Lagrange multipliers, it is possible to show that the solution to this problem has to satisfy:
\[ 2G\alpha = \text{grad}(\alpha^T G\alpha) = \alpha \text{grad}( | | G\alpha - T_a | | ) = \alpha(2G^T G\alpha - 2G T_a) \]  
(18)

Noting that \( G \) is positive definite and \( G^T = G \), it follows that
\[ \tilde{\alpha} = \alpha(G\alpha - T_a) \]  
(19)

then, the problem becomes:
\[ | | G\tilde{\alpha} - T_a | | = \epsilon \text{ and } \tilde{\alpha} = \alpha(G\tilde{\alpha} - T_a) \]  
(20)

From Eq.(20) it follows that:
\[ G\tilde{\alpha} = T_a - \epsilon z \text{ with } | | z | | = 1 \]  
(21)

Substituting Eq.(21) in Eq.(20) gives:
\[ \tilde{\alpha} = \alpha(T_a - \epsilon z - T_a) = -\alpha \epsilon z \]  
(22)

or,
\[ \alpha \epsilon z = \alpha(G\epsilon z + T_a) \Rightarrow \epsilon(1-aG)z = T_a \Rightarrow (1-aG)^{-1}T_a = \epsilon z \]  
(23)

\[ \Rightarrow | | (1-aG)^{-1}T_a | | = \epsilon \]  
(24)

So, after determining \( a \) from Eq.(24), it is possible to derive \( \tilde{\alpha} \) from Eq.(19).

As the problem of determining \( a \) is no longer linear, the solution is not straightforward. Firstly, the matrix \( G \) is diagonalized and, secondly, \( \tilde{\alpha} \) and \( T_a \) are decomposed into
eigenvectors of the matrix $G$.

$$G = BDB^T, \quad \mathbf{a} \equiv \sum_{i=1}^{N} \xi_i \mathbf{e}_i, \quad T_n = \sum_{i=1}^{N} \tau_{ni} \mathbf{e}_i$$

(25)

with $D$ a diagonal matrix with elements given by the eigenvalues $(\lambda_1, \lambda_2, ..., \lambda_N)$ of the matrix $G$ and $B$ an orthogonal matrix whose columns are formed by the eigenvectors $(\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N)$ of $G$.

The Eqs.(19 and 24) now take the following form:

$$\xi_i = a (\lambda_i \xi_i - \tau_{ai})$$

(26)

$$\frac{\sum (\frac{-\tau_{ai}}{\lambda_i})^2 = \epsilon^2}{i=1}$$

(27)

That is a polynomial form in $a$, so that more solutions will be found. To determine which $a$ is correct, the first part of Eq.(16) has to be written as follows:

$$\frac{\sum (\lambda_i \xi_i - \tau_{ai})^2 = \epsilon^2}{i=1}$$

(28)

This shows that Eq.(16) is an ellipsoid. Using the sketch shown in figure 5 where $n$ represents the normal of the ellipse, it can be seen that $a$ corresponding to the minimum length will be negative. It is clear that there is only one negative solution to Eq.(27). So, checking the signs will indicate the correct $a$. Substituting $a$, obtained from Eq.(27), into Eq.(26) and the outcome into Eq.(25) gives the coefficients $a_i$ and the solution of the original inversion problem.

In that way a method has been obtained for regularizing the ill-posed problem. Applying the same method described for the case above where errors were neglected (section 4) would have led to:

$$\mathbf{a} = \sum_{i=1}^{N} \frac{\tau_{ai}}{\lambda_i} \mathbf{e}_i$$

(29)

Now, the solution takes the form of:

$$\mathbf{a} = \sum_{i=1}^{N} \frac{-a \tau_{ai}}{(1-a \lambda_i)} \mathbf{e}_i$$

(30)

A small value for $\lambda_i$ in Eq.(29) can lead to unacceptable errors in the reconstruction, while in Eq.(30), where a limited error is accepted, this will have much less influence.
Figure 5. The determination of the correct $a$.

**Systematic errors due to the radiometer antenna**

In the previous section the statistical error caused by a radiometer receiver was discussed. Systematic errors from the receiver were neglected because they can be seen as an offset voltage which can be compensated by calibration. Systematic errors originating from the finite measurement accuracy with which the antenna pattern is produced can be shown to have little influence. This is shown in detail in Appendix C, but, as can be seen from Eqs.(19) and (24) the problem of determining $a$ and $a$ will be well conditioned if the matrix $G$ has a good condition number. So, the influence of systematic errors on the inversion method is less prominent than the influence of statistical errors.

