# Integral Identities for Passive Systems and Spherical Waves in Scattering and Antenna Problems

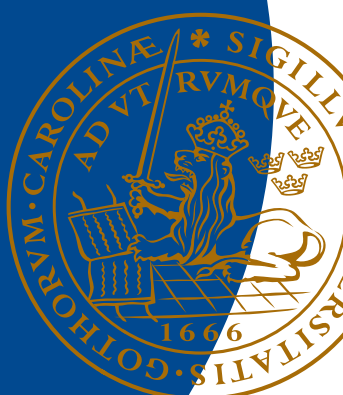
Bernland, Anders

2012

Doctoral thesis

# Integral Identities for Passive Systems and Spherical Waves in Scattering and Antenna Problems

Anders Bernland




Department of Electrical and Information Technology,
Faculty of Engineering, LTH, Lund University, 2012.

# Integral Identities for Passive Systems and Spherical Waves in Scattering and Antenna Problems

Anders Bernland

Doctoral Thesis
Electromagnetic Theory

Lund University
Lund, Sweden
April 2012

Doctoral thesis which, by due permission of the Faculty of Engineering at Lund University, will be publicly defended on Wednesday, May 23, 2012, at 10.15 A.M. in lecture hall E:1406, Department of Electrical and Information Technology, John Ericssons väg 4, Lund, Sweden, for the degree of Doctor of Philosophy in Engineering in Electromagnetic Theory.
Faculty opponent: Professor Johannes Skaar, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

Doktorsavhandling som för avläggande av teknologie doktorsexamen i teoretisk elektroteknik vid tekniska fakulteten vid Lunds Universitet kommer att offentligen försvaras onsdagen den 23 maj 2012, klockan 10.15 i hörsal E:1406, Institutionen för elektro- och informationsteknik, John Ericssons väg 4, Lund.
Fakultetsopponent: Professor Johannes Skaar, Norges teknisk-naturvitenskapelige universitet (NTNU), Trondheim, Norge.

Department of Electrical and Information Technology
Faculty of Engineering, LTH
Lund University
P.O. Box 118, S-221 00 Lund, Sweden

Series of licentiate and doctoral theses
ISSN 1654-790X
No. 40
ISBN 978-91-7473-302-0


Printed in Sweden by Tryckeriet i E-huset, Lund.
April 2012.

*If in other sciences we should arrive at certainty without doubt and truth without error, it behooves us to place the foundations of knowledge in mathematics.*

Roger Bacon (1214-1294)

"Opus Majus", Book 1, Chapter 4

# Abstract

Sum rules and physical limitations within electromagnetic theory and antenna theory have received significant attention in the last few years. However, the derivations are often relying on application specific and sometimes unsupported assumptions, and therefore a mathematically rigorous and generally applicable approach seems timely. Such an approach is presented in this thesis, along with examples and all the necessary proofs. The approach is also applied in the thesis to derive sum rules and physical limitations on electromagnetic spherical wave scattering. This has not been done before, despite the widespread use of spherical wave decompositions. For example, spherical waves and the antenna scattering matrix provide a complete and compact description of all the important properties of an antenna, are crucial parts in spherical near-field antenna measurements, and have been used recently to model antenna-channel interaction and multiple-input multiple-output (MIMO) communication systems. This thesis is also the first to present a method to estimate spherical wave coefficients from propagation channel measurements.

The results of this thesis can roughly be divided into three categories: Firstly, a general approach to derive sum rules and physical limitations on input-output systems based on the assumptions of causality and passivity is presented (Paper I). Secondly, sum rules and physical limitations on the scattering and matching of electromagnetic spherical waves are derived, and the implications for antennas are explored (Paper II–IV). Thirdly, a method to estimate spherical wave coefficients from channel measurements, and the results of a measurement campaign, are presented and analysed (Paper V).

The thesis consists of a General Introduction and five appended papers. In Paper I, the general approach to derive sum rules and physical limitations for passive systems is presented, together with the necessary proofs. The derivations rely on the connection between passive systems and Herglotz functions in conjunction with a set of integral identities for that class of functions. The procedure is described thoroughly, and exemplified with examples from electromagnetic theory.

In Paper II, physical limitations for scattering and absorption of electromagnetic spherical waves are derived, using the approach presented in Paper I. The time domain versions of the spherical waves are used to rigorously describe passive scatterers and find the corresponding Herglotz functions, whereas the frequency domain counterparts are used to derive sum rules and physical limitations. The limitations imply that the diagonal elements of the scattering matrix, which relate the coefficients of the incoming and outgoing waves, cannot be arbitrarily small over a whole frequency interval; the bounds depend on the fractional bandwidth as well as the size, shape and static material properties of the scatterer. Physical interpretations of the results are given, and numerical examples for nanoshells and some small antennas are included.

In Paper III, limitations on broadband matching of spherical waves, i.e. optimal matching bounds of an antenna radiating a specific spherical wave, are presented. It is shown how the optimal matching problem is identical to the closely related, and yet very different problem of finding the scattering limitations for a homogeneous

sphere in its high-contrast limit. The scattering view yields explicitly the necessary low-frequency asymptotics of the reflection coefficient, something that appears to be non-trivial to derive from the classical network point of view.

In Paper IV, the results of Paper II and Paper III are improved on by deriving limitations for scattering of higher order electromagnetic spherical waves (quadrupoles, octopoles and so forth). Moreover, the implications for antennas are investigated further, and it is shown how the limitations place bounds on the antenna scattering matrix. Comparisons to other types of antenna limitations are given, and numerical simulations for two folded spherical helix antennas and a directive Yagi-Uda antenna are included to illuminate and validate the theory.

In Paper V, a method to estimate spherical wave coefficients from channel measurements with a 3D positioner is presented. Results from a measurement campaign are presented and analysed. One conclusion is that using randomly positioned measurements within a volume is less sensitive to noise than using only measurements on the surface.

The General Introduction provides more detailed background for the appended papers: Dispersion relations, sum rules and physical limitations are discussed; electromagnetic spherical waves, the antenna scattering matrix, and their use within antenna measurements and modelling of antenna-channel interaction and MIMO are reviewed; an overview on previous approaches to physical limitations in antenna theory is also included. The contributions of the appended papers are summarized in this context, and some outlook for the future is given.

# Populärvetenskaplig sammanfattning

I avhandlingen presenteras nya teoretiska och experimentella metoder för att analysera spridning av elektromagnetiska vågor, framförallt med fokus på antenner och trådlös kommunikation. Oavsett det handlar om radio- och TV-sändningar, mobiltelefoni, trådlösa datornätverk, eller radar, så är antenner oumbärliga; de är nämligen länken mellan de elektromagnetiska vågorna och elektroniken i apparaterna. En antenningenjör måste uppfylla många krav samtidigt, och till sin hjälp har hon både mätutrustning och beräkningsprogramvara. Icke desto mindre så kan förståelse och intuition baserat på grundläggande fysik också vara till nytta, men på grund av antenners många varierande skepnader och användningsområden kan det vara svårt att dra generella teoretiska slutsatser. Ett sätt att ändå göra detta är genom att härleda så kallade fysikaliska begränsningar, som beskriver vad som går, och vad som inte går, att uppnå. Begränsningarna kan ge en förståelse för vilka faktorer som begränsar prestanda, och även antyda om det finns möjligheter till förbättring eller inte. De senaste åren har mycket forskning tillägnats fysikaliska begränsningar inom antennteori, och mer allmänt inom teori för spridning av elektromagnetiska vågor.

Precis som vågrörelsen på en gitarrsträng kan delas upp i grundton och övertoner, så kan också elektromagnetiska vågor delas upp i mindre beståndsdelar, så kallade sfäriska vågor. I avhandlingen härleds fysikaliska begränsningar för hur olika objekt kan interagera med dessa sfäriska vågor. Resultaten är att bara en begränsad mängd effekt kan absorberas från en sfärisk våg över ett frekvensband, och hur mycket som kan absorberas beror på objektets storlek, form och materialegenskaper. Detta ger begränsningar även på antennprestanda, eftersom man vill att en mottagande antenn ska kunna absorbera effekt över ett brett frekvensband. På motsvarande sätt finns också begränsningar för sändande antenner. Det bör dock nämnas att resultaten inte är begränsade till antenner, utan gäller för helt allmänna objekt. Ett intressant exempel, som också tas upp i avhandlingen, är ljusabsorberande nanopartiklar.

En anledning till att sfäriska vågor studeras är att de används flitigt inom antennteori, till exempel för att modellera multipelantennsystem (eller MIMO-system, från engelskans multiple-input multiple-output). Multipelantennsystem har börjat användas mer och mer på senare tid för att öka prestandan i trådlösa system och tillfredsställa den ökande efterfrågan på höga överföringshastigheter. Sfäriska vågor är också mycket viktiga i bearbetning och tolkning av data från antennmätningar. Däremot har de aldrig tidigare använts för att tolka data från mätningar av trådlösa kanaler (allt det som finns emellan sändar- och mottagarantennerna). Att göra det kan ge en beskrivning av kanaler som passar beskrivningen av antennerna. En sådan mätmetod presenteras i den här avhandlingen; mätningar har genomförts och data analyserats, och en slutsats som dras är att det är bättre att placera mätpunkterna i en volym än på en yta för att effekterna av mätbrus ska minimeras.

I avhandlingen presenteras också en allmän metod att härleda fysikaliska begränsningar. Många tidigare härledningar vilar på alltför specifika, och ibland ogrundade, antaganden. Det antagande som används här är endast att processen man beskriver

är passiv, det vill säga att energi inte produceras. Så är fallet för många fysikaliska processer, till exempel för traditionella antenner. Metamaterial är ett samlingsnamn för konstgjorda material som har extraordinära elektriska egenskaper, och här har metoden med fysikaliska begränsningar använts för att utreda vad som är teoretiskt möjligt och inte. I avhandlingen ges också fler exempel på hur metoden används inom elektroteknik, men även processer inom andra fysik- och teknikområden kan analyseras på samma vis. Kärnan i metoden är en samling integralidentiteter för Herglotzfunktioner, en funktionsklass som är intimt förknippad med passiva processer. Alla nödvändiga matematiska bevis presenteras i avhandlingen, vilket ställer metoden på en stabil grund.

# List of included papers

This thesis consists of a General Introduction and the following appended papers. The order of the authors' names indicates their relative contributions to the papers, and the specific contribution from the author of this thesis is pointed out after each entry. Papers I–III are published, and Paper IV is accepted for publication, in peer-reviewed international journals. Paper V is published in the LUTEDX/TEAT series at Lund University, and has also been submitted for publication in a peer-reviewed international journal.

I. A. Bernland, A. Luger, and M. Gustafsson. Sum rules and constraints on passive systems. *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 14, paper 145205, March 2011.

   The author of this thesis has performed most of the background research, derivations, analysis, the examples, and written the paper.

   Selected for inclusion in IoP Select, a special collection of high-quality articles chosen by the IoP editors.

II. A. Bernland, M. Gustafsson, and S. Nordebo. Physical limitations on the scattering of electromagnetic vector spherical waves. *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 14, paper 145401, March 2011.

   The author of this thesis has performed most of the derivations and analysis, has implemented both the numerical examples, and written the paper.

III. S. Nordebo, A. Bernland, M. Gustafsson, C. Sohl, and G. Kristensson. On the relation between optimal wideband matching and scattering of spherical waves. *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 9, pp. 3358–3369, July 2011.

   The author of this thesis has carried out a large part of the analysis and written parts of the paper.

IV. A. Bernland. Bandwidth limitations for scattering of higher order electromagnetic spherical waves with implications for the antenna scattering matrix. *IEEE Transactions on Antennas and Propagation*, accepted for publication, April 2012.

   The author of this thesis is the lone author of the paper.

V. A. Bernland, M. Gustafsson, C. Gustafson, and F. Tufvesson. Estimation of spherical wave coefficients from 3D positioner channel measurements. Technical Report LUTEDX/TEAT-7215, Lund University, Sweden, pp. 1–11, March 2012. http://www.eit.lth.se. Submitted to *IEEE Antennas and Wireless Propagation Letters*.

   The author of this thesis is responsible for the numerical implementations, have planned and participated in the measurements, performed most of the analysis, and written the paper.

# Other publications by the author

The author of this thesis is also the author or co-author of the following conference papers, which are related to but not considered part of the thesis. The order of the authors' names indicates their relative contributions to the papers, and the first author of each paper presented the paper at the symposium.

VI. C. Sohl, M. Gustafsson, and A. Bernland. Some paradoxes associated with a recent sum rule in scattering theory. In *Proceedings of the XXIXth URSI General Assembly*, Chicago, IL, USA, August 7–16 2008. International Union of Radio Science.

VII. M. Gustafsson, C. Sohl, G. Kristensson, S. Nordebo, C. Larsson, A. Bernland, and D. Sjöberg. An overview of some recent physical bounds in scattering and antenna theory. In *Proceedings of the 3rd European Conference on Antennas and Propagation*, pp. 1795–1798, Berlin, Germany, March 23–27 2009.

VIII. A. Bernland, M. Gustafsson, and S. Nordebo. Summation rules and physical bounds for partial wave scattering in electromagnetics. In *Proceedings of the 9th International Conference on Mathematical and Numerical Aspects of Waves*, pp. 206-207, Pau, France, June 15–19 2009.

IX. M. Gustafsson, G. Kristensson, S. Nordebo, C. Larsson, A. Bernland, and D. Sjöberg. Physical bounds and sum rules in scattering and antenna theory. In *Proceedings of the 11th International Conference on Electromagnetics in Advanced Applications*, pp. 600–603, Turin, Italy, September 14–18 2009.

X. M. Gustafsson, G. Kristensson, A. Bernland, D. Sjöberg, and B. L. G. Jonsson. Physical bounds on the partial realized gain. In *Proceedings of the 26th International Review of Progress in Applied Computational Electromagnetics*, pp. 1–6, Tampere, Finland, April 25–29 2010.

XI. A. Bernland, A. Luger, and M. Gustafsson. Sum rules and constraints on passive systems with applications in electromagnetics. In *Proceedings of the 20th International Symposium on Electromagnetic Theory*, pp. 37–40, Berlin, Germany, August 16–19 2010.

XII. M. Gustafsson, D. Sjöberg, A. Bernland, G. Kristensson, and C. Sohl. Sum rules and physical bounds in electromagnetic theory. In *Proceedings of the 20th International Symposium on Electromagnetic Theory*, pp. 37–40, Berlin, Germany, August 16–19 2010.

XIII. A. Bernland, A. Luger, and M. Gustafsson. Using Herglotz representations to derive limitations on passive physical systems. In *Proceedings of the 7th International Congress on Industrial and Applied Mathematics*, p. 496, Vancouver, BC, Canada, July 18–22 2011.

XIV. A. Bernland. On physical limitations for broadband electromagnetic scattering — high order spherical waves. In *Proceedings of the 10th International Conference on Mathematical and Numerical Aspects of Waves*, pp. 619-622, Vancouver, BC, Canada, July 25–29 2011.

# Preface and acknowledgments

This thesis for the degree of Doctor of Philosophy in Engineering in Electromagnetic Theory summarizes the research I have carried out from the start of my doctoral studies on August 27, 2007 and until the time of print. My work has been part of the project MIMO Antennas and Channels within the High Speed Wireless Communications Center, funded by the Swedish Foundation for Strategic Research. When I started, the formulation of the problem statement was:

> This project studies how the performance of hand-held terminal antennas in multiple-antenna ("MIMO") systems can be optimized with respect to the environment and to physical limitations.

Another member of the project was doctoral student Andrés Alayón Glazunov, who modelled MIMO and the antenna-environment interaction with electromagnetic spherical waves; he presented his doctoral thesis in March 2009. Closely linked to the project was also the work of doctoral student Christian Sohl; he studied dispersion relations and physical limitations in scattering and antenna problems, and presented his doctoral thesis in September 2008.

Coming from a mathematical background, with a Master of Science degree in Engineering Mathematics, my inherent way of tackling a problem was to solve a more general problem, and then show that the original problem was just a special case of that. So, we decided to solve a more general problem: How much interaction can a general scattering object provide with electromagnetic spherical waves over a frequency interval? But that problem was not general enough either, so we decided to solve an even more general problem: What are the constraints imposed on a general physical system due to the assumptions of causality and passivity?

The last question is (more or less) answered in Paper I, and the previous in Paper II and Paper IV. Paper III and Paper IV explores the implications of the scattering limitations for antennas. In Paper V, a method to use spherical waves in propagation channel measurements is presented. With that, the trip from the Borel measures in Paper I to the antenna measurements in Paper V is complete.

There are a number of people who have contributed to this thesis in one way or another, and there are a number of people without whom my four and a half years as a doctoral student would not have been as rewarding and enjoyable as they have been. First and foremost, I express my deepest gratitude to my supervisor Mats Gustafsson, for accepting me as his doctoral student and giving me all the support, guidance and collaboration a supervisor should plus infinitely more. I particularly admire his positive spirit, extraordinary ability to deliver ideas, and remarkable intuition. I am also grateful to my co-supervisor Annemarie Luger for invaluable help in our problems concerning Herglotz functions and integration theory. Many thanks goes to my co-supervisor Gerhard Kristensson, for sharing his knowledge in physics, mathematics and electromagnetic theory, and for guidance on teaching.

An especially warm thank you goes to Anders Melin, for sharing his outstanding knowledge in mathematics and long experience in scientific research. I also thank Sven Nordebo for collaboration on spherical waves and matching. Carl Gustafson

deserves a special thank you (and a Tab?) for collaboration on the measurements, and I also thank Fredrik Tufvesson for valuable discussions on channel modelling and measurements. I would also like to thank Anders Sunesson for taking the time to help me with measurements. I am also indebted to my previous colleagues Andrés Alayón Glazunov and Christian Sohl, for opening up vast fields of interesting research and guiding me through them during my first years. Knowing Andrés and Christian, and building on their work, has truly been to "stand on the shoulders of giants." I am grateful to Christian also for the elegant LaTeX template used to assemble this thesis. I am grateful to Erik Yderborg for collaboration on the Master thesis that eventually led me to this position, and to Alireza Kazemzadeh and Christian Sohl for their encouragement when I was looking for the next step to take after this.

I am grateful to all colleagues, former and present, at the department, for providing a stimulating and pleasant work environment. I've had uncountably many joyful lunches and coffee breaks with you. I am particularly grateful to my colleagues in the Electromagnetic Theory group, to the Wednesday afternoon fika-alliance (Farfar rules!), and to the Friday night pub gang.

Last, but not least, I thank my family and friends, for all the happy moments, and for all your support and encouragement. I look forward to moving to Göteborg, to spend more time with those of you who live there. I especially thank my parents Anna Maria and Lars Göran, and my sisters Karin and Helena, for always believing in me and giving me solid support throughout my life.

I would like to conclude the personal part of this thesis with a quote by Eric Berglund (the quote is reproduced here with his permission); he has provided much of the soundtrack of my time in Lund, since I first encountered his music at a concert at Blekingska in 2004, during my first year as an undergraduate student.

*come with me*
*from a place they call reality*
*come with me*
*neither up nor down*
*it's a special ride you see*
*come with me*
*would you sacrifice this life to make it real?*

*come with me*
*to a place I call reality*

Borlänge, Easter Sunday 2012

Anders Bernland

# Contents

# General Introduction

Anders Bernland

# 1 Introduction

## 1.1 Background

Electromagnetic waves are modelled by Maxwell's equations, which were published by Maxwell in 1865 [67]. The first experimental verification came in 1886, when Hertz was able to use a wire to induce sparks in the gap of a loop, thus realizing a half-wave dipole antenna and a resonant loop antenna [53]. Since then, electromagnetic waves have been employed for various forms of wireless communication, such as communication with ships and other vessels, radio and TV broadcasting, and during the last decades for wireless computer networks and hand-held mobile terminals. Other applications for electromagnetic waves range from everything between cavities in particle accelerators to household use of microwaves for re-heating yesterday's dinner. Various types of imaging techniques relying on electromagnetic waves are widely used for military, industrial and medical purposes, where radar and x-ray imaging are two of the most well-known.

Antennas are a vital part of any wireless communication system, since they constitute the link between the electromagnetic waves and the electronics of the devices. The following definition can be found in *IEEE Standard Definitions of Terms for Antennas* [2]:

**Antenna.** That part of a transmitting or receiving system that is designed to radiate or to receive electromagnetic waves.

See Figure 1 for an illustration. The term "antenna" is used throughout this thesis, instead of the term "aerial," which is common in British English.

An antenna designer faces a heap of practical challenges, such as fine-tuning the resonance frequency and impedance of the antenna at the same time as the antenna must meet power and size constraints and be compatible with the rest of the device and its usage scenario. Numerical simulations and practical measurements are vital tools for an antenna engineer, and both numerical algorithms and computer power as well as measurement procedures and equipment are constantly improved. Nonetheless, theoretical understanding and intuition based on the underlying physics can also be very beneficial. However, because of the complexity of antennas and wireless communication, it can be hard to reach general theoretical conclusions. Even so, one approach to do so is to derive physical limitations, which state what can and what cannot be achieved. The limitations can be used to understand the trade-off between performance and various constraints, such as size and materials, and also indicate if there is room for improvement or not. Limitations on antennas were introduced by Wheeler [106] and Chu [16] in 1947 and 1948, respectively. Closely related are the Fano limitations on broadband matching of arbitrary impedances from 1950 [24]. Since these pioneering results, a significant amount of research has been devoted to this topic.

Physical limitations on antennas are examples of limitations on electromagnetic scattering and absorption; more generally, such limitations state that certain objects can only provide limited interaction with electromagnetic waves. For example, it

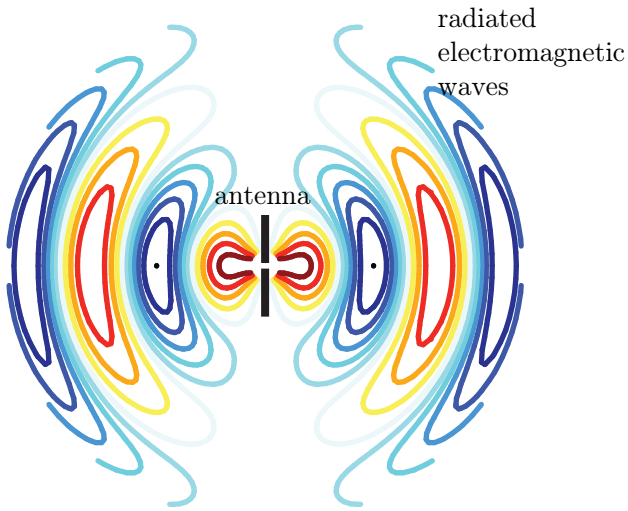



**Figure 1**: Illustration of an antenna and radiated electromagnetic waves.

seems intuitive that objects which are small compared to the wavelength can only disturb the wave to a limited extent. Two early publications discussing these issues are [79] by Purcell and [11] by Bohren and Huffman; their derivations fall back on the Kramers-Kronig dispersion relations, which were derived independently by Kramers [60] and Kronig [19] and specify realizable frequency behaviour of the complex refractive index, see e.g. [63]. Recently, limitations on radar-absorbers [80] and high-impedance surfaces [13] have been derived with similar methods as Fano's matching limitations [24]. In his thesis [96], Sohl derived general limitations on electromagnetic scattering, and also applied the results to antenna theory.

A common feature of the physical limitations mentioned in the previous paragraph is that they are at some point dependent on holomorphic properties of the functions involved. Many physical processes are modelled as input-output systems; there is a cause (the input) and an action (the output). For example, the electric voltage over a resistor causes a current to flow in it, a force applied to an elastic body produces a deformation, and an increase in the temperature of a confined gas results in a higher pressure. Many physical systems satisfy two very basic assumptions; causality (the output cannot precede the input) and passivity (the system cannot produce energy). It is well-known that the transfer functions of causal and passive systems have some beneficial holomorphic properties that allow dispersion relations, sum rules and physical limitations to be derived. However, the derivations are often relying on application specific and sometimes unsupported assumptions. A more generally applicable approach to derive sum rules and physical limitations seems timely.

When studying electromagnetic scattering, it is sometimes beneficial to decompose the fields in electromagnetic vector spherical waves [73] (also referred to as partial waves, TM- and TE-modes or electric and magnetic multipoles). Apart from being used for analysing scattering by spherical particles (Mie theory), spherical waves are also used in the T-matrix method, which can deal with single scatterers, composite scatterers, and systems of scatterers [64, 70]. Electromagnetic vector

spherical waves also provide a complete and compact description of the transmitting, receiving, and scattering properties of an antenna in terms of the antenna scattering matrix. Furthermore, spherical waves are crucial in data-processing in spherical near-field antenna measurements, since they provide the means to determine the far-field radiation pattern of the antenna under test from measured near-field data [44, 62]. Despite the wide-spread use of spherical wave decompositions, however, there are no previously derived physical limitations on spherical wave scattering.

The demand for higher data-rates in wireless communication is growing quickly, and one way to increase performance is to employ multiple antennas at both the transmitter and receiver side, so called multiple-input multiple-output (MIMO) technology. Information capacity can increase linearly with the number of antennas, but this demands that the antennas and propagation environment (the channel) are such that the system can make use of several degrees of freedom to transmit and receive independent data streams [77]. Since the number of degrees of freedom depend both upon the antennas and the propagation channel, it is desirable to separate these two influences in modelling and measurements. One approach to do so is the double-directional channel model, which describes the channel in terms of plane waves, or multi-path components [72]. In his thesis [28], Alayón Glazunov instead used spherical waves to model antenna-channel interaction and the available degrees of freedom in MIMO systems. This has also been done by others, see e.g. [69, 71]. One benefit is that spherical waves form a basis for the electric and magnetic fields, and provide a compact description of the properties of an antenna. However, there are no previous publications where spherical waves are used within propagation channel measurements.

## 1.2 Problem statement and outline

The ambition of this thesis is threefold:

1. Provide a general and mathematically rigorous approach to derive sum rules and physical limitations on input-output systems based on the assumptions of causality and passivity (Paper I).

2. Derive physical limitations on scattering and matching of electromagnetic spherical waves and explore the implications for antennas (Papers II–IV).

3. Implement, carry out, and evaluate a measurement campaign to estimate spherical wave coefficients from channel measurements (Paper V).

The outline of the rest of this General Introduction is the following: Dispersion relations, sum rules and physical limitations are discussed in Section 2, and the contributions of Paper I are summarized. Section 3 briefly reviews electromagnetic vector spherical waves, spherical near-field antenna measurements, and modelling of antenna-channel interaction and MIMO, and Paper V is summarized in this context. Section 4 gives a brief overview of physical limitations in antenna theory, also

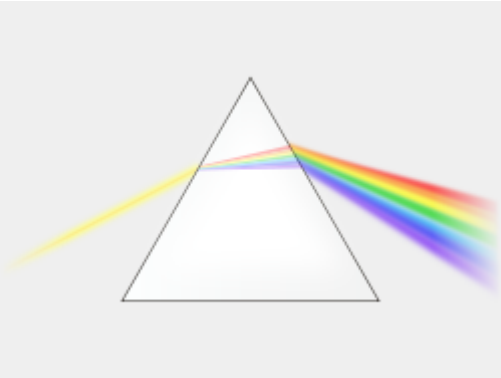


**Figure 2**: The propagation of light through a prism is an example of a dispersive (frequency-dependent) process. Since the refractive index of the prism depends on the frequency (or wavelength) of the light, the white light is decomposed into different colours. The illustration is used under license http://creativecommons.org/licenses/by-sa/3.0/de/deed.en.

including the results of Papers II–IV. Finally, conclusions and outlook is given in Section 5.

# 2 Dispersion relations, sum rules, and physical limitations

In the physical sciences, there is an ambition to model various aspects of nature, and many physical processes are modelled as input-output systems; there is a cause (the input) and an action (the output). The input and output are commonly functions of time, $t$ (with unit seconds, s), but in many applications it is more convenient to analyse physical systems in the frequency domain, where the input and output are instead functions of the angular frequency $\omega$ (with unit Hertz, Hz $= \mathrm{s}^{-1}$). A frequency-dependent physical process is called *dispersive*; consider for example the propagation of light through a prism, see Figure 2. Since the refractive index of the prism depends on the frequency (or wavelength) of the light, the white light is decomposed into different colours.

One way to study dispersive systems is to derive *dispersion relations*, which describe realizable frequency dependence. Dispersion relations are often tightly linked to the assumption of causality, which means that the output of the system cannot precede the input. In some cases, sum rules can be derived from the dispersion relations. In general, a sum rule relates a sum or integral (which is a generalized sum) to some known quantity. However, to derive sum rules requires additional assumptions, apart from causality, on the system.

Many physical systems are passive, which means that the systems cannot produce energy. In Paper I, a general and straightforward approach to derive sum rules for passive systems is presented. It is also shown that the sum rules for passive systems can in turn be used to derive physical limitations, which state what can and cannot

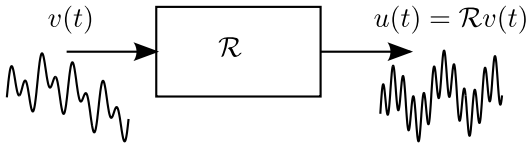



**Figure 3**: An input-output system where the operator $\mathcal{R}$ relates the output signal $u(t)$ to the input signal $v(t)$, see (2.1).

be expected from the systems. The approach relies on the well-known connection between passive systems and Herglotz functions, a class of holomorphic functions with some beneficial properties.

Dispersion relations, sum rules and physical limitations are described in more detail in this section. First, in Section 2.1, general systems in convolution form are discussed. Section 2.2 describes causality and its connection to dispersion relations. Passive systems and Herglotz functions are reviewed in Section 2.3. These three sections provide the necessary background to Section 2.4, which summarizes the general approach put forth in Paper I to derive sum rules and physical limitations for passive systems. The background in Sections 2.1–2.3 is largely inspired by the excellent books [112, 113] by Zemanian, [76] by Nussenzveig, and [56, 57] by King.

## 2.1 Systems in convolution form

Systems in convolution form, which are a subset of more general input-output systems, are discussed in this section. The name comes from the fact that the output is given by convolution of the input with the impulse response, which contains all information about the system in the time domain. The counterpart in the frequency domain is the transfer function, which will play a key role in the dispersion relations in Section 2.2. In this section, it will be seen that the convolution form is closely related to three assumptions: linearity, continuity and time-translational invariance.

As already mentioned, many physical systems are modelled as input-output systems, i.e. rules assigning an output signal $u(t)$ to every input signal $v(t)$:

$$u(t) = \mathcal{R}v(t), \tag{2.1}$$

where $\mathcal{R}$ is an operator. The system may be thought of as a "black box", see Figure 3. It is desirable to allow $u$ and $v$ to be generalized functions, or distributions, i.e. the domain $D(\mathcal{R})$ of the operator $\mathcal{R}$ is some subset of the set of distributions $\mathcal{D}'$. This allows the modelling of functions having point support, i.e. pulses delivering non-zero amounts of energy in a single moment. Furthermore, the distributional setting works well when moving between the time and frequency domains, as discussed below. An introduction to distribution theory can be found e.g. in the books [87] and [33], while e.g. [21, 54, 85, 86, 112] are more thorough treatises.

A completely arbitrary system can of course relate the input signal to the output signal in a completely arbitrary way. However, many physical systems satisfy some basic assumptions:

**Linearity:** The system (2.1) is linear if

$$\mathcal{R}(C_1 v_1 + C_2 v_2)(t) = C_1 \mathcal{R} v_1(t) + C_2 \mathcal{R} v_2(t),$$

for all scalars $C_1$, $C_2$ and all admissible input signals $v_1, v_2 \in D(\mathcal{R})$. Intuitively, linearity means that the output is doubled if the input is doubled.

**Continuity:** An operator is continuous if

$$\lim_{j\to\infty} v_j = v \Rightarrow \lim_{j\to\infty} \mathcal{R} v_j = \mathcal{R} v,$$

where $\{v_j\}_1^\infty$ is a sequence of input signals in $D(\mathcal{R})$. Here the limits must be interpreted in the correct sense and depend on the input $v_j$ and output $\mathcal{R}v_j$, respectively [112]. A simplified interpretation of continuity is that a small perturbation in the input signal only leads to a small perturbation in the output signal.

**Time-translational invariance:** The system (2.1) is time-translational invariant if $\mathcal{R}$ maps $v(t-T)$ to $u(t-T)$, for all $T \in \mathbb{R}$, whenever it maps $v(t)$ to $u(t)$. In other words, delaying the input signal simply delays the output signal. A time-translational invariant system is "non-aging", meaning that an experiment yields the same result regardless of the time when it is conducted.

It can be shown that a system satisfies these assumptions if and only if it is in convolution form, (cf. Theorem 5.8-2 in [112] and pages 134–140 in [87]):

$$u(t) = w * v(t) = \int_{\mathbb{R}} w(t') v(t - t') \, \mathrm{d}t', \tag{2.2}$$

where the second equality holds if $v$ and $w$ are integrable functions. Otherwise, convolution is defined in a more general way, see Chapter 5 in [112]. The generalized function $w$ is called the impulse response of the system, and it contains a complete description of the properties of the system in the time domain.

In many applications, it is desirable to study physical systems in the frequency domain. This implies use of the Fourier transform, which is defined by

$$\tilde{f}(\omega) = (\mathcal{F} f)(\omega) = \int_{\mathbb{R}} f(t) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t \tag{2.3}$$

for integrable functions $f \in \mathrm{L}^1$. In this case the Fourier transform of (2.2) is

$$\tilde{u}(\omega) = \tilde{w}(\omega)\tilde{v}(\omega), \tag{2.4}$$

where the transfer function is the transformed impulse response,

$$\tilde{w}(\omega) = (\mathcal{F} w)(\omega),$$

and $\tilde{v} = \mathcal{F} v$ and $\tilde{u} = \mathcal{F} u$ are the transformed input and output signals, respectively. Throughout the thesis, e denotes Euler's number (the base of the natural

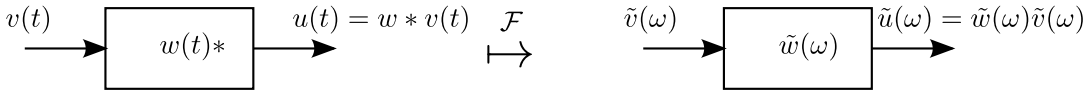



**Figure 4**: An input-output system in convolution form, where the input and output signals are related by the impulse response $w(t)$ in the time domain, see (2.2), and by the transfer function $\tilde{w}(\omega)$ in the frequency domain, see (2.4).

logarithms), and i denotes the imaginary unit, $\mathrm{i}^2 = -1$. Equation (2.4) reveals one reason to study systems in the frequency domain; multiplication is in general preferable over convolution. Just as the impulse response $w$ completely describes the system, so does also the transfer function $\tilde{w}$. The Fourier transform can be extended to cover some, but not all, distributions; those distributions $f(t)$ that grow exponentially for large $|t|$ must be excluded. However, the Fourier transform of a distribution $f(t)$ in $\mathcal{S}' \subseteq \mathcal{D}'$, where $\mathcal{S}'$ denotes distributions of slow growth, is well defined and also a distribution $\tilde{f}(\omega)$ in $\mathcal{S}'$ [33, 112]. If in addition $w$ or $v$ e.g. have compact support, the system (2.2) is mapped to (2.4) under the Fourier transform [112]. An illustration of an input-output system in convolution form can be found in Figure 4.

## 2.2 Causality and dispersion relations

As mentioned in the introduction, a physical process which depends on the frequency is called dispersive. Relations for the realizable frequency dependence for the transfer function (and therefore on the system as a whole) are called dispersion relations, where the Kramers-Kronig relations for the refractive index [63] have already been mentioned. Apart from the assumptions of linearity, continuity and time-translational invariance (which are equivalent to convolution form), dispersion relations also require another assumption: causality. This is the topic of the present section.

For many physical systems it is obvious that causality holds, since the action cannot precede the cause. The precise definition of causality used here is

**Causality:** The system (2.1) is causal if, for all $t_0$,

$$v_1(t) = v_2(t), \quad \text{for } t < t_0 \quad \Rightarrow \quad \mathcal{R}v_1(t) = \mathcal{R}v_2(t), \quad \text{for } t < t_0.$$

For systems in the convolution form (2.2), causality is equivalent to

$$w(t) = 0, \quad \text{for } t < 0.$$

Causality means that the output can only depend on previous values of the input. In other words, the system cannot predict the future.

The starting point to derive dispersion relations for causal systems is often Titchmarsh's theorem [76, Theorem 1.6.1], [56, Section 4.22]:

**Theorem 2.1** (Titchmarsh's theorem)**.** *Let $\tilde{f}(\omega)$ be a square-integrable function ($\tilde{f} \in \mathrm{L}^2$) on the real line, i.e.*

$$\int_{\mathbb{R}} |\tilde{f}(\omega)|^2 \,\mathrm{d}\omega < \infty.$$

*If $\tilde{f}(\omega)$ satisfies one of the four conditions below, then it fulfils all four of them and $\tilde{f}(\omega)$ is called a* causal transform*.*

1. *Its inverse Fourier transform $f(t)$ vanishes for $t < 0$.*

2. *The real and imaginary parts of $\tilde{f}$ satisfy the first Plemelj formula:*

$$\operatorname{Re} \tilde{f}(\omega) = -\frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{|\omega - \xi| > \varepsilon} \frac{\operatorname{Im} \tilde{f}(\xi)}{\omega - \xi} \,\mathrm{d}\xi. \tag{2.5}$$

3. *The real and imaginary parts of $\tilde{f}$ satisfy the second Plemelj formula:*

$$\operatorname{Im} \tilde{f}(\omega) = \frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{|\omega - \xi| > \varepsilon} \frac{\operatorname{Re} \tilde{f}(\xi)}{\omega - \xi} \,\mathrm{d}\xi. \tag{2.6}$$

4. *The function $\tilde{f}(\omega + \mathrm{i}\sigma)$ is a holomorphic function in the open upper half-plane $\mathbb{C}^+ = \{\omega + \mathrm{i}\sigma : \omega \in \mathbb{R}, \sigma > 0\}$. Furthermore, it holds that*

$$\tilde{f}(\omega) = \lim_{\sigma \to 0^+} \tilde{f}(\omega + \mathrm{i}\sigma), \quad \textit{for almost all } \omega \in \mathbb{R},$$

*and*

$$\int_{\mathbb{R}} |\tilde{f}(\omega + \mathrm{i}\sigma)|^2 \,\mathrm{d}\omega < \infty, \quad \textit{for } \sigma > 0.$$

The Plemelj formulae is an example of a Hilbert transform pair; the Hilbert transform $\mathcal{H}$ is defined for functions $F(x)$ on $\mathbb{R}$ that are e.g. square-integrable and locally integrable, as:

$$\mathcal{H}F(x) = \frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{|x - \xi| > \varepsilon} \frac{F(\xi)}{x - \xi} \,\mathrm{d}\xi, \tag{2.7}$$

see e.g. [56, 65]. Its inverse, under the above assumptions on $F$, is

$$F(x) = -\frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{|x - \xi| > \varepsilon} \frac{\mathcal{H}F(\xi)}{x - \xi} \,\mathrm{d}\xi. \tag{2.8}$$

The functions $F(x)$ and $\mathcal{H}F(x)$ in (2.7)–(2.8) are called a Hilbert transform pair.

In this thesis, the name dispersion relations is reserved for the type of relations where a function of frequency is related to the integral of another function, as for example the real and imaginary parts of $\tilde{f}(\omega)$ are related in the two Plemelj formulae (2.5)–(2.6). This is in accordance with the references [57, 76]. Dispersion relations for physical systems can be derived by letting the function $\tilde{f}(\omega)$ be the

transfer function of the system, $\tilde{w}(\omega)$. For example, the Kramers-Kronig relations for the electric susceptibility $\chi(\omega)$ are found by letting $\chi = \tilde{f}(\omega)$ [57]. The electric susceptibility is related to the complex refractive index $n$ as $n = \sqrt{1+\chi}$, for non-magnetic materials. The relations (2.5)–(2.6) can be used as consistency checks for measurements and material models.

Titchmarsh's theorem can be extended to cover other functions and distributions than those in $\mathrm{L}^2$, which is often necessary: A function in $\mathrm{L}^2$ can be interpreted as a function of finite energy, and for this reason the input and output signals are frequently assumed to be in $\mathrm{L}^2$. The transfer function $\tilde{w}$, on the other hand, need in general *not* be in $\mathrm{L}^2$. The extension of Titchmarsh's theorem for many regular functions $\tilde{f} \notin \mathrm{L}^2$ can be handled with *subtractions* [57, 76]. Furthermore, there is also a version of the theorem for distributions of slow growth $\tilde{f} \in \mathcal{S}'$, in which case the right-hand sides of the Plemelj formulae (2.5)–(2.6) are convolutions with the Cauchy principal value distribution (see [76, Section 1.8] or [57, Section 17.3]).

Additional dispersion relations can be derived by exploiting the holomorphic properties of $\tilde{w}(\omega + \mathrm{i}\sigma)$ in the open upper half-plane due to point *4* in the theorem, see [57, 76] and references therein. A holomorphic function (sometimes referred to as an analytic function) is a function of the complex variable $z = \omega + \mathrm{i}\sigma$ that is complex-differentiable. As a result, very powerful tools from complex analysis, such as the Cauchy integral formula and theorem, can be employed. For an introduction to complex analysis, see e.g. [3, 32].

Furthermore, sum rules can be derived in some cases [57, 76]. Instead of relating functions to each other as the dispersion relations do, a sum rule relates a sum or integral (which is a generalized sum) to a fixed value. However, as already mentioned in the introduction, these derivations often rely on assumptions that are specific for each case. When the transfer function is not a regular function, the derivations become very cumbersome. A more straight-forward and generally applicable approach would be useful, and such an approach is presented in Section 2.4 (and in detail in Paper I). It relies solely on the assumption that the process under consideration is passive, which evidently holds for many physical systems. Moreover, the sum rules for passive systems can in turn be used to derive physical limitations.

## 2.3 Passive systems and Herglotz functions

In this section, the notion of passive systems in convolution form is presented. It is shown that the transfer functions of passive systems can be related to Herglotz functions, and that these functions fulfill a representation formula which is similar to the Plemelj formulae in Titchmarsh's theorem. Later, in Section 2.4, the representation is used in the approach to derive sum rules for passive systems.

Passivity means that the system cannot produce energy, and hence the energy content of the output signal is limited to that of the input. Depending on how the power and energy is modelled, the definition of passivity comes in two different forms, with names borrowed from electric circuit theory:

**Admittance-passivity:** The system (2.2) is admittance-passive if the energy ex-

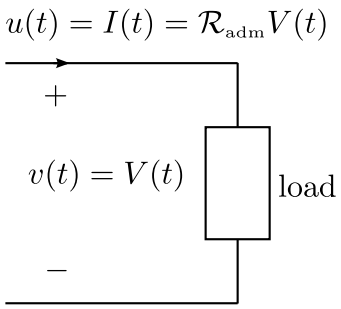



**Figure 5**: Here the input $V(t)$ is the electric voltage over the load, and the output $I(t) = \mathcal{R}_{\mathrm{adm}} V(t)$ is the electric current running through it. They are related by the admittance operator $\mathcal{R}_{\mathrm{adm}}$ of the load. The absorbed energy is given by (2.9).

pression

$$e_{\mathrm{adm}}(T) = \mathrm{Re} \int_{-\infty}^{T} u^*(t) v(t) \, \mathrm{d}t \tag{2.9}$$

is non-negative for all $T \in \mathbb{R}$ and $v \in \mathcal{D}$.

**Scatter-passivity:** The system (2.2) is scatter-passive if the energy expression

$$e_{\mathrm{scat}}(T) = \int_{-\infty}^{T} |v(t)|^2 - |u(t)|^2 \, \mathrm{d}t \tag{2.10}$$

is non-negative for all $T \in \mathbb{R}$ and $v \in \mathcal{D}$.

Here the superscript $^*$ denotes the complex conjugate. Only smooth input signals of compact support, $v \in \mathcal{D}$, are considered in order for the integrals to be well-defined. However, this is often enough to ensure that the corresponding energy expressions are non-negative also for other admissible input signals $v \in D(\mathcal{R})$. In the remainder of this General Introduction, a system or operator that is either admittance-passive or scatter-passive is simply referred to as passive.

If the input $v(t) = V(t)$ is the electric voltage (divided by 1 volt) over a load and the output $u(t) = I(t) = \mathcal{R}_{\mathrm{adm}} V(t)$ is the electric current (divided by 1 ampere) running through it, then the operator $\mathcal{R}_{\mathrm{adm}}$ is the admittance operator, see Figure 5. In this case, the electric energy absorbed by the load until time $T$ is given by (2.9) (for $v \in \mathcal{D}$), and thus the admittance operator of a passive circuit element is an admittance-passive operator. Note that admittance-passivity might as well have been called impedance-passivity, since the current could equally well have been the input and the voltage the output in the example.

Now let the input $v$ be the amplitude of the voltage wave travelling towards the load (measured by the load), $v = (V + I)/2$, and let the output $u$ be the amplitude of the reflected wave, $u = \mathcal{R}_{\mathrm{scat}} v = (V - I)/2$, see Figure 6. Here the operator $\mathcal{R}_{\mathrm{scat}}$ is the scattering (or reflection) operator, and the absorbed energy is given by (2.10) (for $v \in \mathcal{D}$). Hence, passive scattering operators are scatter-passive.

Note that the input and output signals are assumed to be complex-valued in the preceding paragraphs. Observable quantities in the real world are real, but even

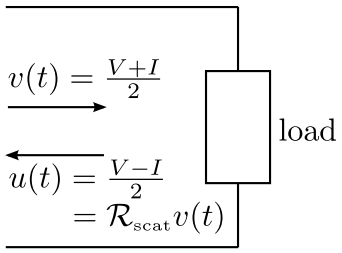



**Figure 6**: In this case, the input $v(t)$ and output $u(t)$ are the amplitudes (measured by the load) of the voltage waves traveling towards and from the load, respectively. They are related by the scattering (or reflection) operator $\mathcal{R}_{\text{scat}}$. The absorbed energy is given by (2.10).

real physical processes can sometimes be modelled with complex-valued functions in the time domain. For example, consider signals that are almost time-harmonic with amplitudes varying over a timescale much larger than the dominating frequency. Another example is the scattering of elliptically polarized electromagnetic plane waves, see Section 4.3.

Passivity has far-reaching implications on the physically realizable behaviour of a system. One consequence is that a passive system must also be causal. Another is that the impulse response $w$ is a distribution of slow growth, $w \in \mathcal{S}'$, and therefore Fourier transformable in the distributional sense, as discussed above. Combined, it guarantees that the transfer function $\tilde{w}(\omega + \mathrm{i}\sigma)$ is well-defined in the open upper half-plane ($\sigma > 0$), and that it is holomorphic there. For conciseness, the notation $z = \omega + \mathrm{i}\sigma$, with $\omega = \operatorname{Re} z$ and $\sigma = \operatorname{Im} z$, is used from now on.[1] Furthermore, the following theorem can be proved (cf. Theorem 3.1 in Paper I and references therein):

**Theorem 2.2.** *Let $\mathcal{R} = w*$ be a convolution operator and let $\tilde{w}$ be the transfer function given by* (2.3)*. If $\mathcal{R}$ is admittance-passive, then $\operatorname{Re}\tilde{w}(z) \geqslant 0$ for $\operatorname{Im} z > 0$. If $\mathcal{R}$ is scatter-passive, then $|\tilde{w}(z)| \leqslant 1$ for $\operatorname{Im} z > 0$. In both cases, $\tilde{w}$ is holomorphic in the open upper half plane $\{z : \operatorname{Im} z > 0\}$.*

The properties of scatter-passive systems were derived by Youla et al. in [110], while the results for admittance-passivity were presented by Zemanian in [111]; the connection between them was discussed by Wohlers and Beltrami in [107]. Both passivity concepts have been generalized to more general Banach-space-valued distributions, see [113] and references therein.

The transfer function of a passive systems can be related to a Herglotz function, which is defined as follows (see also Figure 7):

**Definition 2.1.** *A Herglotz function $h$ is a holomorphic function mapping the open upper half plane $\{z : \operatorname{Im} z > 0\}$ into the closed upper half plane $\{z : \operatorname{Im} z \geqslant 0\}$.*

Herglotz functions are sometimes referred to as Nevanlinna [50], Pick [21], or R-functions [55]. They are closely related to positive harmonic functions and the Hardy

[1] Note that in Paper I, the notation instead allows the angular frequency $\omega = \omega' + \mathrm{i}\omega''$ to take complex values, where $\omega' = \operatorname{Re}\omega$ and $\omega'' = \operatorname{Im}\omega$.