**VII. Performance of the Reconstruction Algorithm**

A skeletal structure for the reconstruction algorithm has been given in the previous sections. In this section, the results obtained with the algorithm are presented. The antenna pattern that will be used to get a first impression of the performance is the pattern of a front-fed parabolic antenna which can be modelled adequately by a function of the following form:

$$G(\theta) = G_0 \left[ \frac{\mathcal{J}_1(a\theta)}{a\theta} \right]^2 \frac{1}{1+(a\theta)^b}$$

(see figure 6.a) where $a$ and $b$ determine the beamwidth and level of the first sidelobe, respectively.
That pattern is transformed into the $G(x)$ pattern (the pattern after integration over the $y$-coordinate, see figure 6.b) followed by time averaging as given by Eq.(5) (see figure 6.b) resulting in the $G'(x)$ pattern. For the examples below, a configuration has been used with $a=2.6$ and $b=0$ which is an approximation of a paraboloid of 50 $\lambda$ diameter with a first sidelobe at $-17.6$ dB. The pattern gives a footprint of 47.1 m at a height of 0.5 km if the elevation angle is $90^\circ$. The integration time for the receiver is set at 1 sec, which is representative for actual radiometers. The results are shown for different values of the resolution, which is defined as the dimension of variation that can be reconstructed (refer to section 5).

Figure 6. a) the function $G$ as given by Eq.(31)
b) $G(x)$ (= $G(x,y)$ averaged over $y$) and $G'(x)$ (= $G(x+vt)$ averaged over time)

Examples

The first example input–object is a Dirac $\delta$–pulse; it can be seen from figure 7 that the reconstructed signal closely resembles a sinc–signal, with its first zero at a distance nearly equal to the resolution. That implies that the algorithm must be a nearly rectangular spatial transfer function.

A good example of the gain in resolution which can be obtained with this method is shown
in the second example input–object. Figure 8 shows what the result will be if the resolution is set to 5m and 15m, respectively. The figure clearly shows the increase in resolution that can be obtained with this method.

The Influence of measurement errors in $T_a$

As stated previously, measurement errors due to the radiometric receiver can result in a completely erroneous reconstruction of the object. In the example below, the signal $T_a$ is obscured by random noise of 0.1 K (equal to the radiometer's sensitivity). If the reconstruction was based on Eq.(15) and $\epsilon$ taken as $10^{-8}$ (so defining the corrupted signal as exact) the reconstruction would fail (see figure 9). If, however, $\epsilon$ would be taken as equal to 0.1 K, the reconstruction would assure good results, even with noisy data.

![Figure 7. A Dirac $\delta$–pulse input signal.](image-url)
Remote Sensing by Passive Imaging

Figure 8. A combination of triangles as input signal.

Figure 9. A combination of triangles as input signal.
The influence of measurements errors in $G$

Another error that was simulated was a systematic error in $G$. The signal $T_s$ was determined with $G$ as given in Eq.(31); while the reconstruction is performed with a $G$ where the parameters $a$ and $b$ are altered or the shape of the main beam is altered. This error could simulate an antenna pattern mismatch resulting from an incorrect measurement in an antenna test range. First, reconstruction was performed with the approximated pattern shown in figure 10; the effect on the reconstructed image is illustrated in figure 11. Figure 12 shows the effect of uncertainty on the parameters $a$ and $b$, demonstrating that the reconstruction technique is rather insensitive to changes in the parameter $a$, i.e. $\theta$-scaling of the pattern. As expected, the value of $b$, representing the sidelobe level, has a greater effect. Here $b=1$ corresponded to a sidelobe level of $-22.4$ dB and $b=15$ to a sidelobe level of $<-70$ dB.

Figure 10. The pattern as given by Eq.(31) (solid) and the approximated main beam (dashed).
Figure 11. $T_a$ determined with $G$ as given by Eq.(31). The reconstruction is performed with the pattern as given in figure 10.

Figure 12. $T_a$ determined with $a=2.6$, $b=0$. While the reconstruction is performed with various combinations of $a$ and $b$. 
VIII. Conclusions
A microwave radiometer inversion method has been presented. The method has been applied to various (noisy) input signals and results have been presented which show the capability of this method for restoring spatial resolution. The effect of statistical errors in the antenna pattern on the results is reduced by spatial and time integration (averaging). The errors in the results due to systematic errors in $G$ were shown to be of the same order of magnitude. The method presented here appeared to have a nearly rectangular spatial transfer function.

IX. Acknowledgements
The authors wish to thank dr. ir. M.H.A.J. Herben and prof. dr. ir. G. Brussaard for their valuable help and discussions.
References:

[1] Stogryn, A.
Estimates of brightness temperatures from scanning radiometer data.

[2] Dittel, R.H.
Deconvolution of microwave radiometry data.
Inverse methods in Electromagnetic Imaging, Part 1, Ed. by W.M. Boerner

[3] Nordius, H.
Physical inversion by means of singular value decomposition of a
ground-based temperature profiling microwave radiometer.
Microwave radiometry and remote sensing applications, Ed. by P. Pamploni,

Antenna pattern correction in scanning radiometry: a singular system
analysis.

Application of Fourier transforms for microwave radiometric inversions.

A spatial reconstruction technique applicable to microwave radiometry.
Eindhoven: Faculty of Electrical Engineering, Eindhoven University of
Appendix A: Writing the Solution as a Combination of Shifted Antenna Patterns

In section IV it was stated that the solution to the problem

\[ G_i \cdot T = T_{ai} \quad i = (1, 2, \ldots, N) \quad \text{and} \quad \min | | T' || \]  

resulted in a \( T'(x) \) which could be written as a linear combination of \( G_i(x) \). The mathematical derivation will be given in this appendix.

Suppose that all \( T's \) that satisfy Eq.(A.1) form a space \( V \) given by:

\[ V = \{ T' | \ G_i \cdot T' = T_{ai} \} \]  

and that \( V' \) is a linear manifold (a shifted version of \( V \) such that it contains the origin):

\[ V' = \{ v | \ G_i \cdot v = 0 \} \]  

and that the space \( W \) perpendicular to \( V' \) is given by:

\[ W = \{ w | \ w \cdot v = 0, \ \forall v \in V' \} \]

If \( T_0 \) is the \( T' \) in \( V \) with the minimum length, the following holds:

\[ | | T_0 || \leq | | T' || \quad (T' \in V) \quad \text{and} \quad T' = T_0 + \gamma v \quad (\gamma \in \mathbb{R}, v \in V') \]  

\( T' \) is an element of \( V \) because:

\[ G_i \cdot T' = G_i \cdot T_0 + \gamma G_i \cdot v = G_i \cdot T_0 = T_{ai} \]  

It remains to prove that if \( T_0 \) satisfies Eq.(A.5) it can be written as a linear combination of \( G_i \). Eq.(A.5) holds for all \( \lambda \in \mathbb{R}, v \in V' \) if:

\[ T_0 \cdot T_0 \leq (T_0 + \gamma v) \cdot (T_0 + \gamma v) = \gamma^2 v \cdot v + 2\gamma T_0 \cdot v + T_0 \cdot T_0 \]

\[ \Rightarrow 2\gamma v \cdot v + 2\gamma T_0 \cdot v \geq 0 \]  

This can be accomplished only if \( 2\gamma T_0 \cdot v = 0 \). So \( T_0 \) has to be an element of \( W \). Combining Eqs.(A.3) and (A.4) shows that

\[ T_0 = \sum_{i=1}^{N} a_i G_i \]  

(A.9)
Appendix B : The Optimization Problem Including Measurement Errors

The problem is stated as:

\[ \| \mathbf{G} \mathbf{a} - \mathbf{T} \| \leq \varepsilon \]  \hspace{1cm} (B.1.a)

and \( \min 4\pi (\mathbf{a}^T \mathbf{G} \mathbf{a}) \)  \hspace{1cm} (B.1.b)

Suppose that the vector \( \mathbf{b} \) minimizes (B.1.b) and \( \| \mathbf{G} \mathbf{b} - \mathbf{T} \| < \varepsilon^2 \)

Let \( \mathbf{a} = \gamma \mathbf{b} \)  \hspace{1cm} (B.2)

The left hand side of Eq.(B.1.a) can be written as:

\[ \| \gamma \mathbf{G} \mathbf{b} - \mathbf{T} \| = \| \mathbf{G} \mathbf{b} - \mathbf{T} + (\gamma - 1) \mathbf{G} \mathbf{b} \| \]

Applying the Cauchy inequality, leads to:

\[ \| \mathbf{G} \mathbf{b} - \mathbf{T} + (\gamma - 1) \mathbf{G} \mathbf{b} \| \leq \| \mathbf{G} \mathbf{b} - \mathbf{T} \| + \| (\gamma - 1) \mathbf{G} \mathbf{b} \| \]  \hspace{1cm} (B.3)

If \( \gamma < 1 \) but nearly equals 1, it is possible that Eq(B.3) will still satisfy Eq.(B.1.a) and:

\[ \| \mathbf{G} \mathbf{b} - \mathbf{T} \| + \| (\gamma - 1) \mathbf{G} \mathbf{b} \| \leq \varepsilon \]

However

\[ \mathbf{a}^T \mathbf{G} \mathbf{a} = \gamma^2 \mathbf{b}^T \mathbf{G} \mathbf{b} < \mathbf{b}^T \mathbf{G} \mathbf{b} \]  (because \( \lambda < 1 \))

which contradicts the fact that \( \mathbf{b} \) was the minimum.

So if \( \mathbf{b} \) satisfies Eq.(B.1.b) then it also satisfies

\[ \| \mathbf{G} \mathbf{b} - \mathbf{T} \| = \varepsilon \]

So, the problem stated in Eq.(B.1) is equivalent to the one where the \( \leq \) sign is replaced by the \( = \) sign.
Appendix C: Influences of Measurement Errors in $G$

Again the problem is stated as in Eq. (16). With the help of the Lagrange Multiplier method it follows that:

$$2G\delta = a (2G^T G\delta - 2G T_a)$$  \hspace{1cm} (C.1)

Because $G$ is positive definite and $G^T = G$, Eq. (16) becomes:

$$||G\delta - T_a|| = \epsilon \text{ and } \delta = a(G\delta - T_a)$$  \hspace{1cm} (C.2)

It is interesting to note that because $a$ is a non-zero vector, $a$ is proportional to $
abla^T = \frac{1}{||G\delta - T_a||} a/\epsilon$. So, $a$ must be a large number.