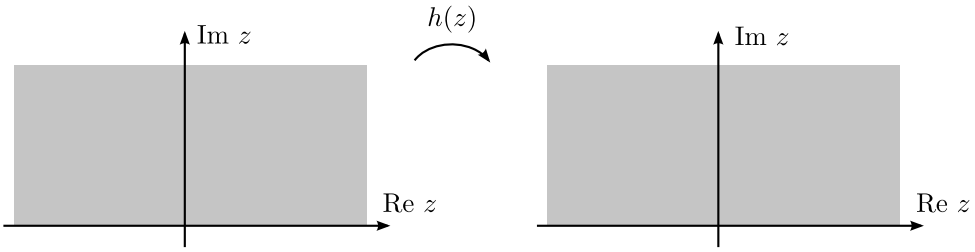



**Figure 7**: A Herglotz function $h$ is a holomorphic function, mapping the open upper half plane $\{z : \operatorname{Im} z > 0\}$ into the closed upper half plane $\{z : \operatorname{Im} z \geqslant 0\}$.

space $H^\infty(\mathbb{C}^+)$ [22], and appear within a wide range of problems, see e.g. [4, 5, 51].

For an admittance-passive system, a Herglotz function is simply given by

$$h(z) = \mathrm{i}\tilde{w}_{\mathrm{adm}}(z). \tag{2.11}$$

For a scatter-passive system, a Herglotz function can be found by applying the inverse Cayley transform $z \mapsto (\mathrm{i}z + \mathrm{i})/(1 - z)$ to $\tilde{w}_{\mathrm{scat}}(z)$, which is nothing else than moving from reflection coefficient to admittance in circuit theory, cf. Figure 5–6. Alternatively, the complex logarithm may be used:

$$h(z) = -\mathrm{i} \log \left( \frac{\tilde{w}_{\mathrm{scat}}(z)}{B(z)} \right), \tag{2.12}$$

which first requires the complex zeros $z_n$ of $\tilde{w}_{\mathrm{scat}}(z)$ in the open upper half plane to be factored out with a Blaschke product:

$$B(z) = \left( \frac{z - \mathrm{i}}{z + \mathrm{i}} \right)^k \prod_{z_n \neq \mathrm{i}} \frac{|z_n^2 + 1|}{z_n^2 + 1} \frac{z - z_n}{z - z_n^*}. \tag{2.13}$$

Here the complex zeros $z_n$ are repeated according to their multiplicity, and $k \geqslant 0$ is the order of the possible zero at $z = \mathrm{i}$, see Paper I for details. Moreover, the composition of two Herglotz functions is also a Herglotz function, a fact that can be useful when deriving sum rules and limitations. It should be mentioned here that some authors prefer the Laplace transform and the related function class of Positive Real (PR) functions over the Fourier transform and Herglotz functions, see Paper III.

There is a representation theorem for all Herglotz functions, which resembles the Plemelj formulae (2.5) and the Hilbert transform (2.7):

**Theorem 2.3.** *A necessary and sufficient condition for a function $h$ to be a Herglotz function is that*

$$h(z) = \beta z + \alpha + \int_{\mathbb{R}} \left( \frac{1}{\xi - z} - \frac{\xi}{1 + \xi^2} \right) \mathrm{d}\mu(\xi), \quad \operatorname{Im} z > 0, \tag{2.14}$$

*where $\beta \geqslant 0$, $\alpha \in \mathbb{R}$ and $\mu$ is a positive Borel measure such that $\int_{\mathbb{R}} \mathrm{d}\mu(\xi)/(1+\xi^2) < \infty$.*

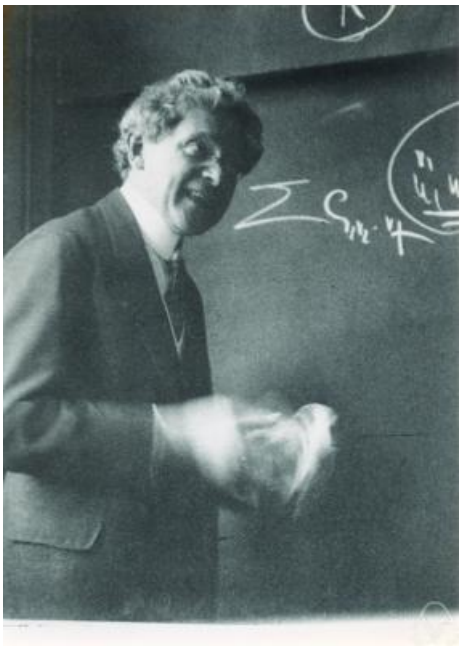


**Figure 8**: The german mathematician Gustav Herglotz (1881–1953) in the year 1930. The photograph is courtesy of Konrad Jacobs, reproduced here under license http://creativecommons.org/licenses/by-sa/3.0/de/deed.en.

The representation theorem is commonly attributed to Nevanlinna's paper [74], but it was presented in its final form by Cauer in [15]. See also [5] for a proof and discussion. The theorem follows from a similar representation theorem for positive harmonic functions on the unit disk due to Herglotz [52], hence the name Herglotz functions. A photograph of Gustav Herglotz can be found in Figure 8.

The representation (2.14) is in some sense a dispersion relation, since the measure $\mu$ can be interpreted as the imaginary part of $h$, see Lemma 4.1 in Paper I and the discussion following the lemma. For example, when the measure is given by a continuous function $\mu'(\xi)$, i.e. $\mathrm{d}\mu(\xi) = \mu'(\xi)\,\mathrm{d}\xi$, then $\lim_{\sigma\to 0^+} \operatorname{Im} h(\omega + \mathrm{i}\sigma) = \mu'(\omega)$ for almost all $\omega \in \mathbb{R}$. The Lebesgue integral over the measure $\mu$ in (2.14) is a generalization of the Riemann integral, and often appears in representation theorems such as Theorem 2.3. The interested reader can find an introduction to measure and integration theory in e.g. [8, 81].

## 2.4 Sum rules and physical limitations for passive systems

In this section, the general approach presented in Paper I to derive sum rules and constraints based on the previous results is summarized. From the representation theorem, it follows that all Herglotz functions have low- and high-frequency asymp-

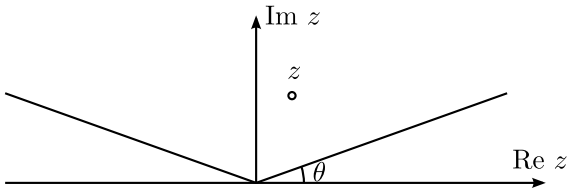



**Figure 9**: The Stoltz domain used in the low- and high-frequency asymptotic expansions (2.15)–(2.16): $\{z : \theta \leqslant \arg z \leqslant \pi - \theta\}$ for some $\theta \in (0, \pi/2]$.

totic expansions of the forms:

$$h(z) = \sum_{n=-1}^{2N-1} a_n z^n + o(z^{2N-1}), \quad \text{as } z\hat{\to}0, \tag{2.15}$$

$$h(z) = \sum_{1-2M}^{m=1} b_m z^m + o(z^{1-2M}), \quad \text{as } z\hat{\to}\infty, \tag{2.16}$$

where all $a_n, b_m \in \mathbb{R}$ and $N, M \geqslant 0$. The Landau symbols $o$ (little-o) and $\mathcal{O}$ (big-O) are used throughout this thesis, cf. [47]. Here $z\hat{\to}0$ is a short-hand notation for $|z| \to 0$ in the Stoltz domain $\theta \leqslant \arg z \leqslant \pi - \theta$ for any $\theta \in (0, \pi/2]$, see Figure 9, and likewise for $z\hat{\to}\infty$. The Stoltz domain ensures that the low-frequency asymptotic expansion only depends on the behaviour of the physical system for arbitrarily large times. Similarly, the high-frequency asymptotic expansion is determined by the response of the physical system for arbitrarily short times, cf. Section 3 of Paper I.

At the core of Paper I are the following integral identities for a Herglotz function $h$:

$$\lim_{\varepsilon\to0^+} \lim_{\sigma\to0^+} \frac{1}{\pi} \int_{\varepsilon<|\omega|<\varepsilon^{-1}} \frac{\operatorname{Im} h(\omega + \mathrm{i}\sigma)}{\omega^p} \,\mathrm{d}\omega = a_{p-1} - b_{p-1}, \quad p = 2-2M, 3-2M, \ldots, 2N. \tag{2.17}$$

The left-hand side of (2.17) is the integral of $\operatorname{Im} h(\omega)/\omega^p$ in the distributional sense, i.e. contributions from possible singularities in the interval $(0, \infty)$ are included, cf. the discussion in Paper I. Therefore, it is convenient to write

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\operatorname{Im} h(\omega)}{\omega^p} \,\mathrm{d}\omega = a_{p-1} - b_{p-1}, \quad p = 2-2M, 3-2M, \ldots, 2N. \tag{2.18}$$

where (2.18) is understood to be replaced by (2.17) whenever necessary.

One property that many physical systems obey has not yet been discussed:

**Reality:** The system (2.1) is real if it maps real-valued input $v$ to real-valued output $u$.

For many physical systems, reality is taken for granted. Reality implies that the impulse response $w(t)$ is real-valued, which in turn implies the symmetry $h(z) =$

$-h^*(-z^*)$ when $h(z)$ is given by (2.11). This restricts the identities (2.18) to even powers $p = 2\hat{p}$ and simplifies them to

$$\frac{2}{\pi}\int_0^\infty \frac{\operatorname{Im} h(\omega)}{\omega^{2\hat{p}}}\,\mathrm{d}\omega = a_{2\hat{p}-1} - b_{2\hat{p}-1}, \quad \hat{p} = 1-M,\ldots,N. \tag{2.19}$$

The derivation of the integral identities (2.17) for $p = 2, 3, \ldots, 2N$ relies on two results; the first (Corollary 4.1 of Paper 1) relates the left-hand sides to moments of the measure $\mu$ in the representation, while the other (Lemma 4.2) relates the convergence and explicit values of these moments to the low-frequency expansion (2.15). A change of variables in the left-hand side of (2.17) enables a proof for $p = 2-2M, 3-2M, \ldots, 1$. It should be noted that if the Herglotz function is e.g. a rational function, the identities (2.17) follow from the Cauchy integral formula (see derivation in [94]).

The integral identities (2.19) constitute a set of sum rules for the transfer function $\tilde{w}(\omega)$ when the Herglotz function $h$ is given by (2.11); the sum rules relate the dynamical behaviour of the system (the left-hand side in (2.19)) to its low- and/or high-frequency properties (the coefficients $a_{2\hat{p}-1}$ and $b_{2\hat{p}-1}$ in the right-hand side). The asymptotic expansions must be found by physical arguments.

Moreover, the identities (2.19) are the starting points to derive physical limitations on a system: Since the integrands are non-negative and the integrals over the positive real line are equal to the right-hand sides, the integrals over any subset of the positive real line must be bounded by the right-hand sides. Let the frequency interval be $\mathcal{B} = [\omega_0(1-B/2), \omega_0(1+B/2)]$, with center frequency $\omega_0$ and fractional bandwidth $B$ ($0 \leqslant B \leqslant 2$), then the following physical limitations may be derived from (2.19):

$$\inf_{\mathcal{B}} \operatorname{Im} h(\omega) \frac{2}{\pi}\int_{\omega_0(1-B/2)}^{\omega_0(1+B/2)} \frac{1}{\omega^{2\hat{p}}}\,\mathrm{d}\omega \leqslant a_{2\hat{p}-1} - b_{2\hat{p}-1}. \tag{2.20}$$

Note that

$$\int_{\omega_0(1-B/2)}^{\omega_0(1+B/2)} \frac{1}{\omega^{2\hat{p}}}\,\mathrm{d}\omega \approx \frac{1}{\omega_0^{2\hat{p}-1}} B, \quad \text{for } B \ll 1,$$

and

$$\int_{\omega_0(1-B/2)}^{\omega_0(1+B/2)} \frac{1}{\omega^{2\hat{p}}}\,\mathrm{d}\omega \leqslant \frac{1}{\omega_0^{2\hat{p}-1}} B, \quad \text{for } p = 0, 1, \ldots$$

The limitations state that the imaginary part of $h$ cannot be arbitrarily large over a frequency interval. Often, the imaginary part of $h$ models the losses of the physical system; for an impedance $\tilde{w}(\omega) = Z(\omega)$ in (2.11), the imaginary part of $h$ is the real part of $Z$. The concept of sum rules and physical limitations is illustrated schematically in Figure 10.

In short, the general approach in Paper I can be described as follows, for the case of a real-valued impulse response $w$:

1. Find the transfer function $\tilde{w}$ of the system and the corresponding Herglotz function $h$. This can always be done for passive systems in convolution form.

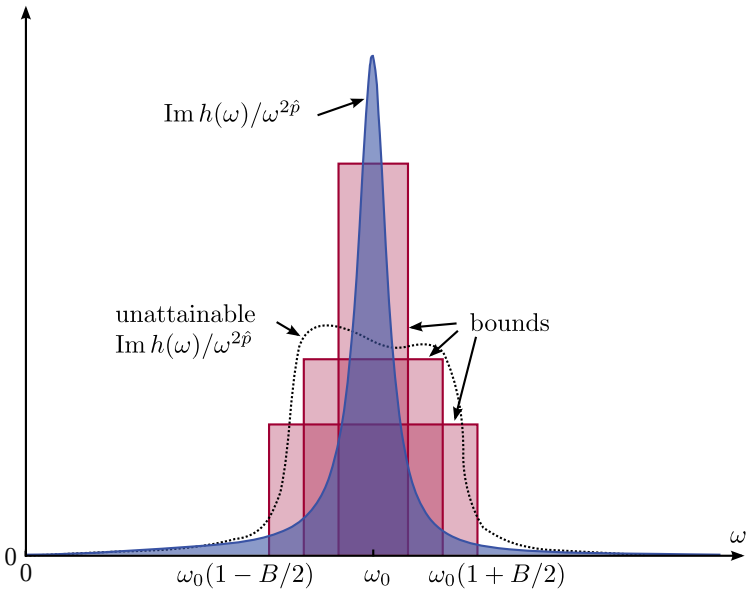



**Figure 10**: Illustration of the sum rules (2.19) and physical limitations (2.20). The integral in the left-hand side of (2.19) corresponds to the area under the curve $\operatorname{Im} h(\omega)/\omega^{2\hat{p}}$, and it is equal to the low-frequency coefficients $a_{2\hat{p}-1}$ and/or high-frequency coefficients $b_{2\hat{p}-1}$ in the right-hand side. Since the integrand is non-negative, the integral over any finite frequency interval is bounded by the right-hand side; therefore the curve has to intersect any box with the same area, which gives the physical limitations (2.20). Also shown in the figure is an unattainable curve of $\operatorname{Im} h(\omega)/\omega^{2\hat{p}}$ (dotted); it does not intersect all the boxes.

2. Determine the low- and/or high-frequency asymptotic expansions in (2.15)–(2.16) from physical arguments. Depending on the asymptotic expansions, a number of sum rules (2.19) follow directly.

3. The physical limitations (2.20) follow from considering finite frequency intervals $\mathcal{B}$.

Furthermore, in step 1, it can be used that the composition of two Herglotz functions is also a Herglotz function, which can lead to new sum rules and limitations in step 2–3.

The reader is referred to Section 5 of Paper 1 for examples of practical uses of the general approach to derive sum rules and physical limitations for passives systems. The method is also used to determine the sum rules and physical limitations for scattering and matching of electromagnetic spherical waves in Papers II–IV, see also Section 4.4 of this General Introduction.

Other applications where the method has been very successful include passive metamaterials in [36] (see also Section 5.4 in Paper I) and high-impedance surfaces in [42]; metamaterials are synthetic materials designed to have extra-ordinary electromagnetic properties, and can potentially be used for cloaking [84]; similarly, high-impedance surfaces are synthetic surfaces that can be used, for example, instead of ordinary ground planes to improve antenna performance. In both cases,

bandwidth limitations of the type (2.20) are derived to show that the extra-ordinary properties are achievable only in a finite frequency interval.

# 3 Electromagnetic spherical waves

Scattering and absorption of electromagnetic waves is the cornerstone in many applications, as discussed in Section 1 of this General Introduction. A large part of this thesis is concerned with electromagnetic spherical waves, which form a basis in which the electric and magnetic fields can be written. Obviously, spherical waves are beneficial when analysing scattering by spherical particles (Mie theory), but they are also used within the computational method known as the T-matrix (or null-field) method; the T-matrix method was introduced by Waterman in [105], and has since then been developed further by many authors and used extensively for analysing electromagnetic, acoustic and elastodynamic scattering by composite objects and systems of particles. The interested reader is referred to [64, 70] and references therein.

In this thesis, the focus is on scattering of spherical waves in antenna theory. Spherical waves are useful in that they provide a complete description of the transmitting, receiving and scattering properties of an antenna in terms of the antenna scattering matrix. This description is especially convenient for small antennas, where only a few terms are needed. Furthermore, spherical waves are crucial in data-processing in spherical near-field antenna measurements, since they provide the means to determine the far-field radiation pattern of the antenna under test from measured near-field data [44, 62].

What's more, spherical waves have also been used to model multiple antenna systems: multiple-input multiple-output (MIMO) technology is one way to increase the capacity in wireless systems and satisfy the increasing demands for higher data rates, since the capacity can increase linearly with the number of elements under the right conditions [77]. The right conditions are met if the antenna elements can transmit and receive independent data streams, and this depends on a complicated interplay between the antennas and the environment (the propagation channel). It is common to use models based on plane waves, or multi-paths, in theoretical and experimental analysis of this interplay [72], but spherical waves can also be used. This was done in the thesis [28] by Alayón Glaunov, where many important theoretical results were derived. Other have also used spherical waves in MIMO-related problems, see e.g. [69, 71].

However, there are no previous publications where spherical waves are used instead of plane waves in channel measurements. In Paper V, a method to do so is proposed, and the results of a measurement campaign is presented. The method shows some similarities with spherical near-field antenna measurements.

The outline of this section is the following: Section 3.1 describes the spherical waves in more detail and introduces the antenna scattering matrix. Section 3.2 briefly describes spherical near-field antenna measurements. MIMO in general, and some important results from Alayón Glazunov's thesis [28] in particular, are reviewed

in Section 3.3. Finally, the channel measurements of Paper V are summarized in Section 3.4.

## 3.1 Background

This section gives more background on electromagnetic spherical waves. The ambition is to give the necessary information without too much technicalities, but some formulas are unavoidable. First it is shown how the spherical waves follow naturally from Maxwell's equations for the electric and magnetic fields in free space when spherical coordinates are used. Later, expressions for the power flux and the far-field in terms of spherical waves are presented. The antenna scattering matrix, which describes the properties of an antenna in terms of spherical waves, is also introduced. The interested reader can find more details on spherical waves in the references [12, 44, 61, 73] and in the appended Papers II–IV. The notation in this General Introduction, as well as in the appended papers, aims to follow that in Kristensson's book [61], but with some minor modifications where appropriate.

The electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{H}$ are modelled with Maxwell's equations, which in free space take the form:

$$\begin{cases} \nabla \times \boldsymbol{E}(\boldsymbol{r},\omega) = \mathrm{i}k\eta_0 \boldsymbol{H}(\boldsymbol{r},\omega) \\ \eta_0 \nabla \times \boldsymbol{H}(\boldsymbol{r},\omega) = -\mathrm{i}k \boldsymbol{E}(\boldsymbol{r},\omega). \end{cases} \tag{3.1}$$

Here $\boldsymbol{r}$ is the spatial coordinate and $\eta_0$ is the wave impedance in free space. The free space wavenumber is $k = \omega/\mathrm{c} = 2\pi f/\mathrm{c} = 2\pi/\lambda$, where $\omega$ is the angular frequency, $f$ is the frequency, c is the speed of light in free space, and $\lambda$ is the free space wavelength. It is convenient to separate the electric and magnetic fields by combining the two equations. In a source-free region, the electric and magnetic fields are divergenceless, viz.

$$\nabla \cdot \boldsymbol{E} = \eta_0 \nabla \cdot \boldsymbol{H} = \boldsymbol{0}, \tag{3.2}$$

which together with (3.1) gives the vector Helmholtz' equation:

$$\nabla^2 \boldsymbol{E}(\boldsymbol{r},\omega) + k^2 \boldsymbol{E}(\boldsymbol{r},\omega) = \boldsymbol{0}. \tag{3.3}$$

The same equation holds also for the magnetic field $\boldsymbol{H}$.

In this section, it is seen that the spherical waves are a set of basis functions which are solutions to the equations (3.2)–(3.3), and thus can be used to describe the electric and magnetic fields. Spherical coordinates $(r, \theta, \phi)$ are used here, and they are related to the Cartesian coordinates $(x, y, z)$ as

$$\begin{cases} x = r\sin\theta\cos\phi \\ y = r\sin\theta\sin\phi \\ z = r\cos\theta, \end{cases} \tag{3.4}$$

where $r = |\boldsymbol{r}|$, see Figure 11. The angles are referred to as the polar angle, $\theta$, and the azimuth angle, $\phi$. The unit vectors are denoted $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$, respectively.

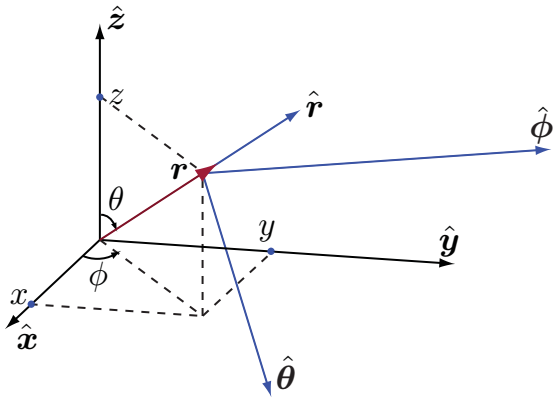



**Figure 11**: The spherical coordinate system $(r, \theta, \phi)$ (with unit-vectors $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$), and the Cartesian coordinate system $(x, y, z)$ (with unit-vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$), related by (3.4).

The basis functions could have been based on another type of coordinate system where the vector Helmholtz' equation (3.3) separate, such as planar or cylindrical coordinates. However, spherical coordinates is the only option that leads to a countable set of basis functions, i.e. a basis in which the electric and magnetic fields can be written as sums over the basis functions. Planar and cylindrical coordinates lead to uncountable sets of basis functions, and the electric and magnetic fields are given by integrals over the basis functions. The reason is that the area of the unit sphere is finite, whereas the area of a cylinder (with infinite height) or a plane is infinite.

### 3.1.1 Scalar spherical harmonics

The first step to construct the spherical waves is a set of basis functions on the unit sphere. The sought for basis is the set of scalar spherical harmonics $Y_{ml}$:

$$Y_{ml}(\theta, \phi) = (-1)^m \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) \mathrm{e}^{\mathrm{i}m\phi}, \tag{3.5}$$

where the range of the indices are $l = 0, 1, \ldots$ and $m = -l, -l+1, \ldots, l$. The basis functions in the cosine of the polar angle $\cos\theta$ are associated Legendre functions $P_l^m$ [6], whereas the basis functions in the azimuth angle $\phi$ simply are complex exponentials $\mathrm{e}^{\mathrm{i}m\phi}$. In some cases, discussed below, it is more convenient to use $\cos(m\phi)$ and $\sin(m\phi)$ instead. To simplify the notation, the unit vector $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$ is used instead of the arguments $(\theta, \phi)$ from now on, see Figure 11. The reason for the normalization in the definition (3.5) is to make the scalar spherical harmonics orthonormal on the unit sphere, viz.

$$\int_{\Omega_{\hat{\boldsymbol{r}}}} Y_{ml}(\hat{\boldsymbol{r}}) Y^*_{m'l'}(\hat{\boldsymbol{r}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \delta_{m,m'} \delta_{l,l'}.$$

where $\Omega_{\hat{\boldsymbol{r}}} = \{(\theta, \phi) : 0 \leqslant \theta < \pi, 0 \leqslant \phi < 2\pi\}$ is the unit sphere and $\mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$ is its surface element. The Kronecker delta $\delta_{m,m'}$ is equal to 1 when $m = m'$ and 0 otherwise.

The scalar spherical harmonics is a basis for square-integrable functions $G(\hat{\boldsymbol{r}})$ on the unit sphere (or any other spherical surface centered at the origin). Recall from Section 2 that a square-integrable function can be interpreted as a function of finite energy. The norm $||\cdot||$ for such functions is defined as

$$||G||^2 = \int_{\Omega_{\hat{\boldsymbol{r}}}} |G(\hat{\boldsymbol{r}})|^2 \,\mathrm{d}\Omega_{\hat{\boldsymbol{r}}}. \tag{3.6}$$

The function function $G$ is written in the basis $Y_{ml}$ as

$$G(\hat{\boldsymbol{r}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} g_{ml} Y_{ml}(\hat{\boldsymbol{r}}), \tag{3.7}$$

where the coefficients $g_{ml}$ are given by

$$g_{ml} = \int_{\Omega_{\hat{\boldsymbol{r}}}} G(\hat{\boldsymbol{r}}) Y_{ml}^*(\hat{\boldsymbol{r}}) \,\mathrm{d}\Omega_{\hat{\boldsymbol{r}}}, \tag{3.8}$$

and the sum converges in the norm $||\cdot||$. A vector-valued function (such as the electric or magnetic fields) can of course also be written in the basis $Y_{ml}$ by letting the coefficients $g_{ml}$ be vectors. However, it is beneficial to let $g_{ml}$ be scalars and instead let the vector character be contained in the basis functions, which leads to the vector spherical harmonics $\boldsymbol{A}_{\tau ml}$.

### 3.1.2 Vector spherical harmonics

To construct a vector-valued basis in three-dimensional space, the three vector spherical harmonics $\boldsymbol{A}_{\tau ml}$ are used:

$$\begin{cases} \boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}) = \dfrac{1}{\sqrt{l(l+1)}} \nabla \times (\boldsymbol{r} Y_{ml}(\hat{\boldsymbol{r}})) \\ \boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}}) = \dfrac{1}{\sqrt{l(l+1)}} r \nabla Y_{ml}(\hat{\boldsymbol{r}}) \\ \boldsymbol{A}_{3ml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}} Y_{ml}(\hat{\boldsymbol{r}}). \end{cases}$$

The vector spherical harmonics are also orthonormal on the unit sphere, viz.

$$\int_{\Omega_{\hat{\boldsymbol{r}}}} \boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{A}_{\tau ml}^*(\hat{\boldsymbol{r}}) \,\mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \delta_{\tau,\tau'} \delta_{m,m'} \delta_{l,l'},$$

and the equations corresponding to (3.6)–(3.8) for square-integrable vector-valued functions $\boldsymbol{G}$ on the unit sphere are

$$||\boldsymbol{G}||^2 = \int_{\Omega_{\hat{\boldsymbol{r}}}} \boldsymbol{G}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{G}^*(\hat{\boldsymbol{r}}) \,\mathrm{d}\Omega_{\hat{\boldsymbol{r}}},$$

$$\boldsymbol{G}(\hat{\boldsymbol{r}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{\tau=1}^{3} g_{\tau ml} \boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}}),$$

and

$$g_{\tau ml} = \int_{\Omega_{\hat{\boldsymbol{r}}}} \boldsymbol{G}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{A}^*_{\tau ml}(\hat{\boldsymbol{r}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}}.$$

Later in Section 3.1.5, it is seen that the vector spherical harmonics are a basis for the far-field pattern of an antenna or scatterer. The definition of the vector spherical harmonics are chosen cleverly, so that

$$\begin{cases} \hat{\boldsymbol{r}} \cdot \boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}} \cdot \boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}}) = 0 \\ \hat{\boldsymbol{r}} \times \boldsymbol{A}_{3ml}(\hat{\boldsymbol{r}}) = \boldsymbol{0} \end{cases} \qquad \begin{cases} \boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}) = \boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}}) \times \hat{\boldsymbol{r}} \\ \boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}} \times \boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}). \end{cases} \tag{3.9}$$

Note especially that only $\boldsymbol{A}_{3ml}(\hat{\boldsymbol{r}})$ has a component in the radial direction, and that $\boldsymbol{A}_{1ml}$ and $\boldsymbol{A}_{2ml}$ switch places when cross multiplied with the radial unit vector $\hat{\boldsymbol{r}}$. These relations will come in handy later on. The vector spherical harmonics for positive and negative $m$ are related as

$$\boldsymbol{A}_{\tau,-m,l}(\hat{\boldsymbol{r}}) = (-1)^m \boldsymbol{A}^*_{\tau ml}(\hat{\boldsymbol{r}}). \tag{3.10}$$

A selection of vector spherical harmonics $\boldsymbol{A}_{\tau ml}$ are depicted in Figure 12.

### 3.1.3 Vector spherical waves

The ambition formulated in the beginning of this section is to find a set of basis functions which are solutions to the equations (3.2)–(3.3), but so far only bases for square-integrable functions on the unit sphere have been considered. To proceed, solutions on the form

$$\boldsymbol{E}(\boldsymbol{r}, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{\tau=1}^{3} e_{\tau ml}(kr) \boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}}) \tag{3.11}$$

are tried in (3.3). This gives conditions on the functions $e_{\tau ml}(kr)$, and eventually it is seen that each pair $l, m$ gives three coupled differential equations involving $e_{\tau ml}(kr)$ for $\tau = 1, 2, 3$. However, only two of them give solutions which satisfy (3.2). The solutions of these inserted into (3.11) give the vector spherical waves, and they depend on the kind of boundary conditions imposed on the differential equations; the most widely used are the incoming, outgoing and regular spherical waves.

In scattering and antenna problems, it is common to use the incoming and outgoing vector spherical waves, which are defined as

$$\begin{cases} \boldsymbol{u}^{(j)}_{1ml}(k\boldsymbol{r}) = \mathrm{h}^{(j)}_l(kr) \boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}) \\ \boldsymbol{u}^{(j)}_{2ml}(k\boldsymbol{r}) = \dfrac{(kr\mathrm{h}^{(j)}_l(kr))'}{kr} \boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}}) + \sqrt{l(l+1)} \dfrac{\mathrm{h}^{(j)}_l(kr)}{kr} \boldsymbol{A}_{3ml}(\hat{\boldsymbol{r}}), \end{cases}$$

where a prime denotes differentiation with respect to the argument $kr$. The outgoing waves ($j = 1$) contain spherical Hankel functions of the first kind and order $l$ [6], denoted $\mathrm{h}^{(1)}_l(kr)$, whereas the incoming waves ($j = 2$) contain spherical Hankel

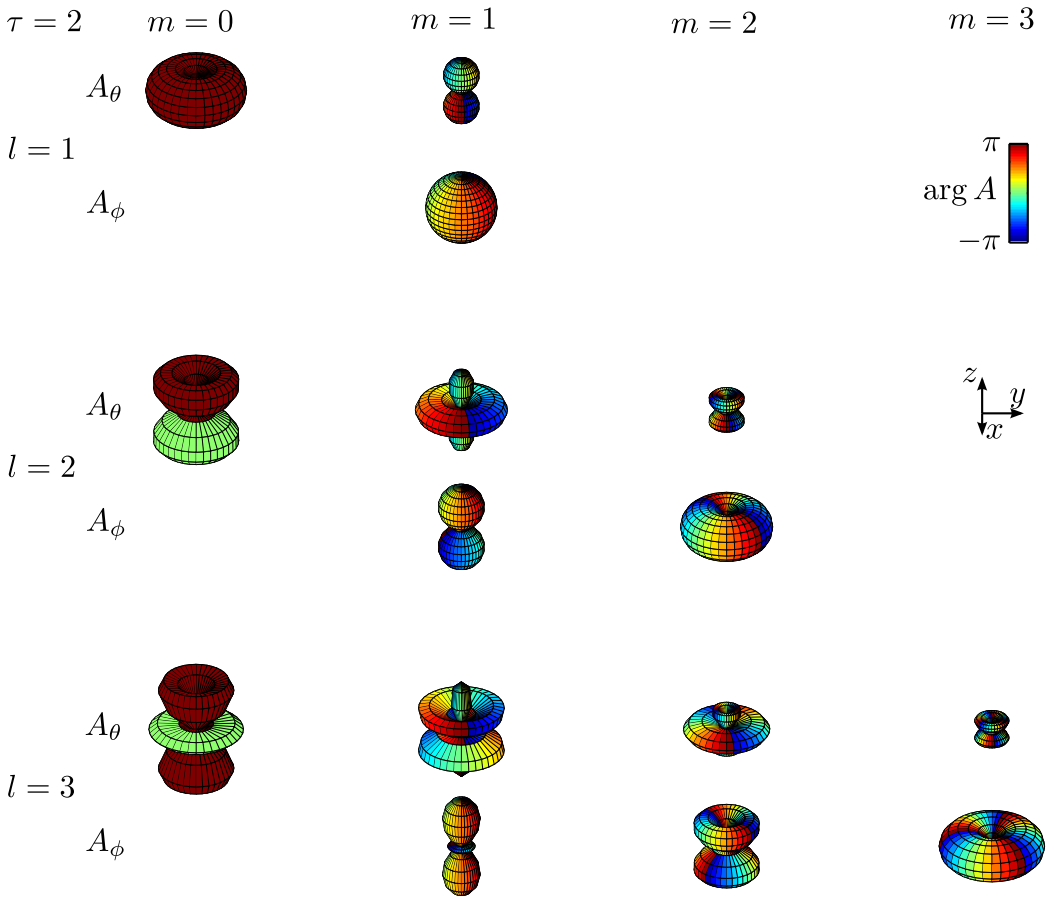



**Figure 12**: Vector spherical harmonics $\boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}})$ for $\tau = 2$, $m \geqslant 0$, and $l = 1, 2, 3$. The graph depicts the theta-component $A_\theta(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{\theta}} \cdot \boldsymbol{A}(\hat{\boldsymbol{r}})$, and the phi-component $A_\phi(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{\phi}} \cdot \boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}})$. For each surface, the radial-component is proportional to the magnitude $|A|$ and the colour is given by the argument $\arg A$. The vector spherical harmonics for $\tau = 1$ are given by exchanging $A_\theta$ and $A_\phi$ (due to (3.9)), and the vector spherical harmonics for $m < 0$ are given by the complex conjugate (due to (3.10)). The vector spherical harmonics for $\tau = 3$ and/or $l = 0$ are not depicted, since they do not appear in the expansion (3.15) of the far-field of a scatterer or antenna. The first graph, $A_\theta$ for $\tau = 2$, $m = 0$, and $l = 1$, is the characteristic doughnut shaped radiation pattern of a $z$-directed electric dipole.

functions of the second kind, $\mathrm{h}_l^{(2)}(kr)$. As the name suggests, the spherical waves are incoming and outgoing with respect to the origin $\boldsymbol{r} = \boldsymbol{0}$, see Section 3.1.4. It is also common to use regular waves $\boldsymbol{v}(k\boldsymbol{r})$, which are equal amounts of incoming and outgoing waves,

$$\boldsymbol{v}(k\boldsymbol{r}) = \frac{\boldsymbol{u}^{(1)}(k\boldsymbol{r}) + \boldsymbol{u}^{(2)}(k\boldsymbol{r})}{2},$$

and therefore regular at the origin $\boldsymbol{r} = \boldsymbol{0}$.

Any electric field $\boldsymbol{E}$ which is the solution to (3.2)–(3.3) can, in a source-free region enclosed by spherical surfaces, be written as a sum of incoming and outgoing vector spherical waves:

$$\boldsymbol{E}(\boldsymbol{r}, k) = k\sqrt{2\eta_0} \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \sum_{\tau=1}^{2} b_{\tau ml}^{(1)} \boldsymbol{u}_{\tau ml}^{(1)}(k\boldsymbol{r}) + b_{\tau ml}^{(2)} \boldsymbol{u}_{\tau ml}^{(2)}(k\boldsymbol{r}). \tag{3.12}$$

Here the range of the $l$-index is $l = 1, 2, \ldots$, since the spherical waves vanish for $l = 0$.

The magnetic field $\boldsymbol{H}$ is also a solution to (3.2)–(3.3), and can thus be written as a sum of spherical waves. Thanks to the clever choice of definitions for the vector spherical harmonics $\boldsymbol{A}_{\tau ml}$, it follows that $\boldsymbol{u}_{1ml}^{(j)}$ and $\boldsymbol{u}_{2ml}^{(j)}$ switch places under rotation $\nabla\times$, viz.

$$\begin{cases} \boldsymbol{u}_{1ml}^{(j)}(k\boldsymbol{r}) = \dfrac{\nabla \times \boldsymbol{u}_{2ml}^{(j)}(k\boldsymbol{r})}{k} \\ \boldsymbol{u}_{2ml}^{(j)}(k\boldsymbol{r}) = \dfrac{\nabla \times \boldsymbol{u}_{1ml}^{(j)}(k\boldsymbol{r})}{k}. \end{cases}$$

These relations together with Maxwell's equations (3.1) give

$$\boldsymbol{H}(\boldsymbol{r}, k) = \frac{k\sqrt{2}}{\mathrm{i}\sqrt{\eta_0}} \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \sum_{\tau=1}^{2} b_{\tau ml}^{(1)} \boldsymbol{u}_{\bar{\tau} ml}^{(1)}(k\boldsymbol{r}) + b_{\tau ml}^{(2)} \boldsymbol{u}_{\bar{\tau} ml}^{(2)}(k\boldsymbol{r}), \tag{3.13}$$

where the dual index $\bar{\tau} = 3 - \tau$ has been introduced. In other words, the electric and magnetic fields are described by the same coefficients $b_{\tau ml}^{(j)}$, where $\tau = 1$ and $\tau = 2$ have switched places.

It is convenient to introduce a multi-index $\nu = 2(l^2 + l - 1 + m) + \tau$ instead of the indices $\{\tau, m, l\}$, since the multi-index allows the three sums in (3.12)–(3.13) to be replaced by one sum, $\sum_{\nu=0}^{\infty}$. The $l$-index denotes the order of the spherical waves, i.e. $l = 1$ denotes dipoles, $l = 2$ quadrupoles, and so on. Furthermore, since only the vector waves $\boldsymbol{u}_{2ml}^{(j)}$ (with $\tau = 2$) have a component in the radial direction due to (3.9), it is clear that $\tau = 1$ (odd $\nu$) corresponds to a transverse-electric wave ($\mathrm{TE}_l$-mode or magnetic $2^l$-pole), while $\tau = 2$ (even $\nu$) identifies a transverse-magnetic wave ($\mathrm{TM}_l$-mode or electric $2^l$-pole), when the electric and magnetic fields are given by (3.12)–(3.13).

It was mentioned above that $\sin(m\phi)$ and $\cos(m\phi)$ can be used instead of the exponentials $\mathrm{e}^{\mathrm{i}m\phi}$ in (3.5). Both ways have their pros and cons: When $\sin(m\phi)$ and $\cos(m\phi)$ are used, the three electric and three magnetic dipoles ($l = 1$) can be

identified with $x$-, $y$- and $z$-directed dipoles (see Paper II). This is beneficial in the theoretical argumentation in Paper II and Paper IV. The exponentials $\mathrm{e}^{\mathrm{i}m\phi}$, however, are more convenient when translations and rotations are used, like for example in the measurement problem considered in Paper V. The exponentials are also chosen in Paper III, where $\mathrm{e}^{\mathrm{i}m\phi}$ corresponds to a time-delay in the circuit analogues of the spherical waves.

### 3.1.4 Power flux

The power flux of the electromagnetic field in (3.12)–(3.13) is readily described in terms of the spherical wave coefficients; the time-average of the power radiating out through a sphere is

$$\langle P(t)\rangle = \int_{\Omega_{\hat{\boldsymbol{r}}}} \hat{\boldsymbol{r}} \cdot \mathrm{Re}\left(\frac{1}{2}\boldsymbol{E}(\boldsymbol{r},k)\times\boldsymbol{H}^*(\boldsymbol{r},k)\right) r^2\,\mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \sum_\nu |b_\nu^{(1)}|^2 - |b_\nu^{(2)}|^2, \quad (3.14)$$

which is consistent with calling $\boldsymbol{u}_\nu^{(1)}$ outgoing and $\boldsymbol{u}_\nu^{(2)}$ incoming. The equation (3.14) is the reason for the normalization chosen in (3.12)–(3.13), which gives the coefficients $b_\nu^{(j)}$ the unit watt$^{1/2}$.

### 3.1.5 Far-field

It is often convenient to analyse the far-field of an antenna or scatterer, instead of the electric or magnetic field in all space. The far-field $\boldsymbol{F}$ is the the field far from the origin, viz.

$$\boldsymbol{F}(\hat{\boldsymbol{r}}) = \lim_{kr\to\infty} \frac{r}{\mathrm{e}^{\mathrm{i}kr}}\boldsymbol{E}(\boldsymbol{r},k) = \lim_{kr\to\infty} \frac{r}{\mathrm{e}^{\mathrm{i}kr}}\hat{\boldsymbol{r}}\times \mathrm{i}\eta_0\boldsymbol{H}(\boldsymbol{r},k).$$

The far-field $\boldsymbol{F}(\hat{\boldsymbol{r}})$ in any direction $\hat{\boldsymbol{r}}$ is a plane wave and has no radial component, and can thus be written as

$$\boldsymbol{F}(\theta,\phi) = F_\theta(\theta,\phi)\hat{\boldsymbol{\theta}} + F_\phi(\theta,\phi)\hat{\boldsymbol{\phi}},$$

where the components are $F_\theta = \hat{\boldsymbol{\theta}}\cdot\boldsymbol{F}$ and $F_\phi = \hat{\boldsymbol{\phi}}\cdot\boldsymbol{F}$, and $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ are the unit vectors associated with the polar angle $\theta$ and azimuth angle $\phi$, respectively, see Figure 11.

The transmitted and scattered electric and magnetic fields of an antenna or scatterer placed at the origin contain only outgoing waves $\boldsymbol{u}_{\tau ml}^{(1)}$, i.e. only the coefficients $b_{\tau ml}^{(1)}$ in (3.12)–(3.13) are different from zero. For the far-field $\boldsymbol{F}$, this gives

$$\boldsymbol{F}(\hat{\boldsymbol{r}}) = \sqrt{2\eta_0}\sum_{l=1}^{\infty}\sum_{m=-l}^{l}\sum_{\tau=1}^{2} \mathrm{i}^{-l-2+\tau} b_{\tau ml}^{(1)}\boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}}), \quad (3.15)$$

where the asymptotic expressions for the Hankel functions

$$\mathrm{h}_l^{(1)}(kr) = \mathrm{i}^{-l-1}\mathrm{e}^{\mathrm{i}kr}\left[(kr)^{-1} + \mathcal{O}((kr)^{-2})\right], \quad \text{as } kr\to\infty,$$

were used. Hence the far-field is a sum of the spherical harmonics $\boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}})$ and $\boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}})$. Note that the spherical harmonics $\boldsymbol{A}_{3ml}$ (with $\tau = 3$) is not part of the sum, since the far-field has no radial component. The vector spherical harmonics $\boldsymbol{A}_{\tau ml}$ can be found in Figure 12. The formula (3.15) can also be inverted to calculate the coefficients $b^{(1)}_{\tau ml}$ from the far-field $\boldsymbol{F}$:

$$b^{(1)}_{\tau ml} = \frac{\mathrm{i}^{l+2-\tau}}{\sqrt{2\eta_0}} \int_{\Omega_{\hat{\boldsymbol{r}}}} \boldsymbol{F}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{A}^*_{\tau ml}(\hat{\boldsymbol{r}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}},$$

which is used in the numerical examples in Section 4.4 in Paper IV.

### 3.1.6 Antenna scattering matrix

The antenna scattering matrix describes how an antenna transmits, receives, and scatters spherical waves. Consider the antenna in Figure 13, which is connected to a local port through a matching network. The antenna scattering matrix $\mathbf{S}^{\mathrm{A}}$ is

$$\underbrace{\begin{pmatrix} \Gamma & \mathbf{R}^{\mathrm{A}} \\ \mathbf{T}^{\mathrm{A}} & \mathbf{S} \end{pmatrix}}_{\mathbf{S}^{\mathrm{A}}} \begin{pmatrix} w^{(2)} \\ \boldsymbol{b}^{(2)} \end{pmatrix} = \begin{pmatrix} w^{(1)} \\ \boldsymbol{b}^{(1)} \end{pmatrix}. \tag{3.16}$$

Here the coefficients of the incoming and outgoing spherical waves have been put into $\infty \times 1$-vectors, viz. $\boldsymbol{b}^{(2)} = (b^{(2)}_1 \; b^{(2)}_2 \; \ldots)^{\mathrm{T}}$ and $\boldsymbol{b}^{(1)} = (b^{(1)}_1 \; b^{(1)}_2 \; \ldots)^{\mathrm{T}}$ (a $^{\mathrm{T}}$ denotes transpose). The transmitted signal is denoted $w^{(2)}$, and $w^{(1)}$ is the received signal. The signals are normalized so that their power content is $|w^{(2)}|^2$ and $|w^{(1)}|^2$, respectively, to match the normalization of the coefficients, see (3.14). The antenna scattering matrix consists of several parts: the scattering matrix $\mathbf{S}$ describes the scattering properties of the antenna, the antenna transmitting coefficients $\mathbf{T}^{\mathrm{A}} = (T^{\mathrm{A}}_1 \; T^{\mathrm{A}}_2 \; \ldots)^{\mathrm{T}}$ and the receiving coefficients $\mathbf{R}^{\mathrm{A}} = (R^{\mathrm{A}}_1 \; R^{\mathrm{A}}_2 \; \ldots)$ describe the transmitting and receiving properties, and the reflection coefficient $\Gamma$ describes the mismatch. Since the spherical waves are a complete representation of the electric and magnetic fields in (3.12)–(3.13), the antenna scattering matrix forms a full description of the antenna properties. Note, though, that the expressions (3.12)–(3.13) for the electric and magnetic fields are only valid outside the hypothetical sphere circumscribing the antenna, see Figure 13.

The antenna scattering matrix can be generalized to cover multi-port antenna systems. For an $M$-port antenna system, the received and transmitted signals are vectors, $\boldsymbol{w}^{(1)}$ and $\boldsymbol{w}^{(2)}$, with dimension $M \times 1$, the transmitting coefficients are contained in the matrix $\mathbf{T}^{\mathrm{A}}$ with dimension $\infty \times M$, and the receiving coefficients are contained in the matrix $\mathbf{R}^{\mathrm{A}}$ with dimension $M \times \infty$.

One advantage of the scattering matrix description is that only a few terms are needed for a small antenna. A rule of thumb given in [44] for the number of significant transmitting coefficients for an antenna is $N = 2(ka + n)(ka + n + 2)$, where $a$ is the radius if the smallest sphere circumscribing the antenna (recall that the wavenumber is $k = \omega/\mathrm{c} = 2\pi/\lambda$). The extra term $n$ is chosen based on the accuracy needed, where typical values are between 0 and 10. A small antenna

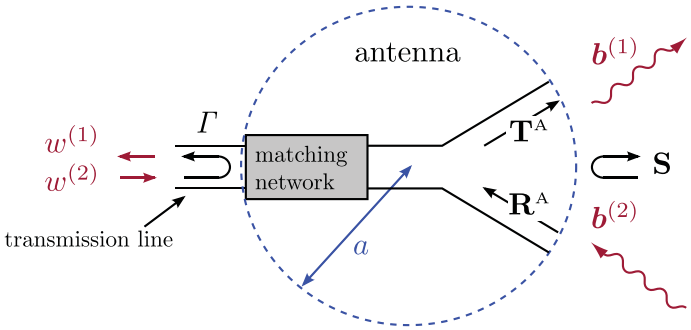



**Figure 13**: The antenna scattering matrix $\mathbf{S}^{\mathrm{A}}$ in (3.16) completely describes the transmitting, receiving and scattering properties of an antenna.

radiating higher order waves will be subject to theoretical upper bounds (physical limitations) on the bandwidth, and these limitations are the topic of Papers II-IV, see also Section 4.4 of this General Introduction.

Spherical near-field antenna measurements utilize spherical wave decompositions and the antenna scattering matrix to determine the far-field pattern of an antenna from measured near-field data, which is the topic of the next section. If desired, the antenna scattering matrix can also be determined from numerical simulations, see Section 4.4 in Paper IV.

## 3.2 Spherical near-field antenna measurements

Even if many antennas can be analysed accurately by numerical simulations, antenna measurements are still vital, both as validations of numerical simulations, and to handle cases where numerical procedures are unfeasible. References on general antenna measurements are e.g. [1, 7]. The antenna parameters of interest can differ from case to case. The reflection coefficient $\Gamma$ is of course of interest, since it determines the ratio of power accepted by the antenna $P_{\mathrm{ant}}$ to available power $P_{\mathrm{tot}}$ as

$$P_{\mathrm{ant}} = (1 - |\Gamma|^2) P_{\mathrm{tot}}.$$

The reflection coefficient is related to the voltage standing wave ratio (VSWR) and the complex antenna input impedance $Z$, and its frequency dependence can be measured with a (scalar/vector) network analyzer (SNA/VNA). The half-power bandwidth $B_{3\,\mathrm{dB}}$ of the antenna is usually defined as the width of the frequency interval where $|\Gamma|^2 \leqslant 1/2$. Standard definitions of antenna terms can be found in [2], see also [43, Appendix E].

Other parameters, such as directivity, efficiency, gain, beam-width of the main lobe, and side-lobes levels, are also of interest. The directivity of an antenna is defined as

$$D(\hat{\boldsymbol{k}}) = \frac{P(\hat{\boldsymbol{k}})}{P_{\mathrm{rad}}/4\pi},$$

where $P(\hat{\boldsymbol{k}})$ is the radiation intensity in the direction given by $\hat{\boldsymbol{k}}$ and $P_{\mathrm{rad}}$ is the

total radiated power. The (absolute) gain of an antenna is

$$G(\hat{\boldsymbol{k}}) = \frac{P(\hat{\boldsymbol{k}})}{P_{\text{ant}}/4\pi} = \eta D(\hat{\boldsymbol{k}}),$$

where the efficiency $\eta$ models ohmic losses, i.e.

$$\eta = \frac{P_{\text{rad}}}{P_{\text{ant}}} \leqslant 1.$$

The realized gain also includes the mismatch, viz.

$$G_{\text{R}}(\hat{\boldsymbol{k}}) = \frac{P(\hat{\boldsymbol{k}})}{P_{\text{tot}}/4\pi} = (1 - |\Gamma|^2)G(\hat{\boldsymbol{k}}) = (1 - |\Gamma|^2)\eta D(\hat{\boldsymbol{k}}).$$

Directivity, gain and realized gain can also be defined for each polarization $\hat{\boldsymbol{e}}$ separately, in which case they get the prefix partial and are denoted with lower case letters, viz. partial directivity $d(\hat{\boldsymbol{k}}, \hat{\boldsymbol{e}})$, partial gain $g(\hat{\boldsymbol{k}}, \hat{\boldsymbol{e}})$, and partial realized gain $g_{\text{R}}(\hat{\boldsymbol{k}}, \hat{\boldsymbol{e}})$.

The parameters discussed above are all inherent in the radiation pattern of the antenna, which is frequency-dependent. Usually, the far-field pattern is desired, as opposed to near-field patterns. Many antennas are reciprocal, which means that the transmitting and receiving characteristics are identical [44]; it also implies that they can be measured equally well in transmitting or receiving mode. There are accurate far-field ranges, where the antenna under test (AUT) and the probe are separated enough to approximate far-field conditions, but the large distances required impose practical problems. The alternative is to use radio anechoic chambers indoors. In a compact antenna test range (CATR), reflectors are used to physically create approximate plane waves impinging on the antenna under test and thus emulate far-field conditions. In a near-field test range (NFT), near-field data is measured on a planar, cylindrical or spherical surface, and then transformed with a near-field to far-field transform. For spherical near-field test ranges (SNFT), spherical waves are used to this end.

The topic of this section is to review some ideas and algorithms used for spherical near-field antenna measurements, based on the spherical waves and antenna scattering matrix in Section 3.1. It is not a complete description, but enough to observe the similarities with the channel measurements presented in Paper V and reviewed in Section 3.4. The definitive reference for spherical near-field antenna measurements is the book [44], edited by J. E. Hansen, and the interested reader is referred there.

The general idea behind spherical near-field antenna measurements is quite straightforward: First, measure the electric field $\boldsymbol{E}(\boldsymbol{r}_i, k)$ at a number of points $\boldsymbol{r}_i$ in the near-field, and invert equation (3.12) to determine the coefficients $b_\nu^{(1)}$, and thereby the antenna transmitting coefficients $T_\nu^{\text{A}}$. After that, it is possible to calculate the radiated far-field for any input signal $w^{(2)}$ using (3.15), and with the far-field the directivity, gain, and so on. There are, however, two problems here: Firstly, it is hard to measure the electric field $\boldsymbol{E}(\boldsymbol{r})$; what is measured is instead the received signal in a probe. Secondly, equation (3.12) must be inverted. The solutions to these problems consist of probe calibration, sampling on a spherical surface, and a clever algorithm using a generalization of Friis' transmission formula [44].

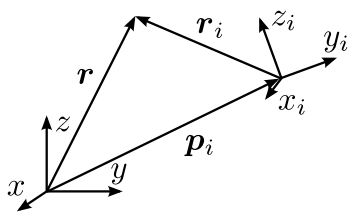



**Figure 14**: The original coordinate system $(x, y, z)$ and the translated and rotated coordinate system $(x_i, y_i, z_i)$.

### 3.2.1 Transmission formula

To derive the required transmission formula involving the antenna under test (AUT) and the probe, expressions for rotations and translations are used to write the spherical waves in the AUT coordinate system $(x, y, z)$ as sums over the spherical waves in the probe coordinate system $(x_i, y_i, z_i)$:

$$\boldsymbol{u}_{\nu}^{(1)}(k\boldsymbol{r}) = \sum_{\nu_i} B_{\nu,\nu_i}^{(1)}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i) \frac{\boldsymbol{u}_{\nu_i}^{(1)}(k\boldsymbol{r}_i) + \boldsymbol{u}_{\nu_i}^{(2)}(k\boldsymbol{r}_i)}{2}, \quad \text{for } |\boldsymbol{p}_i| \geqslant |\boldsymbol{r}_i|,$$

where $\boldsymbol{r}_i$ is the spatial coordinate and $\hat{\boldsymbol{x}}_i$, $\hat{\boldsymbol{y}}_i$, and $\hat{\boldsymbol{z}}_i$ are the Cartesian unit vectors of the translated and rotated coordinate system $(x_i, y_i, z_i)$ with origin at $\boldsymbol{r} = \boldsymbol{p}_i$, see Figure 14. Recall the multi-index $\nu = 2(l^2 + l - 1 + m) + \tau$. The coefficients $B_{\nu,\nu_i}^{(1)}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)$ consist of three parts:

1. A rotation $D$ to make the $z$-axis point in the direction of $\hat{\boldsymbol{p}}_i = \boldsymbol{p}_i/|\boldsymbol{p}_i|$.
2. A translation $C^{(1)}$ a distance $p_i = |\boldsymbol{p}_i|$ along the new $z$-axis.
3. A final rotation $D$ to align the coordinate system according to $(\hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)$.

In matrix notation (where $\mathbf{B}^{(1)}$ is the matrix with elements $B_{\nu,\nu_i}^{(1)}$) it becomes

$$\mathbf{B}^{(1)}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i) = \mathbf{D}(\hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)\ \mathbf{C}^{(1)}(kp_i)\ \mathbf{D}(\hat{\boldsymbol{p}}_i). \tag{3.17}$$

There are closed form, albeit rather complicated, expressions for both the translation $\mathbf{C}^{(1)}$ and the rotations $\mathbf{D}$. The rotations are usually described in terms of the three Euler-angles, $(\alpha, \beta, \gamma)$, see [44, Appendix A.2], and are given by

$$D_{\nu,\nu_i}(\alpha, \beta, \gamma) = \mathrm{e}^{\mathrm{i}m\gamma} d_{l,m,m_i}(\beta) \mathrm{e}^{\mathrm{i}m_i\gamma}, \tag{3.18}$$

where $d_{l,m,m_i}(\beta)$ is the Wigner $d$-function [23]. The translations are given by

$$C_{\nu,\nu_i}^{(1)}(kp_i) = \sum_{q=|l-l_i|}^{l+l_i} c_{\nu,\nu_i,q}(kp_i)\ \mathrm{h}_q^{(1)}(kp_i) \begin{pmatrix} l & l_i & q \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l_i & q \\ m & -m & 0 \end{pmatrix}, \tag{3.19}$$

where $(\cdot)$ denotes the Wigner $3-j$ symbols [23], and an expression for the functions $c_{\nu,\nu_i,q}(kp_i)$ can be found in the footnote.[2]

The generalization of Friis' transmission formula, relating the transmitted signal $w^{(2)}_{\text{AUT}}$ from the antenna under test to the received signal $w^{(1)}_{\text{probe}}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)$ in the probe, can be written:

$$w^{(1)}_{\text{probe}}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i) = \frac{1}{2}\mathbf{R}^{\text{A}}_{\text{probe}}\ \mathbf{B}^{(1)\text{T}}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)\ \mathbf{T}^{\text{A}}_{\text{AUT}}\ w^{(2)}_{\text{AUT}}, \tag{3.20}$$

where $\mathbf{R}^{\text{A}}_{\text{probe}}$ are the probe receiving coefficients and the probe is positioned at $\boldsymbol{p}_i$ and oriented according to $(\hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)$. Note that multiple scattering is neglected here.

### 3.2.2 Measurements

In a spherical near-field test range, the probe is scanning a spherical surface around the antenna under test, measuring the received signal $w^{(1)}$ at a number of points, and clever algorithms are used to invert the transmission formula (3.20) and determine the transmitting coefficients $\mathbf{T}^{\text{A}}_{\text{AUT}}$ of the antenna under test. In many practical set-ups, there are several probes mounted in an arch, and the antenna under test is rotated, see Figure 15. The probe receiving coefficients $\mathbf{R}^{\text{A}}_{\text{probe}}$ are assumed known from a separate measurement or calibration process, see [44]. The algorithm is especially efficient for a first-order probe, which is a probe where only the receiving coefficients with $m = \pm 1$ are different from zero. Conical horns are first-order probes, and rectangular horns are approximately first-order at large distances. Since a spherical scanning surface is used, only one value of the translation matrix $\mathbf{C}^{(1)}(kp_i)$ in (3.17) per wavenumber value needs to be determined. Orthogonality relations for the exponential functions and the Wigner $d$-function in (3.18) are then used to efficiently determine $\mathbf{T}^{\text{A}}_{\text{AUT}}$, and it can be shown that the solution is exact (apart from measurement noise) if the antenna under test has a finite number of non-zero transmitting coefficients (referred to as spatially band-limited in [44]). As already mentioned, the far-field pattern, gain, directivity, and so on, can then be determined from $\mathbf{T}^{\text{A}}_{\text{AUT}}$.

The use of first-order probes makes the algorithms in spherical near-field antenna measurements very efficient, but the downside is that first-order probes are narrow-band. Measurements over a wide frequency interval therefore require changing probes many times, which is time-consuming due to the precision required in alignment and calibration. Higher order probes are more wide-band, but make the algorithms more complicated. However, computer power has increased enough to make higher order probes a good option. An algorithm, involving matrix inversions,

[2]

$$c_{\nu,\nu_i,q}(kp_i) = \sqrt{\frac{(2l+1)(2l_i+1)}{l(l+1)l_i(l_i+1)}}\sqrt{\frac{(l_i+m)!(l-m)!}{(l_i-m)!(l+m)!}}(-1)^m \frac{1}{2}\mathrm{i}^{l-l_i-q}$$
$$\times 2(q+1)\sqrt{\frac{(l+m)!(l_i-m)!}{(l-m)!(l_i+m)!}}\left\{\delta_{\tau,\tau_i}\left[l(l+1)+l_i(l_i+1)-q(q+1)\right]+\delta_{\tau,3-\tau_i}2\mathrm{i}mkp_i\right\}$$

**Figure 15**: A spherical near-field antenna test range (SNFT). The antenna under test (AUT) is positioned on the rotating stand in the middle, and the measurement probes are mounted in an arch. A laser pointer is used to position the antenna accurately. The walls, floor and ceiling of the anechoic chamber are covered with radiation absorbent material. The photograph was taken when the Skycross ultra-wideband (UWB) antenna in Figure 2d in Paper V was characterized in a Satimo Stargate-24 chamber at Lite-On Mobile, Lund, Sweden. The author thanks Anders Sunesson for permission to use the photograph.

was presented by Laitinen et al. in [62]. T. B. Hansen showed that the complexity can be reduced by re-normalizing the matrices so that they are close to the identity matrix for probes that are almost first-order [46]. The papers [82, 83] by Schmidt et al. should also be mentioned here; they use an algorithm based on plane waves instead of spherical waves, and it seems like the complexity is reduced even more.

## 3.3 MIMO and spherical waves

The widespread use of wireless communication has already been touched upon in Section 1 of this General Introduction. It has also been mentioned that the use of multiple antennas in MIMO (multiple-input multiple-output) systems is one approach to satisfy the increasing demands for higher data rates. Clever signal processing is vital in MIMO communication, and there are three overall schemes that can be used to draw advantage of the multiple antenna elements: beamforming, diversity and spatial multiplexing. In beamforming, the signals at the antenna elements are weighted (with complex numbers) to make the individual radiated fields interfere constructively in some directions and destructively in others, and thus create a

beam in the desired direction. Diversity schemes build on the fact that the individual antenna elements experience different fields: if one element experiences a weak field (and therefore low received power), the probability is high that one or more of the other elements experience a stronger field. Beamforming and diversity schemes can be exploited at the transmitter and/or receiver side. In this section, the focus is on the third scheme, spatial multiplexing, which employs the multiple antenna elements to transmit several independent data-streams simultaneously. For a more detailed introduction to MIMO, the reader is referred to the books [72] and [77]; the doctoral theses of Plicanic [78] and Tian [100] are also highly recommended.

In spatial multiplexing, the capacity can increase linearly with the number of antenna elements, but only if the receiving antenna system is able to distinguish between the individual signals transmitted by the transmitting antennas. For this to work, the correlation between the antenna elements needs to be low, and this can be accomplished with different polarizations (polarization diversity), different patterns (angle diversity), and spatial separation (spatial diversity). However, uncorrelated antennas are not necessarily enough, as the received signals depend on a complicated interplay between the antennas and the environment (the propagation channel). Since the environment is dynamic, statistical models are often used for the propagation channel. The most common methods to model the antenna-channel interaction, theoretically and experimentally, are based on plane waves, or multi-path components, in conjunction with the far-field patterns of the antennas, see e.g. [72].

Alayón Glazunov instead used spherical waves and the scattering matrix description (3.16) of antennas in his thesis [28]. Others have also used spherical waves in similar problems, see e.g. [69, 71]. The antenna scattering matrix inherently captures the polarization, angle and spatial diversity of the antenna elements. Spherical wave decompositions are mathematically rigorous descriptions of any electromagnetic field in free space (satisfying (3.1)), as opposed to plane wave decompositions, and, furthermore, only a few terms of the antenna scattering matrix are needed to describe an electrically small antenna system, as noted in Section 3.1.6. Some of the most important results of the thesis [28] are reviewed here, as a motivation of the experimental method to estimate spherical wave coefficients from channel measurements (see Section 3.4 and Paper V), and the theoretical analysis of physical limitations on spherical wave scattering (see Section 4.4 and Papers II–IV).

The two included papers [31] and [30] in Alayón Glazunov's thesis [28] are particularly interesting. In [31], the focus is on analysing the statistical properties of the spherical wave coefficients and to see under which conditions maximum received power and minimum correlation is achieved. A statistical model with a random Gaussian field plus a deterministic component is considered. The random part models rays scattered and diffracted on their way to the receiver, and the deterministic component typically models the line-of-sight (LOS) path. An expression for the covariance matrix of the spherical wave coefficients in terms of the angular distribution of the channel (the power angular spectrum, PAS) is derived [31, Proposition 1]. It is also shown that the spherical wave coefficients are Gaussian variates, and expressions for their mean and variance are derived [31, Lemma 1].

Furthermore, results on the maximum received power and minimum correla-

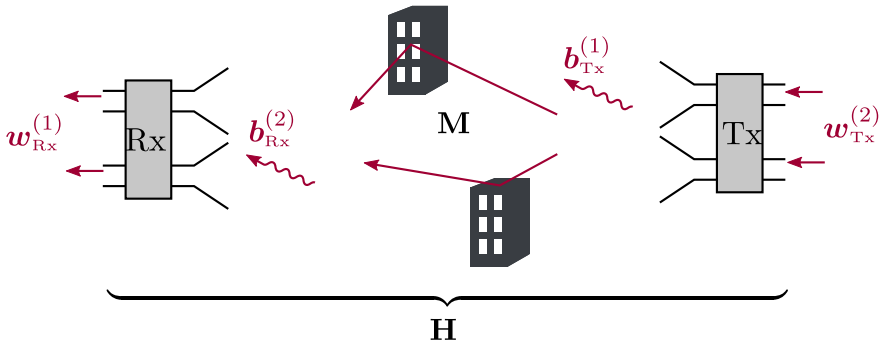



**Figure 16**: Illustration of the channel matrix $\mathbf{H}$ and the mode-to-mode channel matrix $\mathbf{M}$ in (3.22) for a $2 \times 2$ MIMO system.

tion are presented: It is shown that the received power is maximized by conjugate mode-matching, i.e. when the receiving coefficients are proportional to the complex conjugate of the coefficients of the incoming waves, $\mathbf{R}^{\mathrm{A}} \propto \boldsymbol{b}^{(2)*}$ [31, Proposition 2]. This places an upper bound on the mean of received power/incident power (mean effective gain, MEG) in a random field. Minimum correlation (and at the same time maximum received power) for the elements of an $M$-port antenna is achieved when the receiving coefficients are equal to the $M$ strongest eigenvectors of the correlation matrix of the coefficients of the incoming waves $\boldsymbol{b}^{(2)}$ [31, Proposition 5]. In [31], numerical calculations for a microstrip element and an elementary tripole are also included.

In the second paper, [30], the mode-to-mode channel matrix $\mathbf{M}$ is introduced: it maps the outgoing spherical waves excited at the transmitter (Tx) to the incoming spherical waves impinging at the receiver (Rx):

$$\boldsymbol{b}_{\mathrm{Rx}}^{(2)} = \mathbf{M}\, \boldsymbol{b}_{\mathrm{Tx}}^{(1)}. \tag{3.21}$$

The transmitted and received signals are thus related by

$$\boldsymbol{w}_{\mathrm{Rx}}^{(1)} = \mathbf{R}_{\mathrm{Rx}}^{\mathrm{A}}\, \mathbf{M}\, \mathbf{T}_{\mathrm{Tx}}^{\mathrm{A}} \boldsymbol{w}_{\mathrm{Tx}}^{(2)} = \mathbf{H}\, \boldsymbol{w}_{\mathrm{Tx}}^{(2)}, \tag{3.22}$$

where the channel matrix $\mathbf{H}$ is given by $\mathbf{H} = \mathbf{R}_{\mathrm{Rx}}^{\mathrm{A}}\, \mathbf{M}\, \mathbf{T}_{\mathrm{Tx}}^{\mathrm{A}}$, see Figure 16. Explicit expression relating $\mathbf{M}$ and $\mathbf{H}$ to the channel scattering dyadic, which describes the channel in terms of plane waves, are derived. It is shown that the elements of the matrix $\mathbf{M}$ are Gaussian variates, in the case of a Gaussian channel, and expressions for the mean, variance and covariance are presented [31, Proposition 1]. Numerical illustrations of the mode-to-mode channel matrix $\mathbf{M}$ are given for two widely used channel models.

## 3.4 Channel measurements with spherical waves

In propagation channel measurements, the objective is to describe the environment rather than the receiving or transmitting antennas. The antennas can be conveniently described in terms of spherical wave decompositions and the antenna scattering matrix, and it was seen in Section 3.2 that this is a vital part of the algorithms

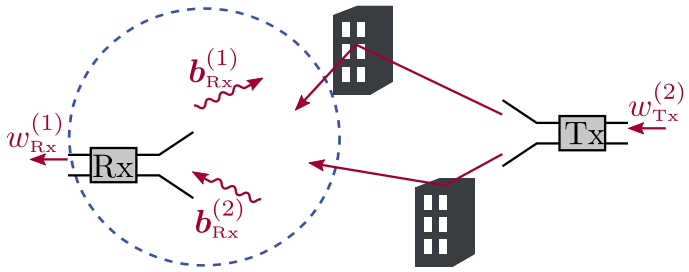



**Figure 17**: The receiving antenna measures the received signal at a number of positions and orientations within the sphere, and the coefficients of the incident field $\boldsymbol{b}_{\mathrm{Rx}}^{(2)}$ with respect to the origin in the center of the sampling region are estimated. It is assumed that the scattered field ($\boldsymbol{b}_{\mathrm{Rx}}^{(1)}$) that is in turn scattered back from nearby objects is negligible.

in spherical near-field antenna measurements. Furthermore, spherical waves are useful also to model the propagation channel and the antenna-channel interaction, as described in Section 3.3. However, there are no previous publications were spherical waves are used within propagation channel measurements.

If it was possible to measure the mode-to-mode channel matrix $\mathbf{M}$ in (3.21), then it could be used to determine the channel-matrix $\mathbf{H}$ in (3.22) for any transmitting and receiving antennas with known characteristics. It could also be used to determine which transmitting and receiving antennas that give the highest received power and lowest correlation (for multi-port antennas), using the results of Alayón Glazunov discussed in Section 3.3.

In Paper V of this thesis, a method to solve a part of this problem is presented, along with the results of a measurement campaign. More precisely, the method deals with estimating the coefficients $\boldsymbol{b}_{\mathrm{Rx}}^{(2)}$ of the incoming spherical waves at the receiver for a given transmitting antenna and a given channel:

$$\boldsymbol{b}_{\mathrm{Rx}}^{(2)} = \mathbf{M}\ \mathbf{T}_{\mathrm{Tx}}^{\mathrm{A}} \boldsymbol{w}_{\mathrm{Tx}}^{(2)},$$

see Figure 17.[3] Note that it is assumed that multiple scattering can be neglected here.

The measurement procedure is outlined here: An in-house patch antenna, kept at a fixed position, was used as the transmit antenna at 5.15 GHz, and a Satimo 5.15 GHz sleeve dipole was chosen as the receiving antenna. The receiving antenna was mounted on the 3D positioner, as depicted in Figure 18, and moved in a $10\times10\times10$ cubical grid with stepsize 15mm ($\approx 0.26\lambda$), measuring $x$, $y$, and $z$ polarization at each point for a total of 3000 measurements. The transmitting and receiving antennas were connected to port 1 and 2 of a vector network analyzer, which was used to measure the transfer function $S_{21} = w_{\mathrm{Rx}}^{(1)}/w_{\mathrm{Tx}}^{(2)}$. For validation purposes, a Skycross ultra-wideband (UWB) antenna was also used as the receiving antenna in separate measurements. The receiving antennas were characterized in a Satimo Stargate-24 chamber to determine their receiving coefficients, see Figure 15.

[3]The notation in Paper V is for regular spherical waves instead of incoming, and the coefficients are related as $\boldsymbol{d}^{(2)} = 2\boldsymbol{b}^{(2)}$.

**Figure 18**: The measurement scenario in Paper V, a small room with many scatterers and obstructed-line-of-sight (OLOS). The transmitting patch antenna is mounted on the stand to the right, and the receiving sleeve dipole is mounted on the 3D positioner to the left.

A transmission formula similar to (3.20) was used:

$$w_{\mathrm{Tx}}^{(1)}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i) = \mathbf{R}_{\mathrm{Rx}}^{\mathrm{A}}\ \mathbf{B}^{\mathrm{T}}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)\ \boldsymbol{b}_{\mathrm{Rx}}^{(2)},$$

where $(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)$ describes the position and orientation of the receiving antenna with respect to the origin in the center of the sampling region, and $\mathbf{B}$ is defined as $\mathbf{B}^{(1)}$ in (3.17) but with the spherical Hankel functions $\mathrm{h}_l^{(1)}$ in (3.19) replaced by spherical Bessel functions $\mathrm{j}_l = (\mathrm{h}_l^{(1)} + \mathrm{h}_l^{(2)})/2$. The matrix $\mathbf{B}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)$ was calculated with in-house Matlab-scripts for each position and orientation of the receiving antenna. The 3000 measurements give a system of 3000 equations for the unknown $\boldsymbol{b}_{\mathrm{Rx}}^{(2)}$, which was inverted using a regularized Tikhonov solution [45]. The results for the scenario in Figure 18 can be found in Paper V.

To investigate if, and how, the number of measurements can be reduced, the spherical wave coefficients $\boldsymbol{b}_{\mathrm{Rx}}^{(2)}$ were also estimated using subsets of the 3000 measurements, and the estimated coefficients were compared to the estimated coefficients when all the measurements were used. It was seen that using only the measurements on the surface gave large errors, using only inner points failed for higher order waves, whereas a randomly chosen subset worked well. Computer generated data was also used to check the accuracy of the method as a function of the signal to noise ratio, SNR. It can be expected that fewer measurement points are sufficient for higher SNR, and in this case they can be placed uniformly, randomly, or on the surface.

In future measurements, it would be desirable to use something else than the rather slow 3D positioner. A real array that measures on a surface does not seem to be feasible for low SNR, but the good results obtained when using randomly chosen

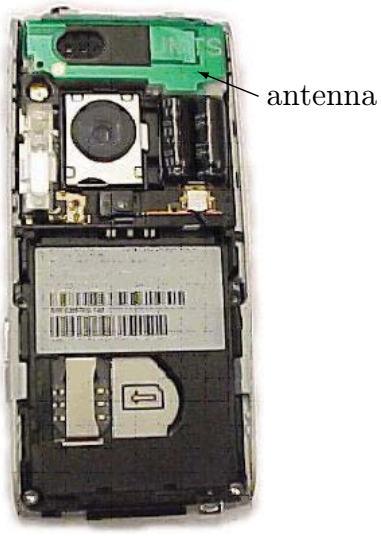



**Figure 19**: The back of a Sony Ericsson K800 mobile phone. Note that the antenna is placed side by side with the batteries, speaker, camera, and so forth. The author thanks Anders Sunesson for the photograph.

points indicate that it is not necessary to use a device that controls the positions precisely, as long as they are measured correctly. Hopefully, this fact can be taken advantage of in order to simplify and speed up the measurements. It would also be desirable to develop a method that separates the influence of the transmitting antenna to estimate the mode-to-mode channel matrix $\mathbf{M}$.

# 4 Physical limitations in antenna theory

At this point, it should be clear that antennas are the components that provide the link between the electromagnetic waves and the electronics of the wireless communication devices. The reader should also have come to appreciate the complexity of antenna characteristics, especially when the surrounding environment is taken into account. The antenna must also coexist with the rest of the device; in a mobile phone, for example, the antenna is placed side by side with the batteries, speaker, camera, and so forth, see Figure 19.

Some important antenna parameters, such as reflection coefficient, bandwidth, radiation pattern, gain, and so on, were introduced in Section 3, where also some of the methods to measure these parameters were reviewed. Numerical simulations can of course also be used. Measurements and numerical simulations notwithstanding, theoretical understanding and intuition based on the underlying physics can also be very beneficial, but hard to achieve due to the complexity of antennas and wireless communication. Analysing physical limitations is one way to try to reach general theoretical conclusions for antennas, by stating what can, and what cannot, be achieved in terms of performance under certain constraints. One significant constraint that limits antenna performance is size; this is intuitively reasonable,

since objects that are small compared to the wavelength can only provide limited interaction with electromagnetic waves.

Wheeler introduced physical limitations in antenna theory in his paper [106] from 1947, where he modelled small antennas as lumped circuits and showed that their maximum achievable bandwidths are limited. Chu introduced the antenna quality factor in his paper [16] from 1948, and derived lower bounds on the minimum $Q$ of small antennas. The quality factor is the quotient of stored energy in the near-field to radiated power, and is often inversely proportional to the antenna operating bandwidth. Since 1948, a significant amount of research has been devoted to physical limitations in antenna theory; much of the work is more or less based on Chu's method, but there are also other approaches. In 1950, Fano published the paper [24] on physical limitations on broadband matching of arbitrary impedances; although the paper is not specifically concerned with antennas, the results have implications for matching of antennas as well. In the paper [92] from 2007, Sohl et al. published a sum rule for the extinction cross section, which is a measure on the power in a plane wave that is scattered or absorbed by an object; the results were also applied to find limitations on antenna parameters, see Sohl's doctoral thesis [96].

This section reviews previous approaches to physical limitations in antenna theory: The antenna quality factor, and associated limitations, are review in in Section 4.1. Fano's theory of optimal broadband matching is reviewed in Section 4.2. Limitations based on the sum rule for the extinction cross section are reviewed in Section 4.3. The limitations on spherical wave scattering and matching from Papers II–IV are summarized in Section 4.4. It is sometimes hard to make clear distinctions between the various approaches, but, nevertheless, the presentation here is divided into these four categories. There is an excellent and thorough review of physical limitations in antenna theory in the book [104] by Volakis et al., and therefore only a brief review is presented here.

## 4.1 Quality factor

In his paper [16], Chu laid the foundation for much of the coming work on physical limitations in antenna theory. Chu, and others following in his path, considered a parameter called the antenna quality factor, or $Q$-factor. The quality factor compares the stored electric and magnetic energies, $W_\mathrm{e}$ and $W_\mathrm{m}$, in the near field to the radiated power $P_\mathrm{rad}$, at the resonance frequency $\omega_0$:

$$Q(\omega_0) = 2\omega_0 \frac{\max(W_\mathrm{e}(\omega_0), W_\mathrm{m}(\omega_0))}{P_\mathrm{rad}(\omega_0)}. \tag{4.1}$$

It is clear that a high quality factor is disadvantageous; large amounts of energy in the near field is in general coupled to high losses. In fact, if $Q(\omega_0)$ is high, its reciprocal can be interpreted as half the half-power bandwidth of the antenna [109, Equation (77)]:

$$B_{3\,\mathrm{dB}} \approx \frac{2}{Q(\omega_0)}, \quad \text{for } Q(\omega_0) \gg 1.$$

A low $Q$-value, on the other hand, is an indication of a broadband antenna. There is, however, no general relationship between $Q$ and bandwidth [41, 97]. To simplify the notation, the argument $\omega_0$ will be omitted from now on, i.e. $Q = Q(\omega_0)$.

Much of the work on the quality factor makes use of the spherical wave decompositions described in Section 3. Chu derived equivalent lumped-element circuits for the spherical waves, which take the form of ladders, and where the lengths of the ladders (and hence the complexities of the circuits) are increased for higher order waves. Instead of using the full circuits, Chu approximated the circuits close to their resonance frequencies. He went on to consider linearly polarized, omni-directional antennas with maximum gain $G$, minimum $Q$, and maximum $G/Q$, and argued that an antenna radiating like an electric dipole ($\mathrm{TM}_1$-mode) yields a minimum $Q$. Harrington [48] generalized Chu's results to cover antennas radiating a combination of electric dipoles and magnetic dipoles ($\mathrm{TE}_1$-modes). Collin and Rothschild [17] considered antennas radiating arbitrary electric ($\mathrm{TM}_l$) or magnetic ($\mathrm{TE}_l$) spherical waves, and used integrals of the electric and magnetic fields to derive exact expressions for the $Q_l$ associated with spherical waves of order $l$ [17, Equation (10)]:

$$\begin{aligned} Q_1 &= \frac{1}{k_0^3 a^3} + \frac{1}{k_0 a} \\ Q_2 &= \frac{18}{k_0^5 a^5} + \frac{6}{k_0^3 a^3} + \frac{3}{k_0 a} \\ Q_3 &= \frac{675}{k_0^7 a^7} + \frac{135}{k_0^5 a^5} + \frac{21}{k_0^3 a^3} + \frac{6}{k_0 a}, \\ &\vdots \end{aligned}$$

where $a$ is the radius of the hypothetical sphere circumscribing the antenna and $k_0 = \omega_0/\mathrm{c}$ is the resonant wavenumber. The generalization to arbitrary combinations of electric and magnetic spherical waves is due to Fante [25]; he concluded that the quality factor for a combination of electric and magnetic spherical waves diminishes by approximately a factor of 2 for small $k_0 a$, see [25, Figure 1]. The minimum $Q$ for a combination of electric and magnetic dipoles ($l = 1$) is:

$$Q_{1,\mathrm{TE+TM}} = \frac{1}{2k_0^3 a^3} + \frac{1}{k_0 a}.$$

Harrington [49] derived the maximum directivity $D$ of an antenna exciting electric and magnetic spherical waves of maximum order $L$ as

$$\max D = L(L+2) = \frac{N}{2}, \quad l \leqslant L,$$

where $N = 2L(L+2)$ is the total number of spherical waves with order $l \leqslant L$. For an antenna radiating only electric *or* magnetic spherical waves, the number of spherical waves and the maximum directivity is divided by a factor of two. For example, an antenna exciting a combination of electric and magnetic dipoles has a maximum directivity of $D = 1(1+2) = 3$, whereas the maximum directivity is $D = 3/2$ for an antenna exciting only electric *or* magnetic dipoles.

Yaghjian and Best [109] proposed an alternative quality factor $Q_Z$ expressed in the antenna input impedance $Z(\omega)$,

$$Q_Z(\omega_0) = \frac{\omega_0}{2R(\omega_0)}|Z'(\omega_0)|, \tag{4.2}$$

where $R(\omega_0) = Z(\omega_0)$ is the real-valued impedance at the resonance frequency, and a prime ($'$) denotes differentiation. For many antennas it holds that $Q(\omega_0) \approx Q_Z(\omega_0)$, and one advantage of $Q_Z$ over $Q$ is that (4.2) is often more straightforward to evaluate than (4.1). Yaghjian and Best also discussed the relation between $Q$, $Q_Z$ and the bandwidth $B$. The relation between $Q$ and bandwidth was also discussed previously by Sten and Hujanen [97], who also noted that there is no generally applicable relationship between $Q$ and bandwidth. Gustafsson and Nordebo (see [41] and Section 4.2) used Fano's theory of broadband matching [24] to study the relationship between $Q_Z$ and $B$, and also noted that although $Q \approx Q_Z$ holds for many antennas, there is no general relationship between $Q$ and $Q_Z$.

Many other authors have also made contributions to limitations in antenna theory based on the quality factor. McLean [68] presented an alternative derivation of lower bounds on $Q$. Foltz and McLean [26] considered prolate spheroidal waves. Thiele et al. [99] determined lower bounds on $Q$ based on the antenna far-field pattern. Geyi [27] considered bounds on directivity and $Q$. Thal [98] predicted stricter bounds on $Q$ for antennas occupying the surface of a sphere, rather than the whole sphere. The folded spherical helix antennas designed by Best [9] approach the Thal-bounds [98]. Alayón Glazunov et al. derived bounds on the mean effective gain over $Q$ for an antenna in a random field in [29] (included in the thesis [28]).

The previously mentioned bounds on $Q$ rely (more or less) on the circumscribing sphere. More strict lower limits on $Q$ for other geometries have been presented recently by Yaghjian and Stuart [108] and Vandenbosch [102]. An optimization procedure for the currents on the antenna, which also leads to lower bounds on $Q$ for arbitrary geometries, is presented by Gustafsson et al. in [40]; the procedure relies on expressions for the stored electric and magnetic energies $W_\mathrm{e}$ and $W_\mathrm{m}$ derived by Vandenbosch [101]. The results mentioned in this paragraph have a lot in common with the antenna limitations based on the extinction cross section, see Section 4.3

## 4.2 Fano matching

Fano is perhaps most known for his work within information theory, but he also presented important research on electric circuits in his doctoral thesis in 1947, also published in [24] in 1950. More specifically, Fano studied broadband matching of a source to a load impedance. When a source is connected to a load impedance $Z$, the power accepted by the load $P_Z$ is usually less than the total power delivered by the source $P_\mathrm{tot}$:

$$P_\mathrm{Z} = (1 - |\Gamma|^2)P_\mathrm{tot}.$$

where $\Gamma$ denotes the reflection coefficient. This fact was already mentioned in Section 3.2 for the special case when the load is an antenna. The reflected power is of course undesired, since it both diminishes efficiency and can damage the source.

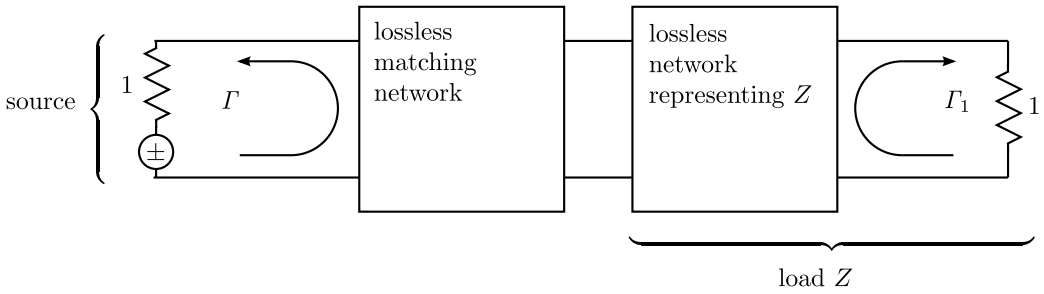



**Figure 20**: The matching problem as described by Fano in [24]. The internal resistance of the source as well as the resistance stemming from the representation of the load $Z$ are normalized to 1.

A given source may be matched perfectly to a load at one specific frequency, but a frequency-dependent load might not be possible to match to the source over a whole frequency interval using only a lossless matching network. Fano both derived theoretical limitations on matching, and also considered synthesis of matching networks. The limits are sometimes called Bode-Fano limits, due to similar work by Bode [10]. Other important contributions are the books [34] by Guillemin (Fano's thesis supervisor), [66] by Matthaei et al., and [14] by Carlin and Civalleri.

Fano considered a source connected to a load via a matching network, normalized so that the source impedance is 1 and the load impedance is $Z$. Furthermore, Fano used a Darlington representation [18] of the load, see Figure 20, where the load $Z$ is represented by a lossless network and a unit resistance. The reflection coefficient $\Gamma$ is of interest, since it determines the power rejected by the load. Fano noted that $|\Gamma_1| = |\Gamma|$ applies, and went on to derive sum rules and physical limitations for $\Gamma_1$.

Fano only considered lumped circuit elements, and thus the impedance $Z(\omega)$ of the load, as well as the reflection coefficient $\Gamma_1(\omega)$, were rational functions. Fano considered cases with known asymptotic expansions at the origin, $\omega = 0$, at infinity, $\omega = \infty$, and at non-zero real, purely imaginary, and complex frequencies $\omega + \mathrm{i}\sigma$, and used the Cauchy integral formula to derive sum rules for the logarithm of the reflection coefficient $\ln|\Gamma_1(\omega)|$. The sum rules are known as Fano's matching equations. For the case of transmission zeros at the origin and infinity, Fano's matching equations take the form (2.19), where the Herglotz function is given by (2.12). The other cases require some more algebra. Fano also derived physical limitations of the type (2.20). Moreover, Fano also addressed the problem of synthesizing the matching network by placing the complex zeros in (2.13). Since the reflection coefficient $\Gamma_1(\omega)$ corresponds to the transfer function of a scatter-passive system, Fano's matching equations can also be derived directly using the integral identities (2.19), see Section 5.3 in Paper I and Section 2.3 in Paper III.

To use Fano's limitations in antenna theory, a model for the antenna input impedance is required. The impedance $Z(\omega)$ of many antennas can be approximated by the resonance circuit in Figure 21 close to the resonance frequency $\omega_0$ [41]. Using

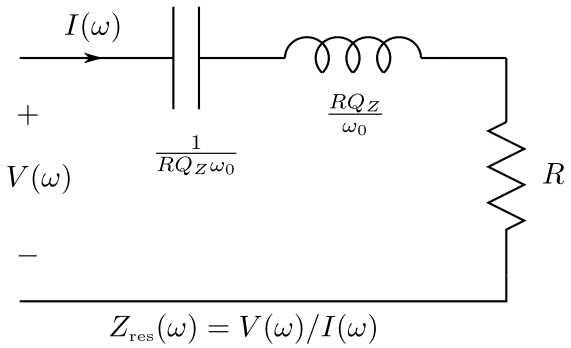



**Figure 21**: For many antennas, the impedance $Z_{\mathrm{res}}(\omega)$ of the resonance circuit is a good approximation for the antenna impedance $Z(\omega)$ close to its resonance frequency $\omega_0$ [41]. The quality factor $Q_Z(\omega_0)$ is given by (4.2).

Fano's limitations, it can be shown that

$$\frac{B}{\pi} \min_{\mathcal{B}} \ln |\Gamma(\omega)|^{-1} \leqslant \frac{1}{Q_Z(\omega_0)} \left(1 - \frac{B^2}{4}\right) \leqslant \frac{1}{Q_Z(\omega_0)},$$

where the frequency interval is $\mathcal{B} = [\omega_0(1-B/2), \omega_0(1+B/2)]$ with center frequency $\omega_0$ and fractional bandwidth $B$, cf. (2.20), and where the quality factor $Q_Z(\omega_0)$ is given by (4.2). More accurate circuit models can also be used, see [41] and references therein. Alternatively, the equivalent circuits for the spherical waves can be used with Fano's matching equations for an antenna radiating a specific spherical wave, see Section 4.4.

## 4.3 Extinction cross section

Sohl et al. adopted a different approach to antenna limitations in 2007; the approach was based on a sum rule for the extinction cross section, valid for almost arbitrary passive scattering objects. They considered a scatterer or antenna illuminated by a plane wave:

$$\boldsymbol{E}(\boldsymbol{r}, k) = \mathrm{e}^{\mathrm{i}\boldsymbol{r}\cdot k\hat{\boldsymbol{k}}} \boldsymbol{E}_0,$$

where $\hat{\boldsymbol{k}}$ is the unit vector that describes the angle of incidence of the wave. The extinction cross section $\sigma_{\mathrm{e}}$ has the unit of area and is a measure on the power $P_{\mathrm{e}}$ in the incident plane wave that is scattered or absorbed, viz.

$$\sigma_{\mathrm{e}} = \frac{P_{\mathrm{e}}}{|\boldsymbol{E}_0|/2\eta_0}.$$

Using the optical theorem (see e.g. [75]) and the Cauchy integral theorem, they derived the following sum rule for the extinction cross section when the incoming wave is linearly polarized:

$$\int_0^\infty \frac{\sigma_{\mathrm{e}}(k, \hat{\boldsymbol{k}}, \hat{\boldsymbol{e}})}{k^2} \,\mathrm{d}k = \frac{\pi}{2} \left[\hat{\boldsymbol{e}} \cdot \boldsymbol{\gamma}_{\mathrm{e}} \cdot \hat{\boldsymbol{e}} + (\hat{\boldsymbol{k}} \times \hat{\boldsymbol{e}}) \cdot \boldsymbol{\gamma}_{\mathrm{m}} \cdot (\hat{\boldsymbol{k}} \times \hat{\boldsymbol{e}})\right]. \tag{4.3}$$

Here $\hat{\boldsymbol{e}} = \boldsymbol{E}_0/|\boldsymbol{E}_0|$ is of unit length, and $\boldsymbol{E}_0$ is real for linear polarization. The electric and magnetic polarizability dyadics, $\boldsymbol{\gamma}_{\mathrm{e}}$ and $\boldsymbol{\gamma}_{\mathrm{m}}$, were defined in [58], and quantify how much the scatterer/antenna responds to static electric and magnetic fields. The electric polarizability dyadic $\boldsymbol{\gamma}_{\mathrm{e}}$ relates the induced electric dipole moment $\boldsymbol{p} = \int \boldsymbol{r}\rho(\boldsymbol{r})\,\mathrm{d}v$ in the scatterer to the applied electrostatic field $\boldsymbol{E}$ as

$$\boldsymbol{p} = \epsilon_0 \boldsymbol{\gamma}_{\mathrm{e}} \cdot \boldsymbol{E},$$

where $\epsilon_0$ denotes the permittivity of free space. Similarly, the magnetic polarizability dyadic $\boldsymbol{\gamma}_{\mathrm{m}}$ gives the induced magnetic dipole moment $\boldsymbol{m} = \frac{1}{2}\int \boldsymbol{r} \times \boldsymbol{J}(\boldsymbol{r})\,\mathrm{d}v$ in the scatterer caused by an applied static magnetic field $\boldsymbol{H}$:

$$\boldsymbol{m} = \boldsymbol{\gamma}_{\mathrm{m}} \cdot \boldsymbol{H}.$$

Here the induced charge and current densities in the scatterer are denoted $\rho$ and $\boldsymbol{J}$, respectively. The physical interpretation of the sum rule (4.3) is that the dynamic interaction of the scatterer with the plane wave integrated over all frequencies is given by the static interaction.

In the companion paper [38], the sum rule (4.3) was used in antenna theory. Two types of theoretical upper bounds were derived: The first type is for bandwidth $B$ and partial realized gain $g_{\mathrm{R}}$ [38, Equation (3.4)]:

$$B \min_{\mathcal{B}} g_{\mathrm{R}}(\hat{\boldsymbol{k}}, \hat{\boldsymbol{e}}) \leqslant \frac{k_0^3}{2}\left[\hat{\boldsymbol{e}} \cdot \boldsymbol{\gamma}_{\mathrm{e}} \cdot \hat{\boldsymbol{e}} + (\hat{\boldsymbol{k}} \times \hat{\boldsymbol{e}}) \cdot \boldsymbol{\gamma}_{\mathrm{m}} \cdot (\hat{\boldsymbol{k}} \times \hat{\boldsymbol{e}})\right], \tag{4.4}$$

where the wavenumber interval is $\mathcal{B} = [k_0(1 - B/2), k_0(1 + B/2)]$ with center wavenumber $k_0$ and fractional bandwidth $B$. The second type is for directivity $D$ and quality factor $Q$ [38, Equation (4.5)]:

$$\frac{D}{Q} \leqslant \frac{k_0^3}{2\pi}\left[\hat{\boldsymbol{e}} \cdot \boldsymbol{\gamma}_{\mathrm{e}} \cdot \hat{\boldsymbol{e}} + (\hat{\boldsymbol{k}} \times \hat{\boldsymbol{e}}) \cdot \boldsymbol{\gamma}_{\mathrm{m}} \cdot (\hat{\boldsymbol{k}} \times \hat{\boldsymbol{e}})\right]. \tag{4.5}$$

Examples of the limitations (4.4)–(4.5) can be found in [38] and [39]. The limitations are applied to ultra-wideband (UWB) antennas in [90], and to loaded dipoles in [20]. The generalization to elliptic polarization can be found in [37].

The polarizability dyadics deserve a longer discussion, since they appear in the right-hand sides of the equations (4.3)–(4.5). The polarizability dyadics also play an important part in the bounds on the quality factor presented in [40, 102, 108], see the last paragraph of Section 4.1. In Section 4.4, it will be seen that the limitations on scattering of the lowest order spherical waves, dipoles, also incorporate the polarizability dyadics, whereas the limitations for higher order spherical waves depend on other, similar, quantities. Closed form expressions for the polarizability dyadics exists for homogeneous spheroidal scatterers, see [92]. For any scatterer or antenna, upper bounds that only depend on the circumscribing geometry can be determined; the polarizability dyadics $\boldsymbol{\gamma}_{\mathrm{e}}$ and $\boldsymbol{\gamma}_{\mathrm{m}}$ are upper bounded by the high-contrast polarizability dyadic $\boldsymbol{\gamma}_{\infty}$, which is the polarizability of the circumscribing geometry with infinite static relative permittivity and permeability [89]. The high

contrast polarizability dyadics of many geometries can be calculated numerically, see [88]. Finally, note that the magnetic polarizability dyadic $\boldsymbol{\gamma}_{\mathrm{m}}$ vanishes for a non-magnetic scatterer/antenna.

Note that the sum rule (4.3) can be applied to other problems within electromagnetic scattering as well, for example metamaterials [93, 95]. There is also a similar sum rule for acoustic scattering [91]. Many of these results are also included in Sohl's thesis [96].

The sum rule (4.3) can be derived directly using the integral identities (2.19), see Example 5 in Paper I and [35]. The case with elliptic polarization is an example of a complex-valued impulse response and a non-symmetric Herglotz function $h$, which means that the identities (2.18) must be used instead of (2.19).

## 4.4 Spherical waves

Spherical wave decompositions and the antenna scattering matrix $\mathbf{S}^{\mathrm{A}}$ are the topics of Papers II-IV, where sum rules and physical limitations are derived using the general approach put forth in Paper I. The results are summarized here. In Paper II and Paper IV, the analysis concerns the scattering matrix $\mathbf{S}$, which is the part of the antenna scattering matrix that describes the scattering properties of an antenna (or scatterer) by relating the coefficients $\boldsymbol{b}^{(1)}$ of the outgoing waves to the coefficients $\boldsymbol{b}^{(2)}$ of the incoming waves, see Figure 13. In Paper III, however, the focus is on broadband matching and the reflection coefficient $\Gamma$. The motivation and inspiration to consider antenna limitations based on spherical waves and the antenna scattering matrix come both from spherical near-field antenna measurements (see Section 3.2) and the connection with MIMO (see Section 3.3). Although many of the limitations based on the quality factor also rely on spherical waves, as discussed in Section 4.1, it was also noted there that there is no general relationship between the quality factor and bandwidth.