The influence of errors in the matrix $G$ (originating from the errors in the antenna pattern $G_1(\chi)$) on the inversion can be determined as follows:

$$a + \delta a = (a + \delta a)((G + \delta G)(a + \delta a) - T_a)$$  \hspace{1cm} (C.3)

Neglecting higher order terms gives:

$$\delta a = (I - aG)^{-1}(G_2 - T_a) \delta a + (I - aG)^{-1} \delta G a$$  \hspace{1cm} (C.4)

$$\delta a = (I - aG)^{-1} \frac{a}{a} \delta a + (\frac{1}{a} - G)^{-1} \delta G a$$  \hspace{1cm} (C.5)

$$\delta a = (I - aG)^{-1} \frac{a}{a} \delta a + (\frac{1}{a} - G)^{-1} \delta G a$$  \hspace{1cm} (C.6)

Where $(\frac{1}{a} - G)^{-1}$ is nearly equal to $G^{-1}$ because $a$ is large and $(I - aG)^{-1}$ causes no problems because $a$ is negative for the minimum.

$$||\delta a|| \leq \frac{\epsilon}{||G^{-1}||} \frac{||\delta a||}{||a||} + \frac{||a^T - G||}{||a||} \frac{||\delta G||}{||G||}$$  \hspace{1cm} (C.7)

$$\frac{||\delta a||}{||a||} = a \mathcal{O}(||\delta G||)$$  \hspace{1cm} (C.8)

Consequently, this error can be made arbitrary small by putting effort in obtaining exact knowledge of the antenna's pattern.
4.5. The Influence of Measurement Errors in the Object Velocity on the Reconstruction

The previous section discussed a new image reconstruction technique that has been applied to various (noisy) input signals and the results are presented in order to illustrate the technique's performance. Special attention has been given to the effect of statistical and systematic errors on the reconstruction, for both measured antenna noise temperature and antenna pattern.

Although, the error analysis in the previous section was quite comprehensive, errors in the reconstruction, resulting from an incorrect estimation of the object's velocity, were not dealt with. It is easy to show, however, that those errors could have been treated in a similar way as the errors in the matrix G (see appendix C of section 4.4). The derivation starts by considering an error in \( v_x \) given by \( \delta v_x \). Given Eq.(8) of Section 4.4, this will result in a \( \delta G \) proportional to:

\[
\delta G_i'(x) = \left[ \int_{\frac{t_1}{2}}^{\frac{t_1 + \tau}{2}} \int_{-\infty}^{\infty} \frac{\partial G(x + v_x \xi, y)}{\partial x} \xi \, dy \, \delta x \right] \delta v_x
\]

where \( \partial \) denotes the partial derivative. Since, the partial derivative of G is bounded, the following holds:

\[
\delta G_i'(x) = \mathcal{O}(\left| \int_{t_1}^{t_1 + \tau} |\xi| d\xi \right| \delta v_x)
\]

\[
\Rightarrow \delta G_i'(x) = \mathcal{O}(|t_1| \delta v_x)
\]

Eq.(13) of Section 4.4 shows that this would lead to an error in the matrix G and in appendix C of Section 4.4 it is shown how to deal with such errors.

Due to the systematic character of the errors, all elements of the vector \( \mathbf{a} \) will be over- or underestimated according to the sign of \( \delta v_x \). Therefore, the accuracy of the absolute value of the image will become uncertain, but its shape will be only marginally affected.

It is interesting to note that the error in G will increase if the total observation time increases. This is trivial if the relation \( x = vt \) is considered.
4.6. The Transfer Function of the Algorithm

A spin-off from section 4.4 that has been considered is the transfer function of the algorithm. Figure 7 in Section 4.4 shows that the reconstructed signal will resemble a sinc signal, if a delta pulse is applied to the algorithm. This implies that the spatial transfer function is nearly rectangular. A way of determining the transfer function is by using a sine-shaped input signal. Figure 4.1 shows that the reconstructed signal will nearly equal the input signal for spatial frequencies lower than 1/resolution. That is to be expected when the sine is a eigenfunction of the reconstruction algorithm. To determine the complete transfer function, the spatial frequency of the sine-shaped input signal, shown in figure 4.1, has to be varied. The resulting transfer function is shown in figure 4.2. The figure demonstrates that an increased resolution will improve the approximation of a rectangular transfer function which is to be expected because higher resolution assumes more slightly shifted versions of the original antenna pattern.

Fig.4.1. A sine-shaped input signal with spatial frequency 0.01 m\(^{-1}\). The resolution is 5m.
Fig. 4.2. The transfer function of the reconstruction algorithm.
4.7. Influence of Shaping the Main Beam on the Trade-off between Resolution and Accuracy of Reconstruction.

Figure 10 of Section 4.4. serves as an introduction to the investigation described in this section. In section 4.4, the performance of the reconstruction technique was discussed using the pattern of a front-fed parabolic antenna modelled by:

\[ G(\theta) = G_0 \left[ \frac{J_0(a\theta)}{a\theta} \right]^2 \frac{1}{1+(a\theta)^b} \]  \hspace{1cm} (4.20)

where \( a \) and \( b \) determine the beamwidth and level of first sidelobe, respectively and \( G_0 \) is a normalization constant.