In Paper II, the time domain versions of the spherical waves are used to describe the passivity of the scattering matrix $\mathbf{S}$, and the low-frequency expansions of the scattering matrix elements are derived to obtain sum rules and limitations. More precisely, it is shown that $\tilde{w}(k) = \mathrm{e}^{\mathrm{i}2ka} S_{\nu,\nu'}(k)$ is the transfer function of a scatter-passive system for a passive scatterer or antenna, since in this case the inverse Fourier transform is the impulse response of a scatter-passive system. Here $S_{\nu,\nu'}$ is an element of the scattering matrix and $a$ is the radius of the hypothetical sphere circumscribing the scatterer/antenna. The exponential $\mathrm{e}^{\mathrm{i}2ka}$ corresponds to a time-shift, and has to be included since the outgoing wave can appear at $r = a$ as soon as the incoming wavefront has reached $r = a$, see Figure 3 in Paper II. The low-frequency expansion (2.15) of the Herglotz function given by (2.12) is derived to order $k_0^3$, and it is shown that two sum rules of the type (2.19) (for $\hat{p} = 1, 2$) follow for the diagonal elements $S_{\nu,\nu}(k)$. Physical limitations of the type (2.20) are also derived from the sum rules, see below.

In Paper III, broadband matching of spherical waves is considered, i.e. Fano matching where the load impedance $Z$ in Figure 20 is the impedance of an outgoing spherical wave $\boldsymbol{u}_\nu^{(1)}$ seen at the surface of the hypothetical sphere of radius $a$.

This corresponds to an antenna radiating the spherical wave $\boldsymbol{u}_\nu^{(1)}$. The associated reflection coefficient $\Gamma_\nu(k)$ is the transfer function $\tilde{w}(k)$ of a scatter-passive system. The low-frequency expansion (2.15) of the Herglotz function given by (2.12) is here derived to order $k_0^{2l+1}$ (recall that $l$ denotes the order of the spherical wave). This seems to be hard to derive from the network point of view, but can be accomplished by noting that the low-frequency asymptotic expansion of the reflection coefficient $\Gamma_\nu(k)$ is the same as that of the diagonal element $S_{\nu,\nu}(k)$ of the scattering matrix of a high-contrast sphere (with infinite relative permittivity and permeability). After that, $l+1$ sum rules for $\Gamma_\nu$ are derived (for $\hat{p} = 1, 2, \ldots, l+1$), and physical limitations are also presented, see below.

In Paper IV, the scattering problem and the scattering matrix $\mathbf{S}$ are considered once more, just like in Paper II. The low-frequency expansion (2.15) of the Herglotz function given by (2.12) is derived to order $k_0^{2l+1}$, this time for the transfer function $\tilde{w}(k) = S_{\nu,\nu}(k)$ for a general, heterogeneous scatterer. Ergo, $l+1$ sum rules for $S_{\nu,\nu}(k)$ follow, and from them physical limitations. The limitations are identical to those presented in Paper II for the dipole case ($l = 1$), but are sharper for higher order waves. Furthermore, the implications of the limitations for the antenna transmitting and receiving coefficients, $\mathbf{T}^{\mathrm{A}}$ and $\mathbf{R}^{\mathrm{A}}$ in (3.16), are explored in Paper IV.

The sum rules presented in Paper IV are repeated here:

$$\begin{cases} \frac{1}{\pi}\int_0^\infty \frac{1}{k^2}\ln\frac{1}{|S_{\nu,\nu}(k)|}\,\mathrm{d}k & = a - \frac{\beta_{\nu,\nu}}{2} + \sum_n \operatorname{Im}\frac{1}{k_n} \\ \frac{1}{\pi}\int_0^\infty \frac{1}{k^{2\hat{p}}}\ln\frac{1}{|S_{\nu,\nu}(k)|}\,\mathrm{d}k & = \frac{1}{2\hat{p}-1}\sum_n \operatorname{Im}\frac{1}{k_n^{2\hat{p}-1}}, \quad \text{for } \hat{p} = 2, 3, \ldots, l \\ \frac{1}{\pi}\int_0^\infty \frac{1}{k^{2l+2}}\ln\frac{1}{|S_{\nu,\nu}(k)|}\,\mathrm{d}k & = c_l T^{[\tau]}_{ml,ml} + \frac{1}{2l+1}\sum_n \operatorname{Im}\frac{1}{k_n^{2l+1}}, \end{cases} \tag{4.6}$$

where $k_n$ are the zeros of $S_{\nu,\nu}(k)$ in the open upper half of the complex plane ($\operatorname{Im} k > 0$). The parameter $\beta_{\nu,\nu} \geqslant 0$ is expected to be zero if the radius $a$ of the circumscribing sphere is chosen as small as possible, see Paper III. The constant $c$ is given by $c_l = [2^{2l}(l+1)!(l-1)!]/[(2l+1)!(2l)!]$. The electrostatic transition matrix elements $T^{[2]}_{ml,ml}$ and the magnetostatic transition matrix elements $T^{[1]}_{ml,ml}$ quantify how the scatterer/antenna responds to static electric and magnetic fields; for dipoles ($l = 1$), the elements $T^{[2]}_{ml,ml}$ and $T^{[1]}_{ml,ml}$ for an uncharged body are (apart from normalization) equal to the elements of the electric and magnetic polarizability dyadics $\boldsymbol{\gamma}_{\mathrm{e}}$ and $\boldsymbol{\gamma}_{\mathrm{m}}$ introduced in Section 4.3, see Paper IV. In Paper II, only the two sum rules for $\hat{p} = 1, 2$ were derived. To get the sum rules presented in Paper III, $S_{\nu,\nu}$ must be replaced by $\Gamma_\nu$, and $T^{[2]}_{ml} = T^{[1]}_{ml} = a^{2l+1}$ (high-contrast sphere).

Physical limitations of the type (2.20) can be derived by considering finite wavenumber intervals and solving the resulting optimization problem for the complex zeros $k_n$:

$$\frac{B \min_{\mathcal{B}} \ln S_{\nu,\nu}^{-1}}{\pi} \leqslant f_\nu(T^{[\tau]}_{ml,ml}; k_0 a), \tag{4.7}$$

where $f_\nu(T^{[\tau]}_{ml,ml}; k_0 a)$ is the solution and, as before, the wavenumber interval is $\mathcal{B} = [k_0(1 - B/2), k_0(1 + B/2)]$ with center wavenumber $k_0$ and fractional bandwidth $B$. There is a closed-form solution $f_\nu$ for the dipole case ($l = 1$), but numerical solutions are required for higher order waves ($l > 1$). In this thesis, a relaxation of

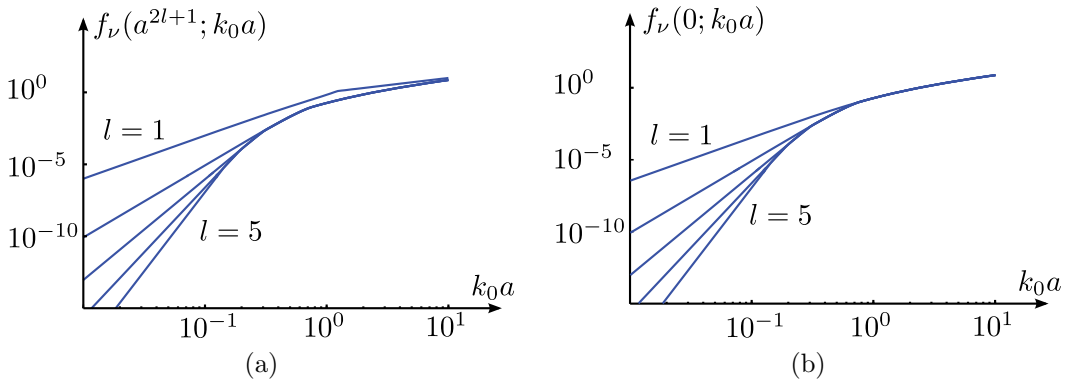



**Figure 22**: Upper bounds on the functions $f_\nu(T^{[\tau]}_{ml,ml}; k_0a)$ in (4.7)-(4.8), for $l = 1 \ldots 5$. (a): Bounds for a high-contrast sphere, $T^{[\tau]}_{ml,ml} = a^{2l+1}$, which apply for matching of spherical waves (see Paper III). (b): Bounds for $T^{[\tau]}_{ml,ml} = 0$, which apply e.g. for the combination non-magnetic scatterers/antennas and TE-modes.

the optimization problem is used to calculate upper bounds on the function $f_\nu$ in the right-hand side, see Figure 22. In the figure, note that the order of the bound for small $k_0a$ increases with the order $l$ of the spherical waves. More advanced numerical algorithms have been used by Villalobos et al. [103] and Kogan [59], but are considered out of scope here.

An interpretation of the limitations (4.7) is that the scattering matrix elements $S_{\nu,\nu}$ cannot be arbitrarily small over the whole wavenumber interval $\mathcal{B}$; how small they can be is determined by the relative bandwidth $B$, as well as the electrical size of the scatterer (center wavenumber $k_0$ times radius $a$ of the circumscribing sphere) and its shape and static material properties (described by the static transition matrix elements $T^{[\tau]}_{ml,ml}$). Another interpretation is that the absorption of power over the wavenumber interval is limited. For the matching case, the same interpretation applies to the reflection coefficients, $\Gamma_\nu$, but in this case the bounds only depend on the electrical size $k_0a$, and not on the shape and material properties.

For antennas, the limitations on the scattering matrix elements $S_{\nu,\nu}$ can be used to derive limitations also on the transmitting and receiving coefficients $T^{\mathrm{A}}_\nu$ and $R^{\mathrm{A}}_\nu$; the receiving coefficients are bounded by

$$\min_{\mathcal{B}} |R^{\mathrm{A}}_\nu(k)| \leqslant \min_{\mathcal{B}} (1 - |S_{\nu,\nu}|^2) \leqslant 1 - \mathrm{e}^{-2\pi f_\nu(T^{[\tau]}_{ml,ml}; k_0a)/B}. \tag{4.8}$$

For reciprocal antennas, it holds that the moduli of the receiving and transmitting coefficients are equal, $|R^{\mathrm{A}}_\nu| = |T^{\mathrm{A}}_\nu|$, and therefore (4.8) applies also with $R^{\mathrm{A}}_\nu$ replaced by $T^{\mathrm{A}}_\nu$ in this case. A comparison between the scattering limitations for antennas in Paper IV and the matching limitations in Paper III can be found in Figure 23.

In Paper IV, comparisons to limitations based on the quality factor and extinction cross section are given. A method to calculate the scattering matrix numerically is also presented in Paper IV, and used to illustrate the limitations (4.7) and (4.8) for some antennas. Numerical illustrations for dipole-like antennas are also given

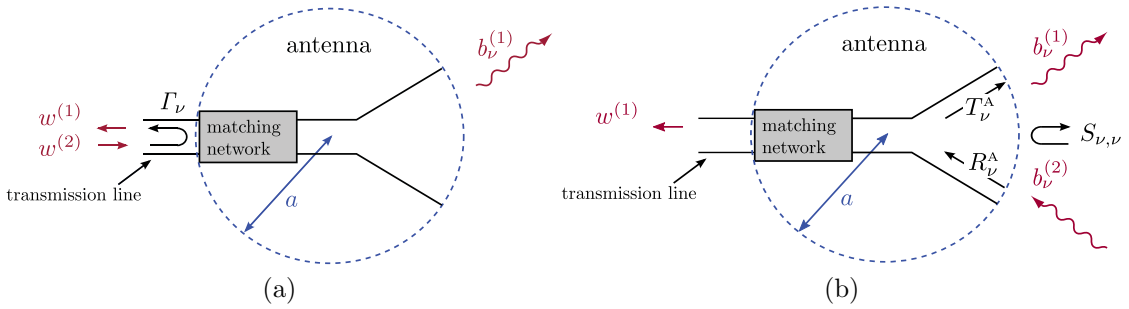



**Figure 23**: Comparison of the broadband matching limitations presented in Paper III and the scattering limitations presented in Paper IV. (a): The matching limitations place a lower bound on the antenna reflection coefficient $\Gamma_\nu$ when the antenna is transmitting a spherical wave with index $\nu$. (b): The scattering limitations place lower bounds on the antenna scattering coefficients $S_{\nu,\nu}$, and upper bounds on the antenna receiving and transmitting coefficients $R_\nu^{\mathrm{A}}$ and $T_\nu^{\mathrm{A}}$.

in Paper II. Finally, it should be mentioned that the sum rules (4.6) and scattering limitations (4.7) are valid for other passive scattering objects as well; an example concerning absorbing nanoshells can be found in Paper II.

# 5 Conclusions and outlook

The ambition with this General Introduction was to give a more detailed background to the problems studied in this thesis, and also to summarize the main contributions of the appended papers in the right context.

In Section 2, physical systems, dispersion relations, and the approach to derive sum rules and physical limitations put forth in Paper I were reviewed. Models of physical processes as input-output systems were discussed; it was seen that a system is in convolution form under very general assumptions, and that the transfer function is a complete description of the frequency dependence in this case. For many causal systems, dispersion relations for the transfer function in the form of a Hilbert transform pair can be derived from Titchmarsh's theorem. For a passive system, the transfer function can be related to a Herglotz function. A set of integral identities, or sum rules, for Herglotz functions was derived in Paper I. The sum rules relate dynamical properties of the system to its low- and/or high-frequency asymptotic expansions. Furthermore, physical limitations, which state what can and cannot be achieved under certain constraints, follow from the sum rules. To the author's best knowledge, this is the first time that a general, mathematically rigorous, and cromulent approach to derive sum rules and physical limitations has been presented. The approach shows great potential to be used in a wide-range of problems; so far, several successful applications to electromagnetic theory have been published, but the method is not restricted to this area of physics.

In Section 3, the general topic was electromagnetic spherical waves, which follow naturally as solutions to Maxwell's equations in free space when spherical coordi-

nates are used. Antenna measurements, and in particular spherical near-field antenna measurements (SNFT), were discussed, and it was seen that clever algorithms based on spherical waves can be used to determine the radiation pattern of the antenna under test from measurements with a probe scanning the near-field. Furthermore, theoretical modelling of antenna-channel interaction and multiple-input multiple-output (MIMO) systems with spherical waves, and some important results from the thesis [28] by Alayón Glazunov, were reviewed. Paper V presents the first estimation of spherical wave coefficients from channel measurements. Hopefully, the procedure can be improved in the future, and enable estimation of the mode-to-mode channel matrix $\mathbf{M}$ in (3.21). It would also be interesting to make several measurements of different propagation scenarios, and to analyse the statistical properties of the estimated spherical wave coefficients.

In Section 4, physical limitations in antenna theory were reviewed. Much of the previously published limitations are based on the antenna quality factor $Q$, or on Fano's theory of optimal broadband matching. The approach introduced in Sohl's thesis [96], based on a sum rule for the extinction cross section, was also reviewed. Paper II and Paper IV are the first where limitations on spherical wave scattering and the antenna scattering matrix are derived, and the relation to matching of spherical waves is made clear in Paper III. The general approach from Paper I is used in Papers II-IV, and it was also mentioned that Fano's matching equations and the sum rule for the extinction cross section can be derived similarly. There has been a large amount of work devoted to physical limitations in antenna theory, and it is still the focus of extensive research; the various approaches reach slightly different conclusions, and therefore complement each other well. However, most of the work is restricted to single antenna systems, but it would be very interesting to explore the limitations on MIMO further. Hopefully, the limitations on spherical wave scattering, in conjunction with the work by Alayón Glazunov and others, can be one approach to realize this. Furthermore, it might be possible to improve the limitations on spherical wave scattering; one way could be to analyse the complex zeros $k_n$ in the right-hand sides of the sum rules in (4.6), and here it could be useful with better understanding of the physical meaning of the zeros. Combinations of spherical waves should also be considered.

Both theoretical understanding and practical use of electromagnetic waves and wireless communication have evolved remarkably since the pioneering works by Maxwell, Hertz, and others. The author hopes that this thesis brings increased understanding of passive physical systems in general, and scattering and antenna problems in particular, and that our collective knowledge about the world we live in is embiggened with it.

# References

[1] IEEE standard test procedures for antennas. *ANSI/IEEE Std 149-1979*, pages 1–129, 1979.

[2] IEEE standard definitions of terms for antennas. *IEEE Std 145-1993*, pages

1–36, 1993.

[3] L. V. Ahlfors. *Complex Analysis*. McGraw-Hill, New York, second edition, 1966.

[4] N. I. Akhiezer. *The classical moment problem*. Oliver and Boyd, Edinburgh and London, 1965.

[5] N. I. Akhiezer and I. M. Glazman. *Theory of linear operators in Hilbert space*, volume 2. Frederick Ungar Publishing Co, New York, 1963.

[6] G. B. Arfken and H. J. Weber. *Mathematical Methods for Physicists*. Academic Press, New York, fifth edition, 2001.

[7] C. A. Balanis. *Antenna Theory*. John Wiley & Sons, New York, third edition, 2005.

[8] Y. M. Berezansky, Z. G. Sheftel, and G. F. Us. *Functional Analysis*. Birkhäuser, Boston, 1996.

[9] S. R. Best. Low Q electrically small linear and elliptical polarized spherical dipole antennas. *IEEE Trans. Antennas Propagat.*, **53**(3), 1047–1053, 2005.

[10] H. W. Bode. *Network analysis and feedback amplifier design*. John Wiley & Sons, New York, 1945.

[11] C. F. Bohren and D. R. Huffman. *Absorption and Scattering of Light by Small Particles*. John Wiley & Sons, New York, 1983.

[12] A. Boström, G. Kristensson, and S. Ström. Transformation properties of plane, spherical and cylindrical scalar and vector wave functions. In V. V. Varadan, A. Lakhtakia, and V. K. Varadan, editors, *Field Representations and Introduction to Scattering*, Acoustic, Electromagnetic and Elastic Wave Scattering, chapter 4, pages 165–210. Elsevier Science Publishers, Amsterdam, 1991.

[13] C. R. Brewitt-Taylor. Limitation on the bandwidth of artificial perfect magnetic conductor surfaces. *Microwaves, Antennas & Propagation, IET*, **1**(1), 255–260, 2007.

[14] H. J. Carlin and P. P. Civalleri. *Wideband circuit design*. CRC Press, Boca Raton, 1998.

[15] W. Cauer. The Poisson integral for functions with positive real part. *Bulletin of the American Mathematical Society*, **38**(1919), 713–717, 1932.

[16] L. J. Chu. Physical limitations of omni-directional antennas. *J. Appl. Phys.*, **19**, 1163–1175, 1948.

[17] R. E. Collin and S. Rothschild. Evaluation of antenna Q. *IEEE Trans. Antennas Propagat.*, **12**, 23–27, January 1964.

[18] S. Darlington. Synthesis of reactance 4-poles. *Journal of Mathematics and Physics*, **18**, 275–353, September 1939.

[19] R. de L. Kronig. On the theory of dispersion of X-rays. *J. Opt. Soc. Am.*, **12**(6), 547–557, 1926.

[20] A. Derneryd, M. Gustafsson, G. Kristensson, and C. Sohl. Application of gain-bandwidth bounds on loaded dipoles. *IET Microwaves, Antennas & Propagation*, **3**(6), 959–966, 2009.

[21] W. F. Donoghue, Jr. *Distributions and Fourier Transforms*. Academic Press, New York, 1969.

[22] P. L. Duren. *Theory of $H^p$ Spaces*. Academic Press, New York, 1970.

[23] A. R. Edmonds. *Angular Momentum in Quantum Mechanics*. Princeton University Press, Princeton, New Jersey, second edition, 1960.

[24] R. M. Fano. Theoretical limitations on the broadband matching of arbitrary impedances. *Journal of the Franklin Institute*, **249**(1,2), 57–83 and 139–154, 1950.

[25] R. L. Fante. Quality factor of general antennas. *IEEE Trans. Antennas Propagat.*, **17**(2), 151–155, March 1969.

[26] H. D. Foltz and J. S. McLean. Limits on the radiation Q of electrically small antennas restricted to oblong bounding regions. In *Proc. 1999 IEEE Antennas and Propagation Society International Symposium*, volume 4, pages 2702–2705, 1999.

[27] W. Geyi. Physical limitations of antenna. *IEEE Trans. Antennas Propagat.*, **51**(8), 2116–2123, August 2003.

[28] A. A. Glazunov. *On the Antenna-Channel Interactions: A Spherical Vector Wave Expansion Approach*. Doctoral thesis, Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2009. http://www.eit.lth.se.

[29] A. A. Glazunov, M. Gustafsson, and A. Molisch. On the physical limitations of the interaction of a spherical aperture and a random field. *IEEE Trans. Antennas Propagat.*, **59**(1), 119–128, January 2011.

[30] A. A. Glazunov, M. Gustafsson, A. Molisch, and F. Tufvesson. Physical modeling of multiple-input multiple-output antennas and channels by means of the spherical vector wave expansion. *IET Microwaves, Antennas & Propagation*, **4**(6), 778–791, 2010.

[31] A. A. Glazunov, M. Gustafsson, A. Molisch, F. Tufvesson, and G. Kristensson. Spherical vector wave expansion of gaussian electromagnetic fields for antenna-channel interaction analysis. *IEEE Trans. Antennas Propagat.*, **3**(2), 214–227, 2009.

[32] R. E. Greene and S. G. Krantz. *Function Theory of One Complex Variable*. American Mathematical Society, Providence, Rhode Island, third edition, 2006.

[33] D. H. Griffel. *Applied functional analysis*. Ellis Horwood Limited, Chichester, 1981.

[34] E. A. Guillemin. *Synthesis of passive networks*. John Wiley & Sons, New York, 1957.

[35] M. Gustafsson. Time-domain approach to the forward scattering sum rule. *Proc. R. Soc. A*, **466**, 3579–3592, 2010.

[36] M. Gustafsson and D. Sjöberg. Sum rules and physical bounds on passive metamaterials. *New Journal of Physics*, **12**, 043046, 2010.

[37] M. Gustafsson and C. Sohl. New physical bounds on elliptically polarized antennas. In *Proceedings of the Third European Conference on Antennas and Propagation*, pages 400–402, Berlin, Germany, March 23–27 2009. The Institution of Engineering and Technology.

[38] M. Gustafsson, C. Sohl, and G. Kristensson. Physical limitations on antennas of arbitrary shape. *Proc. R. Soc. A*, **463**, 2589–2607, 2007.

[39] M. Gustafsson, C. Sohl, and G. Kristensson. Illustrations of new physical bounds on linearly polarized antennas. *IEEE Trans. Antennas Propagat.*, **57**(5), 1319–1327, May 2009.

[40] M. Gustafsson, M. Cismasu, and B. L. G. Jonsson. Physical bounds and optimal currents on antennas. *IEEE Trans. Antennas Propagat.*, 2012. in press.

[41] M. Gustafsson and S. Nordebo. Bandwidth, Q factor, and resonance models of antennas. *Progress in Electromagnetics Research*, **62**, 1–20, 2006.

[42] M. Gustafsson and D. Sjöberg. Physical bounds and sum rules for high-impedance surfaces. *IEEE Trans. Antennas Propagat.*, **59**(6), 2196–2204, 2011.

[43] M. Gustafsson, C. Sohl, and G. Kristensson. Physical limitations on antennas of arbitrary shape. Technical Report LUTEDX/(TEAT-7153)/1–37/(2007), Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2007. http://www.eit.lth.se.

[44] J. E. Hansen, editor. *Spherical Near-Field Antenna Measurements*. Number 26 in IEE electromagnetic waves series. Peter Peregrinus Ltd., Stevenage, UK, 1988. ISBN: 0-86341-110-X.

[45] P. C. Hansen. Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, **6**, 1–35, 1994.

[46] T. B. Hansen. Spherical near-field scanning with higher-order probes. *IEEE Trans. Antennas Propagat.*, **59**(11), 4049–4059, November 2011.

[47] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Clarendon Press, Oxford, fifth edition, 1979.

[48] R. F. Harrington. Effect of antenna size on gain, bandwidth and efficiency. *Journal of Research of the National Bureau of Standards – D. Radio Propagation*, **64D**, 1–12, January – February 1960.

[49] R. F. Harrington. *Time Harmonic Electromagnetic Fields*. McGraw-Hill, New York, 1961.

[50] S. Hassi and A. Luger. Generalized zeros and poles of $\mathcal{N}_\kappa$-functions: On the underlying spectral structure. *Methods of Functional Analysis and Topology*, **12**(2), 131–150, 2006.

[51] P. Henrici. *Applied and Computational Complex Analysis*, volume 2. John Wiley & Sons, New York, 1977.

[52] G. Herglotz. Über Potenzreihen mit positivem, reellem Teil im Einheitskreis. *Leipziger Berichte*, **63**, 501–511, 1911.

[53] H. Hertz. *Electric Waves: Being Researches on the Propagation of Electric Action with Finite Velocity Through Space*. Macmillan, London, 1893.

[54] L. Hörmander. *The Analysis of Linear Partial Differential Operators I*. Grundlehren der mathematischen Wissenschaften 256. Springer-Verlag, Berlin Heidelberg, second edition, 1990.

[55] I. S. Kac and M. G. Kreĭn. R-functions — analytic functions mapping the upper halfplane into itself. *Amer. Math. Soc. Transl.(2)*, **103**, 1–18, 1974.

[56] F. W. King. *Hilbert Transforms, Volume 1*. Cambridge University Press, 2009.

[57] F. W. King. *Hilbert Transforms, Volume 2*. Cambridge University Press, 2009.

[58] R. E. Kleinman and T. B. A. Senior. Rayleigh scattering. In V. V. Varadan and V. K. Varadan, editors, *Low and high frequency asymptotics*, volume 2 of *Handbook on Acoustic, Electromagnetic and Elastic Wave Scattering*, chapter 1, pages 1–70. Elsevier Science Publishers, Amsterdam, 1986.

[59] B. Kogan. Comments on "Broadband matching limitations for higher order spherical modes". *IEEE Trans. Antennas Propagat.*, **58**(5), 1826, 2010.

[60] M. H. A. Kramers. La diffusion de la lumière par les atomes. *Atti. Congr. Int. Fis. Como*, **2**, 545–557, 1927.

[61] G. Kristensson. *Spridningsteori med antenntillämpningar*. Studentlitteratur, Lund, 1999. (In Swedish).

[62] T. Laitinen, S. Pivnenko, J. M. Nielsen, and O. Breinbjerg. Theory and practice of the FFT/matrix inversion technique for probe-corrected spherical near-field antenna measurements with high-order probes. *IEEE Trans. Antennas Propagat.*, **58**(8), 2623–2631, August 2010.

[63] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskiĭ. *Electrodynamics of Continuous Media*. Pergamon, Oxford, second edition, 1984.

[64] P. A. Martin. *Multiple Scattering: Interaction of Time-Harmonic Waves with N Obstacles*, volume 107 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, U.K., 2006.

[65] J. Mashreghi. *Representation Theorems in Hardy Spaces*. Cambridge University Press, Cambridge, U.K., 2009.

[66] G. L. Matthaei, L. Young, and E. M. T. Jones. *Microwave filters, impedance-matching networks, and coupling structures*. McGraw-Hill, New York, 1964.

[67] J. C. Maxwell. A dynamical theory of the electromagnetic field. *Phil. Trans. Royal Soc. London*, **155**, pp. 459–512, 1865.

[68] J. S. McLean. A re-examination of the fundamental limits on the radiation $Q$ of electrically small antennas. *IEEE Trans. Antennas Propagat.*, **44**(5), 672–676, May 1996.

[69] M. D. Migliore. An intuitive electromagnetic approach to MIMO communication systems. *IEEE Trans. Antennas Propagat.*, **48**(3), 128–137, June 2006.

[70] M. I. Mishchenko, L. D. Travis, and D. W. Mackowski. T-matrix method and its applications to electromagnetic scattering by particles: A current perspective. *J. Quant. Spectrosc. Radiat. Transfer*, **111**(11), 1700–1703, 2010.

[71] M. Mohajer, S. Safavi-Naeini, and S. K. Chaudhuri. Spherical vector wave method for analysis and design of MIMO antenna systems. *IEEE Antennas and Wireless Propagation Letters*, **9**, 1267–1270, 2010.

[72] A. F. Molisch. *Wireless Communications*. John Wiley & Sons, New York, second edition, 2011.

[73] P. M. Morse and H. Feshbach. *Methods of Theoretical Physics*, volume 2. McGraw-Hill, New York, 1953.

[74] R. H. Nevanlinna. Asymptotische Entwicklungen beschränkter Funktionen und das Stieltjes'sche Momentenproblem. *Annales Academiae Scientiarum Fennicae A*, **18**(5), 1–53, 1922.

[75] R. G. Newton. *Scattering Theory of Waves and Particles*. Dover Publications, New York, second edition, 2002.

[76] H. M. Nussenzveig. *Causality and dispersion relations*. Academic Press, London, 1972.

[77] A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space-Time Wireless Communications*. Cambridge University Press, Cambridge, U.K., 2003.

[78] V. Plicanic. *Characterization and Enhancement of Antenna System Performance in Compact MIMO Terminals*. Doctoral thesis, Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2011. http://www.eit.lth.se.

[79] E. M. Purcell. On the absorption and emission of light by interstellar grains. *J. Astrophys.*, **158**, 433–440, 1969.

[80] K. N. Rozanov. Ultimate thickness to bandwidth ratio of radar absorbers. *IEEE Trans. Antennas Propagat.*, **48**(8), 1230–1234, August 2000.

[81] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 1987.

[82] C. H. Schmidt and T. F. Eibert. Multilevel plane wave based near-field far-field transformation for electrically large antennas in free-space or above material halfspace. *IEEE Trans. Antennas Propagat.*, **57**(5), 1382–1390, May 2009.

[83] C. H. Schmidt, M. M. Leibfritz, and T. F. Eibert. Fully probe-corrected near-field far-field transformation employing plane wave expansion and diagonal translation operators. *IEEE Trans. Antennas Propagat.*, **56**(3), 737–746, March 2008.

[84] D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, and D. R. Smith. Metamaterial electromagnetic cloak at microwave frequencies. *Science*, **314**(5801), 977–980, November 2006.

[85] L. Schwartz. *Theorié des distributions*, volume 1. Hermann, Paris, 1957.

[86] L. Schwartz. *Theorié des distributions*, volume 2. Hermann, Paris, 1959.

[87] L. Schwartz. *Mathematics for the physical sciences*. Hermann, Paris, 1966.

[88] A. Sihvola, P. Ylä-Oijala, S. Järvenpää, and J. Avelin. Polarizabilities of Platonic solids. *IEEE Trans. Antennas Propagat.*, **52**(9), 2226–2233, 2004.

[89] D. Sjöberg. Variational principles for the static electric and magnetic polarizabilities of anisotropic media with perfect electric conductor inclusions. *J. Phys. A: Math. Theor.*, **42**, 335403, 2009.

[90] C. Sohl and M. Gustafsson. A priori estimates on the partial realized gain of Ultra-Wideband (UWB) antennas. *Quart. J. Mech. Appl. Math.*, **61**(3), 415–430, 2008.

[91] C. Sohl, M. Gustafsson, and G. Kristensson. The integrated extinction for broadband scattering of acoustic waves. *J. Acoust. Soc. Am.*, **122**(6), 3206–3210, 2007.

[92] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on broadband scattering by heterogeneous obstacles. *J. Phys. A: Math. Theor.*, **40**, 11165–11182, 2007.

[93] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on metamaterials: Restrictions on scattering and absorption over a frequency interval. *J. Phys. D: Applied Phys.*, **40**, 7146–7151, 2007.

[94] C. Sohl, M. Gustafsson, G. Kristensson, and S. Nordebo. A general approach for deriving bounds in electromagnetic theory. In *Proceedings of the XXIXth URSI General Assembly*, page B01p4, Chicago, IL, USA, August 7–16 2008. International Union of Radio Science.

[95] C. Sohl, C. Larsson, M. Gustafsson, and G. Kristensson. A scattering and absorption identity for metamaterials: experimental results and comparison with theory. *J. Appl. Phys.*, **103**(5), 054906, 2008.

[96] C. Sohl. *Dispersion Relations in Scattering and Antenna Problems*. Doctoral thesis, Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2008. http://www.eit.lth.se.

[97] J. Sten and A. Hujanen. Notes on the quality factor and bandwidth of radiating systems. *Electrical Engineering (Archiv fur Elektrotechnik)*, **84**(4), 189–195, 2002.

[98] H. L. Thal. New radiation Q limits for spherical wire antennas. *IEEE Trans. Antennas Propagat.*, **54**(10), 2757–2763, October 2006.

[99] G. A. Thiele, P. L. Detweiler, and R. P. Penno. On the lower bound of the radiation Q for electrically small antennas. *IEEE Trans. Antennas Propagat.*, **51**(6), 1263–1269, June 2003.

[100] R. Tian. *Design and Evaluation of Compact Multi-antennas for Efficient MIMO Communications*. Doctoral thesis, Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2011. http://www.eit.lth.se.

[101] G. A. E. Vandenbosch. Reactive energies, impedance, and Q factor of radiating structures. *IEEE Trans. Antennas Propagat.*, **58**(4), 1112–1127, 2010.

[102] G. A. E. Vandenbosch. Simple procedure to derive lower bounds for radiation Q of electrically small devices of arbitrary topology. *IEEE Trans. Antennas Propagat.*, **59**(6), 2217–2225, 2011.

[103] M. C. Villalobos, H. D. Foltz, and J. S. McLean. Broadband matching limitations for higher order spherical modes. *IEEE Trans. Antennas Propagat.*, **57**(4), 1018–1026, 2009.

[104] J. Volakis, C. C. Chen, and K. Fujimoto. *Small Antennas: Miniaturization Techniques & Applications*. McGraw-Hill, New York, 2010.

[105] P. Waterman. Matrix formulation of electromagnetic scattering. *Proc. IEEE*, **53**(8), 805–812, August 1965.

[106] H. A. Wheeler. Fundamental limitations of small antennas. *Proc. IRE*, **35**(12), 1479–1484, 1947.

[107] M. Wohlers and E. Beltrami. Distribution theory as the basis of generalized passive-network analysis. *IEEE Transactions on Circuit Theory*, **12**(2), 164–170, 1965.

[108] A. D. Yaghjian and H. R. Stuart. Lower bounds on the Q of electrically small dipole antennas. *IEEE Trans. Antennas Propagat.*, **58**(10), 3114–3121, 2010.

[109] A. D. Yaghjian and S. R. Best. Impedance, bandwidth, and $Q$ of antennas. *IEEE Trans. Antennas Propagat.*, **53**(4), 1298–1324, 2005.

[110] D. Youla, L. Castriota, and H. Carlin. Bounded real scattering matrices and the foundations of linear passive network theory. *IRE Transactions on Circuit Theory*, **6**(1), 102–124, 1959.

[111] A. H. Zemanian. An n-port realizability theory based on the theory of distributions. *IEEE Transactions on Circuit Theory*, **10**(2), 265–274, 1963.

[112] A. H. Zemanian. *Distribution theory and transform analysis: an introduction to generalized functions, with applications*. McGraw-Hill, New York, 1965.

[113] A. H. Zemanian. *Realizability theory for continuous linear systems*. Academic Press, New York, 1972.

# Sum Rules and Constraints on Passive Systems

Paper I

Anders Bernland, Annemarie Luger, and Mats Gustafsson

**Based on:** A. Bernland, A. Luger, and M. Gustafsson. Sum rules and constraints on passive systems. *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 14, paper 145205, March 2011.

**Abstract**

A passive system is one that cannot produce energy, a property that naturally poses constraints on the system. A system in convolution form is fully described by its transfer function, and the class of Herglotz functions, holomorphic functions mapping the open upper half plane to the closed upper half plane, is closely related to the transfer functions of passive systems. Following a well-known representation theorem, Herglotz functions can be represented by means of positive measures on the real line. This fact is exploited in this paper in order to rigorously prove a set of integral identities for Herglotz functions that relate weighted integrals of the function to its asymptotic expansions at the origin and infinity.

The integral identities are the core of a general approach introduced here to derive sum rules and physical limitations on various passive physical systems. Although similar approaches have previously been applied to a wide range of specific applications, this paper is the first to deliver a general procedure together with the necessary proofs. This procedure is described thoroughly, and exemplified with examples from electromagnetic theory.

# 1 Introduction

The concept of passivity is fundamental in many applications. Intuitively, a passive system is one that does not in itself produce energy (if the system does not consume energy either, it is called lossless); hence the energy-content of the output signal is limited to that of the input. Passivity poses severe constraints, or physical limitations, on a system. The aim of this paper is to investigate these constraints. In particular, a general approach to derive sum rules and physical limitations is presented along with the necessary proofs.

A system in convolution form is fully described by its impulse response, $w$. The convolution form is intimately related to the assumptions of linearity, continuity and time-translational invariance. With the added assumptions of causality and passivity, the Fourier transform of $w$ is related to a Herglotz function [23, 28] (sometimes referred to as a Nevanlinna [18], Pick [9], or R-function [21]). The Laplace transform and the related function class of positive real (PR) functions are commonly preferred by some authors [38, 40, 42].

As holomorphic mappings between half-planes, Herglotz functions are closely related to positive harmonic functions and the Hardy space $H^\infty(\mathbb{C}^+)$ via the Cayley transform [10, 26]. Herglotz functions appear in literature concerning continued fractions and the problem of moments [1, 19, 32], but also within functional analysis and spectral theory for self-adjoint operators [2, 18]. There is a powerful representation theorem for Herglotz functions, relating them to positive measures on $\mathbb{R}$. Under certain assumptions on a Herglotz function $h$ it is possible to derive a set of integral identities, relating weighted integrals of $h$ over infinite intervals to its expansion coefficients at the origin and infinity.

The integral identities can be used to derive sum rules for various physical systems, effectively relating dynamic behaviour to static and/or high-frequency prop-

erties. This is very beneficial, since static properties are often easier to determine than dynamical behaviour. The representation in itself can also provide information on a system in the form of dispersion relations; consider e.g. the Kramers-Kronig relations [23, 25] discussed in Section 5.4. One way to take advantage of the sum rules is to derive constraints, or physical limitations, by considering finite frequency intervals. The physical limitations indicate what can and cannot be expected from a system.

Some previous examples of sum rules and physical limitations within electromagnetic theory are in the analysis of matching networks [11], temporal dispersion for metamaterials [13], broadband electromagnetic interaction with objects [34], bandwidth and directivity for antennas of certain sizes [15], extra ordinary transmission through sub-wavelength apertures [16], radar absorbers [30], high-impedance surfaces [4] and frequency selective surfaces [17]. The physical limitations can be very helpful, both from a theoretical point of view where one wishes to understand what factors limit the performance, but also from a designer view-point where the physical limitations can signal if there is room for improvement or not. As the examples show, similar methods to the one presented in this paper have been widely used to derive sum rules for systems in convolution form. For many causal systems, Titchmarsh's theorem can be used to derive dispersion relations in the form of a Hilbert transform pair [22, 23, 28]. However, some more assumptions are needed in order to obtain sum rules, see e.g. [23] and references therein. If, for instance, the transfer function is rational, the Cauchy integral formula may be used, see e.g. [11, 35].

This paper presents an approach to derive sum rules and physical limitations under the assumption that the system under consideration is causal and passive. There does not seem to be a previous account on such an approach. At the core are the integral identities for Herglotz functions, which are proved rigorously in this paper. Many physical systems obey passivity, and so the results presented here are applicable to a wide range of problems. The paper is divided into different sections: First, the class of Herglotz functions along with some of its important properties are reviewed in order to pave the way for the integral identities. After this section there is a discussion about passive systems and their connection to Herglotz functions. The proof of the integral identities comes next, and after that follow some examples which serve to illuminate the theory. Last come some concluding remarks.

## 2 Herglotz functions and integral identities

The aim of this section is to introduce the class of Herglotz functions and recall some well known properties of this class. This naturally leads to the introduction of the integral identities, presented in the end of the section. First a few words on the notation adopted throughout this paper (cf. [3, 31]): If $\mu$ is a positive measure on the Borel subsets $\mathcal{E}$ of $\mathbb{R}$ and $E \in \mathcal{E}$, denote $\mu(E) = \int_E \mathrm{d}\mu(\xi)$. The measure is referred to as $\mu$ or $\mathrm{d}\mu$. The Lebesgue integral of $f$ with respect to $\mu$ is denoted by $\int_{\mathbb{R}} f(\xi)\,\mathrm{d}\mu(\xi)$ whenever $f$ is a complex-valued measurable function on $\mathbb{R}$. The positive measure that maps $E$ to $\int_E u(\xi)\,\mathrm{d}\mu(\xi)$ for some non-negative measurable

function $u$ on $\mathbb{R}$ is denoted by $u\,\mathrm{d}\mu$.

Here is the definition of a Herglotz function:

**Definition 2.1.** *A Herglotz function is defined as a holomorphic function $h : \mathbb{C}^+ \to \mathbb{C}^+ \cup \mathbb{R}$ where $\mathbb{C}^+ = \{z : \operatorname{Im} z > 0\}$.*

There is a powerful representation theorem for the set of Herglotz functions $\mathcal{H}$ due to Nevanlinna [27], presented in the following form by Cauer [6] (see also [2]):

**Theorem 2.1.** *A necessary and sufficient condition for a function $h$ to be a Herglotz function is that*

$$h(z) = \beta z + \alpha + \int_{\mathbb{R}} \left( \frac{1}{\xi - z} - \frac{\xi}{1+\xi^2} \right) \mathrm{d}\mu(\xi), \quad \operatorname{Im} z > 0, \tag{2.1}$$

*where $\beta \geqslant 0$, $\alpha \in \mathbb{R}$ and $\mu$ is a positive Borel measure such that $\int_{\mathbb{R}} \mathrm{d}\mu(\xi)/(1+\xi^2) < \infty$.*

Note the resemblance of (2.1) to the Hilbert transform [22, 26]. The representation follows e.g. from a representation theorem for positive harmonic functions on the unit disk due to Herglotz [20]. This representation theorem in turn relies on the Riesz representation theorem for continuous, linear functionals on a compact metric space. The equation (2.1) may be cast into the slightly different form

$$h(z) = \beta z + \alpha + \int_{\mathbb{R}} \frac{1+\xi z}{\xi - z} \,\mathrm{d}\nu(\xi), \quad \operatorname{Im} z > 0, \tag{2.2}$$

where $\mathrm{d}\nu(\xi) = \mathrm{d}\mu(\xi)/(1+\xi^2)$ is a positive and finite measure. Note that the only way in which a Herglotz function can be real-valued in $\mathbb{C}^+$ is if $h \equiv \alpha$ for some $\alpha \in \mathbb{R}$.

From the representation (2.1) it follows that $h(z)/z \to \beta$, as $z \hat{\to} \infty$, where $z \hat{\to} \infty$ is a short-hand notation for $|z| \to \infty$ in the Stoltz domain $\theta \leqslant \arg z \leqslant \pi - \theta$ for any $\theta \in (0, \pi/2]$ (see Appendix A.1). Hence it makes sense to consider Herglotz functions with the asymptotic expansion

$$h(z) = \sum_{1-2M}^{m=1} b_m z^m + o(z^{1-2M}), \quad \text{as } z \hat{\to} \infty, \tag{2.3}$$

where $b_m \in \mathbb{R}$. Since $b_1 = \beta$, this expansion is always possible for some integer $M \geqslant 0$. It will simplify notation to define $b_m = 0$ for $m > 1$. The representation also implies that $zh(z) \to -\mu(\{0\})$, as $z \hat{\to} 0$ (once more, see Appendix A.1), and so an asymptotic expansion

$$h(z) = \sum_{n=-1}^{2N-1} a_n z^n + o(z^{2N-1}), \quad \text{as } z \hat{\to} 0, \tag{2.4}$$

where $a_{-1} = -\mu(\{0\})$ and all $a_n$ are real, is available for some integer $N \geqslant 0$. The coefficients $a_n$ are defined to be zero for $n < -1$. It will turn out that it suffices to

consider the asymptotic expansions along the imaginary axis, i.e. for $\arg z = \pi/2$ (see Lemma 4.2).

At the core of the approach presented in this paper to derive sum rules for passive systems are the integral identities

$$\lim_{\varepsilon\to 0^+}\lim_{y\to 0^+}\frac{1}{\pi}\int_{\varepsilon<|x|<\varepsilon^{-1}}\frac{\operatorname{Im} h(x+\mathrm{i}y)}{x^p}\,\mathrm{d}x = a_{p-1}-b_{p-1},\quad p=2-2M,3-2M,\ldots,2N. \tag{2.5}$$

Throughout this paper i denotes the imaginary unit ($\mathrm{i}^2=-1$), and $x=\operatorname{Re} z$ and $y=\operatorname{Im} z$ are implicit. Note that the origin is no more special than any other point on the real line; a Herglotz function shifted to the left or right is still a Herglotz function. Compositions of Herglotz functions with each other yields new Herglotz functions (barring the trivial case when $h\equiv\alpha$), a property that may be exploited to determine a family of sum rules. See Section 5.1 and Section 5.4.

One more point deserves a discussion here: In physical applications it is often desirable to interpret the left-hand side of (2.5) as an integral over the real line. In that case the integral must be interpreted in the distributional sense; the generalized function $h(x)=\lim_{y\to 0^+}h(x+\mathrm{i}y)$, where the right-hand side is interpreted as a limit of distributions, is a distribution of slow growth. It can also be shown that, for almost all $x\in\mathbb{R}$, the limit $\lim_{y\to 0^+}\operatorname{Im} h(x+\mathrm{i}y)$ exists as a finite number. The left-hand side of (2.5) is precisely the integral over the finite part of the limit plus possible contributions from singularities in $\{x : 0<|x|<\infty\}$, cf. (4.3), Section 5.1 and Section 5.2.

In some special cases the integral identities follow directly from the Cauchy integral formula [11, 35]. This requires some extra assumptions, e.g. that the Herglotz function is the restriction to $\mathbb{C}^+$ of a rational function. Alternative approaches to obtain integral identities from the Hilbert transform under different assumptions are discussed by King [23].

# 3 Sum rules for passive systems

The integral identities (2.5) offer an approach to construct sum rules and associated physical limitations on various systems. The first step is to ensure that the system can be modelled with a Herglotz function. Secondly, the asymptotic expansions (2.3) and (2.4), here referred to as the high- and low-frequency asymptotic expansions, have to be determined. This step commonly uses physical arguments, and is specific to each application. Finally, the integrals in (2.5) are bounded to construct the physical limitations.

Herglotz functions appear in the context of linear, time translational invariant, continuous, causal and passive systems, see e.g. the paper [40] by Youla et. al., [41] and [42] by Zemanian, and [38] by Wohlers and Beltrami. These treatises are in the context of distributions, while a study in a more general setting is given in [43] and references therein. A short summary of some important results are given in this section. See also the book [28] by Nussenzveig.

Let $\mathcal{D}'$ denote the space of distributions of one variable, and let $\mathcal{D}'_0$ denote distributions with compact support [42]. Consider an operator $\mathcal{R} : D(\mathcal{R}) \subseteq \mathcal{D}' \to \mathcal{D}'$. It is a convolution operator if and only if it is linear, time translational invariant, and continuous [42, Theorem 5.8-2]:

$$u(t) = \mathcal{R}v(t) = w * v(t), \tag{3.1}$$

where $t$ denotes time, $*$ denotes temporal convolution and $w \in \mathcal{D}'$ is the impulse response. The exact definitions of linearity, time translational invariance and continuity can be found in [42].

The operator is causal if $w$ is not supported in $t < 0$, i.e. $\operatorname{supp} w \subseteq [0, \infty)$. The last crucial property of the operator is that of passivity, which is considered in two different forms. The terminology is borrowed from electric circuit theory. Let $v$ correspond to the electric voltage over some port, and let $u$ correspond to the current into the said port. Assume that the voltage and current are almost time-harmonic with amplitudes varying over a timescale much larger than the dominating frequency, and model this with complex-valued distributions $u$ and $v$. The power absorbed by the system at the time $t$ is $\operatorname{Re} u^*(t)v(t)$ (if $u$ and $v$ are regular functions), where the superscript $^*$ denotes the complex conjugate. The operator $\mathcal{R}$ defined by $u = \mathcal{R}v$ is called the admittance operator. If instead the input signal is $q = (v + u)/2$ and the output is $r = \mathcal{W}q = (v - u)/2$, the corresponding operator $\mathcal{W}$ is the scattering operator, and the absorbed power is $|q(t)|^2 - |r(t)|^2$. Let $\mathcal{D}$ denote the space of smooth functions with compact support and make the following definition [38, 42, 43]:

**Definition 3.1.** *Let $\mathcal{R}$ be a convolution operator with input $v$ and output $u = \mathcal{R}v$. Define the energy expressions*

$$e_{\mathrm{adm}}(T) = \operatorname{Re} \int_{-\infty}^{T} u^*(t)v(t)\,\mathrm{d}t \quad \textit{and} \quad e_{\mathrm{scat}}(T) = \int_{-\infty}^{T} \left(|v(t)|^2 - |u(t)|^2\right) \mathrm{d}t.$$

*The operator is admittance-passive (scatter-passive) if $e_{\mathrm{adm}}(T)$ ($e_{\mathrm{scat}}(T)$) is non-negative for all $T \in \mathbb{R}$ and $v \in \mathcal{D}$.*

Note that admittance-passive might as well have been called impedance-passive, if the electric current was assumed to be input and the voltage output in the example from which the name stems. An operator which is admittance-passive or scatter-passive is called passive in this paper. As it turns out, passivity implies causality for operators in convolution form. Furthermore, in this case the impulse response $w$ must be a distribution of slow growth, i.e. $w \in \mathcal{S}'$ [38, 42], and thus (3.1) is defined for smooth input signals of rapid descent, $v \in \mathcal{S}$. The set of distributions of slow growth is denoted by $\mathcal{S}'$, and $\mathcal{S}$ stands for smooth functions of rapid descent in the remainder of the paper. Note that (3.1) is also defined for all input signals $v$ with support bounded on the left, since $\operatorname{supp} w \subseteq [0, \infty)$ [42].

Since the impulse response is in $\mathcal{S}'$, its Fourier transform may be defined as $\langle \mathcal{F}w, \varphi \rangle = \langle w, \mathcal{F}\varphi \rangle$ , for all $\varphi \in \mathcal{S}$, where $\langle f, \varphi \rangle$ is the value in $\mathbb{C}$ that $f \in \mathcal{S}'$ assigns

to $\varphi \in \mathcal{S}$ [42]. The Fourier transform of $\varphi$ is defined as $\mathcal{F}\varphi(\omega) = \int_{\mathbb{R}} \varphi(t)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t$. The Fourier transform of $w$ is the transfer function $\tilde{w}$ of the system, viz.

$$\tilde{w}(\omega) = \mathcal{F}w(\omega). \tag{3.2}$$

The convolution in (3.1) is mapped to multiplication if e.g. $v \in \mathcal{D}_0'$ or $v \in \mathcal{S}$ [42]. In that case the frequency domain system is modelled by $\tilde{u}(\omega) = \tilde{w}(\omega)\tilde{v}(\omega)$, where $\tilde{v} = \mathcal{F}v$ and $\tilde{u} = \mathcal{F}u$ are the input and output signals, respectively. The transfer function $\tilde{w}(\omega)$ is in $\mathcal{S}'$ for real $\omega$, but since the support of $w$ is bounded on the left the region of convergence for $\tilde{w}$ contains $\mathbb{C}^+$ and $\tilde{w}$ is holomorphic there. The Laplace transform is commonly used in system theory, generating the corresponding transfer function $\tilde{w}_{\mathrm{Laplace}}(s) = \tilde{w}(\mathrm{i}s)$. Scrutinizing the transfer function, the following theorem is proved (cf. Theorem 10.4-1 in [42], Theorem 2 in [38] and Theorems 7.4-3 and 8.12-1 in [43]):

**Theorem 3.1.** *Let $\mathcal{R} = w*$ be a convolution operator and let $\tilde{w}$ be given by* (3.2)*. If $\mathcal{R}$ is admittance-passive, then $\operatorname{Re}\tilde{w}(\omega) \geqslant 0$ for all $\omega \in \mathbb{C}^+$. If $\mathcal{R}$ is scatter-passive, then $|\tilde{w}(\omega)| \leqslant 1$ for all $\omega \in \mathbb{C}^+$. In both cases $\tilde{w}$ is holomorphic in $\mathbb{C}^+$.*

The converse statement to the theorem can also be made, i.e. that every transfer function on one of the forms described in the theorem generates an admittance-passive or scatter-passive operator, respectively [42, Theorem 10.6-1], [43, Theorems 7.5-1 and 8.12-1].

Evidently, the transfer function of an admittance-passive operator multiplied with the imaginary unit is a Herglotz function, $h = \mathrm{i}\tilde{w}$. For scatter-passive operators a Herglotz function can be constructed from $\tilde{w}$ via the inverse Cayley transform $z \mapsto (\mathrm{i}z + \mathrm{i})/(1 - z)$. Alternatively, factorize $\tilde{w}(\omega) = H(\omega)B(\omega)$, where $H(\omega)$ is a zero free holomorphic function such that $|H(\omega)| \leqslant 1$ for all $\omega \in \mathbb{C}^+$ and

$$B(\omega) = \left(\frac{\omega - \mathrm{i}}{\omega + \mathrm{i}}\right)^k \prod_{\omega_n \neq \mathrm{i}} \frac{|\omega_n^2 + 1|}{\omega_n^2 + 1} \frac{\omega - \omega_n}{\omega - \omega_n^*} \tag{3.3}$$

is a Blaschke product [10, 26]. Here the zeros $\omega_n$ of $\tilde{w}$ in $\mathbb{C}^+$ are repeated according to their multiplicity and $k \geqslant 0$ is the order of the possible zero at $\omega = \mathrm{i}$. The convergence factors $|\omega_n^2 + 1|/(\omega_n^2 + 1)$ may be omitted if all $|\omega_n|$ are bounded by the same constant or if $\tilde{w}$ satisfies the symmetry (3.7) discussed below. Since $\tilde{w}$ belongs to the Hardy space $H^\infty(\mathbb{C}^+)$, this factorization is always possible due to a theorem of F. Riesz [10, 26]. Moving on, the function $H$ may be represented as $H(\omega) = \mathrm{e}^{\mathrm{i}h(\omega)}$ since it is holomorphic and zero-free on the simply connected domain $\mathbb{C}^+$. Here the holomorphic function $h$ must have a non-negative imaginary part. Note that the converse to the factorization also holds; a function $\tilde{w}$ is holomorphic and bounded in magnitude by one in $\mathbb{C}^+$ if and only if it is of the form

$$\tilde{w}(\omega) = B(\omega)\mathrm{e}^{\mathrm{i}h(\omega)}, \tag{3.4}$$

where $B$ is a Blaschke product given by (3.3) and $h$ is a Herglotz function.

The formula (3.4) may be inverted: $h(\omega) = -\mathrm{i}\log(\tilde{w}(\omega)/B(\omega))$, if the logarithm is defined as

$$\log H(z) = \ln|H(z_0)| + \mathrm{i}\arg H(z_0) + \int_{\gamma_{z_0}^z} \frac{\mathrm{d}H/\,\mathrm{d}\zeta}{H(\zeta)}\,\mathrm{d}\zeta. \tag{3.5}$$

Here $\gamma_{z_0}^z$ is any piecewise $\mathrm{C}^1$ curve from $z_0$ to $z$ in $\mathbb{C}^+$. The left-hand side of (2.5) takes the form

$$\lim_{\varepsilon\to 0^+}\lim_{y\to 0^+}\int_{\varepsilon<|x|<\varepsilon^{-1}} \frac{-\ln|\tilde{w}(x+\mathrm{i}y)/B(x+\mathrm{i}y)|}{x^p}\,\mathrm{d}x.$$

The modulus $|B(z)|$ tends to 1 as $z\hat{\to}x$ for almost all $x\in\mathbb{R}$ (the exceptions are the $x$ which are accumulation points of the zeros of $\tilde{w}$ [26]). If the origin is not an accumulation point of the zeros of $\tilde{w}$, the low-frequency asymptotic expansion of $h$ is

$$h(\omega) = -\mathrm{i}\log\tilde{w}(\omega) - \arg B(0) - \mathrm{i}\sum_{m=1}^{\infty}\frac{\omega^m}{m}\sum_{\omega_n}\omega_n^{-m} - \omega_n^{*-m}, \quad \text{as } \omega\hat{\to}0. \tag{3.6}$$

A similar argument may be applied to the high-frequency asymptotic expansion. The asymptotic expansions of $\log\tilde{w}$ must be found by physical arguments.

For operators $\mathcal{R}$ mapping real-valued input to real-valued output, the impulse response $w$ is real-valued. This implies the symmetry

$$\tilde{w}(\omega) = \tilde{w}^*(-\omega^*), \tag{3.7}$$

which is transferred to the Herglotz function as

$$h(\omega) = -h^*(-\omega^*) \tag{3.8}$$

if it is defined by $h = \mathrm{i}\tilde{w}(\omega)$ (for admittance-passive systems) or by the inverse Cayley transform of $\pm\tilde{w}$ (for scatter-passive systems). The Herglotz function $h$ in (3.4) must be of the form $h = h_1 + \alpha$, where $h_1(\omega) = -h_1^*(-\omega^*)$, and $\alpha\in\mathbb{R}$ is the argument of $\mathrm{e}^{\mathrm{i}h(\omega)}$ for purely imaginary $\omega$. The symmetry restricts the identities (2.5) to even powers and simplifies them to

$$\lim_{\varepsilon\to 0^+}\lim_{y\to 0^+}\frac{2}{\pi}\int_{\varepsilon}^{\varepsilon^{-1}}\frac{\operatorname{Im}h(x+\mathrm{i}y)}{x^{2\hat{p}}}\,\mathrm{d}x = a_{2\hat{p}-1} - b_{2\hat{p}-1}, \quad \hat{p} = 1-M,\ldots,N. \tag{3.9}$$

In general, the integral identities (2.5) for even $p$ are the starting point to derive constraints on the system as the non-negative integrand can be bounded by a finite frequency interval.

Summing up, there are two essentially equivalent ways to evaluate if a system can be modelled with a Herglotz function and potentially be constrained according to (2.5): First, just based on a priori knowledge of linearity, continuity and time-translational invariance (i.e. the convolution form (3.1)) together with passivity. This approach can often be applied directly to various physical systems. The second,

frequency domain case is often more involved and requires direct verification that $h(\omega)$ is holomorphic and $\operatorname{Im} h(\omega) \geqslant 0$ for $\operatorname{Im}\omega > 0$. Alternative characterizations in the frequency domain are given in [38].

The high-frequency expansions (2.3) are sometimes hard to evaluate for physical systems. The high-frequency behaviours of $\tilde{w}(\omega)$ and $h(\omega)$ are determined by the behaviour of $w(t)$ for arbitrarily short times. To see this, first assume that $w$ is a regular, integrable function. Then $\tilde{w}$ is defined as

$$\tilde{w}(\omega) = \int_0^\infty w(t)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t = \int_0^\varepsilon w(t)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t + \int_\varepsilon^\infty w(t)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t.$$

The second term in the right-hand side goes to zero as $\omega\hat{\to}\infty$ (but not as $|\omega| \to \infty$ on the real line) for any $\varepsilon > 0$. This verifies the statement for $w \in \mathrm{L}^1$. For a general $w \in \mathcal{S}'$, consider the equivalent definition of $\tilde{w}(\omega)$ for $\operatorname{Im}\omega > 0$ [42]:

$$\tilde{w}(\omega) = \left\langle w(t), \lambda(t)\mathrm{e}^{\mathrm{i}\omega t}\right\rangle = \left\langle w(t), \lambda_1(t)\mathrm{e}^{\mathrm{i}\omega t}\right\rangle + \left\langle w(t), \lambda_2(t)\mathrm{e}^{\mathrm{i}\omega t}\right\rangle.$$

Here $\lambda(t)$ is a smooth function with support bounded on the left, and such that $\lambda(t) \equiv 1$ for $t \geqslant 0$. It is decomposed into two non-negative smooth functions, $\lambda = \lambda_1 + \lambda_2$, where $\lambda_2 \equiv 0$ for $t \leqslant \varepsilon$ for some $\varepsilon > 0$. The second term in the right-hand side vanishes as $\omega\hat{\to}\infty$. A similar argument may be carried out for the low-frequency expansion (2.3), essentially relating it to the behaviour of $w(t)$ for arbitrarily large $t$.

# 4 Proof of the integral identities

The main theorem (Theorem 4.1) of this paper contains the integral identities (2.5). For $p = 2, 3, \ldots, 2N$ they rely on two results: The first (Corollary 4.1) states that the left-hand sides of (2.5) are equal to moments of the measure $\mathrm{d}\mu(\xi)$. The second (Lemma 4.2) relates the convergence and explicit value of these moments to expansion (2.4). A change of variables in the left-hand side of (2.5) enables a proof for $p = 2 - 2M, 3 - 2M, \ldots, 1$.

The following lemma is is a well known result, see e.g. Lemma S1.2.1 in [21] and Theorem 11.9 in [26]. It is included here for clarity.

**Lemma 4.1.** *Let $h$ denote a Herglotz function. Suppose that the function $\varphi : \mathbb{R} \to \mathbb{R}$ is piecewise $\mathrm{C}^1$, and that there is a constant $D \geqslant 0$ such that $|\varphi(x)| \leqslant D/(1 + x^2)$ for all $x \in \mathbb{R}$. Then it follows that*

$$\lim_{y\to 0^+} \frac{1}{\pi}\int_{\mathbb{R}} \varphi(x)\operatorname{Im} h(x + \mathrm{i}y)\,\mathrm{d}x = \int_{\mathbb{R}} \breve{\varphi}(\xi)\,\mathrm{d}\mu(\xi), \tag{4.1}$$

*where $\mu(\xi)$ is the measure in the representation (2.1) of $h$, and*

$$\breve{\varphi}(\xi) = \begin{cases} \varphi(\xi), & \text{if } \varphi \text{ is continuous at } \xi \\ \frac{\varphi(\xi^-)+\varphi(\xi^+)}{2}, & \text{otherwise.} \end{cases} \tag{4.2}$$

*Here $\varphi(\xi^\pm) = \lim_{\zeta\to\xi^\pm} \varphi(\zeta)$.*

The proof can be found in Appendix A.2. It is readily shown that the limit may be replaced by any non-tangential limit, i.e. the left-hand side of (4.1) may be replaced by $\lim_{u\xrightarrow{}0}\int_{\mathbb{R}}\varphi(x)\operatorname{Im}h(x+u)\,\mathrm{d}x$.

Note that the lemma is in some sense an inversion formula; whereas the representation (2.1) gives the Herglotz function $h$ from the measure $\mu$, (4.1) instead makes possible the retrieval of $\mu$ when $h$ is known. In fact, the lemma is the Stieltjes inversion formula in a different form [1, 21, 32]. The inversion is clarified by decomposing the measure as $\mu = \mu_{\mathrm{a}} + \mu_{\mathrm{s}}$, where $\mu_{\mathrm{a}}$ is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}\xi$ and $\mu_{\mathrm{s}}$ is singular in the same sense [3]. Recall that $\mathcal{E}$ denotes the set of Borel subsets of $\mathbb{R}$. Then

$$\mu_{\mathrm{a}}(E) = \int_E \mu_{\mathrm{a}}'(\xi)\,\mathrm{d}\xi, \quad \text{for all } E \in \mathcal{E},$$

where the Radon-Nikodym derivative $\mu_{\mathrm{a}}'$ of $\mu_{\mathrm{a}}$ with respect to $\mathrm{d}x$ is a finite, locally integrable function, for almost all $x \in \mathbb{R}$ uniquely defined as [3]

$$\mu_{\mathrm{a}}'(x) = \lim_{s\to 0}\frac{\mu_{\mathrm{a}}([x-s, x+s])}{2s}.$$

"Almost all" is with respect to $\mathrm{d}x$. Furthermore [26],

$$\lim_{s\to 0}\frac{\mu_{\mathrm{s}}([x-s, x+s])}{2s} = 0, \quad \text{for almost all } x \in \mathbb{R}.$$

Hence Lemma 4.1 implies that

$$\lim_{z\xrightarrow{}x}\frac{1}{\pi}\operatorname{Im}h(z) = \lim_{s\to 0}\frac{\mu([x-s, x+s])}{2s}, \quad \text{for almost all } x \in \mathbb{R}.$$

See also [21].

In physical applications it is often desirable to move the limit inside the integral in the left-hand side of (4.1). Clearly, this is possible if $\mu = \mu_{\mathrm{a}}$. Otherwise, set $g(x) = \lim_{y\to 0^+}\operatorname{Im}h(x+\mathrm{i}y)$, whenever the limit exists finitely, to get

$$\lim_{y\to 0^+}\frac{1}{\pi}\int_{\mathbb{R}}\varphi(x)\operatorname{Im}h(x+\mathrm{i}y)\,\mathrm{d}x = \frac{1}{\pi}\int_{\mathbb{R}}\varphi(x)g(x)\,\mathrm{d}x + \int_{\mathbb{R}}\breve{\varphi}(\xi)\,\mathrm{d}\mu_s(\xi), \tag{4.3}$$

where the second term on the right-hand side represents contributions from singularities on the real line. Equivalently, the left-hand side of (4.1) may be interpreted as an integral over the real line in the distributional sense.

The first result needed for the main theorem is this corollary to Lemma 4.1:

**Corollary 4.1.** *For all Herglotz functions $h$ given by* (2.1) *it holds that*

$$\lim_{\varepsilon\to 0^+}\lim_{\breve{\varepsilon}\to 0^+}\lim_{y\to 0^+}\frac{1}{\pi}\int_{-\breve{\varepsilon}^{-1}}^{-\varepsilon}\frac{\operatorname{Im}h(x+\mathrm{i}y)}{x^p}\,\mathrm{d}x + \lim_{\varepsilon\to 0^+}\lim_{\breve{\varepsilon}\to 0^+}\lim_{y\to 0^+}\frac{1}{\pi}\int_{\varepsilon}^{\breve{\varepsilon}^{-1}}\frac{\operatorname{Im}h(x+\mathrm{i}y)}{x^p}\,\mathrm{d}x = \int_{\mathbb{R}}\frac{\mathrm{d}\mu_0(\xi)}{\xi^p}, \quad p = 0, \pm 1, \pm 2, \ldots$$

*Here $\mu_0 = \mu - \mu(\{0\})\delta_0$, i.e. the measure in the representation* (2.1) *with the point mass at the origin removed. The terms in the left-hand side are not necessarily finite. The right-hand side is not defined in the case the left-hand side equals $-\infty + \infty$.*

The proof can be found in Appendix A.3. Before presenting the second result needed for the main theorem, it is noted that $h$ may be decomposed as

$$h(z) = \beta z + \alpha - \frac{\mu(\{0\})}{z} + \int_{\mathbb{R}} \left( \frac{1}{\xi - z} - \frac{\xi}{1+\xi^2} \right) \mathrm{d}\mu_0(\xi), \tag{4.4}$$

where once again $\mu_0 = \mu - \mu(\{0\})\delta_0$. This decomposition follows directly from the fact that $zh(z) \to -\mu(\{0\})$ as $z \hat{\to} 0$.

**Lemma 4.2.** *Let $h$ be a Herglotz function given by (2.1) and $N \geqslant 0$ an integer. Then the following statements are equivalent:*

1. *The function $h$ has the asymptotic expansion (2.4), i.e.*

$$h(z) = \sum_{n=-1}^{2N-1} a_n z^n + o(z^{2N-1}), \quad \text{as } |z| \to 0,$$

   *for $z$ in the Stoltz domain $\theta \leqslant \arg z \leqslant \pi - \theta$ for any $\theta \in (0, \pi/2]$. Here all $a_n$ are real.*

2. *Statement 1 is true for $\theta = \pi/2$.*

3. *The measure $\mu_0 = \mu - \mu(\{0\})\delta_0$ satisfies*

$$\int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^{2N}(1+\xi^2)} < \infty.$$

*The expansion coefficients in (2.4) equal:*

$$a_0 = \alpha + \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi(1+\xi^2)}, \quad \textit{and} \quad a_{p-1} = \delta_{p,2}\beta + \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^p}, \quad p = 2, 3, \ldots, 2N, \tag{4.5}$$

*where $\delta_{i,j}$ denotes the Kronecker delta.*

A similar result is a well-known theorem due to Hamburger and Nevanlinna [1, Theorem 3.2.1], [32, Theorem 2.2]. See also Lemma 6.1 in [18]. Note that the case $N = 0$ is trivial, since then all three statements are true for all Herglotz functions. The proof for $N \geqslant 1$ can be found in Appendix A.4. The convergence of $\int_{\mathbb{R}} \mathrm{d}\mu_0(\xi)/(|\xi^{2N+1}|(1+\xi^2))$ does guarantee an expansion with real coefficients up to a rest term $o(z^{2N})$, but the converse is not true. A counterexample for $N = 0$ is given by the measure $\mathrm{d}\mu_0(\xi) = \mu_0'(\xi)\,\mathrm{d}\xi$ where $\mu_0'(\xi) = -(\ln|\xi|)^{-1}$ when $\xi < 1$ and $\mu_0'(\xi) = 0$ otherwise.

The integral identities for $p = 2, 3, \ldots 2N$ follow directly from Corollary 4.1 and Lemma 4.2 (recall that $b_1 = \beta$ and that $b_{p-1} = 0$ for $p = 3, 4, \ldots$). To prove the identities for $p = 2 - 2M, 3 - 2M, \ldots, 1$, consider the Herglotz function $\breve{h}(z) = h(-1/z)$. With obvious notation, its high- and low-frequency asymptotic expansions

are related to those of $h$ as $\breve{b}_n = (-1)^n a_{-n}$ and $\breve{a}_n = (-1)^n b_{-n}$. Evidently, $\breve{M} = N$ and $\breve{N} = M$ applies. Following (2.2) and (4.4), $\breve{h}$ admits the representation

$$\breve{h}(z) = \frac{-\beta}{z} + \alpha + \mu(\{0\})z + \int_{\mathbb{R}} \frac{1 - \xi z^{-1}}{\xi + z^{-1}} \,\mathrm{d}\nu_0(\xi), \quad \operatorname{Im} z > 0,$$

where $\mathrm{d}\nu_0(\xi) = \mathrm{d}\mu_0(\xi)/(1+\xi^2)$. It would be desirable to make a change of variables $\xi \mapsto -1/\xi$ in the integral. Therefore, consider the continuous bijection $j : \mathbb{R}\backslash\{0\} \to \mathbb{R}\backslash\{0\}$ defined by $j\xi = -1/\xi$. It is its own inverse, i.e. $j^2\xi = \xi$. Furthermore, it maps Borel sets to Borel sets, which makes the following a valid definition:

**Definition 4.1.** *Let $j : \mathbb{R}\backslash\{0\} \to \mathbb{R}\backslash\{0\}$ be the mapping that takes $\xi$ to $-1/\xi$. Let $\mathcal{E}(\mathbb{R}\backslash\{0\})$ be the Borel sets of $\mathbb{R}\backslash\{0\}$ and $\mathrm{M}(\mathrm{R}\backslash\{0\})$ be the set of measures on $\mathcal{E}(\mathbb{R}\backslash\{0\})$. Define the mapping $J : \mathrm{M}(\mathrm{R}\backslash\{0\}) \to \mathrm{M}(\mathrm{R}\backslash\{0\})$ through*

$$J\sigma(E) = \sigma(jE),$$

*for all $\sigma \in \mathrm{M}(\mathrm{R}\backslash\{0\})$ and $E \in \mathcal{E}(\mathbb{R}\backslash\{0\})$.*

From this definition it is clear that $J^2\sigma = \sigma$ and moreover

$$\int_{\mathbb{R}\backslash\{0\}} f(\xi) \,\mathrm{d}\sigma(\xi) = \int_{\mathbb{R}\backslash\{0\}} f(j\xi) \,\mathrm{d}\,(J\sigma)(\xi)$$

for all measurable functions $f$ on $\mathbb{R}\backslash\{0\}$, since it holds if $f$ is a simple measurable function [31]. The representation of $\breve{h}$ can now be rewritten:

$$\breve{h}(z) = \frac{-\beta}{z} + \alpha + \mu(\{0\})z + \int_{\mathbb{R}} \frac{1 + \xi z}{\xi - z} \,\mathrm{d}\,(J\nu_0)(\xi), \quad \operatorname{Im} z > 0.$$

The function $\breve{h}$ is thus represented by the measure $\mathrm{d}\breve{\nu}_0 = \mathrm{d}\,(J\nu_0)$, or equivalently $\mathrm{d}\breve{\mu}_0 = \xi^2 \,\mathrm{d}\,(J\mu_0)$. Therefore

$$\begin{aligned}\lim_{y\to 0^+} \frac{1}{\pi} \int_{\varepsilon}^{\breve{\varepsilon}^{-1}} \frac{\operatorname{Im} h(x+\mathrm{i}y)}{x^p} \,\mathrm{d}x &= \int_{\mathbb{R}} \breve{\varphi}_{p,\varepsilon,\breve{\varepsilon}}(\xi) \,\mathrm{d}\mu_0(\xi) = \int_{\mathbb{R}} \breve{\varphi}_{p,\varepsilon,\breve{\varepsilon}}(-1/\xi) \frac{\mathrm{d}\breve{\mu}_0(\xi)}{\xi^2} \\ = \lim_{y\to 0^+} (-1)^p \frac{1}{\pi} \int_{-\varepsilon^{-1}}^{-\breve{\varepsilon}} \frac{\operatorname{Im} \breve{h}(x+\mathrm{i}y)}{x^{2-p}} \,\mathrm{d}x, &\quad \text{for } p = 0, \pm 1, \pm 2, \ldots \text{ and } 0 < \varepsilon < \breve{\varepsilon}^{-1},\end{aligned} \tag{4.6}$$

and likewise for the corresponding integral over $(-\breve{\varepsilon}^{-1}, -\varepsilon)$. Here $\breve{\varphi}_{p,\varepsilon,\breve{\varepsilon}}$ is given by (A.2) and (4.2). The proof of the integral identities (2.5) for $p = 2-2M, 3-2M, \ldots, 0$ have now been returned to the case $p = 2, 3, \ldots, 2N$. Here at last is the sought for theorem:

**Theorem 4.1** (Main Theorem)**.** *Let $h$ be a Herglotz function. Then it has the asymptotic expansions* (2.3) *and* (2.4) *if and only if the corresponding left-hand sides in* (2.5) *are absolutely convergent. In this case the integral identities* (2.5) *apply.*

The integrals in the left-hand side of (2.5) may be taken over the set $\{x : \varepsilon < |x| < \infty\}$ when $p = 2, 3, \ldots, 2N$ and $\{x : 0 < |x| < \varepsilon^{-1}\}$ when $p = 2 - 2M, 3 - 2M, \ldots, 0$, see Appendix A.3. In this case there is an extra term $-\delta_{p,0}\ a_{-1}$ in the right-hand side. This fact is used in the examples below to obtain neater expressions.

*Proof.* The theorem for $p = 2, 3, \ldots 2N$ follows directly from Corollary 4.1 and Lemma 4.2. For $p = 2 - 2M, 3 - 2M, \ldots, 0$ it also requires (4.6) and the relation between the asymptotic expansions of $h$ and $\breve{h}$. The case $p = 1$ is special as it requires both high- and low-frequency expansions. Assume that the asymptotic expansions (2.3) and (2.4) are valid for $N = M = 1$ and use equation (4.5) for $h$ and $\breve{h}$:

$$\begin{aligned} a_0 - b_0 &= \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi(1+\xi^2)} - \int_{\mathbb{R}} \frac{\mathrm{d}\breve{\mu}_0(\xi)}{\xi(1+\xi^2)} = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi} \\ &= \lim_{\varepsilon\to 0^+} \lim_{\breve{\varepsilon}\to 0^+} \lim_{y\to 0^+} \int_{\varepsilon<|x|<\breve{\varepsilon}^{-1}} \frac{\operatorname{Im} h(x+\mathrm{i}y)}{x}\,\mathrm{d}x. \end{aligned}$$

Here all integrals are absolutely convergent. If on the other hand the left-hand sides of (2.5) are absolutely convergent for $p = 0, 1, 2$, then the asymptotic expansions (2.3) and (2.4) clearly hold for $N = 1$ and $M = 1$, respectively. □

The main theorem is essential to the approach to derive sum rules for passive systems; it implies that high- and low-frequency asymptotic expansions of the form (2.3) and (2.4) for a Herglotz function $h$ are both necessary and sufficient conditions for the integral identities (2.5) to apply. Note especially that only the low-frequency asymptote is required for the integral identities to be valid for $p = 2, 3, \ldots, 2N$. Recall also that the identities are simplified to (3.9) for most physical systems, due to the symmetry (3.8). These observations conclude the proofs section.

# 5 Examples

For all the examples, the notation $\omega = \omega' + \mathrm{i}\omega''$, with $\omega' = \operatorname{Re}\omega$ and $\omega'' = \operatorname{Im}\omega$, is used.

## 5.1 Elementary Herglotz functions

Examples of elementary Herglotz functions are

$$\beta z, \quad C, \quad \frac{-\beta}{z}, \quad \sqrt{z}, \quad \log(z), \quad \mathrm{i}\log(1-\mathrm{i}z),$$

with $\beta \geqslant 0$, $\operatorname{Im} C \geqslant 0$, and appropriate branch cuts for $\sqrt{\ }$ and log.

Herglotz functions are related to the unit ball of the Hardy space $\mathrm{H}^\infty(\mathbb{C}^+)$ via the Cayley transform. An example is $\mathrm{e}^{\mathrm{i}z}$ which shows that

$$h_{\mathrm{e}}(z) = \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}z} + \mathrm{i}}{1 - \mathrm{e}^{\mathrm{i}z}}$$

is a Herglotz function. Therefore $\tan z = -1/h_{\mathrm{e}}(2z)$ is a Herglotz function as well. It satisfies the symmetry (3.8), and its asymptotic expansions are $\tan z = \mathrm{i} + o(1)$, as $z\hat{\to}\infty$, and

$$\tan z = z + \frac{z^3}{3} + \frac{2z^5}{15} + \ldots, \quad \text{as } z \to 0,$$

respectively. Note that the integer-order terms in the low-frequency asymptotic expansion are infinite in number since $\tan z$ is holomorphic in a neighbourhood of the origin. Thus there are identities of the type (3.9) for $\hat{p} = 1, 2, \ldots$:

$$\lim_{\varepsilon\to 0^+} \lim_{y\to 0} \frac{2}{\pi} \int_\varepsilon^\infty \frac{\operatorname{Im}\tan(x+\mathrm{i}y)}{x^{2\hat{p}}}\,\mathrm{d}x = \begin{cases} 1 & \text{for } \hat{p}=1 \\ 1/3 & \text{for } \hat{p}=2 \\ 2/15 & \text{for } \hat{p}=3 \\ \vdots \end{cases}$$

On the real axis, except for $x = n\pi$, where $n = 0, \pm 1, \pm 2, \ldots$, $\tan(x)$ is $\mathrm{C}^\infty$ and $\operatorname{Im}\tan(x) = 0$. It is not locally integrable around $x = n\pi$, where $\tan z$ has simple poles. There is an essential singularity at $\infty$, and the limit as $x \to \infty$ of $\tan(x)/x^{2\hat{p}}$ is not defined for any $\hat{p}$. This is thus an illustration of a case where it is difficult to use Cauchy integrals or Hilbert transform techniques to derive integral identities of the form (2.5).

If $h_1$ and $h_2$ are Herglotz functions, then so is the composition $h_2 \circ h_1$ (unless $h_1 \equiv \alpha \in \mathbb{R}$). This may be used to derive families of integral identities. Continue the example with $h_1 = \tan z$ and construct the new Herglotz function

$$\mathrm{i}\log(1 - \mathrm{i}\tan z) = \begin{cases} z + \mathcal{O}(1), & \text{as } z \to 0 \\ \mathcal{O}(1), & \text{as } z\hat{\to}\infty, \end{cases}$$

yielding an identity of the type (3.9):

$$\lim_{\varepsilon\to 0^+} \lim_{y\to 0^+} \frac{2}{\pi} \int_\varepsilon^\infty \frac{\ln|1 - \mathrm{i}\tan(x+\mathrm{i}y)|}{x^2}\,\mathrm{d}x = 1.$$

It is also illustrative to consider a case with odd weighting factors in (2.5). The function $\ln(1 + \tan(z))$ has the asymptotic expansions

$$\ln(1+\tan(z)) = \begin{cases} z - z^2/2 + 2z^3/3 + \ldots, & \text{as } z \to 0 \\ \mathcal{O}(1), & \text{as } z\hat{\to}\infty. \end{cases}$$

This gives the (2.5)-identities

$$\lim_{\varepsilon\to 0^+} \lim_{y\to 0^+} \frac{1}{\pi} \int_{|x|>\varepsilon} \frac{\arg(1+\tan(x+\mathrm{i}y))}{x^p}\,\mathrm{d}x = \begin{cases} 1 & \text{for } p=2 \\ -1/2 & \text{for } p=3 \\ 2/3 & \text{for } p=4 \\ \vdots \end{cases}$$

where it is observed that the negative part of the integrand dominates for $p = 3$. There are other manipulations of Herglotz functions that generate new Herglotz functions as well, e.g. $h_1 + h_2$ and $\sqrt{h_1 h_2}$.

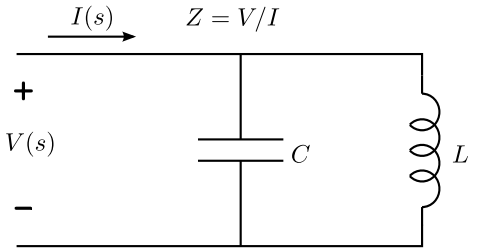



**Figure 1**: The lossless resonance circuit of Section 5.2.

## 5.2 Lossless resonance circuit

Consider a parallel resonance circuit consisting of a lumped inductance, $L$, and a lumped capacitance, $C$, see Fig. 1. This is an example of an admittance-passive system, where the impedance $Z(s) = sL/(1+s^2LC)$ is the Laplace-transfer function of the system in which the electric current over $Z$ is the input and the voltage is the output, and where $s = -\mathrm{i}\omega$ denotes the Laplace parameter. Therefore the transfer function given by (3.2) multiplied by i is a Herglotz function:

$$h(\omega) = \mathrm{i}Z(-\mathrm{i}\omega) = -\frac{\omega_0^2 L}{2}\left(\frac{1}{\omega-\omega_0}+\frac{1}{\omega+\omega_0}\right) = \begin{cases} \sqrt{\frac{L}{C}}\sum_{n=0}^{\infty}\frac{\omega^{2n+1}}{\omega_0^{2n+1}}, & \text{as } \omega \to 0 \\ -\sqrt{\frac{L}{C}}\sum_{n=0}^{\infty}\frac{\omega_0^{2n+1}}{\omega^{2n+1}}, & \text{as } \omega \to \infty, \end{cases}$$

where $\omega_0 = 1/\sqrt{LC}$ is the resonance frequency of the circuit. The real part $\operatorname{Re} Z(-\mathrm{i}\omega) = \operatorname{Im} h(\omega)$ determines the power absorbed by the impedance $Z$.

Use of the identities in (3.9) gives the sum rules

$$\lim_{\varepsilon\to 0^+}\lim_{\omega''\to 0^+}\frac{2}{\pi}\int_{\varepsilon}^{\varepsilon^{-1}}\frac{\operatorname{Re} Z(-\mathrm{i}\omega'+\omega'')}{\omega'^{2p}}\,\mathrm{d}\omega' = \sqrt{\frac{L}{C}}\omega_0^{-2p+1}, \quad \text{for } p = 0, \pm 1, \pm 2, \ldots \tag{5.1}$$

Note that on the imaginary axis, $\operatorname{Re} Z(-\mathrm{i}\omega') = 0$ applies for $\omega' \neq \pm\omega_0$. All of the contribution to the integral comes from the singularity, which becomes clear if the left-hand side of (5.1) is calculated explicitly. A physical interpretation is that even though the circuit is lossless for any frequency $\omega' \neq \omega_0$, input signals of frequency $\omega' = \omega_0$ are "trapped" in its resonance and thus absorbed by $Z$.

## 5.3 Fano's matching equations revisited

Sum rules and physical limitations on broadband matching of a source to a load impedance were published by Fano in [11]. Consider a source connected to a load via a matching network, and use the normalization where the source impedance is 1 and the load impedance is $Z(s)$, where $s = -\mathrm{i}\omega$ denotes the Laplace parameter. To study the problem, Fano used a Darlington representation [8] of the load, see Figure 2, where the load $Z(s)$ is represented by a lossless network and a unit resistance. The reflection coefficient $\rho_2(s)$ is of interest, since it determines the power rejected by the load. Fano noted that $|\rho_1(s)| = |\rho_2(s)|$ applies, and went on to derive sum rules and physical limitations for $\rho_1(s)$.

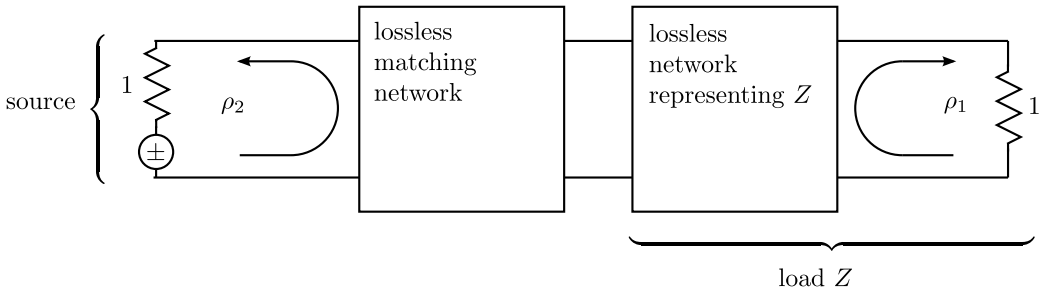



**Figure 2**: The matching problem as described in [11].

This is clearly a scatter-passive system, so the Fourier transfer function $\tilde{w}(\omega) = \rho_1(-\mathrm{i}\omega)$ is holomorphic and bounded in magnitude by one in $\mathbb{C}^+$, and it also satisfies the symmetry (3.7). Assume the asymptotic expansion

$$-\mathrm{i}\log(\tilde{w}(\omega)) = \arg\tilde{w}(0)+c_1\omega+c_3\omega^3+\ldots+c_{2N-1}\omega^{2N-1}+o(\omega^{2N-1}), \quad \text{as } \omega\hat{\to}0, \tag{5.2}$$

where $\arg\tilde{w}(0) = \lim_{\omega\hat{\to}0}\arg\tilde{w}(\omega)$ and all $c_i$ are real. This is the case e.g. if the network representing $Z(-\mathrm{i}\omega)$ in Figure 2 has a transmission zero of order $N$ at $\omega = 0$ [11]. The low-frequency asymptotic expansion of the Herglotz-function in (3.4) is

$$\begin{aligned} h(\omega) = \arg\tilde{w}(0) + c_1\omega + c_3\omega^3 + \ldots + c_{2N-1}\omega^{2N-1} + o(\omega^{2N-1}) \\ - \arg B(0) + 2\sum_{m=1,3,\ldots}^{\infty}\frac{\omega^m}{m}\sum_{\omega_n}\operatorname{Im}\omega_n^{-m}, \quad \text{as } \omega\hat{\to}0, \end{aligned}$$

according to (3.6), where $\omega_n$ are the zeros of $\tilde{w}(\omega)$ in $\mathbb{C}^+$. In this case only odd terms appear in the sum originating from the Blaschke product due to the symmetry (3.7). This implies the (3.9)-identities

$$\begin{aligned} \lim_{\varepsilon\to0^+}\lim_{\omega''\to0^+}\frac{2}{\pi}\int_{\varepsilon}^{\infty}\frac{-\ln|\rho_1(-\mathrm{i}\omega'+\omega'')|}{\omega'^{2\hat{p}}}\,\mathrm{d}\omega' \\ = c_{2\hat{p}-1} - \delta_{\hat{p},1}\beta + \frac{2}{2\hat{p}-1}\sum_{\omega_n}\operatorname{Im}\omega_n^{1-2\hat{p}}, \quad \text{for } \hat{p} = 1,2,\ldots,N, \end{aligned} \tag{5.3}$$

where $\beta = \lim_{\omega\hat{\to}\infty}h(\omega)/\omega$ comes from the high-frequency asymptote of $h(\omega)$. The limit as $\omega'' \to 0^+$ may be moved inside the integral if $\rho_1(-\mathrm{i}\omega)$ has no zeros at the imaginary axis.

The equations in (5.3) are the original Fano matching equations, derived with the Cauchy integral formula in [11]. This can be done if the matching network and load impedance $Z$ are realizable with a finite number of lumped circuit elements, since then the reflection coefficients $\rho_1(s)$ and $\rho_2(s)$ are rational functions. In this case, the parameter $\beta$ vanishes. In [11], Fano used the equations to derive the best possible match of a source to a load over a frequency interval, and also discussed how the lossless matching network should be constructed to obtain this best match.

Fano also considered cases with known asymptotic expansions of $\rho_1$ at infinity, in which case the integral identities (3.9) for $\hat{p} = 1 - M, 2 - M, \ldots, 0$ are used, and at non-zero real, purely imaginary, and complex frequencies, which require some more algebra. It should be mentioned that Fano's results have been treated more generally by other authors, see e.g. [5] and references therein.

For a reflection coefficient $\tilde{w}$ which is not a rational function (consider e.g. the scattering of electromagnetic waves by a permittive object), the Cauchy integral formula-approach falls short. Theorem 4.1 guarantees integral identities for passive systems as long as asymptotic expansions of the type (5.2) are valid as $\omega \hat{\to} 0$ and/or $\omega \hat{\to} \infty$, respectively.

## 5.4 Kramers-Kronig relations and $\epsilon$ near-zero materials

Suppose there is an isotropic constitutive relation in convolution form relating the electric field $\boldsymbol{E} = E\hat{\boldsymbol{e}}$ to the electric displacement $\boldsymbol{D} = D\hat{\boldsymbol{e}}$, where $\hat{\boldsymbol{e}}$ is a vector of unit length [25]:

$$D(t) = \epsilon_0 \chi * E(t). \tag{5.4}$$

The permittivity of free space is denoted $\epsilon_0$, and a possible instantaneous response is included in $\chi(t)$ as a term $\epsilon_\infty \delta(t)$, where $\epsilon_\infty \geqslant 0$. Let the input be $v(t) = \epsilon_0 E(t)$ and the output be $u(t) = \partial D/\partial t$. The impulse response of this system is $w(t) = \partial \chi/\partial t$. The system is admittance-passive if the material is passive, since that means that the energy expression [25]

$$e(T) = \int_{-\infty}^{T} E(t) \frac{\partial D}{\partial t} \,\mathrm{d}t$$

is non-negative for all $E \in \mathcal{D}$ and $T \in \mathbb{R}$. The Herglotz function given by $h = \mathrm{i}\tilde{w}$ is $h(\omega) = \omega\epsilon(\omega)$, where $\epsilon(\omega) = \mathcal{F}\chi(\omega)$. It satisfies the symmetry (3.8), since $w(t)$ is real-valued.

Lemma 4.1 may be applied to the representation (2.1), since $|1/(\xi - z) - \xi/(1 + \xi^2)| \leqslant D_z/(1 + \xi^2)$ for any fixed $z \in \mathbb{C}^+$. This gives

$$\omega\epsilon(\omega) = \omega\epsilon_\infty + \lim_{\psi \to 0^+} \frac{1}{\pi} \int_{\mathbb{R}} \left[ \frac{1}{\xi - \omega} - \frac{\xi}{1 + \xi^2} \right] \Big[ \psi \operatorname{Re} \epsilon(\xi + \mathrm{i}\psi) + \xi \operatorname{Im} \epsilon(\xi + \mathrm{i}\psi) \Big] \,\mathrm{d}\xi, \tag{5.5}$$

for $\operatorname{Im}\omega > 0$. This is one of the two Kramers-Kronig relations [23, 25] in a general form, where no assumptions other than those of convolution form and passivity has been made for the constitutive relation in the time domain. It may be simplified if $\epsilon(\omega') = \lim_{\omega'' \to 0^+} \epsilon(\omega' + \mathrm{i}\omega'')$ is sufficiently well-behaved. If for instance $\epsilon(\omega')$ is a continuous and bounded function, the limit may be moved inside the integral in (5.5):

$$\omega\epsilon(\omega) = \omega\epsilon_\infty + \frac{1}{\pi} \int_{\mathbb{R}} \left[ \frac{1}{\xi - \omega} - \frac{\xi}{1 + \xi^2} \right] \xi \operatorname{Im} \epsilon(\xi) \,\mathrm{d}\xi, \quad \operatorname{Im}\omega > 0.$$

Assuming that $\operatorname{Im}\epsilon(\omega') = \mathcal{O}(1/\omega')$ as $\omega' \to \pm\infty$ and employing the fact that $\operatorname{Im}\epsilon(\omega')$ is odd gives (after division with $\omega$)

$$\epsilon(\omega) = \epsilon_\infty + \frac{1}{\pi} \int_{\mathbb{R}} \frac{1}{\xi - \omega} \operatorname{Im} \epsilon(\xi) \,\mathrm{d}\xi, \quad \operatorname{Im}\omega > 0.$$

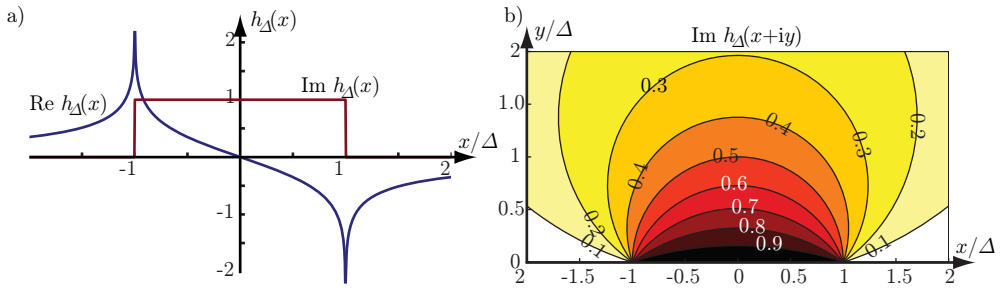



**Figure 3**: The function $h_\Delta(x+\mathrm{i}y)$ given by (5.7) illustrated by its limit as $y \to 0$ to the left and the contours of $\operatorname{Im} h_\Delta(x+\mathrm{i}y)$ to the right.

Letting $\omega'' \to 0$ and using the distributional limit $\lim_{\omega''\to 0}(\xi - \omega' - \mathrm{i}\omega'')^{-1} = \mathcal{P}(\xi - \omega')^{-1} + \mathrm{i}\pi\delta(\xi - \omega')$, where $\mathcal{P}$ is the Cauchy principal value, yields

$$\epsilon(\omega') = \epsilon_\infty + \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|\xi-\omega'|>\varepsilon}\frac{\operatorname{Im}\epsilon(\xi)}{\xi-\omega'}\,\mathrm{d}\xi + \mathrm{i}\operatorname{Im}\epsilon(\omega').$$

The real part is the Kramers-Kronig relation (5.5) as presented in e.g. [25]:

$$\operatorname{Re}\epsilon(\omega') = \epsilon_\infty + \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|\xi-\omega'|>\varepsilon}\frac{\operatorname{Im}\epsilon(\xi)}{\xi-\omega'}\,\mathrm{d}\xi.$$

The assumption that $\epsilon(\omega')$ is continuous rules out the possibility of static conductivity, which however can be included with a small modification of the arguments. Assuming that $h(\omega) = \omega\epsilon(0) + o(\omega)$, as $\omega\hat{\to}0$, there is a sum rule of the type (3.9) for $\hat{p} = 1$ (also presented in e.g. [25]):

$$\lim_{\varepsilon\to 0^+}\frac{2}{\pi}\int_\varepsilon^\infty\frac{\operatorname{Im}\epsilon(\omega')}{\omega'}\,\mathrm{d}\omega' = \epsilon(0) - \epsilon_\infty.$$

It shows that the losses are related to the difference between the static and instantaneous responses of the medium. The Kramers-Kronig relations and their connection to Herglotz functions are also discussed in [23, 28, 37, 39].