In this section, the results obtained with the pattern of a defocused parabolic reflector antenna (see figure 4.3) are given. For the sake of simplicity, the main lobe was approximated by an idealized pattern as given in figure 4.3.b. The reason for that was purely academic, because it made it easier to see the effect of the values of \( \theta_a \) and \( \theta_b \) (see figure 4.3.b) on the results.

The defocused pattern has been included because it was possible that shaping of the main lobe to a more rectangular pattern could affect the trade-off between the resolution and accuracy of reconstruction (see Section 4.4.V).

For the following examples, \( \theta_a \) and \( \theta_b \) were chosen to approximate a defocused version of the antenna as given in section 4.4; where, the feed was axially displaced by a distance \( 2\lambda \) from the focal point of the antenna towards the reflector. Their values were \( \theta_a = 1^\circ \) and \( \theta_b = 3^\circ \). The corresponding patterns \( G(x) \) and \( G'(x) \) are shown in figure 4.4.

To facilitate comparison between the results obtained with the pattern of a focused and a defocused antenna system, the same input signals as in section 4.4 are considered.

In the first example, the input-object was again a Dirac \( \delta \)-pulse. As can be seen in figure 4.5 the reconstructed signal resembles a sinc-signal, with its first zero at a distance equal to the resolution.
Fig. 4.3.  

a) The pattern of a defocused parabolic reflector antenna system ($D/\lambda=73.3$ and $F/D = 0.4$).

b) The approximated main beam.

Fig. 4.4. $G(x)$ ($=G(x,y)$ averaged over $y$) and $G'(x)$ ($=G(x)$ averaged over time).
Fig. 4.5. A Dirac δ—pulse input signal.

In the second example, the input—object was again a combination of triangles. The results are shown for a reconstruction that was made with resolutions of 5 and 15 m, respectively. Again the gain in resolution is clearly visible.

Fig. 4.6. A combination of triangles as input signal.
4.7.1. The Influence of Measurement Errors in $T_a$

The signal $T_a$ could be obscured in the same way as that described in section 4.4. If $\epsilon$ is made equal to the radiometer sensitivity, the reconstruction will ensure good results even with noisy data and a defocused antenna pattern (see figure 4.7).

![Figure 4.7](image)

Fig. 4.7. A combination of triangles as input signal.

4.7.2. The Influence of Measurement Errors in $G$

The convolution was performed with $G$ as defined in section 4.6, but the reconstruction was performed with $G$ where the parameters $\theta_a$ and $\theta_b$ are altered. Only this systematic error was considered, because the influence of statistical errors could be neglected. Figure 4.8 and 4.9 show the effects of various combinations of $\theta_a$ and $\theta_b$.

As can be seen from the figures, the reconstruction was rather insensitive for small mismatches with respect to the shape of the main beam.
The Microwave Radiometer as a Remote Sensing Device

Fig. 4.8. \( T_a \) determined with \( \theta_a=1^\circ, \theta_b=3^\circ \). The reconstruction is performed with \( \theta_a=0.9 \times 1^\circ \) and \( 1.1 \times 1^\circ \), while \( \theta_b=3^\circ \).

Fig. 4.9. \( T_a \) determined with \( \theta_a=1^\circ, \theta_b=3^\circ \). The reconstruction is performed with \( \theta_b=0.9 \times 3^\circ \) and \( 1.1 \times 3^\circ \), while \( \theta_a=1^\circ \).
4.7.3. Comparing the Results Obtained with a Focused and Defocused Antenna System

In the previous sections, tests on the reconstruction technique were presented, using different input objects. Furthermore, the performance of the technique was demonstrated when using a focused and a defocused antenna pattern. As mentioned before, the reason for including a defocused antenna pattern was the possibility that shaping of the main lobe could affect the trade-off between resolution and accuracy of reconstruction. In order to be able to make a quantitative comparison, a square error was calculated as follows:

\[
\varepsilon_{sq} = \frac{\sum_{i=1}^{N} [T(x_i) - T'(x_i)]^2}{\sum_{i=1}^{N} [T(x_i)]^2} \cdot \frac{\text{length of image of } T}{N} \quad (4.21)
\]

Where \( T \) is the input signal (in this case the input signal was equal to the one used in figure 4.7), and \( T' \) the reconstructed signal. The error can be treated as a sort of signal to noise ratio \((S/N)^{-1}\).

The error has been calculated for the pattern given by Eq.(4.20) in section 4.4., as well as for different defocused patterns as a function of resolution. The result is presented in Figure 4.10. It can be seen from this figure, that the results obtained with a defocused antenna system have larger errors than those obtained with a focused antenna system.

So, defocussing the antenna system does not appear to offer any advantages.