In applications such as high-impedance surfaces and waveguides, it is desirable to have so called $\epsilon$ near-zero materials [33], i.e. materials with $\epsilon(\omega') \approx 0$ in a frequency interval around some center frequency $\omega_0$. Define the Herglotz function

$$h_1(\omega) = \frac{\omega}{\omega_0}\epsilon(\omega) = \begin{cases} o(\omega^{-1}), & \text{as } \omega\hat{\to}0 \\ \frac{\omega}{\omega_0}\epsilon_\infty + o(\omega), & \text{as } \omega\hat{\to}\infty. \end{cases} \tag{5.6}$$

Compositions of Herglotz functions may be used to derive limitations different from those that $h_1$ would produce on its own. In the present case the area of interest is the frequency region where $h_1(\omega) \approx 0$. A promising function is

$$h_\Delta(z) = \frac{1}{\pi}\int_{-\Delta}^{\Delta}\frac{1}{\xi - z}\,\mathrm{d}\xi = \frac{1}{\pi}\ln\frac{z-\Delta}{z+\Delta} = \begin{cases} \mathrm{i} + o(1), & \text{as } z \to 0 \\ \frac{-2\Delta}{\pi z} + o(z^{-1}), & \text{as } z \to \infty, \end{cases} \tag{5.7}$$

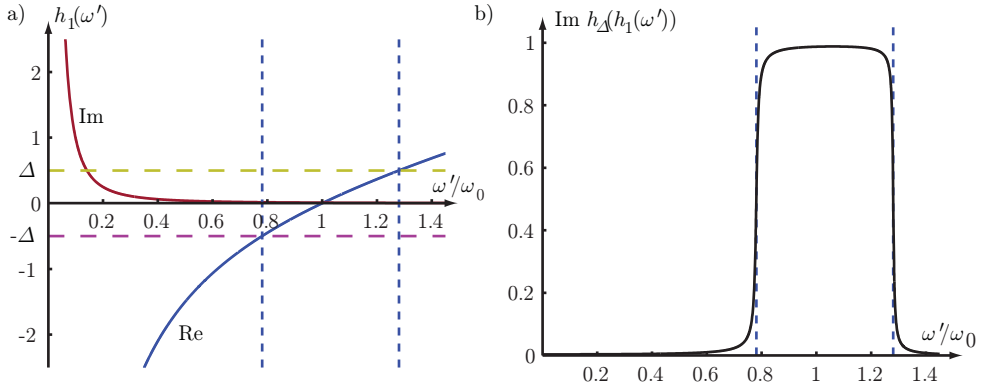



**Figure 4**: The left figure depicts the real and imaginary part of $h_1(\omega') = \lim_{\omega''\to 0} h_1(\omega' + \mathrm{i}\omega'')$, where $h_1$ is given by (5.6) and the permittivity is described by the Drude model $\epsilon(\omega) = 1 - (\omega/\omega_0\,(\omega/\omega_0 - 0.01\mathrm{i}))^{-1}$. The right figure depicts the integrand $\operatorname{Im} h_\Delta(h_1(\omega')) = \lim_{\omega''\to 0} \operatorname{Im} h_\Delta(h_1(\omega' + \mathrm{i}\omega''))$ in (5.8) for this choice of $\epsilon(\omega)$ and $\Delta = 1/2$.

chosen because $\operatorname{Im} h_\Delta(z) \approx 1$ for $\operatorname{Im} z \approx 0$ and $|\operatorname{Re} z| \leqslant \Delta$, see Fig. 3. Here the logarithm has its branch cut along the negative imaginary axis. The asymptotic expansions of the composition are

$$h_\Delta(h_1(\omega)) = \begin{cases} \mathcal{O}(1), & \text{as } \omega \hat{\to} 0 \\ \frac{-2\omega_0\Delta}{\omega\pi\epsilon_\infty} + o(\omega^{-1}), & \text{as } \omega \hat{\to} \infty, \end{cases}$$

yielding the following sum rule for $\hat{p} = 0$:

$$\lim_{\varepsilon\to 0^+} \lim_{\omega''\to 0^+} \int_0^{\varepsilon^{-1}} \operatorname{Im} h_\Delta(h_1(\omega' + \mathrm{i}\omega''))\,\mathrm{d}\omega' = \lim_{\varepsilon\to 0^+} \lim_{\omega''\to 0^+} \int_0^{\varepsilon^{-1}} \arg\left(\frac{(\omega' + \mathrm{i}\omega'')\epsilon_\infty - \Delta\omega_0}{(\omega' + \mathrm{i}\omega'')\epsilon_\infty - \Delta\omega_0}\right) \mathrm{d}\omega' = \frac{\omega_0\Delta}{\epsilon_\infty}. \quad (5.8)$$

An illustration of $\lim_{\omega''\to 0} \operatorname{Im} h_\Delta(h_1(\omega' + \mathrm{i}\omega''))$ for a permittivity function $\epsilon$ described by a Drude model can be found in Fig. 4.

Let the frequency interval be $\mathcal{B} = [\omega_0(1 - B_\mathrm{F}/2), \omega_0(1 + B_\mathrm{F}/2)]$, where $B_\mathrm{F}$ denotes the fractional bandwidth, and let $\Delta = \sup_{\omega'\in\mathcal{B}} |h_1(\omega')|$, where $h_1(\omega') = \lim_{\omega''\to 0^+} h_1(\omega' + \mathrm{i}\omega'')$. Then $\inf_{\omega'\in\mathcal{B}} \lim_{\omega''\to 0^+} \operatorname{Im} h_\Delta(h_1(\omega' + \mathrm{i}\omega'')) \geqslant 1/2$, which yields the bound

$$\sup_{\omega'\in\mathcal{B}} |h_1(\omega')| \geqslant \frac{B_\mathrm{F}}{2}\epsilon_\infty \quad \text{or} \quad \sup_{\omega'\in\mathcal{B}} |\epsilon(\omega')| \geqslant \frac{B_\mathrm{F}}{2 + B_\mathrm{F}}\epsilon_\infty.$$

This shows that $\epsilon$ near-zero materials are dispersive and that the deviation from zero is proportional to the fractional bandwidth when $B_\mathrm{F} \ll 1$.

## 5.5 Extinction cross section

This example revisits a set of sum rules for the extinction cross sections of certain passive scattering objects. The sum rules were first presented for linearly polarized waves in [34], and later generalized to elliptical polarizations in [14]. A time domain approach to derive them was adopted in [12]. Here they are reviewed in the special case of a spherically symmetric scatterer; the material properties of the scatterer considered is only dependent on the distance $r$ from the origin in the center of the sphere. Furthermore, the isotropic constitutive relation for the electric flux density in the object is in convolution form as described in (5.4), and the material is passive. For simplicity, the sphere is assumed to be non-magnetic and surrounded by free space.

Let a plane electromagnetic wave, propagating in the $\hat{\boldsymbol{k}}$-direction, impinge on the sphere. The electric field of such a plane wave in the time domain is $\boldsymbol{E}_{\mathrm{i}}(t, \boldsymbol{r}) = \boldsymbol{E}_0(t - \boldsymbol{r} \cdot \hat{\boldsymbol{k}}/\mathrm{c})$. Here $\boldsymbol{r}$ denotes the spatial coordinate, $\hat{\boldsymbol{k}}$ is of unit length, and c denotes the speed of light in free space. The electric field in the frequency domain may be written $\widetilde{\boldsymbol{E}}_{\mathrm{i}}(k) = \mathrm{e}^{\mathrm{i}\boldsymbol{r}\cdot k\hat{\boldsymbol{k}}}\widetilde{\boldsymbol{E}}_0(k)$, where the wavenumber $k = \omega/\mathrm{c}$ is used instead of the angular frequency $\omega$.

The extinction cross section $\sigma_{\mathrm{e}}(k)$ is a measure of the amount of energy in the incoming wave that is scattered or absorbed when the wave interacts with the sphere:

$$e_{\mathrm{e}}(\infty) = \frac{\mathrm{c}}{2\pi\eta_0} \int_{-\infty}^{\infty} \sigma_{\mathrm{e}}(k) |\widetilde{\boldsymbol{E}}_{\mathrm{i}}(k)|^2 \, \mathrm{d}k.$$

Here $\eta_0$ is the wave impedance of free space. The extinct energy, and hence also the extinction cross section, must be non-negative when the material of the sphere is passive. In fact, $\sigma_{\mathrm{e}}(k)$ is given by the imaginary part of a Herglotz function $h(k)$ due to the optical theorem [12, 14, 34]:

$$\sigma_{\mathrm{e}}(k) = \operatorname{Im} h(k), \quad \text{where} \quad h(k) = \frac{4\pi}{k} \tilde{S}(k; 0).$$

Here $\tilde{S}(k; 0)$ describes the scattered field in the forward direction. This Herglotz function satisfies the symmetry (3.8).

For most materials, it can be argued that $h(k) = \mathcal{O}(1)$ as $k \hat{\to} \infty$ [12, 34]. If the sphere is coated with metal (or some other material with static conductivity), then the low-frequency behaviour of $h(k)$ is described by

$$h(k) = 4\pi a^3 k + \mathcal{O}(k^2), \quad \text{as } k \to 0,$$

where $a$ is the outer radius. Note that the dominating term does not depend on the type of metal used. Consequently, the following sum rule applies for the extinction cross section of a sphere coated with metal:

$$\lim_{\varepsilon \to 0^+} \lim_{k'' \to 0^+} \int_{\varepsilon}^{\infty} \frac{\sigma_{\mathrm{e}}(k' + \mathrm{i}k'')}{k'^2} \, \mathrm{d}k' = 2\pi^2 a^3.$$

Alternatively, express the extinction cross section as a function of the wavelength, $\lambda = 2\pi/k$:

$$\lim_{\varepsilon \to 0^+} \lim_{\lambda'' \to 0^+} \int_0^{\varepsilon^{-1}} \sigma_{\mathrm{e},\lambda}(\lambda' - \mathrm{i}\lambda'') \, \mathrm{d}\lambda' = 4\pi^3 a^3. \tag{5.9}$$

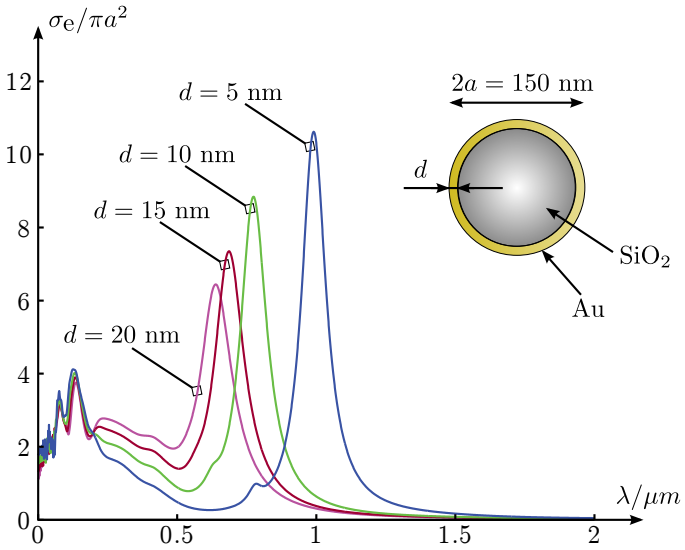



**Figure 5**: The normalized extinction cross section for four nanoshells, consisting of spherical silicon dioxide ($SiO_2$) cores with coats of gold. The outer radius is $a = 75\,\text{nm}$ and the shell thicknesses is $d = 5, 10, 15$ and $20\,\text{nm}$, respectively. The extinction cross section $\sigma_e$ was calculated from a closed form expression, using a Matlab-script for a Lorentz-Drude model for gold by Ung et al. [36]. The silicon dioxide core is modelled as being lossless with a constant complex permittivity $\epsilon(\omega) \equiv 2.25$, which is a good model at least for wavelengths 0.4–1.1 $\mu$m [24]. Following the sum rule (5.9), the integrated extinction for all four nanoshells is $4\pi^3 a^3$, which is confirmed by a numerical integration.

To exemplify the sum rule (5.9), consider the spherical nanoshells depicted in Fig. 5. A nanoshell is a dielectric core covered by a thin coat of metal, used for instance for biomedical imaging or treatment of tumours. Depending on the application, the core radius, shell thickness, and materials are varied to make the nanoshells scatter or absorb different parts of the visible light and near-infrared (NIR) spectra. In [7, 29], the nanoshells are spherical cores of silicon dioxide ($SiO_2$) covered with gold. The radius of the core is typically around 60 nm, and the gold shell is $5 - 20\,\text{nm}$ thick. The extinction cross sections for four such spheres are plotted in Fig. 5. Following the sum rule (5.9), the integrated extinction for any nanoshells is $4\pi^3 a^3$. This is confirmed by a numerical integration.

# 6 Conclusions

Many physical systems are modelled as a rule that assigns an output signal to every input signal. It is often natural to let the space of admissible input signals be some subset of the space of distributions, since generalized functions such as the delta function should be allowed. Under the general assumptions of linearity, continuity and time-translational invariance, such a system is in convolution form, and thus fully described by its impulse response in the time domain, and by its

transfer function in the frequency domain. The assumption of passivity (and thereby causality, as described in Section 3), implies that the transfer function is related to a Herglotz function [38, 40, 42].

A set of integral identities for Herglotz functions is presented and proved in this paper, showing that weighted integrals of Herglotz functions over infinite intervals are determined by their high- and low-frequency asymptotic expansions. The identities rely on a well-known representation theorem for Herglotz functions [2], and furthermore make use of results from the classical problem of moments [1, 32].

The integral identities make possible a general approach to derive sum rules for passive systems. The first step is to use the assumptions listed above to assure that the transfer function is related to a Herglotz function, $h$. Secondly, the low- and/or high-frequency asymptotic expansions of $h$ must be determined. Finally, physical limitations may be derived by considering finite frequency intervals. The sum rules effectively relate dynamic behaviour to static and/or high frequency properties, which must be found by physical arguments. However, since static properties are often easier to determine than dynamical behaviour in various applications, this is beneficial. The physical limitations indicate what can and cannot be expected from certain physical systems.

Sum rules, or more general dispersion relations, and physical limitations, have been widely used in e.g. electromagnetic theory. Two famous examples are the Kramers-Kronig relations for the frequency dependence of the electric permittivity [23, 25], discussed in Section 5.4, and Fano's matching equations [11], considered in Section 5.3. There are more recent examples as well, see e.g. [4, 13, 15–17, 30, 34].

For many causal systems in convolution form, dispersion relations in the form of a Hilbert transform pair follow from Titchmarsh's theorem [22, 23, 28]. Sometimes, sum rules can be derived from the dispersion relations [23]. Many previous papers use the Cauchy integral formula, see e.g. [11, 35]. This approach demands e.g. that the transfer function $\tilde{w}$ is rational. The present paper seems to be the first to describe and rigorously prove a general approach to obtain sum rules for systems in convolution form under the assumption of passivity. It should be stressed that since the different approaches works under different assumptions, they are complementary rather than in competition. One advantage of the Herglotz function-approach presented in this paper is that a wide range of physical systems obey passivity. Another advantage is that it gives an insight into how compositions of Herglotz functions may be used to derive new physical limitations, see Section 5.4.

# Acknowledgments

The financial support by the High Speed Wireless Communications Center of the Swedish Foundation for Strategic Research (SSF) is gratefully acknowledged. The authors would like to thank Anders Melin for invaluable assistance and fruitful discussions during the work with this paper.

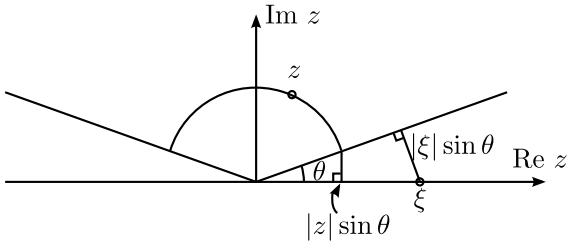



**Figure 6**: The Stoltz domain, $\{z : \theta \leqslant \arg z \leqslant \pi - \theta\}$ for some $\theta \in (0, \pi/2]$.

# Appendix

## A.1 Calculation of the limits $\lim_{z\hat{\to}\infty} h(z)/z$ and $\lim_{z\hat{\to}0} zh(z)$

For all $z$ in the Stoltz domain $\theta \leqslant \arg z \leqslant \pi - \theta$, $|\xi - z|$ is greater than or equal to both $|z|\sin\theta$ and $|\xi|\sin\theta$, see Figure 6. Thus

$$\frac{|1 + \xi z|}{|z(\xi - z)|} \leqslant \frac{1 + 1/|z|^2}{\sin\theta},$$

and (2.2) implies that

$$\lim_{z\hat{\to}\infty} \frac{h(z)}{z} = \beta + \lim_{z\hat{\to}\infty} \int_{\mathbb{R}} \frac{1 + \xi z}{z(\xi - z)} \,\mathrm{d}\nu(\xi) = \beta,$$

where Theorem A.2 has been used to move the limit inside the integral. Likewise, $|z(1 + \xi z)|/|\xi - z| \leqslant (1 + |z|^2)/\sin\theta$, which together with Theorem A.2 gives

$$\lim_{z\hat{\to}0} zh(z) = \lim_{z\hat{\to}0} \int_{\mathbb{R}} \frac{z(1 + \xi z)}{\xi - z} \,\mathrm{d}\nu(\xi) = -\nu(\{0\}) = -\mu(\{0\}).$$

## A.2 Proof of Lemma 4.1

The left-hand side of (4.1) is

$$\lim_{y\to 0^+} \int_{\mathbb{R}} \varphi(x) \left( \beta y + \int_{\mathbb{R}} \frac{y}{(x - \xi)^2 + y^2} \,\mathrm{d}\mu(\xi) \right) \mathrm{d}x = \lim_{y\to 0^+} \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x) \frac{y}{(x - \xi)^2 + y^2} \,\mathrm{d}x \,\mathrm{d}\mu(\xi),$$

where Fubini's Theorem [31, p. 164] was used to change the order of integration.

Lebesgue's Dominated Convergence Theorem (Theorem A.2) is used to show that the order of the limit and the integrals may be interchanged. First set

$$f_y(\xi) = \int_{\mathbb{R}} \varphi(x) \frac{y}{(x - \xi)^2 + y^2} \,\mathrm{d}x.$$

To find an integrable majorant $g \in \mathrm{L}^1(\mu)$ such that $|f_y(\xi)| \leqslant g(\xi)$ for all $\xi \in \mathbb{R}$ and $y \geqslant 0$, handle the cases $|\xi| < 2$ and $|\xi| \geqslant 2$ separately. For $|\xi| < 2$ it holds that

$$|f_y(\xi)| \leqslant \int_{\mathbb{R}} D \frac{y}{(x - \xi)^2 + y^2} \,\mathrm{d}x = D\pi.$$

For $|\xi| \geqslant 2$, divide the integral into $|x-\xi|<1$ and $|x-\xi| \geqslant 1$:

$$\left|\int_{|x-\xi|<1} \varphi(x)\frac{y}{(x-\xi)^2+y^2}\,\mathrm{d}x\right| \leqslant \frac{2D}{\xi^2+1}\int_{\mathbb{R}}\frac{y}{(x-\xi)^2+y^2}\,\mathrm{d}x = \frac{2\pi D}{\xi^2+1}$$

and

$$\left|\int_{|x-\xi|\geqslant 1} \varphi(x)\frac{y}{(x-\xi)^2+y^2}\,\mathrm{d}x\right| \leqslant \int_{|x-\xi|\geqslant 1}\frac{D}{1+x^2}\frac{y}{(x-\xi)^2}\,\mathrm{d}x$$
$$= Dy\left[\frac{\xi}{(\xi^2+1)^2}\ln\left|\frac{(\xi-1)^2+1}{(\xi+1)^2+1}\right| + \frac{2}{1+\xi^2} + \frac{\xi^2-1}{(\xi^2+1)^2}\frac{\pi}{2}\right] \leqslant \frac{D_1 y}{\xi^2+1}.$$

It is clear now that there is an integrable majorant. Since $\lim_{y\to 0^+} f_y(\xi)$ exists for all $\xi \in \mathbb{R}$ (shown below), the conditions of Theorem A.2 are fulfilled, and the limit may be moved inside the first integral.

Now let

$$f_{y,\xi}(x) = (\varphi(x)-\varphi(\xi))\,\frac{y}{(x-\xi)^2+y^2}.$$

First suppose that $\xi$ is not a point of discontinuity for $\varphi(\xi)$, so that there is some $K>0$ such that $\varphi(x)$ is continuous for $x \in [\xi-K, \xi+K]$. The constant $K$ may be chosen so that $\varphi$ is continuously differentiable in the interval, except possibly at the point $x=\xi$. For $x \in [\xi-K, \xi+K]$,

$$|f_{y,\xi}(x)| \leqslant \max_{|\zeta-\xi|\leqslant K}|\varphi'(\zeta)||x-\xi|\frac{y}{(x-\xi)^2+y^2} \leqslant D_3,$$

for some constant $D_3 \geqslant 0$. An integrable majorant for $f_{y,\xi}(x)$ is

$$|f_{y,\xi}(x)| \leqslant g_\xi(x) = \begin{cases} D_3, & \text{for } |x-\xi| \leqslant K \\ \frac{2D}{(x-\xi)^2}, & \text{otherwise,} \end{cases} \qquad \text{for all } y \leqslant 1.$$

Furthermore, the limit $\lim_{y\to 0^+} f_{y,\xi}(x)$ exists and is zero for all $x \in \mathbb{R}$. Thus Theorem A.2 applies and states that

$$\lim_{y\to 0^+}\int_{\mathbb{R}}\varphi(x)\frac{y}{(x-\xi)^2+y^2}\,\mathrm{d}x = \pi\varphi(\xi).$$

This proves the lemma for continuous $\varphi$.

Now suppose that $\xi$ is a point where $\varphi(\xi)$ has a discontinuity. Divide $\varphi(x)$ into two parts:

$$\varphi(x) = \underbrace{\frac{1}{2}\left(\varphi(x)+\varphi(2\xi-x)\right)}_{\varphi_{\mathrm{even}}(x)} + \underbrace{\frac{1}{2}\left(\varphi(x)-\varphi(2\xi-x)\right)}_{\varphi_{\mathrm{odd}}(x)},$$

where $\varphi_{\mathrm{even}}$ is even in $x$ with respect to an origin at the point $x=\xi$, and likewise $\varphi_{\mathrm{odd}}$ is odd in the same sense. Therefore

$$\int_{\mathbb{R}}\varphi_{\mathrm{odd}}(x)\frac{y}{(x-\xi)^2+y^2}\,\mathrm{d}x = 0, \quad \text{for all } y \geqslant 0. \tag{A.1}$$

Since the discontinuities of $\varphi$ are isolated points, $\varphi_{\text{even}}$ is continuous in a neighbourhood of $\xi$ and continuously differentiable except possibly at the point $x = \xi$. Furthermore, $\varphi_{\text{even}}(\xi) = \breve{\varphi}(\xi)$. The same reasoning as for continuous $\varphi$ results in

$$\lim_{y\to 0^+} \int_{\mathbb{R}} \varphi_{\text{even}}(x) \frac{y}{(x-\xi)^2 + y^2} \,\mathrm{d}x = \pi \breve{\varphi}(\xi).$$

Together with (A.1) this concludes the proof of the lemma.

## A.3 Proof of Corollary 4.1

Let $p = 0, \pm 1, \pm 2, \ldots$ and set

$$\varphi_{p,\varepsilon,\breve{\varepsilon}}(x) = \begin{cases} 0, & x < \varepsilon \\ x^{-p}, & \varepsilon < x < \breve{\varepsilon}^{-1} \\ 0, & x > \breve{\varepsilon}^{-1}. \end{cases} \tag{A.2}$$

This function satisfies the conditions of Lemma 4.1 for each fixed pair $\varepsilon, \breve{\varepsilon} > 0$, so

$$\lim_{y\to 0^+} \frac{1}{\pi} \int_{\varepsilon}^{\breve{\varepsilon}^{-1}} \frac{\operatorname{Im} h(x + \mathrm{i}y)}{x^p} \,\mathrm{d}x = \int_{\mathbb{R}} \breve{\varphi}_{p,\varepsilon,\breve{\varepsilon}}(\xi) \,\mathrm{d}\mu(\xi),$$

where $\breve{\varphi}_{p,\varepsilon,\breve{\varepsilon}}(\xi)$ is given by (4.2). The function $\breve{\varphi}_{p,\varepsilon,\breve{\varepsilon}}$ is monotonically increasing as $\varepsilon \to 0^+$ and/or $\breve{\varepsilon} \to 0^+$. The limit is:

$$\lim_{\varepsilon\to 0^+} \lim_{\breve{\varepsilon}\to 0^+} \breve{\varphi}_{p,\varepsilon,\breve{\varepsilon}}(\xi) = \begin{cases} 0, & \xi \leqslant 0 \\ \xi^{-p}, & \xi > 0. \end{cases}$$

Implement Lebesgue's Monotone Convergence Theorem (Theorem A.1), to get

$$\lim_{\varepsilon\to 0^+} \lim_{\breve{\varepsilon}\to 0^+} \int_{\mathbb{R}} \breve{\varphi}_{p,\varepsilon,\breve{\varepsilon}}(\xi) \,\mathrm{d}\mu(\xi) = \int_{\xi>0} \frac{\mathrm{d}\mu(\xi)}{\xi^p}, \quad p = 0, \pm 1, \pm 2, \ldots$$

The integral over $(-\breve{\varepsilon}^{-1}, -\varepsilon)$ is treated in the same manner. This proves the lemma, seeing that

$$\int_{\xi<0} \frac{\mathrm{d}\mu(\xi)}{\xi^p} + \int_{\xi>0} \frac{\mathrm{d}\mu(\xi)}{\xi^p} = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^p},$$

unless the left-hand side is $-\infty + \infty$. In this case the right-hand side is not defined.

For $p = 2, 3, \ldots$, the order of the limits $\breve{\varepsilon} \to 0^+$ and $y \to 0^+$ may be interchanged. Likewise, for $p = 0, -1, -2, \ldots$ the order of the limits $\varepsilon \to 0^+$ and $y \to 0^+$ may be interchanged. In that case there is an extra term $\delta_{p,0}\ \mu(\{0\})$ in the right-hand side. This is readily proved by considering the functions $\lim_{\breve{\varepsilon}\to 0^+} \varphi_{p,\varepsilon,\breve{\varepsilon}}(x)$ and $\lim_{\varepsilon\to 0^+} \varphi_{p,\varepsilon,\breve{\varepsilon}}(x)$, respectively.

## A.4 Proof of Lemma 4.2

Evidently, statement *(i)* always implies *(ii)*. Here it will be shown that *(ii)* implies *(iii)* and that *(iii)* implies *(i)*. Start with the case $N = 1$ and assume that *(iii)* holds. Consider the Herglotz function $h_0(z) = h(z) + \mu\{0\}/z$, represented by the measure $\mu_0$. Set

$$a_0 = \lim_{z\hat{\to}0} h_0(z) = \alpha + \lim_{z\hat{\to}0} \int_{\mathbb{R}} \frac{1+\xi z}{(\xi - z)(1+\xi^2)}\,\mathrm{d}\mu_0(\xi) = \alpha + \int_{\mathbb{R}} \frac{1}{\xi(1+\xi^2)}\,\mathrm{d}\mu_0(\xi).$$

Here Theorem A.2 could be used to move the limit under the integral sign, since for $z$ restricted to the Stoltz domain $\theta \leqslant \arg z \leqslant \pi - \theta$ it holds that $|\xi - z| \geqslant |\xi| \sin\theta$ (see Fig. 6) and $\int_{\mathbb{R}} \xi^{-2}\,\mathrm{d}\mu_0(\xi)$ is finite by assumption. Use this expression for $a_0$:

$$\lim_{z\hat{\to}0} \frac{h_0(z) - a_0}{z} = \beta + \lim_{z\hat{\to}0} \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{(\xi - z)\xi} = \beta + \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^2} = a_1,$$

where Theorem A.2 was used once more. Summing up, statement *(i)* is true.

Now assume that statement *(ii)* is valid (still $N = 1$), i.e. $h_0(\mathrm{i}y) = a_0 + a_1 \mathrm{i}y + o(y)$, as $y \to 0^+$, where $a_0, a_1 \in \mathbb{R}$. From this condition it follows that

$$\lim_{y\to 0^+} \frac{h_0(\mathrm{i}y) - h_0^*(\mathrm{i}y)}{2\mathrm{i}y} = \lim_{y\to 0^+} \left( a_1 + \frac{o(y)}{\mathrm{i}y} \right) = a_1.$$

But on the other hand,

$$\lim_{y\to 0^+} \frac{h_0(\mathrm{i}y) - h_0^*(\mathrm{i}y)}{2\mathrm{i}y} = \beta + \lim_{y\to 0^+} \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^2 + y^2} = \beta + \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^2}.$$

The change of order of the limit and integral is motivated by Theorem A.1. Ergo, $\int_{\mathbb{R}} \xi^{-2}\,\mathrm{d}\mu_0(\xi) = a_1 - \beta < \infty$, and thus statement *(iii)* is true.

The equivalence of the statements for all $N = 0, 1, 2, \ldots$ is proved by induction. For this reason, suppose that the equivalence has been proven for some $N \geqslant 1$, and that statement *(iii)* holds for $N + 1$, i.e. $\int_{\mathbb{R}} \xi^{-2N-2}\,\mathrm{d}\mu_0(\xi) < \infty$. Consider the function

$$h_1(z) = \frac{h_0(z) - a_0 - a_1 z}{z^2} = \frac{1}{z^2}\Bigg[ \beta z + \alpha + \int_{\mathbb{R}} \left( \frac{1}{\xi - z} - \frac{\xi}{1+\xi^2} \right) \mathrm{d}\mu_0(\xi) - \left( \alpha + \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi(1+\xi^2)} \right) - z \left( \beta + \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^2} \right) \Bigg] = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_1(\xi)}{\xi - z},$$

where $\mathrm{d}\mu_1(\xi) = \xi^{-2}\,\mathrm{d}\mu_0(\xi)$, and hence $h_1$ is a Herglotz function. Since $\int_{\mathbb{R}} \xi^{-2N}\,\mathrm{d}\mu_1(\xi)$ is finite, $h_1$ has the asymptotic expansion

$$h_1(z) = \sum_{n=0}^{2N-1} a_{n+2} z^n + o(z^{2N-1}) \quad \text{as } z\hat{\to}0,$$

where all $a_n$ are real. This proves statement *(i)* for $N + 1$.

On the other hand, assume that statement *(ii)* is valid for $N+1$, where $N \geqslant 1$. Consider the function $h_1$ once more. The induction assumption ensures that $\int_{\mathbb{R}} \xi^{-2N} \,\mathrm{d}\mu_1(\xi) = \int_{\mathbb{R}} \xi^{-2N-2} \,\mathrm{d}\mu_0(\xi) < \infty$, which proves statement *(iii)* for $N+1$.

Finally, note that from the representation of $h_1$ it is clear that

$$a_2 = \lim_{z \hat{\to} 0} h_1(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_1(\xi)}{\xi} = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^3}, \quad \text{and} \quad a_3 = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_1(\xi)}{\xi^2} = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0(\xi)}{\xi^4}.$$

This procedure may be continued for $a_4, a_5, \ldots, a_{2N-1}$ to prove (4.5), concluding the proof of the lemma.

## A.5 Auxiliary theorems

The following theorem can be found in e.g. [31], page 21:

**Theorem A.1** (Lebesgue's Monotone Convergence Theorem)**.** *Let $\{f_n\}$ be a sequence of real-valued measurable functions on $X$, and suppose that*

$$0 \leqslant f_1(x) \leqslant f_2(x) \leqslant \ldots \leqslant \infty, \quad \textit{for all } x \in X$$

*and*

$$f_n(x) \to f(x), \quad \textit{as } n \to \infty \textit{ for all } x \in X.$$

*Then $f$ is measurable, and*

$$\lim_{n\to\infty} \int_X f_n(x) \,\mathrm{d}\mu(x) = \int_X f(x) \,\mathrm{d}\mu(x).$$

The next theorem is also available in e.g. [31], page 26:

**Theorem A.2** (Lebesgue's Dominated Convergence Theorem)**.** *Suppose that $\{f_n\}$ is a sequence of complex-valued measurable functions on $X$ such that*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

*exists for every $x \in X$. If there is a function $g \in \mathrm{L}^1(\mu)$ such that*

$$|f_n(x)| \leqslant g(x), \quad \textit{for all } n = 1, 2, \ldots \textit{ and } x \in X,$$

*then $f \in \mathrm{L}^1(\mu)$,*

$$\lim_{n\to\infty} \int_X |f_n(x) - f(x)| \,\mathrm{d}\mu(x) = 0$$

*and*

$$\lim_{n\to\infty} \int_X f_n(x) \,\mathrm{d}\mu(x) = \int_X f(x) \,\mathrm{d}\mu(x).$$

# References

[1] N. I. Akhiezer. *The classical moment problem.* Oliver and Boyd, Edinburgh and London, 1965.

[2] N. I. Akhiezer and I. M. Glazman. *Theory of linear operators in Hilbert space*, volume 2. Frederick Ungar Publishing Co, New York, 1963.

[3] Y. M. Berezansky, Z. G. Sheftel, and G. F. Us. *Functional Analysis.* Birkhäuser, Boston, 1996.

[4] C. R. Brewitt-Taylor. Limitation on the bandwidth of artificial perfect magnetic conductor surfaces. *Microwaves, Antennas & Propagation, IET*, **1**(1), 255–260, 2007.

[5] H. J. Carlin and P. P. Civalleri. *Wideband circuit design.* CRC Press, Boca Raton, 1998.

[6] W. Cauer. The Poisson integral for functions with positive real part. *Bulletin of the American Mathematical Society*, **38**(1919), 713–717, 1932.

[7] M. R. Choi, K. J. Stanton-Maxey, J. K. Stanley, C. S. Levin, R. Bardhan, D. Akin, S. Badve, J. Sturgis, J. P. Robinson, R. Bashir, et al. A cellular Trojan horse for delivery of therapeutic nanoparticles into tumors. *Nano Letters*, **7**(12), 3759–3765, 2007.

[8] S. Darlington. Synthesis of reactance 4-poles. *Journal of Mathematics and Physics*, **18**, 275–353, September 1939.

[9] W. F. Donoghue, Jr. *Distributions and Fourier Transforms.* Academic Press, New York, 1969.

[10] P. L. Duren. *Theory of $H^p$ Spaces.* Dover Publications, New York, 2000.

[11] R. M. Fano. Theoretical limitations on the broadband matching of arbitrary impedances. *Journal of the Franklin Institute*, **249**(1,2), 57–83 and 139–154, 1950.

[12] M. Gustafsson. Time-domain approach to the forward scattering sum rule. *Proc. R. Soc. A*, **466**, 3579–3592, 2010.

[13] M. Gustafsson and D. Sjöberg. Sum rules and physical bounds on passive metamaterials. *New Journal of Physics*, **12**, 043046, 2010.

[14] M. Gustafsson and C. Sohl. New physical bounds on elliptically polarized antennas. In *Proceedings of the Third European Conference on Antennas and Propagation*, pages 400–402, Berlin, Germany, March 23–27 2009. The Institution of Engineering and Technology.

[15] M. Gustafsson, C. Sohl, and G. Kristensson. Physical limitations on antennas of arbitrary shape. *Proc. R. Soc. A*, **463**, 2589–2607, 2007.

[16] M. Gustafsson. Sum rule for the transmission cross section of apertures in thin opaque screens. *Opt. Lett.*, **34**(13), 2003–2005, 2009.

[17] M. Gustafsson, C. Sohl, C. Larsson, and D. Sjöberg. Physical bounds on the all-spectrum transmission through periodic arrays. *EPL Europhysics Letters*, **87**(3), 34002 (6pp), 2009.

[18] S. Hassi and A. Luger. Generalized zeros and poles of $\mathcal{N}_\kappa$-functions: On the underlying spectral structure. *Methods of Functional Analysis and Topology*, **12**(2), 131–150, 2006.

[19] P. Henrici. *Applied and Computational Complex Analysis*, volume 2. John Wiley & Sons, New York, 1977.

[20] G. Herglotz. Über Potenzreihen mit positivem, reellem Teil im Einheitskreis. *Leipziger Berichte*, **63**, 501–511, 1911.

[21] I. S. Kac and M. G. Kreĭn. R-functions — analytic functions mapping the upper halfplane into itself. *Amer. Math. Soc. Transl.(2)*, **103**, 1–18, 1974.

[22] F. W. King. *Hilbert Transforms, Volume 1.* Cambridge University Press, 2009.

[23] F. W. King. *Hilbert Transforms, Volume 2.* Cambridge University Press, 2009.

[24] R. Kitamura, L. Pilon, and M. Jonasz. Optical constants of silica glass from extreme ultraviolet to far infrared at near room temperature. *Applied optics*, **46**(33), 8118–8133, 2007.

[25] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskiĭ. *Electrodynamics of Continuous Media.* Pergamon, Oxford, second edition, 1984.

[26] J. Mashreghi. *Representation Theorems in Hardy Spaces.* Cambridge University Press, Cambridge, U.K., 2009.

[27] R. H. Nevanlinna. Asymptotische Entwicklungen beschränkter Funktionen und das Stieltjes'sche Momentenproblem. *Annales Academiae Scientiarum Fennicae A*, **18**(5), 1–53, 1922.

[28] H. M. Nussenzveig. *Causality and dispersion relations.* Academic Press, London, 1972.

[29] S. J. Oldenburg, R. D. Averitt, S. L. Westcott, and N. J. Halas. Nanoengineering of optical resonances. *Chemical Physics Letters*, **288**(2-4), 243–247, 1998.

[30] K. N. Rozanov. Ultimate thickness to bandwidth ratio of radar absorbers. *IEEE Trans. Antennas Propagat.*, **48**(8), 1230–1234, August 2000.

[31] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 1987.

[32] J. A. Shohat and J. D. Tamarkin. *The problem of moments*. American Mathematical Society, 1943.

[33] M. G. Silveirinha and N. Engheta. Theory of supercoupling, squeezing wave energy, and field confinement in narrow channels and tight bends using $\varepsilon$ near-zero metamaterials. *Phys. Rev. B*, **76**(24), 245109, 2007.

[34] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on broadband scattering by heterogeneous obstacles. *J. Phys. A: Math. Theor.*, **40**, 11165–11182, 2007.

[35] C. Sohl, M. Gustafsson, G. Kristensson, and S. Nordebo. A general approach for deriving bounds in electromagnetic theory. In *Proceedings of the XXIXth URSI General Assembly*, page B01p4, Chicago, IL, USA, August 7–16 2008. International Union of Radio Science.

[36] B. Ung and Y. Sheng. Interference of surface waves in a metallic nanoslit. *Optics Express*, **15**(3), 1182–1190, 2007.

[37] R. L. Weaver and Y. H. Pao. Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media. *J. Math. Phys.*, **22**, 1909–1918, 1981.

[38] M. Wohlers and E. Beltrami. Distribution theory as the basis of generalized passive-network analysis. *IEEE Transactions on Circuit Theory*, **12**(2), 164–170, 1965.

[39] T. T. Wu. Some properties of impedance as a causal operator. *Journal of Mathematical Physics*, **3**, 262, 1962.

[40] D. Youla, L. Castriota, and H. Carlin. Bounded real scattering matrices and the foundations of linear passive network theory. *IRE Transactions on Circuit Theory*, **6**(1), 102–124, 1959.

[41] A. H. Zemanian. An n-port realizability theory based on the theory of distributions. *IEEE Transactions on Circuit Theory*, **10**(2), 265–274, 1963.

[42] A. H. Zemanian. *Distribution theory and transform analysis: an introduction to generalized functions, with applications*. McGraw-Hill, New York, 1965.

[43] A. H. Zemanian. *Realizability theory for continuous linear systems*. Academic Press, New York, 1972.

# Physical Limitations on the Scattering of Electromagnetic Vector Spherical Waves

**Paper II**

Anders Bernland, Mats Gustafsson, and Sven Nordebo

**Based on:** A. Bernland, M. Gustafsson, and S. Nordebo. Physical limitations on the scattering of electromagnetic vector spherical waves. *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 14, paper 145401, March 2011.

**Abstract**

Understanding the interaction between electromagnetic waves and matter is vital in applications ranging from classical optics to antenna theory. This paper derives physical limitations on the scattering of electromagnetic vector spherical waves. The assumptions made are that the heterogeneous scatterer is passive, and has constitutive relations which are in convolution form in the time domain and anisotropic in the static limit. The resulting bounds limit the reflection coefficient of the modes over a frequency interval, and can thus be interpreted as limitations on the absorption of power from a single mode. They can be used within a wide range of applications, and are particularly useful for electrically small scatterers. The derivation follows a general approach to derive sum rules and physical limitations on passive systems in convolution form. The time domain versions of the vector spherical waves are used to describe the passivity of the scatterer, and a set of integral identities for Herglotz functions are applied to derive sum rules from which the physical limitations follow.

# 1 Introduction

Understanding how electromagnetic fields interact with matter is not only vital in classical science, like optics and scattering theory, but also in modern applications like wireless communication, cloaking and metamaterials. When interacting with various objects, electromagnetic waves may be scattered and/or absorbed. If the objects are small compared to the wavelength, this interaction is limited. An early paper addressing these limits is Purcell's [19], discussing radiation emission and absorption by interstellar dust. Results similar to Purcell's can also be found in [3]. Limitations on antenna performance where introduced by Chu in [6]. Sohl et al. derives limitations on the extinction cross sections of arbitrary heterogeneous, anisotropic objects in [24], results that are directly applicable to antenna theory [9]. A summary of some important results on physical limitations on antennas can be found in Volakis' book [26], see also Hansen's book [12]. More general dispersion relations for electromagnetic as well as quantum-mechanical scattering are discussed in e.g. [13, 17] and references therein.

Electromagnetic fields can be decomposed into orthogonal vector spherical waves [4, 11], also referred to as partial waves, (electric and magnetic) multipoles, or (TM and TE) modes. Such a decomposition is very beneficial in scattering theory. In wireless communication, these orthogonal modes are closely related to the orthogonal communication channels of multiple-input multiple-output (MIMO) systems [8].

The present paper seems to be the first to derive physical limitations on the scattering and absorption of electromagnetic vector spherical waves. To do so, a general approach to obtain sum rules and physical limitations for passive systems in convolution form put forth in [2] is used. At the core of this approach is a set of integral identities for Herglotz functions, a class of functions that is intimately linked to the transfer functions of passive systems.

The main results of this paper are physical limitations on the reflection coefficients of the modes for arbitrary heterogeneous, passive scatterers with constitutive relations in convolution form, and anisotropic in the static limit. The limitations state that the reflection coefficient cannot be arbitrarily small over a frequency interval of non-zero length; how small it can be depends upon the smallest sphere circumscribing the scatterer, its static material properties and the fractional bandwidth. An interpretation of the bounds on the reflection coefficients is as bounds on the maximum absorption of power from a single mode. The limitations are particularly useful for electrically small scatterers, and so they are well suited to analyse sub-wavelength particles designed to be resonant in one or more frequency bands, such as antennas and metamaterials.

This paper is divided into the following sections: First, in Section 2, the general approach to derive sum rules and physical limitations for passive systems presented in [2] is reviewed. In order to use this method and obtain the bounds in this paper, expressions for the vector spherical waves in both the time and frequency domains are needed. This is the topic of Section 3. In Section 4, the scattering matrix is introduced, and the physical limitations are derived. After this comes two examples in Section 5, one which discusses absorption of power in nanoshells, and another considers limitations on antenna performance. Last come some concluding remarks in Section 6.

## 2 A general approach to obtain sum rules and physical limitations on passive systems

The derivation of the physical limitations on scattering of vector spherical waves in this paper follows a general approach to obtain sum rules and physical limitations for passive systems in convolution form presented in [2]. This section summarizes this general approach in order to put the following sections in the right context. The general approach is described more thoroughly in [2], where all the necessary proofs can be found.

There are three major steps to obtain sum rules and associated limitations for a physical system: First, the transfer function of the system is related to a Herglotz function $h$. Secondly, the low-frequency asymptotic expansion of the transfer function is determined. This step commonly uses physical arguments, and is specific to each application. Then a set of integral identities for Herglotz functions, relating weighted integrals of $h$ to its low-frequency asymptotic expansion, is used. Essentially, this relates the dynamical behaviour of the physical system to its static properties. In the third step, physical limitations are derived by estimating the integral. Variational principles can sometimes be applied to the static parameters if they are unknown.

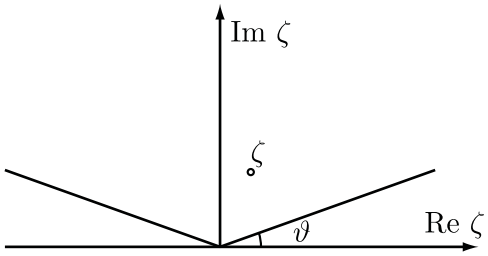



**Figure 1**: The cone $\{\zeta : \vartheta \leqslant \arg\zeta \leqslant \pi - \vartheta\}$ for some $\vartheta \in (0, \pi/2]$.

## 2.1 Herglotz functions and integral identities

Here the class of Herglotz functions is reviewed briefly, and the integral identities used to obtain sum rules for passive systems are presented. A Herglotz function $h$ is defined as a function holomorphic in $\mathbb{C}^+ = \{\zeta : \operatorname{Im}\zeta > 0\}$, satisfying $\operatorname{Im} h(\zeta) \geqslant 0$ there. Furthermore, many Herglotz functions appearing in various applications are of the form $h(\zeta) = \alpha + h_1(\zeta)$, where $h_1$ exhibits the symmetry $h_1(\zeta) = -h_1^*(-\zeta^*)$ and $\alpha \in \mathbb{R}$ [2]. Such a function $h$ is called symmetric in this paper, and it satisfies the low-frequency expansion

$$h(\zeta) = \alpha + \sum_{n=0}^{N} A_{2n-1}\zeta^{2n-1} + \mathrm{o}(\zeta^{2N-1}), \quad \text{as } \zeta \hat{\to} 0, \tag{2.1}$$

for some integer $N \geqslant 0$. Here $A_{-1} \leqslant 0$ and all $A_n$ are real. The limit $\zeta \hat{\to} 0$ is a short-hand notation for $|\zeta| \to 0$ for $\zeta$ in the cone $\vartheta \leqslant \arg\zeta \leqslant \pi - \vartheta$ for any $\vartheta \in (0, \pi/2]$, see Figure 1. The asymptotic expansion (2.1) is clearly valid as $\zeta \to 0$ for any argument in the case $h$ is holomorphic in a neighbourhood of the origin.

There is a set of integral identities for a symmetric Herglotz function $h$ [2]:

$$\frac{2}{\pi}\int_0^\infty \frac{\operatorname{Im} h(\omega)}{\omega^{2p}}\,\mathrm{d}\omega = A_{2p-1} - \delta_{p,1}\beta, \qquad p = 1, 2\ldots, N. \tag{2.2}$$

Here $\delta_{p,q}$ denotes the Kronecker delta and $\beta = \lim_{\zeta \hat{\to} \infty} h(\zeta)/\zeta \geqslant 0$, which always exists finitely. The notation $\omega = \operatorname{Re}\zeta$ and $\sigma = \operatorname{Im}\zeta$ is used throughout this paper. If the imaginary part $\operatorname{Im} h(\omega)$ is not regular on the real line, the above integral should be interpreted as

$$\lim_{\varepsilon\to 0^+}\lim_{\sigma\to 0^+}\frac{2}{\pi}\int_\varepsilon^\infty \frac{\operatorname{Im} h(\omega + \mathrm{i}\sigma)}{\omega^{2p}}\,\mathrm{d}\omega = A_{2p-1} - \delta_{p,1}\beta, \qquad p = 1, 2\ldots, N. \tag{2.3}$$

This is equivalent to interpreting the left-hand side of (2.2) in the distributional sense; the generalized function $h(x) = \lim_{\sigma\to 0^+} h(\omega + \mathrm{i}\sigma)$, where the right-hand side is interpreted as a limit of distributions, is a distribution of slow growth [2]. Equation (2.2) is understood to be replaced by (2.3) whenever necessary throughout this paper.

## 2.2 Sum rules for passive systems

Having introduced Herglotz functions, it remains to discuss the link between this class of functions and the transfer functions of passive systems in convolution form, i.e. the first step of the general approach. The results presented here relies mainly on the work by Youla et. al. [28], Zemanian [29–31] and Wohlers and Beltrami [27]. See also the book [17] by Nussenzveig. How the integral identities (2.2) can be used to derive sum rules for such systems is also explained in this section.

Consider a physical system in convolution form in the time domain,

$$u(t) = w * v(t), \tag{2.4}$$

where $v$ and $u$ are the input and output signals, respectively, related to each other by the impulse response $w$. The context of distributions is natural, since generalized functions such as the delta function should be allowed; hence, the impulse response is assumed to be a distribution, $w \in \mathcal{D}'$, and the convolution is interpreted in the distributional sense [30]. Many physical systems obey causality, which intuitively means that the output cannot precede the input. For the mathematical model (2.4) it means that $\operatorname{supp} w \subseteq [0, \infty)$ [30].

Another crucial assumption is that of passivity; if the power of the input (output) signal at the time $t$ is $|v(t)|^2$ ($|u(t)|^2$), the power absorbed by the system is $|v(t)|^2 - |u(t)|^2$. In this paper, a system is defined to be passive if the energy expression

$$e(T) = \int_{-\infty}^{T} \left( |v(t)|^2 - |u(t)|^2 \right) \mathrm{d}t \tag{2.5}$$

is non-negative for all $T \in \mathbb{R}$ and $v \in \mathcal{D}$, where $\mathcal{D}$ denotes smooth functions of compact support [27, 31].[1] Only input signals $v \in \mathcal{D}$ are considered in order for the integral to be well-defined. However, this is often enough to ensure that the corresponding energy expressions are non-negative for other admissible input signals $v$.

It turns out passivity implies causality for systems in convolution form [28, 31]. Passivity also implies that the impulse response is a distribution of slow growth [27, 31], $w \in \mathcal{S}'$ and hence Fourier transformable in the distributional sense. In this paper, the Fourier transform for all such distributions $f$ is defined through $\langle \mathcal{F}f, \varphi \rangle = \langle f, \mathcal{F}\varphi \rangle$ for all $\varphi \in \mathcal{S}$. Here $\mathcal{S}$ denotes the set of smooth functions of rapid descent, $\langle f, \varphi \rangle$ is the value in $\mathbb{C}$ that $f \in \mathcal{S}'$ assigns to $\varphi \in \mathcal{S}$ [30], and the Fourier transform of $\varphi$ is defined as $\mathcal{F}\varphi(\omega) = \int_{\mathbb{R}} \varphi(t) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t$. The frequency domain version of (2.4) is $\tilde{u}(\omega) = \tilde{w}(\omega)\tilde{v}(\omega)$, where the transfer function of the system is given by

$$\tilde{w}(\omega) = (\mathcal{F}w)(\omega),$$

and $\tilde{v} = \mathcal{F}v$ and $\tilde{u} = \mathcal{F}u$ are the input and output signals, respectively [2].

Passivity implies that the region of convergence for $\tilde{w}(\zeta)$ contains $\mathbb{C}^+$ and $\tilde{w}(\zeta)$ is holomorphic there. In other words, $\tilde{w}(\zeta)$ is holomorphic for $\operatorname{Im} \zeta > 0$. Furthermore,

[1] This definition of passivity is sometimes cast in a different form, see [2, 28, 30].

the transfer-function $\tilde{w}(\zeta)$ is bounded in magnitude by one for $\operatorname{Im}\zeta > 0$ [27, 31]. The transfer function $\tilde{w}$ is not necessarily holomorphic in a neighbourhood of the real axis, but the function $\tilde{w}(\omega) = \lim_{\sigma\to 0^+} \tilde{w}(\omega + \mathrm{i}\sigma)$ is well-defined for almost all $\omega \in \mathbb{R}$ and bounded in magnitude by one [16].

One more assumption on the physical system (2.4) is convenient (but not necessary): It is assumed that it maps real-valued input signals to real-valued output signals, which means that $w$ is real. This implies the symmetry

$$\tilde{w}(\zeta) = \tilde{w}^*(-\zeta^*), \qquad \operatorname{Im}\zeta > 0, \tag{2.6}$$

where the superscript $^*$ is used to denote the complex conjugate.

A Herglotz function can be constructed from $\tilde{w}$ in two ways, either with the inverse Cayley transform of $\pm\tilde{w}$, or by taking the complex logarithm of $\tilde{w}$ [2]. The latter way is chosen here. It requires that the zeros of $\tilde{w}$ are removed, which is done with a Blaschke-product. The Herglotz function is therefore

$$h(\zeta) = -\mathrm{i}\log\left(\frac{\tilde{w}(\zeta)}{B(\zeta)}\right), \tag{2.7}$$

where

$$B(\zeta) = \prod_n \frac{1-\zeta/\zeta_n}{1-\zeta/\zeta_n^*} \tag{2.8}$$

is a Blaschke product [16], repeating the possible zeros $\zeta_n$ of $\tilde{w}$ in $\mathbb{C}^+$ according to their multiplicity. The symmetry (2.6) implies that $h(\zeta)$ is symmetric in the sense discussed in Section 2.1, with $\alpha = \arg\tilde{w}(\mathrm{i}\sigma)$.

The integral identities (2.2) applied to the function in (2.7) yield [2]

$$\frac{2}{\pi}\int_0^\infty \frac{1}{\omega^{2p}} \ln\frac{1}{|\tilde{w}(\omega)|}\,\mathrm{d}\omega = A_{2p-1} - \delta_{p,1}\beta, \qquad p = 1, 2\ldots, N, \tag{2.9}$$

where it has been used that $|B(\omega+\mathrm{i}\sigma)| \to 1$ as $\sigma \to 0$ for almost all $\omega \in \mathbb{R}$ [16]. The low-frequency asymptotic expansion in (2.1) may be related to the behaviour of $\tilde{w}(\zeta)$ as $\zeta\hat{\to}0$, where as before $\zeta\hat{\to}0$ is short-hand notation for $|\zeta| \to 0$ for $\zeta$ in the cone $\vartheta \leqslant \arg\zeta \leqslant \pi - \vartheta$ for any $\vartheta \in (0, \pi/2]$. If $\tilde{w}(\zeta)$ is holomorphic in a neighbourhood of the origin, the low frequency limit is identical whatever the argument of $\zeta$, so the asymptotic expansion can be determined along the real axis ($\lim_{\omega\to 0}\tilde{w}(\omega)$). The cone assures that the low-frequency limit is only dependent on the behaviour of $w(t)$ for arbitrarily large times $t$ [2]. The asymptotic behaviour of $\tilde{w}(\zeta)$ as $\zeta\hat{\to}0$ must be found by physical arguments specific to each application, and constitutes the second step of the general three-step approach [2]. In the third step, physical limitations may be derived by considering integrals over finite frequency intervals, since the integrand in (2.9) is non-negative. In some cases, variational principles are used to bound the expansion coefficients $A_p$ of $h$ when they are unknown.

# 3 Vector spherical waves in the time and frequency domains

Expansions of the electric and magnetic fields in vector spherical waves are widely employed in the frequency domain, see e.g. [4, 11]. Their counterparts in the time domain have been treated by Shlivinski and Heyman [20, 21]. Both the time and frequency domain vector spherical waves are considered in this section, since they are both required in Section 4 to derive physical limitations for passive scatterers according to the general approach described in Section 2; the time domain waves are necessary in order to rigorously describe passive scatterers and find their corresponding Herglotz functions, whereas the frequency domain counterparts are needed to derive the low-frequency behaviour of the scatterer and determine sum rules and physical limitations. A tilde ($\tilde{}$) is used in the remainder of this paper to denote functions in the frequency domain, and it is also convenient to employ the wavenumber $k = \omega/\mathrm{c}$, so that $\tilde{f}(k) = \mathcal{F}f(\omega)$. Here c is the speed of light in free space.

Consider an uncharged object in free space, and let $a$ be the radius of a sphere (centered at the origin) containing the object, see Figure 2. Outside this sphere, the electric field is expanded in outgoing and incoming vector spherical waves, denoted $\boldsymbol{u}_\nu^{(1)}$ and $\boldsymbol{u}_\nu^{(2)}$, respectively:

$$\widetilde{\boldsymbol{E}}(\boldsymbol{r}, k) = k\sqrt{\eta_0} \sum_\nu \mathrm{i}^{l+2-\tau} \left[ \tilde{b}_\nu^{(1)}(k) \boldsymbol{u}_\nu^{(1)}(k\boldsymbol{r}) + \tilde{b}_\nu^{(2)}(k) \boldsymbol{u}_\nu^{(2)}(k\boldsymbol{r}) \right], \tag{3.1}$$

where the expansion coefficients $\tilde{b}_\nu^{(1)}(k)$ and $\tilde{b}_\nu^{(2)}(k)$ are functions of the wavenumber $k$. Here $\eta_0$ is the wave impedance in free space. The spatial coordinate is denoted $\boldsymbol{r}$, and in the rest of the paper the notation $r = |\boldsymbol{r}|$ and $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$ is employed. For a definition of the vector spherical waves, see Appendix A.1. The multi-index $\nu = \{\tau, s, m, l\}$ is introduced to simplify the notation, and the factors $k\sqrt{\eta_0}\mathrm{i}^{l+2-\tau}$ are included for consistency with the time domain expansion described below. The corresponding magnetic field is

$$\widetilde{\boldsymbol{H}}(\boldsymbol{r}, k) = \frac{k}{\sqrt{\eta_0}} \sum_\nu \mathrm{i}^{l+1-\tau} \left[ \tilde{b}_\nu^{(1)}(k) \boldsymbol{u}_{\bar{\nu}}^{(1)}(k\boldsymbol{r}) + \tilde{b}_\nu^{(2)}(k) \boldsymbol{u}_{\bar{\nu}}^{(2)}(k\boldsymbol{r}) \right], \tag{3.2}$$

where the dual multi-index $\bar{\nu} = \{\bar{\tau}, s, m, l\}$ with $\bar{\tau} = 3 - \tau$ has been introduced.

Outgoing vector spherical waves in the time domain are described thoroughly in [20, 21]. A short description, also covering incoming waves, is included here for clarity. Assuming that the fields vanish as $t \to -\infty$, the inverse Laplace transform may be applied to (3.1)–(3.2) with $k = \mathrm{i}s/\mathrm{c}$ and the integration curve over $s$ sufficiently far into the right half-plane. Using the explicit expressions (A.1) for the vector spherical waves yields the transverse electric field $\widetilde{\boldsymbol{E}}_\mathrm{T} = \widetilde{\boldsymbol{E}} - \hat{\boldsymbol{r}}(\hat{\boldsymbol{r}} \cdot \widetilde{\boldsymbol{E}})$:

$$\widetilde{\boldsymbol{E}}_\mathrm{T}(\boldsymbol{r}, k) = \frac{\sqrt{\eta_0}}{r} \sum_\nu \left[ \tilde{b}_\nu^{(1)}(k) \mathrm{e}^{-sr/\mathrm{c}} R_{\tau,l}^{(1)}(sr/\mathrm{c}) + \tilde{b}_\nu^{(2)}(k) \mathrm{e}^{sr/\mathrm{c}} R_{\tau,l}^{(2)}(sr/\mathrm{c}) \right] \boldsymbol{A}_\nu(\hat{\boldsymbol{r}}), \tag{3.3}$$

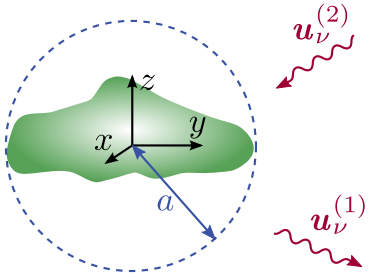



**Figure 2**: The scatterer is placed in free space, and contained in a sphere of radius $a$ centered at the origin. Outside this sphere, the electric and magnetic fields are expanded in outgoing and incoming vector spherical waves, $\boldsymbol{u}_\nu^{(1)}$ and $\boldsymbol{u}_\nu^{(2)}$, with index $\nu$.

where

$$\begin{cases} R_{1,l}^{(1)}(s) = \displaystyle\sum_{n=0}^{l} D_{n,l} s^{-n} \\ R_{1,l}^{(2)}(s) = (-1)^{l-1} R_{1,l}^{(1)}(-s) \\ R_{2,l}^{(j)}(s) = R_{1,l-1}^{(j)}(s) + \dfrac{l}{s} R_{1,l}^{(j)}(s), \qquad j = 1, 2, \end{cases}$$

and $D_{n,l} = (l+n)!/(2^n n!(l-n)!)$ according to (A.6). The vector spherical harmonics $\boldsymbol{A}_\nu$ are defined in Appendix A.1. Applying the inverse Laplace transform yields

$$\boldsymbol{E}_{\mathrm{T}}(\boldsymbol{r}, t) = \frac{\sqrt{\eta_0}}{r} \sum_\nu \left[ \mathcal{R}_{\tau,l}^{(1)} b_\nu^{(1)}(t - r/\mathrm{c}) + \mathcal{R}_{\tau,l}^{(2)} b_\nu^{(2)}(t + r/\mathrm{c}) \right] \boldsymbol{A}_\nu(\hat{\boldsymbol{r}}).$$

Here the operators $\mathcal{R}_{\tau,l}^{(j)} : \mathcal{D} \to \mathcal{D}$ in the time domain are defined by

$$\begin{cases} \mathcal{R}_{1,l}^{(j)} f(t) = (\pm 1)^{l-1} \displaystyle\sum_{n=0}^{l} D_{n,l} \left( \pm \frac{\mathrm{c}}{r} \mathrm{d}_t^{-1} \right)^n f(t) \\ \mathcal{R}_{2,l}^{(j)} f(t) = \mathcal{R}_{1,l-1}^{(j)} f(t) \pm l \dfrac{\mathrm{c}}{r} \mathrm{d}_t^{-1} \mathcal{R}_{1,l}^{(j)} f(t), \end{cases}$$

where the upper (lower) signs are for $j = 1$ ($j = 2$). The inverse to differentiation $\mathrm{d}_t^{-1}$ is chosen so that $\mathrm{d}_t^{-1} f(t)$ is the distributional primitive to $f$ that vanishes at $t = -\infty$, i.e. $\mathrm{d}_t^{-1} f(t) = \int_{-\infty}^{t} f(t')\, \mathrm{d}t'$ for regular functions $f$. A similar representation is used for the magnetic field, giving

$$\boldsymbol{H}_{\mathrm{T}}(\boldsymbol{r}, t) = \frac{1}{r\sqrt{\eta_0}} \sum_\nu \left[ \mathcal{R}_{\bar{\tau},l}^{(1)} b_\nu^{(1)}(t - r/\mathrm{c}) + \mathcal{R}_{\bar{\tau},l}^{(2)} b_\nu^{(2)}(t + r/\mathrm{c}) \right] (-1)^{\tau-1} \boldsymbol{A}_{\bar{\nu}}(\hat{\boldsymbol{r}}).$$

Recall that $b_\nu^{(j)}(t)$ are assumed to be distributions in general. In the case they are regular functions, the electromagnetic power passing in the negative $r$-direction

through a spherical shell of radius $r$ at the time $t$ is

$$P(r,t) = \int_{\Omega_{\hat{\boldsymbol{r}}}} r^2 \ (-\hat{\boldsymbol{r}}) \cdot [\boldsymbol{E}_{\mathrm{T}}(\boldsymbol{r},t) \times \boldsymbol{H}_{\mathrm{T}}(\boldsymbol{r},t)] \,\mathrm{d}\Omega_{\hat{\boldsymbol{r}}}$$

$$= -\int_{\Omega_{\hat{\boldsymbol{r}}}} r^2 \ \boldsymbol{E}_{\mathrm{T}}(\boldsymbol{r},t) \cdot [\boldsymbol{H}_{\mathrm{T}}(\boldsymbol{r},t) \times \hat{\boldsymbol{r}}] \,\mathrm{d}\Omega_{\hat{\boldsymbol{r}}}$$

$$= -\int_{\Omega_{\hat{\boldsymbol{r}}}} \left[ \sum_{\nu} \sum_{j=1}^{2} \mathcal{R}_{\tau,l}^{(j)} b_{\nu}^{(j)}(t \mp r/\mathrm{c}) \boldsymbol{A}_{\nu}(\hat{\boldsymbol{r}}) \right] \cdot \left[ \sum_{\nu} \sum_{j=1}^{2} \mathcal{R}_{\bar{\tau},l}^{(j)} b_{\nu}^{(j)}(t \mp r/\mathrm{c}) \boldsymbol{A}_{\nu}(\hat{\boldsymbol{r}}) \right] \mathrm{d}\Omega_{\hat{\boldsymbol{r}}},$$

where $\Omega_{\hat{\boldsymbol{r}}} = \{(\theta,\phi) : 0 \leqslant \theta \leqslant \pi, 0 \leqslant \phi \leqslant 2\pi\}$ is the unit sphere and $\mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi$. Here (A.2) has been employed, and the upper (lower) signs are for $j = 1$ ($j = 2$). The orthogonality relation (A.3) ensures that the two sums over $\nu$ may be replaced by one. Also, all cross-terms in $j$ cancel each other:

$$P(r,t) = -\sum_{\nu} \sum_{j=1}^{2} \left[ \mathcal{R}_{\tau,l}^{(j)} b_{\nu}^{(j)}(t \mp r/\mathrm{c}) \right] \left[ \mathcal{R}_{\bar{\tau},l}^{(j)} b_{\nu}^{(j)}(t \mp r/\mathrm{c}) \right].$$

The power $P(r,t)$ may be divided into one radiating part, and another part appertaining to the reactive near-field:

$$P(r,t) = P_{\mathrm{rad}}(r,t) + P_{\mathrm{react}}(r,t), \tag{3.4}$$

where

$$P_{\mathrm{rad}}(r,t) = \sum_{\nu} |b_{\nu}^{(2)}(t + r/\mathrm{c})|^2 - |b_{\nu}^{(1)}(t - r/\mathrm{c})|^2 \tag{3.5}$$

is only dependent on $r$ via $t \mp r/\mathrm{c}$. The reactive power $P_{\mathrm{react}}(r,t)$ tends to zero as $r \to \infty$, and furthermore it has a zero mean for all $r \geqslant a$, i.e.

$$\int_{-\infty}^{\infty} P_{\mathrm{react}}(r,t) \,\mathrm{d}t = 0. \tag{3.6}$$

This result is derived in [21], where also $P_{\mathrm{rad}}$ and $P_{\mathrm{react}}$ are described in more detail. An illustration of the radiative and reactive power flow for TM-modes of orders $l = 2$ and $l = 5$ can be found in Figure 1 and Figure 2 in [21].

# 4 The scattering matrix $\tilde{\mathbf{S}}$

This section introduces the scattering matrix $\tilde{\mathbf{S}}$, which for a given scatterer relates the outgoing wave amplitudes $\tilde{b}_{\nu}^{(1)}(k)$ to the incoming $\tilde{b}_{\nu}^{(2)}(k)$. The equivalent to the scattering matrix in the time domain is also covered. The elements of the scattering matrix are related to passive systems (as described in Section 2.2) in case the scatterer is passive. This is described in more detail below, using the time domain expressions for the vector spherical waves presented in Section 3. Herglotz functions corresponding to (2.7) and their low-frequency expansions of the type (2.1)

are derived next. In the end of the section all this is used to obtain sum rules and physical limitations on the diagonal elements of $\tilde{\mathbf{S}}$.

Assume that the scatterer is linear, continuous, and time-translational invariant, i.e. that the constitutive relations relating the electric and magnetic flux densities $\boldsymbol{D}(t)$ and $\boldsymbol{B}(t)$ to the electric and magnetic fields $\boldsymbol{E}(t)$ and $\boldsymbol{H}(t)$ are in convolution form. In this case the relation between the outgoing and incoming amplitudes $b_\nu^{(1)}(t)$ and $b_\nu^{(2)}(t)$ must also be in convolution form, $b_\nu^{(1)}(t) = \sum_{\nu'} S_{\nu,\nu'} * b_{\nu'}^{(2)}(t)$. With matrix notation,

$$\boldsymbol{b}^{(1)}(t) = \mathbf{S} * \boldsymbol{b}^{(2)}(t), \tag{4.1}$$

where $\boldsymbol{b}^{(1)} = [b_1^{(1)}\ b_2^{(1)} \ldots]^{\mathrm{T}}$ and $\boldsymbol{b}^{(2)}$ is defined analogously. The order of the multi-index is specified in Appendix A.1. In the frequency domain, (4.1) reads

$$\tilde{\boldsymbol{b}}^{(1)}(k) = \tilde{\mathbf{S}}(k)\tilde{\boldsymbol{b}}^{(2)}(k), \tag{4.2}$$

where $\tilde{\mathbf{S}}(k)$ is the infinite dimensional scattering matrix.

## 4.1 Implications of passivity on $\tilde{\mathbf{S}}$

It is now shown that the elements of $\mathbf{S}(t-2a/\mathrm{c})$ are the impulse responses of passive systems in case the scatterer is passive; in this case the total radiative power that has passed through a sphere of radius $r \geqslant a$ before the time $T$ must be non-negative. This means

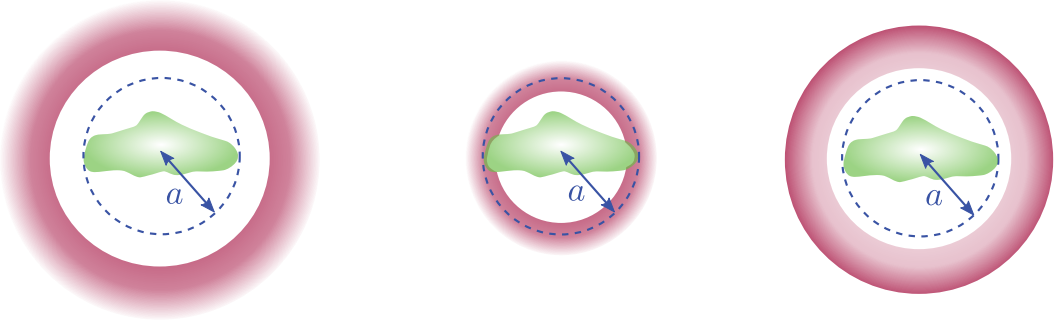
$$\int_{-\infty}^{T} P_{\mathrm{rad}}(r,t)\,\mathrm{d}t = \int_{-\infty}^{T} \sum_{\nu} \left( |b_\nu^{(2)}(t+r/\mathrm{c})|^2 - |b_\nu^{(1)}(t-r/\mathrm{c})|^2 \right) \mathrm{d}t \geqslant 0,$$

for all $T \in \mathbb{R}$ and $r \geqslant a$, where (3.5) has been used. Recall that it is only necessary to consider smooth, compactly supported incoming wave amplitudes $b_\nu^{(2)} \in \mathcal{D}$, as discussed in Section 2.2. Using (4.1) and letting the incoming field consist of only one vector spherical wave give

$$\int_{-\infty}^{T} P_{\mathrm{rad},\nu'}(r,t)\,\mathrm{d}t = \int_{-\infty}^{T} \left( b_{\nu'}^{(2)}(t+r/\mathrm{c}) - \sum_{\nu} |S_{\nu,\nu'} * b_{\nu'}^{(2)}(t-r/\mathrm{c})|^2 \right) \mathrm{d}t \geqslant 0,$$

for all $T \in \mathbb{R}$, $r \geqslant a$ and $\nu'$. Note that the above energy expression closely resembles that in (2.5), except for the time shifts $-2r/\mathrm{c}$ in the outgoing waves. Hence $S_{\nu,\nu'}(t-2a/\mathrm{c})$ is the impulse response of a passive system for all $\nu, \nu'$, and so its Fourier-transform $\mathrm{e}^{\mathrm{i}2ka}\tilde{S}_{\nu,\nu'}(k)$ is holomorphic and bounded in magnitude by one for $k \in \mathbb{C}^+$, see Section 2.2 and [2, 27, 31]. Furthermore, $\mathrm{e}^{\mathrm{i}2ka}\tilde{S}_{\nu,\nu'}(k)$ satisfies the symmetry (2.6).

The time shift $-2a/\mathrm{c}$ can be understood intuitively in the sense that the outgoing wave can appear at $r = a$ as soon as the incoming wavefront has reached $r = a$, see Figure 3. This is discussed from a somewhat different perspective in [17].




(a) (b) (c)

**Figure 3**: An incoming spherical wave $b^{(2)}_{\nu'}(t+r/\mathrm{c})$: a) impinges on the scatterer, b) interacts with the scatterer, and c) creates outgoing waves $\sum_\nu S_{\nu,\nu'} * b^{(2)}_\nu(t-r/\mathrm{c})$. Note that the picture is over-simplified, but it makes it believable that $S_{\nu,\nu'}(t-2a/\mathrm{c})$ is the impulse response of a passive system for all $\nu$ and $\nu'$.

## 4.2 Low-frequency asymptotic behaviour of $\tilde{\mathbf{S}}$

To derive equalities of the type (2.9), the low-frequency asymptotic expansion of the $\tilde{\mathbf{S}}$-matrix is required. For this reason, consider the alternative decomposition of the electric field in outgoing and regular vector spherical waves:

$$\widetilde{\boldsymbol{E}}(\boldsymbol{r},k) = k\sqrt{\eta_0}\sum_\nu \mathrm{i}^{l+2-\tau}\left[\tilde{d}^{(1)}_\nu(k)\boldsymbol{u}^{(1)}_\nu(k\boldsymbol{r}) + \tilde{d}^{(v)}_\nu(k)\boldsymbol{v}_\nu(k\boldsymbol{r})\right]. \tag{4.3}$$

Here $\boldsymbol{v}_\nu(k\boldsymbol{r})$ denotes regular vector spherical waves, defined as $\boldsymbol{v}_\nu(k\boldsymbol{r}) = (\boldsymbol{u}^{(1)}_\nu(k\boldsymbol{r}) + \boldsymbol{u}^{(2)}_\nu(k\boldsymbol{r}))/2$ (see Appendix A.1). The relation corresponding to (4.2) is $\tilde{\boldsymbol{d}}^{(1)}(k) = \tilde{\mathbf{T}}(k)\tilde{\boldsymbol{d}}^{(v)}(k)$, where $\tilde{\mathbf{T}}$ is the so called transition, or $T-$, matrix. Evidently, $\tilde{\mathbf{S}} = 2\tilde{\mathbf{T}} + \mathbf{I}$, where $\mathbf{I}$ is the infinite dimensional identity matrix.

The advantage of a decomposition in regular and outgoing waves is that a plane wave $\widetilde{\boldsymbol{E}}_\mathrm{i}$ impinging on the scatterer is regular everywhere, while the produced scattered field $\widetilde{\boldsymbol{E}}_\mathrm{s}$ has to satisfy the radiation condition. Accordingly, in this situation $\widetilde{\boldsymbol{E}}_\mathrm{i}$ equals the sum over $\boldsymbol{v}_\nu$, while $\widetilde{\boldsymbol{E}}_\mathrm{s}$ is the sum over $\boldsymbol{u}^{(1)}_\nu$. Consider a plane wave $\boldsymbol{E}_0(t - \boldsymbol{r}\cdot\hat{\boldsymbol{k}}/c)$ propagating in the $\hat{\boldsymbol{k}}$-direction, corresponding to $\mathrm{e}^{\mathrm{i}\boldsymbol{r}\cdot\boldsymbol{k}}\widetilde{\boldsymbol{E}}_0(k)$ in the frequency domain. Here $\boldsymbol{k} = k\hat{\boldsymbol{k}}$ and as usual $\widetilde{\boldsymbol{E}}_0(k) = (\mathcal{F}\boldsymbol{E}_0)(\omega)$ with $k = \omega/\mathrm{c}$.

The radiating part of the scattered field is described by the far-field amplitude $\boldsymbol{F}$, viz.

$$\boldsymbol{E}_\mathrm{s}(t,\boldsymbol{r}) = \frac{\boldsymbol{F}(t-r/\mathrm{c},\hat{\boldsymbol{r}})}{r} + \mathcal{O}(r^{-2}), \qquad \widetilde{\boldsymbol{E}}_\mathrm{s}(k,\boldsymbol{r}) = \frac{\mathrm{e}^{\mathrm{i}kr}\widetilde{\boldsymbol{F}}(k,\hat{\boldsymbol{r}})}{r} + \mathcal{O}(r^{-2}), \tag{4.4}$$

as $r\to\infty$. Due to the assumption of convolution form for the constitutive relations, a scattering dyadic $\tilde{\mathbf{S}}_\mathrm{d}$ may be defined:

$$\boldsymbol{F}(t,\hat{\boldsymbol{r}}) = \Big(\mathbf{S}_\mathrm{d}(t,\hat{\boldsymbol{r}},\hat{\boldsymbol{k}}) * \boldsymbol{E}_0(t)\Big)(t), \qquad \widetilde{\boldsymbol{F}}(k,\hat{\boldsymbol{r}}) = \tilde{\mathbf{S}}_\mathrm{d}(k,\hat{\boldsymbol{r}},\hat{\boldsymbol{k}})\cdot\widetilde{\boldsymbol{E}}_0(k). \tag{4.5}$$

The elements of the $T$-matrix can be deduced from the scattering dyadic $\tilde{\mathbf{S}}_{\mathrm{d}}$:

$$\tilde{T}_{\nu,\nu'}(k) = \frac{\mathrm{i}k}{4\pi} \int\int \boldsymbol{A}_{\nu}(\hat{\boldsymbol{r}}) \cdot \tilde{\mathbf{S}}_{\mathrm{d}}(k; \hat{\boldsymbol{r}}, \hat{\boldsymbol{k}}) \cdot \boldsymbol{A}_{\nu'}(\hat{\boldsymbol{k}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}} \, \mathrm{d}\Omega_{\hat{\boldsymbol{k}}}. \tag{4.6}$$

See Appendix A.2 for details.

Assume that the medium of the scatterer is anisotropic in the static limit ($k = 0$), so that the constitutive relations are

$$\begin{aligned} \widetilde{\boldsymbol{D}}(0, \boldsymbol{r}) &= \epsilon_0 \boldsymbol{\epsilon}(0, \boldsymbol{r}) \cdot \widetilde{\boldsymbol{E}}(0, \boldsymbol{r}) \\ \widetilde{\boldsymbol{B}}(0, \boldsymbol{r}) &= \mu_0 \boldsymbol{\mu}(0, \boldsymbol{r}) \cdot \widetilde{\boldsymbol{H}}(0, \boldsymbol{r}). \end{aligned}$$

Here $\widetilde{\boldsymbol{D}}(k, \boldsymbol{r})$ denotes the electric flux density and $\widetilde{\boldsymbol{B}}(k, \boldsymbol{r})$ the magnetic flux density at the point $\boldsymbol{r}$ and wavenumber $k$. The relative permittivity and permeability dyadics are denoted $\boldsymbol{\epsilon}(k, \boldsymbol{r})$ and $\boldsymbol{\mu}(k, \boldsymbol{r})$, respectively, and $\epsilon_0$ and $\mu_0$ are the permittivity and permeability of free space, respectively. The low frequency expansion of $\tilde{\mathbf{S}}_{\mathrm{d}}(k, \hat{\boldsymbol{r}}, \hat{\boldsymbol{k}})$ is then [15, p. 18]

$$\tilde{\mathbf{S}}_{\mathrm{d}}(k, \hat{\boldsymbol{r}}, \hat{\boldsymbol{k}}) \cdot \boldsymbol{E} = \frac{k^2}{4\pi} \Big\{ \hat{\boldsymbol{r}} \times \Big[ (\boldsymbol{\gamma}_{\mathrm{e}} \cdot \boldsymbol{E}) \times \hat{\boldsymbol{r}} \Big] + \Big[ \boldsymbol{\gamma}_{\mathrm{m}} \cdot (\hat{\boldsymbol{k}} \times \boldsymbol{E}) \Big] \times \hat{\boldsymbol{r}} \Big\} + \mathcal{O}(k^3), \tag{4.7}$$

as $k \to 0$, where $\boldsymbol{E}$ is a constant vector. The electric polarizability dyadic $\boldsymbol{\gamma}_{\mathrm{e}}$ relates the electric dipole moment induced in the scatterer to an applied static homogeneous electric field $\widetilde{\boldsymbol{E}}(0)$, namely $\boldsymbol{p} = \epsilon_0 \boldsymbol{\gamma}_{\mathrm{e}} \cdot \widetilde{\boldsymbol{E}}(0)$. Similarly, the magnetic dipole moment induced by an applied static homogeneous magnetic field $\widetilde{\boldsymbol{H}}(0)$ is given by $\boldsymbol{m} = \boldsymbol{\gamma}_{\mathrm{e}} \cdot \widetilde{\boldsymbol{H}}(0)$. The polarizability dyadics are thoroughly discussed in [15] and [24]. Now let $\boldsymbol{E} = \boldsymbol{A}_{\nu'}(\hat{\boldsymbol{k}})$. From (4.6) and (4.7) it follows that

$$\tilde{S}_{\nu,\nu'}(k) = \delta_{\nu,\nu'} + \mathrm{i}2\rho_{\nu,\nu'} k^3 a^3 + \mathcal{O}(k^4), \quad \text{as } k \to 0, \tag{4.8}$$

where

$$\rho_{\nu,\nu'} = \frac{1}{16\pi^2 a^3} \iint \Big[ \boldsymbol{A}_{\nu}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{\gamma}_{\mathrm{e}} \cdot \boldsymbol{A}_{\nu'}(\hat{\boldsymbol{k}}) + (-1)^{\tau+\tau'} \boldsymbol{A}_{\bar{\nu}}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{\gamma}_{\mathrm{m}} \cdot \boldsymbol{A}_{\bar{\nu}'}(\hat{\boldsymbol{k}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}} \Big] \, \mathrm{d}\Omega_{\hat{\boldsymbol{k}}}.$$

Here (A.2) was used, and recall that the dual multi-index is $\bar{\nu} = \{\bar{\tau}, s, m, l\}$ with $\bar{\tau} = 3 - \tau$.

Explicit expressions for $\rho_{\nu,\nu'}$ are derived by using the identities (A.5) for the Cartesian unit vectors $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$:

$$\rho_{\nu,\nu'} = \frac{1}{6\pi a^3} \delta_{l,1} \delta_{l',1} \hat{\boldsymbol{n}}_{sm} \cdot (\delta_{\tau,1} \delta_{\tau',1} \boldsymbol{\gamma}_{\mathrm{m}} + \delta_{\tau,2} \delta_{\tau',2} \boldsymbol{\gamma}_{\mathrm{e}}) \cdot \hat{\boldsymbol{n}}_{s'm'}, \tag{4.9}$$

where

$$\hat{\boldsymbol{n}}_{sm} = \begin{cases} \hat{\boldsymbol{x}}, & \text{for s=2, m=1} \\ \hat{\boldsymbol{y}}, & \text{for s=1, m=1} \\ \hat{\boldsymbol{z}}, & \text{for s=2, m=0.} \end{cases}$$

Note that $\rho_{\nu,\nu'} = 0$ for non-dipole modes ($l \geqslant 2$ or $l' \geqslant 2$), and that $\rho_{\nu,\nu'} = 0$ for $\tau = 1$ ($\tau = 2$) when the scatterer is non-magnetic (non-electric).

## 4.3 The polarizability dyadics and bounds on $\rho_{\nu,\nu}$

It is clear now that the polarizability dyadics are of vital importance. Until now, the only assumptions made on the constitutive relations of the scatterer is that they are in convolution form in the time domain and passive, and furthermore anisotropic in the static limit. If the scatterer is heterogeneous, these assumptions are made for all points $\boldsymbol{r}$ within the scatterer. It is common to assume that the permittivity and permeability dyadics are symmetric in the static limit, i.e. $\boldsymbol{\epsilon}(0,\boldsymbol{r}) = \boldsymbol{\epsilon}(0,\boldsymbol{r})^{\mathrm{T}}$ and $\boldsymbol{\mu}(0,\boldsymbol{r}) = \boldsymbol{\mu}(0,\boldsymbol{r})^{\mathrm{T}}$. This implies that the polarizability dyadics are also symmetric [24], and hence diagonal for a suitable choice of coordinates. Closed form expressions for the polarizability dyadics exists for anisotropic homogeneous spheroidal scatterers, see [24] and references therein. For the simple case of an isotropic sphere of radius $a$, they are

$$\boldsymbol{\gamma}_{\mathrm{e}} = 4\pi a^3 \frac{\epsilon(0)-1}{\epsilon(0)+2}\,\mathbf{I} \quad \text{and} \qquad \boldsymbol{\gamma}_{\mathrm{m}} = 4\pi a^3 \frac{\mu(0)-1}{\mu(0)+2}\,\mathbf{I},$$

where $\mathbf{I}$ is the identity dyadic.

Furthermore, under the assumption of symmetry it can be shown that $\boldsymbol{\gamma}_{\mathrm{e}}$ and $\boldsymbol{\gamma}_{\mathrm{m}}$ are non-decreasing as functions of $\boldsymbol{\epsilon}(0,\boldsymbol{r})$ and $\boldsymbol{\mu}(0,\boldsymbol{r})$ [23]. More specifically, consider two objects with permittivity $\boldsymbol{\epsilon}(0,\boldsymbol{r})$ and $\boldsymbol{\epsilon}'(0,\boldsymbol{r})$, respectively. If $\boldsymbol{\epsilon}'(0,\boldsymbol{r}) - \boldsymbol{\epsilon}(0,\boldsymbol{r})$ is a positive semidefinite dyadic for all $\boldsymbol{r}$ in the object, then $\boldsymbol{\gamma}'_{\mathrm{e}} - \boldsymbol{\gamma}_{\mathrm{e}}$ is positive semidefinite as well. The same holds for $\boldsymbol{\gamma}_{\mathrm{m}}$, with $\boldsymbol{\epsilon}(0,\boldsymbol{r})$ replaced by $\boldsymbol{\mu}(0,\boldsymbol{r})$. The diagonal elements of $\boldsymbol{\gamma}_{\mathrm{e}}$ and $\boldsymbol{\gamma}_{\mathrm{m}}$ for any scatterer (satisfying the aforementioned assumptions) contained in the sphere of radius $a$ are therefore bounded by $4\pi a^3$ for the high contrast sphere. Following (4.9), the parameters $\rho_{\nu,\nu}$ are non-decreasing as functions of $\boldsymbol{\epsilon}(0,\boldsymbol{r})$ and $\boldsymbol{\mu}(0,\boldsymbol{r})$, and thus bounded from above by $\rho_{\nu,\nu} = 2/3$.

If the scatterer is contained within a non-spherical geometry, the diagonal elements of $\boldsymbol{\gamma}_{\mathrm{e}}$ and $\boldsymbol{\gamma}_{\mathrm{m}}$ are bounded by the largest eigenvalue $\gamma_1 \leqslant 4\pi a^3$ of the high-contrast polarizability dyadic $\boldsymbol{\gamma}_\infty$ of that geometry. Therefore a sharper bound on $\rho_{\nu,\nu}$, given by $\rho_{\nu,\nu} \leqslant \gamma_1/(6\pi a^3) \leqslant 2/3$, can be determined. The high-contrast polarizability dyadics $\boldsymbol{\gamma}_\infty$ of many geometries can be calculated numerically, see [22] for some examples.

A widely used material model is the perfect electric conductor (PEC). The electric field vanish inside a perfect electric conductor, which implies $\epsilon(0) = \infty$. In this paper, a PEC body is defined as one where also magnetic fields vanish (in other words, the Meissner effect is perfect), in accordance with e.g. [15, 23]. This means that $\mu(0) = 0$ [15, pp. 39–40]. Consequently, for a scatterer with a PEC inclusion, $\boldsymbol{\gamma}_{\mathrm{e}}$ ($\boldsymbol{\gamma}_{\mathrm{m}}$) is non-decreasing (non-increasing) as the volume of the PEC inclusion increases [23].

## 4.4 Sum rules and physical limitations on $\tilde{\mathbf{S}}$

Now it has been shown that $\mathrm{e}^{\mathrm{i}2ka}\tilde{S}_{\nu,\nu'}(k)$ is a holomorphic function bounded in magnitude by one in $\mathbb{C}^+$ for all $\nu$ and $\nu'$, due to the passivity assumption. Furthermore, its low frequency asymptotic expansion has been determined in (4.8) and (4.9). It

remains to define a Herglotz function and derive sum rules of the type (2.9). The Herglotz function corresponding to (2.7) is

$$h_{\nu,\nu'}(k) = -\mathrm{i}\log\left(\frac{\mathrm{e}^{\mathrm{i}2ka}\tilde{S}_{\nu,\nu'}(k)}{B_{\nu,\nu'}(k)}\right).$$

Here $B_{\nu,\nu'}$ is a Blaschke products of the form (2.8) for each pair $(\nu,\nu')$. Since $\mathrm{e}^{\mathrm{i}2ka}\tilde{S}_{\nu,\nu}(k) \to 1$ as $k \to 0$ when $\tilde{S}_{\nu,\nu}$ is a diagonal element of the scattering matrix, the low-frequency expansion may be calculated separately for that factor and the Blaschke product (cf. [2]):

$$h_{\nu,\nu}(k) = 2ka + 2\rho_{\nu,\nu}k^3a^3 + \mathcal{O}(k^4) + 2\sum_n \sum_{q=1,3,\ldots}^{\infty} \frac{k^q}{q}\operatorname{Im}\frac{1}{k_n^q}, \quad \text{as } k \to 0. \tag{4.10}$$

This is not necessarily possible for the off diagonal terms $h_{\nu,\nu'}$, where $\nu \neq \nu'$, since then $\tilde{S}_{\nu,\nu'}(k)$ tends to zero as $k \to 0$. Only terms with odd $q$ appear in the sum in (4.10) due to the symmetry (2.6).

Note that the low-frequency asymptotic expansions (4.7) and (4.10) are valid as $k \to 0$ for all arguments of $k$, and especially as $k\hat{\to}0$. With the notation of Section 2.1, $N = 2$ and hence two sum rules of the type (2.9) (using $p = 1, 2$) can be deduced:

$$\frac{1}{\pi}\int_0^\infty \frac{1}{k^2}\ln\frac{1}{|\tilde{S}_{\nu,\nu}(k)|}\,\mathrm{d}k = a - \frac{\beta_{\nu,\nu}}{2} + \sum_n \operatorname{Im}\frac{1}{k_n} \tag{4.11}$$

and

$$\frac{1}{\pi}\int_0^\infty \frac{1}{k^4}\ln\frac{1}{|\tilde{S}_{\nu,\nu}(k)|}\,\mathrm{d}k = a^3\rho_{\nu,\nu} + \frac{1}{3}\sum_n \operatorname{Im}\frac{1}{k_n^3}. \tag{4.12}$$

Both sum rules incorporate the radius $a$ of the circumscribing sphere, and the second depends on the material and shape of the scatterer via $\rho_{\nu,\nu}$ given by (4.9). The parameter $\beta_{\nu,\nu} = \lim_{k\hat{\to}\infty} h_{\nu,\nu}(k)/k$ is greater than or equal to zero. Evidently, $\beta_{\nu,\nu} > 0$ applies if the chosen circumscribing sphere is larger than the smallest circumscribing sphere, but it is expected that $\beta_{\nu,\nu} = 0$ if $a$ is chosen as small as possible. This is true for isotropic spherical scatterers with material described by e.g. the Debye or Lorentz models. It is hard to prove this statement for an arbitrary scatterer, so it is assumed that $\beta_{\nu,\nu}$ can be larger than zero.

In order to derive physical limitations, consider a finite wavenumber interval, $\mathcal{K} = [k_0(1 - B_K/2), k_0(1 + B_K/2)]$, with center wavenumber $k_0$ and fractional bandwidth $B_K < 2$. Letting $S_{0,\nu} = \sup_{k\in\mathcal{K}} |\tilde{S}_{\nu,\nu}(k)|$, it follows that

$$\frac{B_K \ln S_{0,\nu}^{-1}}{\pi} \leqslant k_0 a + \sum_n \operatorname{Im}\frac{k_0}{k_n} \tag{4.13}$$

and

$$\frac{B_K \ln S_{0,\nu}^{-1}}{\pi} \leqslant k_0^3 a^3 \rho_{\nu,\nu} + \frac{1}{3}\sum_n \operatorname{Im}\frac{k_0^3}{k_n^3}. \tag{4.14}$$

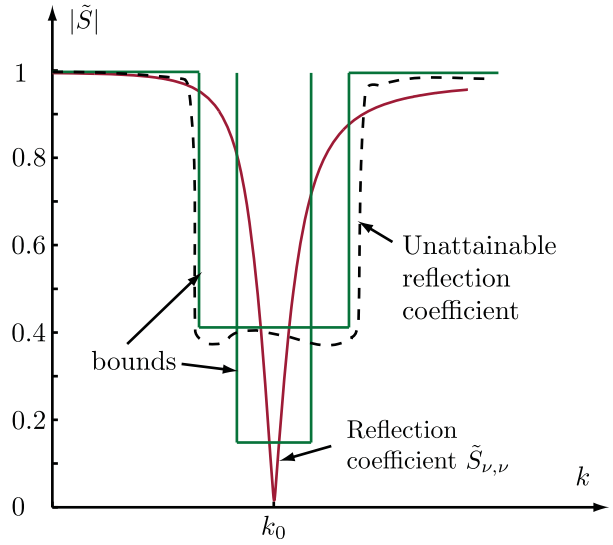



**Figure 4**: Interpretation of the limitation (4.16). In the figure, bounds for a given center wavenumber $k_0$ are depicted for two different values of $S_{0,\nu}$ (and thus two different values of $B_K$). The limitation states that the magnitude of all reflection coefficients $\tilde{S}_{\nu,\nu}$ have to intersect the boxes when the scatterer satisfies the aforementioned assumptions; also shown in the figure is one attainable reflection coefficient $\tilde{S}_{\nu,\nu}$ and one unattainable reflection coefficient.

Here it has been used that $k_0^{2p-1}\int_{\mathcal{K}} 1/k^{2p}\,\mathrm{d}k \geqslant B_K$ for $p = 0, 1, \ldots$ Note also that $k_0^{2p-1}\int_{\mathcal{K}} 1/k^{2p}\,\mathrm{d}k \approx B_K$ when $B_K \ll 1$.

The sum in the right-hand side of (4.13) is non-positive (since $\operatorname{Im} k_n > 0$ for all $k_n$), and so

$$\frac{B_K \ln S_{0,\nu}^{-1}}{\pi} \leqslant k_0 a. \tag{4.15}$$

An alternative bound not containing the sum over all zeros can also be derived (see Appendix A.3):

$$\begin{aligned}\frac{B_K \ln S_{0,\nu}^{-1}}{\pi} &\leqslant f_\nu(k_0 a) = k_0 a - \sqrt[3]{\iota+\xi} + \sqrt[3]{\iota-\xi} \\ &= \left(\frac{1}{3} + \rho_{\nu,\nu}\right)\left(k_0^3 a^3 - k_0^5 a^5\right) + \mathcal{O}(k_0^7), \qquad \text{as } k_0 \to 0.\end{aligned} \tag{4.16}$$

Here the material and geometry of the scatterer are contained in $\rho_{\nu,\nu}$ via $\xi = 3k_0 a(1-\rho_{\nu,\nu}k_0^2 a^2)/2$ and $\iota = \sqrt{1+\xi^2}$. The term $k_0^3 a^3/3$ in the bound stems from the circumscribing sphere. The limitation (4.16) states that, somewhere on the wavenumber interval $\mathcal{K}$, the reflection coefficient $\tilde{S}_{\nu,\nu}(k)$ for mode $\nu$ must be larger in magnitude than some value prescribed by the fractional bandwidth $B_K$, the radius of the smallest circumscribing sphere $a$, and the material properties of the scatterer via $\rho_{\nu,\nu}$, see Figure 4. Note that the bound in (4.16) is sharper than that in (4.15) for $k_0 a \leqslant \rho_\nu^{-1/2}$, and vice versa.

An interpretation of the limitation is as a bound on the absorption of a vector spherical wave over a bandwidth. To see this, consider the energy $e(\infty)$ absorbed by the scatterer for all times $t \in \mathbb{R}$ when the incoming field consists of only the mode $\nu$, and $b_\nu^{(2)}(t)$ is assumed to be in $\mathrm{L}^2$. By (3.4)–(3.6), it is (with $r \geqslant a$)

$$e(\infty) = \int_{-\infty}^{\infty} P_\nu(r,t)\,\mathrm{d}t = \int_{-\infty}^{\infty} P_{\mathrm{rad},\nu}(r,t)\,\mathrm{d}t$$
$$= \int_{-\infty}^{\infty} \left( |b_\nu^{(2)}(t)|^2 - \sum_{\nu'} |S_{\nu',\nu} * b_\nu^{(2)}(t)|^2 \right) \mathrm{d}t.$$

The expression for $e(\infty)$ may be rewritten with Parseval's equation:

$$e(\infty) = \frac{1}{2\pi \mathrm{c}} \int_{-\infty}^{\infty} |\tilde{b}_\nu^{(2)}(k)|^2 \left( 1 - \sum_{\nu'} |\tilde{S}_{\nu',\nu}(k)|^2 \right) \mathrm{d}k.$$

Hence the absorption efficiency

$$\chi_\nu(k) = 1 - \sum_{\nu'} |\tilde{S}_{\nu',\nu}(k)|^2 \tag{4.17}$$

is the normalized energy of the incoming mode $\nu$ that is absorbed by the scatterer at wavenumber $k$; all of the incoming energy is absorbed if $\chi_\nu(k) = 1$, while no energy is absorbed in the case $\chi_\nu(k) = 0$.

From the limitation (4.16) it follows that

$$\inf_{k \in \mathcal{K}} \chi_\nu(k) \leqslant 1 - \mathrm{e}^{-2\pi f_\nu(k_0 a)/B_K}, \tag{4.18}$$

where as before the wavenumber interval $\mathcal{K} = [k_0(1 - B_K/2), k_0(1 + B_K/2)]$ has center wavenumber $k_0$ and fractional bandwidth $B_K$ and $f_\nu(k_0 a)$ is given in (4.16). The limitation on the absorption efficiency $\chi_\nu$ states that it cannot be arbitrarily high over a whole wavelength interval.