![Graph showing square error as a function of resolution for different antenna patterns](image)

Fig. 4.10. Square error as a function of resolution for different antenna patterns (focused and defocused).
4.8. Conclusions

In this chapter, a numerical deconvolution technique has been described which is able to process data from a passive microwave radiometer. When applied to a radiometer system, such a technique ensures a performance that is equivalent to that of a radiometer with an antenna of larger dimensions. Consequently, it is possible to extend the performance of a radiometer's antenna beyond its physical dimensions which are often the limiting factors that control the spatial resolution of a radiometer system in practice.

Extensive tests have indicated that a tenfold increase in resolution can be attained. Although other deconvolution techniques exist, the new technique performs quite satisfactory when confronted with the effects of receiver noise. The proposed algorithm will ensure good results even with noisy data, when most deconvolution techniques will fail. The reason is that the method makes use of a-priori knowledge of the existence of the radiometric receiver's uncertainty.

Another benefit of this technique is that it makes optimal use of the available measurement data. More data guarantees a higher resolution, while the resolution is limited by the amount of data available.

When assessing the performance of the deconvolution process, it has to be borne in mind that the measurement device is a radiometer and that generally involves some kind of time integration. A study of removing the time integration was not included, although it could be possible that there is an optimal integration time and an integration law which is related to the spatial integration of the antenna.

As an overall conclusion, it can be stated that the new proposed method is robust, reliable and contains no fuzzy constants.
5. SUMMARY AND CONCLUSIONS

Many radiometer systems incorporate an antenna which originated as part of a communications system, however, the specific design criteria for a remote sensing antenna are not compatible with those for a communications system antenna.

The number of relevant radiometer parameters is large. Due to the diversity of the parameters and due to the various ways to combine them, it is difficult to find uniformity in the design proposals for radiometer antennas. However, there is a consensus that the commonest design parameters are beam efficiency and integrated pattern function $h$. Their quantitative values of these differ from proposal to proposal and it is better to say that mostly they are specified to be as "good as practicable possible". A pragmatic design approach does not exist yet.

In chapter 2, an optimization method is presented that is able to deal with different radiometer antenna parameters simultaneously, with and without constraints to the (integrated) pattern. That optimization method results in a theoretical optimum which helps evaluation of each practical design.

In order to produce general functional and operational specifications, the interaction of the antenna with its surroundings was modelled. The model described as many of the foreseeable geometrical and meteorological situations as possible. The dynamic surroundings of the antenna lead to uncertainty in the antenna's output signal due to a finite angular extension of the main beam, the near-in sidelobes, the spillover lobe, the far-out sidelobes, the backlobe, and the finite accuracy of pointing the antenna. It has been shown that the spillover lobe is a major contributor to uncertainty and that noise originating from the region nearest the Earth's surface has a great influence on the total uncertainty. Based on that uncertainty, a design procedure for radiometer antennas was developed. Both a front-fed and an offset parabolic reflector antenna configurations were compared in terms of the desired performance; while, their practical performance was checked against a theoretical optimal. It can be concluded that an offset antenna is a particularly promising configuration for radiometry applications.

As already stated, the beam efficiency and the integrated pattern function were the most important parameters. In principle, they can be considered as a single parameter, because the integrated pattern function can be seen as the corollary of beam efficiency. Practical values tend to be in the vicinity of 90% within a solid angle of 5°. Taking this as a starting point, it was shown that an uncertainty of approximately 3–4 K occurs at 30 GHz for the derived brightness temperature for time percentages of 90%. When reflected against the typical values of 0.5 to 1 K for calibration accuracy of the radiometer receiver,
there would seem to be an interfacing problem. That automatically forms the introduction to chapter 3, which concentrates on the receiver.

A classical design for radiometer receivers needs a continuous temperature stabilization, and verification with a reference load in order to give a reliable output. Temperature stabilization, which usually consists of an enclosure in which the temperature is kept very stable, may account for a high percentage of a receiver's total cost. Despite considerable technical effort and cost, temperature gradients inside a stabilized environment cannot be avoided entirely in practice, because the active components generate heat locally. However, a temperature stabilized environment has become the generally accepted solution and, in this way, stabilization has been used since the earliest radiometer receiver designs.

In this thesis, it has been shown that a more compact and less expensive design can be feasible by omitting (part of) the temperature stabilization. The validity of that approach was demonstrated with a breadboard model. The novel stabilization method incorporates small thermocouples and on-line data preprocessing. Measurements produced functional relationships which describe radiometer's performance as a function of the ambient and local temperatures. Those relationships serve as the input to a software-based temperature stabilization procedure. In such a way, readily available computing power can be used to keep track of the system's performance and the output signal from the radiometer can be constantly corrected according to the updated performance data. The advantages of such an approach are most striking: when a personal computer is already being used for data-acquisition, no extra temperature stabilization hardware is needed and, therefore, the complete configuration will be simple and inexpensive. The main disadvantage of the novel approach is the difficulty of obtaining temperature curves which require some experience, judgment and time. Practice has shown the value of using a tunnel diode as the detection device in order to ensure accurate compensation. The uncertainty penalty of such a radiometer configuration was found to be 2.5 K. Referring back to the uncertainty by the radiometer antenna, the novel stabilization method appears to provide the right balance between antenna and receiver.

Furthermore, due to its low costs, the new approach can lead to a low-cost radiometer. That could have important implications for the scale of operational use of this type of instrument; particularly as the existing radiometers are either unwieldy or expensive. The traditional disciplines where ground-based microwave radiometry has proven its potential are the collection of statistics concerning attenuation of single frequency operations and the retrieval of atmospheric parameters for multi-frequency operations. Dual-frequency radiometry (20 and 30 GHz) has been used experimentally for retrieving atmospheric
precipitable water vapour and liquid water content. Much effort has been made to develop theoretical models for improving the accuracy in the retrieval of geophysical parameters from radiometer data. In general, the current trend is towards the inclusion of more frequencies (in the 50–60 GHz band). Potentially, that could increase the accuracy of retrieval, but at the same time the system becomes complex. The low-cost receiver proposed here may prove effective in this field.

Another point that has attracted attention is that, especially for satellite-based radiometers, resolution might be an insurmountable problem for accurate retrieval. For passive systems, a compromise between the radiometer receiver's sensitivity and spatial resolution must be accepted. That introduces chapter 4 which gives a description of spatial reconstruction.

In contrast to quantitatively oriented remote sensing techniques, which concentrate on the long-term statistical properties, another branch of remote sensing techniques exists. That branch can be referred to as being image-oriented because it emphasizes the image aspects of the object and relies on the generation of an image. For these remote sensing techniques, IR-observations were mainly used; however, the use of microwaves for this purpose has developed rapidly. This is among others motivated by the lesser extinction of microwaves in contrast to IR-waves if clouds are present. A major disadvantage of microwave radiometry is its substantially lower spatial resolution when compared to that of infrared systems. A partial enhancement of spatial resolution can be obtained by exploiting the relative motion between the observed scene and the microwave radiometer. That motion can be produced with a scanning antenna or by the natural movement of the cloudy atmosphere. The first type of movement has the advantage that it can be controlled. Since no scanning antenna configuration was present at the EUT antenna site, chapter 4 had to be focused on exploiting the natural movement.

Several techniques have been proposed to enhance spatial resolution; however, most of them use doubtful approximations. Furthermore, those methods involve parameters, whose values are critical for the accuracy of the inversion, which make the process to converge and the computer time to decrease. As a consequence, the methods involve trial—and error constants or need proper truncation of a series.

In this thesis, a numerical deconvolution technique is described which is able to process data from a passive microwave radiometer. When applied to a radiometer system, such a technique will produce performance that is equivalent to that of a radiometer with an antenna of larger dimensions. Consequently, it is possible to extend the performance of a radiometer antenna beyond its physical dimensions, which are often the limiting factor in achievable spatial resolution of a radiometer system.
Extensive tests have indicated that resolution can be improved by a factor of ten.

Although other deconvolution techniques exist, this technique is pleasingly elegant if it is evaluated against the effects of receiver noise. The proposed algorithm will ensure good results even with noisy data, where most other deconvolution techniques fail. This is because the new method makes use of the a-priori knowledge of the existence of the radiometric receiver uncertainty.

Another benefit of this technique is the optimal use of measurement data that is available. More data guarantees a higher resolution, and so the limit of resolution, in the first instance, is connected to the amount of data available.

When judging the performance of the deconvolution process, it has to be kept in mind that the measurement device is a radiometer and this generally involves some kind of time integration. The effect of removing the time integration has not been included, although it is possible that an optimal integration time and integration law can be proposed which are related to the spatial integration of the antenna.

As an overall conclusion, it can be said that the novel method proposed is robust, reliable and contains no fuzzy constants.

It is believed that this method could have a beneficial impact on satellite-based radiometer systems because their physical dimension can often be the limiting factor to the spatial resolution that can be achieved.
Verantwoording

In dit proefschrift zijn resultaten beschreven die ten dele behaald zijn door bijdragen van anderen. Deze bijdragen worden hieronder nader omschreven.

- An Optimization Method for Radiometer Antennas. (par.2.3.)
  J.W. Wittekamp heeft een computerprogramma, ontwikkeld door de auteur van dit proefschrift, aangepast zodanig dat het geschikt was voor het optimaliseren van radiometer antennes. Dit als onderdeel van zijn afstudeerwerk.

- A Review and Comparison of Some Asymptotic Techniques for Calculating the Wide-angle Radiation Pattern of Paraboloid Reflector Antennas. (par.2.4.)
  J.Chen heeft de asymptotische methoden geïmplementeerd in software, in het kader van zijn "visiting research fellowship" onder begeleiding van de auteur en dr.ir. M.H.A.J. Herben.

- On the Design of Radiometer Antennas. (par.2.6.)
  Het onderzoek beschreven in dit artikel is mede uitgevoerd door J.W. Wittekamp in het kader van zijn afstudeerwerk, onder begeleiding van de auteur van dit proefschrift en dr.ir. M.H.A.J. Herben.

- The Temperature Dependence of Schottky versus Backward Diodes for Radiometry Applications. (par.3.4.)
  T.F. Buss schreef de software in het kader van zijn stage. De auteur van dit proefschrift en dr.ir. T.G. van de Roer hebben de heer Buss bij zijn stagewerk begeleid.

- A Novel Radiometer Receiver Stabilization Method. (par.3.5.)

- Image Reconstruction using a Passive Microwave Radiometer. (par.4.4.)
  dr.ir. H.G. ter Morsche heeft de in dit artikel beschreven wiskundige bewijzen van de methode verduidelijkt.
Acknowledgements

The work described in this thesis was performed during my stay at the Telecommunications Division of the Eindhoven University of Technology. The completion of this thesis was only possible thanks to the close cooperation of students and colleagues of this department. The support of the late mr. B. Stal in the initial phase when he initiated me into the secrets of radiometer receivers should be mentioned in particular. Furthermore, I want to express my indebtedness to my promoter prof.dr.ir. G. Brussaard and copromoter dr.ir. M.H.A.J. Herben for their stimulating support, instructive discussions and encouragement.

Members of other departments of the University were also helpful. In particular dr. H.G. ter Morsche and prof.dr. J. Boersma of the Department of Mathematics and prof.dr. F.W. Sluijter of the Department of Applied Physics should be mentioned.

Last but not least, I am very grateful to all close relatives for their support in all daily matters and for their moral and financial support.
Curriculum Vitae

Peter de Maagt was born in Pauluspolder, on February 26, 1964. From the Jansenius Secondary School he obtained the school-leaving certificate ("VWO-diploma") in 1982. In the same year he started studying Electrical Engineering at the Eindhoven University of Technology, The Netherlands, from which he graduated in 1988. His master's thesis described the design of a triple-frequency antenna feed system. In 1988, he started a second-phase designer course on Information and Communication Technology. This study programme concerned the design principles of radiometer systems. The research programme was extended and included the work recorded in this thesis on "The Radiometer as a Remote Sensing Device". The research work described was performed in the Telecommunications Division of the Department of Electrical Engineering at Eindhoven University of Technology, The Netherlands.

STELLINGEN

BEHORENDE BIJ HET PROEFSCHRIFT:

THE MICROWAVE RADIOMETER AS A REMOTE SENSING DEVICE:
DESIGN AND APPLICATION.

van

P.J.I. de Maagt

1) Een 4-jarige Assistent in Opleiding wordt ten onrechte gezien als een gediplomeerd student.

2) De Gouldamadines (Chloebia Gouldiae) kunnen gezien worden als een overgangsvorm tussen de papagaaiamadines (Genus Erythrura) en de grasvinken (genus Chloebia). Dit zou ook uit de positie in de systematiek moeten blijken.

3) Het gejoel tegen gekleurde voetballers in voetbalstadions moet veeleer gezien worden als een poging om spelers van de tegenpartij uit balans te brengen dan als rassenhaat.

4) Het toepassen van een multi-frequentie belichtersysteem ten behoeve van de Olympus propagatiemetingen bespaart niet alleen op de kosten voor meerdere reflectoren maar maakt ook meerdere soorten metingen mogelijk.

   P.J.I. de Maagt and H.H.A.J. Hechen,
   ON THE DESIGN OF A TRIPLE-FREQUENCY ANTENNA SYSTEM FOR THE
   OLYMPUS PROPAGATION EXPERIMENT,
   Proc. of the ITG-Fachtagung, 20–23 March 1990, Wiesbaden, p.79–84.

5) De ontwikkelde ontwerpprocedure voor radiometerantennes rechtvaardigt de twijfel of de tot nu toe gebruikte radiometerantennes ook daadwerkelijk voor dit doel geschikt zijn.

   Dit proefschrift, par.2.6.

6) De aanwezigheid van radiometerontvangerruis hoeft niet automatisch te leiden tot een foutieve reconstructie van het waargenomen object.

   Dit proefschrift, par.4.4.

7) Het nut van dure calibratiemethodieken voor radiometer ontvangers is discutabel, als de onzekerheid van het antenna uitgangssignaal in ogenschouw genomen wordt.
8) De voorkeur voor Schottky diodes als "square law detector" in radiometer ontvangers mag niet gebaseerd worden op hun vermeende hogere spanningsgevoeligheid.

P.J.J. de Maagt, and T.F. Buis, and T.G. van de Roer.
TEMPERATURE DEPENDENCE OF SCHOTTKY VERSUS BACKWARD DIODES FOR RADIOMETRY APPLICATIONS.

9) Het toepassen van thermocouples en "on-line data-preprocessing" is een goedkoop alternatief voor temperatuurstabilisatie.

Dit proefschrift, par.3.3.

10) Het is een wijd verspreid misverstand dat de in de literatuur vaak gebruikte GTD "slope-diffractie termen" het volledige 2e orde diffractie veld vertegenwoordigen. Dit is slechts het geval als het invallende veld op de rand van de reflectorantenne gelijk is aan 0.

INTRODUCTION TO THE UNIFORM GEOMETRICAL THEORY OF DIFFRACTION.
Artech House, Boston, 1990.