It should be stressed that the scatterer was assumed to be placed in free space (cf. Figure 2). Since the reflection coefficients $\tilde{S}_\nu$ and absorption efficiencies $\chi_{\nu,\nu}$ are functions of the scatterers properties only, and do not depend on the surroundings, the bounds on the reflection coefficients apply also if there are objects nearby. But, if parts of the scattered field is reflected back towards the scatterer from surrounding objects, more power can be absorbed. The explanation is that in this case the coefficients of the incoming waves $\tilde{b}_\nu^{(2)}(k)$ also incorporate these multiple scattering effects. To find limitations that also take the surroundings into account is an intriguing problem, but out of scope of this paper.

## 4.5 Algorithm to make use of the limitation

This section describes the algorithm to make use of the limitations given by (4.16) and (4.18) to find bounds on the reflection coefficients $\tilde{S}_{\nu,\nu}$ and absorption efficiencies $\chi_\nu$ for a scatterer contained in a given sphere of radius $a$. This procedure is

used in Example 5.1. The necessary assumptions on the scatterer are that it is heterogeneous, passive, and has constitutive relations which are in convolution form in the time domain and anisotropic in the static limit.

Firstly, an upper bound on the material parameter $\rho_{\nu,\nu}$ (given by (4.9)) appearing in the right-hand sides of (4.16) and (4.18) must be found. Recall that $\rho_{\nu,\nu} = 0$ applies if the multi-index $\nu$ corresponds to a non-dipole mode ($l \geqslant 2$). Furthermore, $\rho_{\nu,\nu}$ vanishes also if the scatterer is non-magnetic (non-electric) and the multi-index $\nu$ is that of a magnetic (electric) dipole. For a general scatterer, the material parameter $\rho_{\nu,\nu}$ is bounded from above by $2/3$ for the dipole modes ($l = 1$). If the scatterer is contained in a non-spherical geometry, a sharper bound on $\rho_{\nu,\nu}$ can be determined numerically, as discussed in Section 4.3.

Secondly, the wavenumber interval $\mathcal{K} = [k_0(1 - B_K/2), k_0(1 + B_K/2)]$ of interest is chosen. Inserting the center wavenumber $k_0$, fractional bandwidth $B_K$, bound on material parameter $\rho_{\nu,\nu}$ and radius $a$ of the circumscribing sphere into (4.16) gives a lower bound on $S_{0,\nu} = \sup_{k\in\mathcal{K}} |S_{\nu,\nu}(k)|$, see Figure 4. Alternatively, the largest acceptable value of $S_{0,\nu}$ is chosen, and an upper bound on the achievable fractional bandwidth $B_K$ follows from (4.16). Equivalently, (4.18) can be used to find a bound on the absorption efficiency $\chi_\nu$, defined in (4.17).

# 5 Examples

## 5.1 Nanoshells

A nanoshell is a dielectric core covered by a thin coat of metal. By varying the core radius, shell thickness, and materials, they can be constructed to scatter or absorb large parts of incoming electromagnetic waves in the visible light and near-infrared (NIR) spectra. Applications include e.g. biomedical imaging and treatment of tumours.

In cancer treatment, the nanoshells are shuttled into the tumour using a so called "Trojan horse"-method [5]. Hereafter they are illuminated by laser light, causing most of the cancer cells to die, see Figure 1 in [5]. It is thus desirable to design nanoshells that absorb large parts of the laser energy. In [5, 18], the nanoshells are spherical cores of silicon dioxide ($SiO_2$) covered with gold. The radius of the core is typically around 60 nm, and the gold shell is $5 - 20$ nm thick.

The bound in (4.18) is well suited to study this problem. To have something to compare the bound to, consider two nanoshells of outer radius $a = 75$ nm, consisting of spherical silicon dioxide cores covered by layers of gold of thickness $d = 5$ nm and $d = 10$ nm, respectively. The absorption efficiencies are approximated as $\chi_\nu = 1 - |\tilde{S}_{\nu,\nu}|^2$, which can be calculated relatively easily since the nanoshells are layered spheres. They are calculated for the electric dipole modes ($\tau = 2, l = 1$), to be compared with the bound in (4.18). The nanoshells are resonant at wavelengths $\lambda_0 = 986$ nm and $\lambda_0 = 765$ nm, respectively. These resonant wavenumbers $k_0 = 2\pi/\lambda_0$ are chosen as center wavenumbers in (4.18). The upper bound on the material parameter $\rho_{\nu,\nu} \leqslant 2/3$ is inserted, along with a proposed desired fractional bandwidth $B_K = 0.4$. The resulting bounds on the absorption efficiency is shown in Figure 5. The

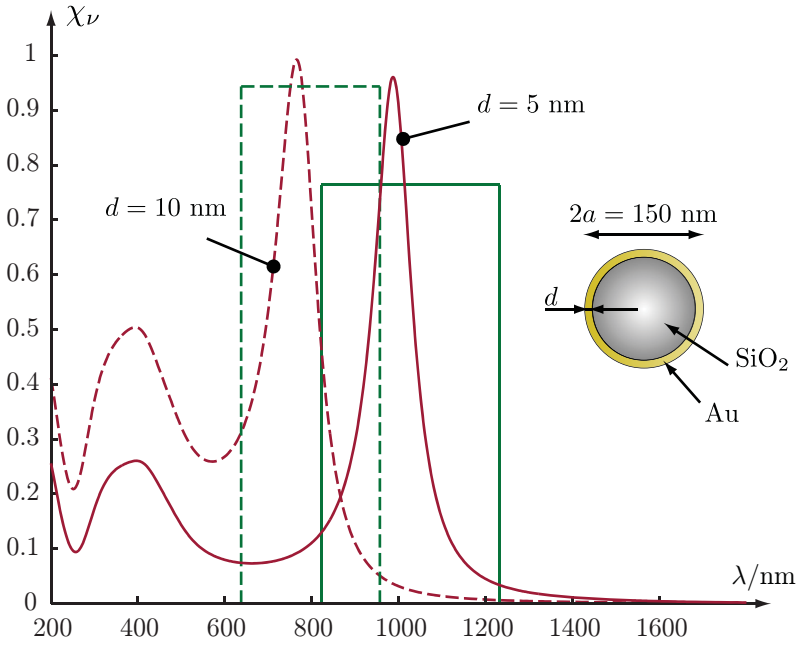



**Figure 5**: The absorption efficiencies $\chi_\nu$ (approximated as $1 - |\tilde{S}_{\nu,\nu}|^2$) of the electric dipole modes ($\tau = 2, l = 1$) for nanoshells of outer radius $a = 75\,\text{nm}$, consisting of spherical silicon dioxide cores covered by layers of gold of thickness $d = 5\,\text{nm}$ and $d = 10\,\text{nm}$, respectively. Here $\lambda = 2\pi/k$ denotes the wavelength. The bound is (4.18) with $\rho_{\nu,\nu} = 2/3$, $B_K = 0.4$ and center wavelengths $\lambda_0 = 986\,\text{nm}$ and $\lambda_0 = 765\,\text{nm}$, respectively. The limitation states that somewhere on the wavelength intervals, the absorption efficiencies must be lower than 76% and 94%, respectively. In other words, the curves have to intersect the boxes. The absorption efficiencies $\chi_\nu$ was calculated from the closed form expression, using a Matlab-script for a Lorentz-Drude model for gold by Ung et al. [25]. Silicon dioxide is modelled as having negligible losses and a refractive index of $n = 1.5$, which is a good model at least for wavelengths 400–1100 nm [14].

limitation states that somewhere on the wavenumber interval $[k_0(1 - B_K/2), k_0(1 + B_K/2)]$, the absorption efficiency must be lower than 76% and 94%, respectively. This applies to *all* nanoshells of given outer radius $a = 75\,\text{nm}$, as well as for all spherical and non-spherical objects contained in the same sphere.

## 5.2 Physical limitations on antennas

As discussed in Section 4.4, (4.16) places a bound on the absorption of a spherical wave over a bandwidth, which makes it a good candidate to find limits on the performance of antennas. Furthermore, the communications channels of multiple-input multiple-output (MIMO) antenna systems are coupled to orthogonal sets of vector spherical waves [8]. It is unusual to compute the $\tilde{\mathbf{S}}$-matrix elements of an antenna. Instead, consider the setup depicted in Figure 6. The antenna is fed the power $P_{\text{in}}(k)$ by a transmission line, and a matching network is employed in order

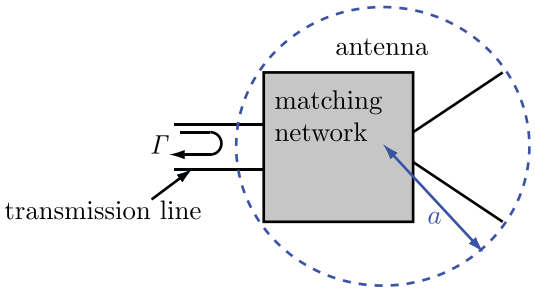



**Figure 6**: The antenna and matching network considered in Example 5.2.

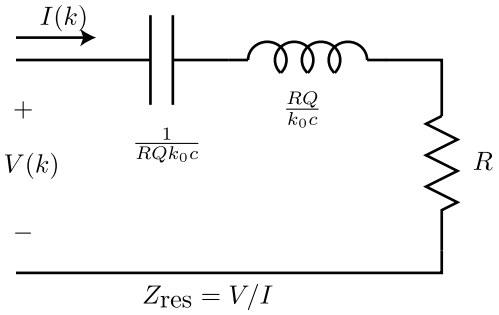



**Figure 7**: For many antennas, the impedance $Z_{\text{res}}(k)$ of the resonance circuit is a good approximation for the antenna input impedance $Z(k)$ close to its resonance wavenumber $k_0$ [10]. The quality factor $Q$ is given by (5.1).

to minimize the reflection coefficient $\Gamma(k)$. The power rejected due to mismatch is $|\Gamma(k)|^2 P_{\text{in}}(k)$, and obviously the radiated power is bounded as

$$P_{\text{rad}}(k) \leqslant (1 - |\Gamma(k)|^2) P_{\text{in}}(k),$$

with equality if there are no ohmic losses in the antenna.

Many antennas can be modelled by the resonance circuit in Figure 7 in a frequency interval close to their respective resonance frequencies [10]. Here the quality factor is

$$Q = k_0 \mathrm{c} \frac{Z'(k_0)}{2R}, \tag{5.1}$$

where $k_0$ is the resonance wavenumber of the antenna, $Z$ its input impedance, and $R = Z(k_0)$ the real-valued input impedance at the resonance. A prime denotes differentiation with respect to the argument. Using Fano's bounds on optimal matching [7], it is straightforward to show that [10]

$$\frac{B_K \ln \Gamma_0^{-1}}{\pi} \leqslant \frac{1}{Q}, \tag{5.2}$$

applies whatever the matching network is. Here $\Gamma_0 = \max_{k \in \mathcal{K}} |\Gamma(k)|$. The wavenumber interval is $\mathcal{K} = [k_0(1 - B_K/2), k_0(1 + B_K/2)]$, with center wavenumber $k_0$ and fractional bandwidth $B_K$.

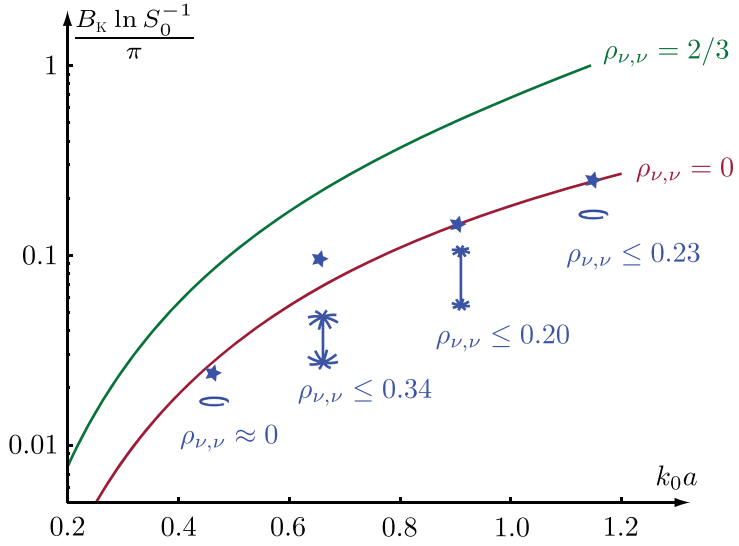



**Figure 8**: The lines are the bound in (4.16) for $\rho_{\nu,\nu} = 2/3$ and $\rho_{\nu,\nu} = 0$, respectively. Four wire antennas were used in the example, with wires modelled as perfect electric conductors of diameter 2 mm. The radii of the loops are 60 mm (giving $a = 61$ mm), and the heights of the umbrellas are 100 mm (so that $a = 51$ mm). Here $a$ is the radius of the smallest circumscribing sphere. The loop with resonance at $k_0 a \approx 0.46$ is in series with a 100 F capacitance, causing it to radiate much like a pure magnetic dipole close to the resonance. The input impedances and resonance wavenumbers for the antennas were calculated using the commercial software E-Field (http://www.efieldsolutions.com). The inverse of $Q$ given by (5.1) is depicted for the four antennas at their respective resonance wavenumbers $k_0$. The electric polarizability dyadics $\boldsymbol{\gamma}_{\mathrm{e}}$ were calculated using a Method of Moments code, and from them the bounds on $\rho_{\nu,\nu}$ shown in the figure could be determined.

The input impedance $Z(k)$ of an antenna, and hence also the quality factor $Q$ in (5.1), may be calculated numerically. Equation (5.2) provides a means to compare the bound in (4.16) to the quality factor of an antenna; since $1-|\Gamma|^2$ places a bound on the radiated power in terms of the input power, and $1-|\tilde{S}_{\nu,\nu}|^2$ limits the absorbed power from a single mode $\nu$, $\Gamma$ and $\tilde{S}_{\nu,\nu}$ are on equal footing. In Figure 8, the bound in (4.16) is compared to the inverse of the numerically determined quality factor $Q$ of four wire antennas.

# 6 Conclusions

Electromagnetic waves may be scattered and/or absorbed when they interact with various objects. Understanding this interaction is vital in many applications, from classical optics to antenna theory. One way to analyse it is to apply physical limitations to it; in essence, the physical limitations state what can and cannot be expected from a certain physical system.

There are several publications addressing physical limitations in scattering and

antenna theory, see e.g. [3, 6, 9, 19, 24, 26]. However, the present paper seems to be the first to derive physical limitations on the scattering of electromagnetic vector spherical waves. The vector spherical waves constitute a means to expand a given electromagnetic wave in orthogonal waves, and are commonly used [4, 11, 20, 21]. In wireless communication, they are intimately linked to the orthogonal communication channels of multiple-input multiple-output (MIMO) systems [8].

The derivation makes use of a general approach to obtain sum rules and physical limitations on passive physical systems in convolution form presented in [2]. The limitations in this paper are valid for all heterogeneous passive scatterers with constitutive relations in convolution form in the time domain, and anisotropic in the static limit. They state that the reflection coefficients cannot be arbitrarily small over a whole wavenumber interval; how small is determined by the center wavenumber and fractional bandwidth, the radius of the smallest sphere circumscribing the scatterer, and its static material properties.

The limitations can be interpreted as bounds on the absorption of power from the respective modes. They are particularly useful for the electrically small scatterers, and can therefore be employed to analyse sub-wavelength structures designed to be resonant in one or more frequency bands. Two examples are nanoshells and antennas, discussed in the examples in this paper.

# Acknowledgments

The financial support by the High Speed Wireless Communications Center of the Swedish Foundation for Strategic Research (SSF) is gratefully acknowledged.

# Appendix

## A.1 Definition of vector spherical waves

The incoming ($j = 2$) and outgoing ($j = 1$) vector spherical waves are defined as in [4] by Boström et al:

$$\begin{cases} \boldsymbol{u}^{(j)}_{1sml}(k\boldsymbol{r}) = \mathrm{h}^{(j)}_l(kr)\boldsymbol{A}_{1sml}(\hat{\boldsymbol{r}}) \\ \boldsymbol{u}^{(j)}_{2sml}(k\boldsymbol{r}) = \dfrac{(kr\mathrm{h}^{(j)}_l(kr))'}{kr}\boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) + \sqrt{l(l+1)}\dfrac{\mathrm{h}^{(j)}_l(kr)}{kr}\boldsymbol{A}_{3sml}(\hat{\boldsymbol{r}}). \end{cases} \tag{A.1}$$

Here $\mathrm{h}^{(j)}_l$ denotes the spherical Hankel function of the $j$:th kind and order $l$, and a prime denotes differentiation with respect to the argument $kr$. The regular vector spherical waves $\boldsymbol{v}_\nu$ are almost identical; for them the spherical Hankel functions have been replaced by spherical Bessel functions $\mathrm{j}_l = (\mathrm{h}^{(1)}_l + \mathrm{h}^{(2)}_l)/2$. The vector spherical

harmonics $\boldsymbol{A}_{\tau sml}$ are defined by

$$\begin{cases} \boldsymbol{A}_{1sml}(\hat{\boldsymbol{r}}) = \dfrac{1}{\sqrt{l(l+1)}} \nabla \times (\boldsymbol{r} Y_{sml}(\hat{\boldsymbol{r}})) \\ \boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) = \dfrac{1}{\sqrt{l(l+1)}} r \nabla Y_{sml}(\hat{\boldsymbol{r}}) \\ \boldsymbol{A}_{3sml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}} Y_{sml}(\hat{\boldsymbol{r}}). \end{cases}$$

Here $Y_{sml}$ are the (scalar) spherical harmonics

$$Y_{sml}(\theta, \phi) = \sqrt{\frac{2 - \delta_{m0}}{2\pi}} \sqrt{\frac{2l+1}{2} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) \left\{ \begin{array}{c} \cos m\phi \\ \sin m\phi \end{array} \right\}$$

and $P_l^m$ are associated Legendre polynomials [1]. The polar angle is denoted $\theta$ while $\phi$ is the azimuth angle. The upper (lower) expression is for $s = 2$ ($s = 1$), and the range of the indices are $l = 1, 2, \dots$ , $m = 0, 1, \dots, l$, $\tau = 1, 2$, $s = 2$ when $m = 0$ and $s = 1, 2$ otherwise. The multi-index $\nu = \{\tau, s, m, l\}$ is introduced to simplify the notation. It is ordered such that $\nu = 2(l^2 + l - 1 + (-1)^s m) + \tau$.

Note that

$$\begin{cases} \hat{\boldsymbol{r}} \cdot \boldsymbol{A}_{1sml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}} \cdot \boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) = 0 \\ \hat{\boldsymbol{r}} \times \boldsymbol{A}_{3sml}(\hat{\boldsymbol{r}}) = \boldsymbol{0}, \end{cases}$$

for which reason $\tau = 1$ (odd $\nu$) identifies a TE mode (magnetic $2^l$-pole) while $\tau = 2$ (even $\nu$) identifies a TM mode (electric $2^l$-pole) when the electric and magnetic fields are defined by (3.1) and (3.2), respectively. Furthermore,

$$\begin{cases} \boldsymbol{A}_{1sml}(\hat{\boldsymbol{r}}) = \boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) \times \hat{\boldsymbol{r}} \\ \boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}} \times \boldsymbol{A}_{1sml}(\hat{\boldsymbol{r}}). \end{cases} \tag{A.2}$$

The vector spherical harmonics are orthonormal on the unit sphere. More specifically, they satisfy

$$\int_{\Omega_{\hat{\boldsymbol{r}}}} \boldsymbol{A}_{\tau sml}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{A}_{\tau' s' m' l'}(\hat{\boldsymbol{r}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \delta_{\tau,\tau'} \delta_{s,s'} \delta_{m,m'} \delta_{l,l'}, \tag{A.3}$$

where $\Omega_{\hat{\boldsymbol{r}}} = \{(\theta, \phi) : 0 \leqslant \theta \leqslant \pi, 0 \leqslant \phi \leqslant 2\pi\}$ is the unit sphere and $\mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$. Define the $\mathrm{L}^2$-norm $||\cdot||$ for vector-valued functions on $\Omega_{\hat{\boldsymbol{r}}}$:

$$||\boldsymbol{G}||^2 = \int_{\Omega_{\hat{\boldsymbol{r}}}} \boldsymbol{G}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{G}^*(\hat{\boldsymbol{r}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}}.$$

If the norm of $\boldsymbol{G}$ is finite, it may be expanded in vector spherical harmonics:

$$\boldsymbol{G}(\hat{\boldsymbol{r}}) = \sum_{l=0}^{\infty} \sum_{m=0}^{l} \sum_{s=1}^{2} \sum_{\tau=1}^{3} c_{\tau sml} \boldsymbol{A}_{\tau sml}(\hat{\boldsymbol{r}}), \tag{A.4}$$

where the coefficients $c_\nu$ are given by

$$c_{\tau sml} = \int_{\Omega_{\hat{\boldsymbol{r}}}} \boldsymbol{G}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{A}_{\tau sml}(\hat{\boldsymbol{r}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}},$$

and the sum in the right-hand side of (A.4) converges in the norm $||\cdot||$.

The following expressions for the Cartesian unit vectors are used in (4.9):

$$\begin{cases} \hat{\boldsymbol{x}} = \sqrt{\dfrac{4\pi}{3}} \boldsymbol{A}_{3211}(\hat{\boldsymbol{r}}) + \sqrt{\dfrac{8\pi}{3}} \boldsymbol{A}_{2211}(\hat{\boldsymbol{r}}) \\ \hat{\boldsymbol{y}} = \sqrt{\dfrac{4\pi}{3}} \boldsymbol{A}_{3111}(\hat{\boldsymbol{r}}) + \sqrt{\dfrac{8\pi}{3}} \boldsymbol{A}_{2111}(\hat{\boldsymbol{r}}) \\ \hat{\boldsymbol{z}} = \sqrt{\dfrac{4\pi}{3}} \boldsymbol{A}_{3201}(\hat{\boldsymbol{r}}) + \sqrt{\dfrac{8\pi}{3}} \boldsymbol{A}_{2201}(\hat{\boldsymbol{r}}). \end{cases} \tag{A.5}$$

There are expansions for the Hankel functions, used to determine the polynomials $R_{\tau,l}^{(j)}$ in (3.3):

$$\begin{cases} h_l^{(1)}(z) = \dfrac{\mathrm{e}^{\mathrm{i}z}}{\mathrm{i}^{l+1} z} \displaystyle\sum_{n=0}^{l} \frac{(l+n)!}{n!(l-n)!} (-2\mathrm{i}z)^{-n} \\ h_l^{(2)}(z) = \dfrac{\mathrm{i}^{l+1}\mathrm{e}^{-\mathrm{i}z}}{z} \displaystyle\sum_{n=0}^{l} \frac{(l+n)!}{n!(l-n)!} (2\mathrm{i}z)^{-n}. \end{cases} \tag{A.6}$$

## A.2 Derivation of (4.6)

The scattered field $\widetilde{\boldsymbol{E}}_\mathrm{s}$ is the sum over $\boldsymbol{u}_\nu^{(1)}$ in (4.3), viz.

$$\widetilde{\boldsymbol{E}}_\mathrm{s}(k, \boldsymbol{r}) = \sqrt{\eta_0} \sum_\nu \tilde{d}_\nu^{(1)}(k) \boldsymbol{A}_\nu(\hat{\boldsymbol{r}}) \frac{\mathrm{e}^{ikr}}{r} \left(1 + \mathcal{O}(r^{-1})\right), \quad \text{as } r \to \infty,$$

where (A.1) and (A.6) have been used. From the above equation it is clear that the far-field amplitude $\widetilde{\boldsymbol{F}}(k, \hat{\boldsymbol{r}})$ in (4.4) is given by

$$\widetilde{\boldsymbol{F}}(k, \hat{\boldsymbol{r}}) = \sqrt{\eta_0} \sum_\nu \tilde{d}_\nu^{(1)}(k) \boldsymbol{A}_\nu(\hat{\boldsymbol{r}}).$$

Using (4.5), multiplying with $\boldsymbol{A}_{\nu'}(\hat{\boldsymbol{r}})$ and integrating over the unit sphere yield

$$\int \boldsymbol{A}_{\nu'}(\hat{\boldsymbol{r}}) \cdot \tilde{\mathbf{S}}_\mathrm{d}(k, \hat{\boldsymbol{r}}, \hat{\boldsymbol{k}}) \cdot \widetilde{\boldsymbol{E}}_0(k) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \sqrt{\eta_0} \tilde{d}_{\nu'}^{(1)}(k), \tag{A.7}$$

due to (A.3).

The coefficients $\tilde{d}_{\nu'}^{(1)}(k)$ are given by

$$\sqrt{\eta_0} \tilde{d}_{\nu'}^{(1)}(k) = \sqrt{\eta_0} \sum_{\nu''} \tilde{T}_{\nu',\nu''}(k) \tilde{d}_{\nu''}^{(v)}(k) = \frac{4\pi}{\mathrm{i}k} \sum_{\nu''} \tilde{T}_{\nu',\nu''}(k) \widetilde{\boldsymbol{E}}_0(k) \cdot \boldsymbol{A}_{\nu''}(\hat{\boldsymbol{k}}),$$

where the expansion coefficients $\tilde{d}^{(v)}_{\nu''}(k)$ of a plane wave $\mathrm{e}^{\mathrm{i}\boldsymbol{r}\cdot\boldsymbol{k}}\widetilde{\boldsymbol{E}}_0(k)$ have been used [11, p. 341]. Inserting this into (A.7) gives

$$\int \boldsymbol{A}_{\nu'}(\hat{\boldsymbol{r}}) \cdot \tilde{\mathbf{S}}_{\mathrm{d}}(k, \hat{\boldsymbol{r}}, \hat{\boldsymbol{k}}) \cdot \widetilde{\boldsymbol{E}}_0(k)\, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \frac{4\pi}{\mathrm{i}k} \sum_{\nu''} \tilde{T}_{\nu',\nu''}(k) \widetilde{\boldsymbol{E}}_0(k) \cdot \boldsymbol{A}_{\nu''}(\hat{\boldsymbol{k}}),$$

which must be valid for all $\hat{\boldsymbol{k}}$ and $\widetilde{\boldsymbol{E}}_0$. Letting $\widetilde{\boldsymbol{E}}_0(k, \hat{\boldsymbol{k}}) = \boldsymbol{A}_{\nu'''}(\hat{\boldsymbol{k}})\varphi(k)$ for some $\varphi \in \mathcal{S}$ and integrating once more over the unit sphere leads to

$$\int\int \boldsymbol{A}_{\nu'}(\hat{\boldsymbol{r}}) \cdot \tilde{\mathbf{S}}_{\mathrm{d}}(k, \hat{\boldsymbol{r}}, \hat{\boldsymbol{k}}) \cdot \boldsymbol{A}_{\nu'''}(\hat{\boldsymbol{k}})\varphi(k)\, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}}\, \mathrm{d}\Omega_{\hat{\boldsymbol{k}}} = \frac{4\pi}{\mathrm{i}k} \tilde{T}_{\nu',\nu'''}(k)\varphi(k),$$

and (4.6) is proven.

## A.3 Derivation of (4.16)

First set $k_0/k_n = \theta'_n - i\theta''_n$, where $\theta'_n \in \mathbb{R}$ and $\theta''_n > 0$. With $\theta_0 = \sum_n \theta''_n$, (4.13) takes the form

$$\frac{B_K \ln S^{-1}_{0,\nu}}{\pi} \leqslant k_0 a - \theta_0. \tag{A.8}$$

Furthermore, it follows that $\sum_n \operatorname{Im} k_0^3/k_n^3 \leqslant \sum_n \theta''^3_n \leqslant \theta_0^3$, since

$$\operatorname{Im} \frac{k_0^3}{k_n^3} = \frac{k_0^3}{|k_n|^6} \left[ (\operatorname{Im} k_n)^3 - (\operatorname{Re} k_n)^2 \operatorname{Im} k_n \right] \leqslant \frac{k_0^3}{|k_n|^6} (\operatorname{Im} k_n)^3 = \theta''^3_n.$$

Hence (4.14) becomes

$$\frac{B_K \ln S_0^{-1}}{\pi} \leqslant k_0^3 a^3 \rho_{\nu,\nu} + \frac{\theta_0^3}{3}. \tag{A.9}$$

Combining (A.8) and (A.9) yields

$$\frac{B_K \ln S^{-1}_{0,\nu}}{\pi} \leqslant k_0^3 a^3 \rho_{\nu,\nu} + \frac{1}{3} \left( k_0 a - \frac{B_K \ln S^{-1}_{0,\nu}}{\pi} \right)^3,$$

with solution (4.16).

# References


[1] G. B. Arfken and H. J. Weber. *Mathematical Methods for Physicists*. Academic Press, New York, fifth edition, 2001.

[2] A. Bernland, A. Luger, and M. Gustafsson. Sum rules and constraints on passive systems. *J. Phys. A: Math. Theor.*, **44**(14), 145205, 2011.

[3] C. F. Bohren and D. R. Huffman. *Absorption and Scattering of Light by Small Particles*. John Wiley & Sons, New York, 1983.

[4] A. Boström, G. Kristensson, and S. Ström. Transformation properties of plane, spherical and cylindrical scalar and vector wave functions. In V. V. Varadan, A. Lakhtakia, and V. K. Varadan, editors, *Field Representations and Introduction to Scattering*, Acoustic, Electromagnetic and Elastic Wave Scattering, chapter 4, pages 165–210. Elsevier Science Publishers, Amsterdam, 1991.

[5] M. R. Choi, K. J. Stanton-Maxey, J. K. Stanley, C. S. Levin, R. Bardhan, D. Akin, S. Badve, J. Sturgis, J. P. Robinson, R. Bashir, et al. A cellular Trojan horse for delivery of therapeutic nanoparticles into tumors. *Nano Letters*, **7**(12), 3759–3765, 2007.

[6] L. J. Chu. Physical limitations of omni-directional antennas. *J. Appl. Phys.*, **19**, 1163–1175, 1948.

[7] R. M. Fano. Theoretical limitations on the broadband matching of arbitrary impedances. *Journal of the Franklin Institute*, **249**(1,2), 57–83 and 139–154, 1950.

[8] A. A. Glazunov, M. Gustafsson, A. Molisch, and F. Tufvesson. Physical modeling of multiple-input multiple-output antennas and channels by means of the spherical vector wave expansion. *IET Microwaves, Antennas & Propagation*, **4**(6), 778–791, 2010.

[9] M. Gustafsson, C. Sohl, and G. Kristensson. Physical limitations on antennas of arbitrary shape. *Proc. R. Soc. A*, **463**, 2589–2607, 2007.

[10] M. Gustafsson and S. Nordebo. Bandwidth, Q factor, and resonance models of antennas. *Progress in Electromagnetics Research*, **62**, 1–20, 2006.

[11] J. E. Hansen, editor. *Spherical Near-Field Antenna Measurements*. Number 26 in IEE electromagnetic waves series. Peter Peregrinus Ltd., Stevenage, UK, 1988. ISBN: 0-86341-110-X.

[12] R. C. Hansen. *Electrically small, superdirective, and superconductive antennas*. John Wiley & Sons, New Jersey, 2006.

[13] F. W. King. *Hilbert Transforms, Volume 2*. Cambridge University Press, 2009.

[14] R. Kitamura, L. Pilon, and M. Jonasz. Optical constants of silica glass from extreme ultraviolet to far infrared at near room temperature. *Applied optics*, **46**(33), 8118–8133, 2007.

[15] R. E. Kleinman and T. B. A. Senior. Rayleigh scattering. In V. V. Varadan and V. K. Varadan, editors, *Low and high frequency asymptotics*, volume 2 of *Handbook on Acoustic, Electromagnetic and Elastic Wave Scattering*, chapter 1, pages 1–70. Elsevier Science Publishers, Amsterdam, 1986.

[16] J. Mashreghi. *Representation Theorems in Hardy Spaces*. Cambridge University Press, Cambridge, U.K., 2009.

[17] H. M. Nussenzveig. *Causality and dispersion relations*. Academic Press, London, 1972.

[18] S. J. Oldenburg, R. D. Averitt, S. L. Westcott, and N. J. Halas. Nanoengineering of optical resonances. *Chemical Physics Letters*, **288**(2-4), 243–247, 1998.

[19] E. M. Purcell. On the absorption and emission of light by interstellar grains. *J. Astrophys.*, **158**, 433–440, 1969.

[20] A. Shlivinski and E. Heyman. Time-domain near-field analysis of short-pulse antennas. I. Spherical wave (multipole) expansion. *IEEE Trans. Antennas Propagat.*, **47**(2), 271–279, February 1999.

[21] A. Shlivinski and E. Heyman. Time-domain near-field analysis of short-pulse antennas. II. Reactive energy and the antenna Q. *IEEE Trans. Antennas Propagat.*, **47**(2), 280–286, February 1999.

[22] A. Sihvola, P. Ylä-Oijala, S. Järvenpää, and J. Avelin. Polarizabilities of Platonic solids. *IEEE Trans. Antennas Propagat.*, **52**(9), 2226–2233, 2004.

[23] D. Sjöberg. Variational principles for the static electric and magnetic polarizabilities of anisotropic media with perfect electric conductor inclusions. *J. Phys. A: Math. Theor.*, **42**, 335403, 2009.

[24] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on broadband scattering by heterogeneous obstacles. *J. Phys. A: Math. Theor.*, **40**, 11165–11182, 2007.

[25] B. Ung and Y. Sheng. Interference of surface waves in a metallic nanoslit. *Optics Express*, **15**(3), 1182–1190, 2007.

[26] J. Volakis, C. C. Chen, and K. Fujimoto. *Small Antennas: Miniaturization Techniques & Applications*. McGraw-Hill, New York, 2010.

[27] M. Wohlers and E. Beltrami. Distribution theory as the basis of generalized passive-network analysis. *IEEE Transactions on Circuit Theory*, **12**(2), 164–170, 1965.

[28] D. Youla, L. Castriota, and H. Carlin. Bounded real scattering matrices and the foundations of linear passive network theory. *IRE Transactions on Circuit Theory*, **6**(1), 102–124, 1959.

[29] A. H. Zemanian. An n-port realizability theory based on the theory of distributions. *IEEE Transactions on Circuit Theory*, **10**(2), 265–274, 1963.

[30] A. H. Zemanian. *Distribution theory and transform analysis: an introduction to generalized functions, with applications*. McGraw-Hill, New York, 1965.

[31] A. H. Zemanian. *Realizability theory for continuous linear systems*. Academic Press, New York, 1972.

# On the Relation Between Optimal Wideband Matching and Scattering of Spherical Waves

## Paper III

Sven Nordebo, Anders Bernland, Mats Gustafsson,
Christian Sohl, and Gerhard Kristensson

**Based on:** S. Nordebo, A. Bernland, M. Gustafsson, C. Sohl, and G. Kristensson. On the relation between optimal wideband matching and scattering of spherical waves. *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 9, pp. 3358–3369, July 2011.

**Abstract**

Using an exact circuit analogy for the scattering of vector spherical waves, it is shown how the problem of determining the optimal scattering bounds for a homogeneous sphere in its high-contrast limit is identical to the closely related, and yet very different problem of finding the broadband tuning limits of the spherical waves. Using integral relations similar to Fano's broadband matching bounds, the optimal scattering limitations are determined by the static response as well as the high-frequency asymptotics of the reflection coefficient. The scattering view of the matching problem yields explicitly the necessary low-frequency asymptotics of the reflection coefficient that is used with Fano's broadband matching bounds for spherical waves, something that appears to be non-trivial to derive from the classical network point of view.

# 1 Introduction

Integral identities based on the properties of Herglotz functions [4], or positive real (PR) functions [37], constitute the basis for deriving Fano's broadband matching bounds [8] and have been used recently to describe a series of new sum rules for the scattering of electromagnetic waves [3, 4, 26, 27]. Hence, under the assumptions of linearity, continuity, time-translational invariance and passivity, sum rules can be derived from the analytic properties of the forward scattering dyadic, see e.g. [26, 27], and have also applications in antenna theory, see e.g. [11, 12, 25, 30, 32]. In [24], similar relations are used to determine the ultimate thickness to bandwidth ratio of radar absorbers. Limitations on the scattering of vector spherical waves have been considered in [3].

The sum rules rely on the well-known connection between the transfer functions of causal and passive systems and Herglotz functions, or positive real (PR) functions, as well as the analytic properties of these functions, see e.g. [35, 37, 38]. Consequently, sum rules and limitations on arbitrary reflection coefficients stemming from passive systems can be derived, as described in [4]. The procedure is reviewed briefly in this paper.

By using Fano's approach, optimum broadband tuning limits of the higher-order spherical waves are considered in [30, 32], giving important physical insight into the matching limitations for Ultra-Wideband (UWB) antennas, see also [9, 14, 15, 18, 33, 34, 36]. Previously, the Fano broadband matching bounds have been applied mainly to the lowest order spherical waves. However, methods of finding solutions to the matching limitations for the spherical waves of higher orders have recently received further development, see e.g. [20, 30, 31] with references. Hence, there is a need to further develop analytical results as an aid in the related numerical analysis.

In [30, 32], it is conjectured that the low-frequency asymptotics of the positive real function $-\log\rho_{\tau l}$ is of the form

$$-\log(\pm\rho_{\tau l}) = 2\frac{a}{c_0}s + 2(-1)^l\left(\frac{a}{c_0}\right)^{2l+1}c_{\tau l}s^{2l+1} + \cdots, \tag{1.1}$$

where $\rho_{\tau l}$ is the reflection coefficient corresponding to a TE ($\tau = 1$) or TM ($\tau = 2$) spherical wave of order $l$, $s$ the Laplace variable, $a$ the radius of a circumscribing

sphere, $c_0$ the speed of light in free space, and $c_{\tau l}$ constants to be determined from network analysis and the circuit analogy of the spherical wave impedance. The aim of this paper is to give an analytical solution to the conjecture (1.1) made in [30, 32], by using the recent developments in the application of sum rules for passive scatterers.

The rest of the paper is organized as follows: In section 2 is given a brief outline on the general approach to obtain sum rules and physical limitations for reflection coefficients stemming from passive systems, and the Fano broadband matching bounds for spherical waves is put in this context. The related conjecture (1.1) is also given a precise formulation.

In section 3 is treated the problem of finding the optimal limitations for scattering of vector spherical waves. Here the geometry of the spherical object is known but the dispersion is unknown. A detailed study of the high-frequency asymptotics of the reflection coefficient is performed including e.g. the Debye and the Lorentz dispersion models, and is given in the Appendix A. Using the integral relations derived in [4], which are similar to the relations in the derivation of Fano's broadband matching bounds [8], the optimal scattering limitations are determined by the static response as well as the high-frequency asymptotics of the reflection coefficient. As with the Fano approach, the integral relations yield a non-convex global optimization problem which in general is difficult to handle.

In section 4, the two previous sections are tied together. An exact circuit analogy for the scattering of spherical waves is used similar to [6, 29], to show how the problem of determining the scattering limitations for a homogeneous sphere in its high-contrast limit becomes identical to the closely related, and yet very different problem of finding the broadband tuning limits of the spherical waves [30, 32]. Furthermore, the scattering view of the matching problem yields explicitly the necessary low-frequency asymptotics of the reflection coefficient (1.1), i.e. the coefficients $c_{\tau l}$ that are used with Fano's broadband matching bounds for spherical waves. The coefficients $c_{\tau l}$ are given by the equation (4.16) in this paper. This is something that appears to be non-trivial to derive from the classical network point of view.

Finally, in section 5 is given a numerical example where a relaxation of the Fano equations is considered which is easily solved, and which is especially useful in the regime of Rayleigh scattering.

## 2 Limitations on passive reflection coefficients

This section reviews the general approach presented in [4] to find sum rules and physical limitations for reflection coefficients stemming from linear, continuous, time-translational invariant, and passive physical systems. The approach, which is used for the matching and scattering problems in the following sections, relies on the well-known connection between the transfer functions of causal and passive systems and Herglotz (or positive real) functions, as described in e.g. [4, 35, 37, 38]. A set of integral identities for Herglotz functions was proved in [4]. Applied to a reflection coefficient $\rho$, they give a set of sum rules. The sum rules relate integrals

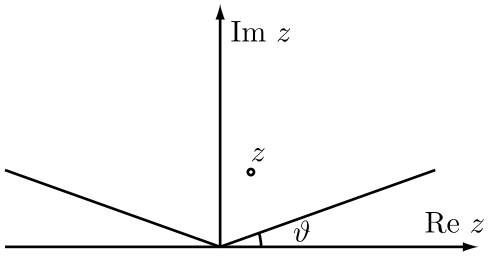



**Figure 1**: The cone $\{z : \vartheta \leqslant \arg z \leqslant \pi - \vartheta\}$ for some $\vartheta \in (0, \pi/2]$.

of $\rho$ over infinite frequency intervals to the static and high-frequency properties of the system, and so are much like Fano's matching equations. Physical limitations for the reflection coefficient are derived by considering finite frequency intervals. The general approach is presented in more detail in [4], where also all the necessary proofs are given.

## 2.1 Herglotz functions and integral identities

Here the class of Herglotz functions is reviewed briefly, and the integral identities used to obtain sum rules and limitations for reflection coefficients are presented. A Herglotz function $h(z)$ is defined as an analytic function for $z \in \mathbb{C}_+ = \{z : \operatorname{Im} z > 0\}$ with the property that $\operatorname{Im} h(z) \geqslant 0$, cf. [4, 21]. It is assumed that $h$ obeys the symmetry

$$h(z) = -h^*(-z^*) \tag{2.1}$$

where $(\cdot)^*$ denotes the complex conjugate [4]. Herglotz functions stemming from reflection coefficients in real physical systems exhibit this symmetry property [35, 37, 38]. For all Herglotz functions it holds that $\lim_{z\hat{\to}0} zh(z) = a_{-1} \leqslant 0$ and $\lim_{z\hat{\to}\infty} h(z)/z = b_1 \geqslant 0$. Throughout this paper, $z\hat{\to}0$ means $|z| \to 0$ in the cone $\vartheta \leqslant \arg z \leqslant \pi - \vartheta$ for any $\vartheta \in (0, \pi/2]$, and likewise for $z\hat{\to}\infty$, see Figure 1.

It is assumed that the low- and high-frequency asymptotic expansions are given by

$$\begin{cases} h(z) = \displaystyle\sum_{m=0}^{N} a_{2m-1} z^{2m-1} + o(z^{2N-1}) & \text{as } z\hat{\to}0 \\ h(z) = \displaystyle\sum_{m=0}^{M} b_{1-2m} z^{1-2m} + o(z^{1-2M}) & \text{as } z\hat{\to}\infty, \end{cases} \tag{2.2}$$

where the little-o notation $o$ is defined as in [22], and $M$ and $N$ are non-negative integers (or possibly infinity), chosen so that all the coefficients $a_m$ and $b_m$ are real, and hence that all the even indexed coefficients are zero [4]. The asymptotic expansions are clearly valid as $z \to 0$ ($z \to \infty$) for any argument in the case $h$ is analytic in a neighbourhood of the origin (infinity).

The following integral identities have been derived in [4], and they are the starting

point to derive limitations on reflection coefficients:

$$\frac{2}{\pi}\int_0^\infty \frac{\operatorname{Im} h(x)}{x^{2p}}\,\mathrm{d}x = a_{2p-1} - b_{2p-1}, \quad \text{for } p = 1-M, 2-M, \ldots, N. \tag{2.3}$$

It should be noted that the integral identities in (2.3) do not apply in the case when the largest possible integers $N$ and $M$ are $N = M = 0$ in (2.2). In case the imaginary part $\operatorname{Im} h(x)$ is not regular on the real axis, the integral should be interpreted as

$$\lim_{\varepsilon\to 0^+}\lim_{y\to 0^+}\frac{2}{\pi}\int_{\varepsilon<|x|<\varepsilon^{-1}} \frac{\operatorname{Im} h(x+\mathrm{i}y)}{x^{2p}}\,\mathrm{d}x = a_{2p-1}-b_{2p-1}, \quad \text{for } p = 1-M, 2-M, \ldots, N, \tag{2.4}$$

and where i denotes the imaginary unit, $\mathrm{i}^2 = -1$. This is equivalent to interpreting (2.3) in the distributional sense [4]. Equation (2.3) is assumed to be replaced by (2.4) whenever necessary throughout this paper. The identities (2.3) can be used to derive Fano's matching equations [8]. They have also been used recently to derive a series of new sum rules for the scattering of electromagnetic waves [3, 26, 27], with applications in antenna theory [11, 30, 32]. There are other applications for the identities (2.3) as well, see e.g. [5, 10, 13, 16, 24].

## 2.2 Limitations on passive reflection coefficients

Let $\rho(\omega)$ denote a reflection coefficient of a system where the reflected signal $u(\omega)$ is related to the incoming signal $v(\omega)$ as

$$u(\omega) = \rho(\omega)v(\omega),$$

where $\omega$ is the angular frequency. It is assumed that the reflection coefficient is the Fourier transform of a real-valued convolution kernel $\rho_{\mathrm{r}}(t)$. The Fourier transform is defined as $\rho(\omega) = \int_{-\infty}^{\infty}\rho_{\mathrm{r}}(t)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t$ when $\rho_{\mathrm{r}}(t)$ is sufficiently regular, and it is otherwise defined in the appropriate distributional sense [4, 38].

If $\rho(\omega)$ corresponds to a passive system, it is bounded with $|\rho(\omega)| \leqslant 1$. The system is causal if the reflection coefficient corresponds to a causal convolution kernel $\rho_{\mathrm{r}}^{\mathrm{c}}(t)$ which vanishes for $t < 0$. It is a well-known result that the reflection coefficient $\rho^{\mathrm{c}}(\omega)$ of a passive and causal system is an analytic function bounded in magnitude by one in the open upper half plane, i.e. $\rho^{\mathrm{c}}(\omega + \mathrm{i}\sigma)$ is analytic and $|\rho^{\mathrm{c}}(\omega + \mathrm{i}\sigma)| \leqslant 1$ for $\sigma > 0$ [35, 37]. The scattering of electromagnetic waves is always causal. However, sometimes the scattering of electromagnetic waves, such as the scattering of incoming and outgoing spherical waves, may be perceived as non-causal depending on the definition of the scattering coefficient [3, 21]. Hence, for a non-causal system a time delay $t_0$ can be introduced so that $\mathrm{e}^{\mathrm{i}\omega t_0}\rho(\omega)$ corresponds to a causal convolution kernel $\rho_{\mathrm{r}}(t - t_0)$. The reflection coefficient $\rho(\omega + \mathrm{i}\sigma)$ is thus an analytic function for $\sigma > 0$, and it is bounded according to $|\mathrm{e}^{(\mathrm{i}\omega-\sigma)t_0}\rho(\omega + \mathrm{i}\sigma)| \leqslant 1$.

A Herglotz function can be constructed by taking the complex logarithm of $\rho$ [4]. It requires that the zeros of $\rho$ are removed, which is done with a Blaschke-product [7]. The Herglotz function is therefore (with $z = \omega + \mathrm{i}\sigma$):

$$h(z) = -\mathrm{i}\log\left(\mathrm{e}^{\mathrm{i}zt_0}\rho(z)\prod_n \frac{1 - z/z_n^*}{1 - z/z_n}\right), \tag{2.5}$$

where the zeros $z_n$ of $\rho$ in $\mathbb{C}^+$ are repeated according to their multiplicity. The convolution kernel $\rho_{\mathrm{r}}(t)$ is real-valued, and so $\rho(\mathrm{i}\sigma)$ is real-valued on the imaginary axis with the symmetry $\rho(z) = \rho^*(-z^*)$ for $z \in \mathbb{C}^+$. Without loss of generality it may be assumed that $\rho(\mathrm{i}\sigma) > 0$, in which case $h(z)$ obeys the symmetry (2.1). If $\rho(\mathrm{i}\sigma) < 0$, consider the function $-\rho(z)$ instead.

Suppose that the low-frequency asymptotics of $-\mathrm{i}\log\rho(z)$ is given by

$$-\mathrm{i}\log\rho(z) = \sum_{m=0}^{N} a_{2m-1}^{(0)} z^{2m-1} + o(z^{2N-1}), \quad \text{as } z\hat{\to}0.$$

The low-frequency asymptotics of $h(z)$ is then

$$h(z) = zt_0 + \sum_{m=0}^{N} a_{2m-1}^{(0)} z^{2m-1} + o(z^{2N-1}) + \sum_{m=1,3,\ldots}^{\infty} \frac{2}{m} \sum_n \mathrm{Im}\left\{\frac{1}{z_n^m}\right\} z^m, \quad \text{as } z\hat{\to}0. \tag{2.6}$$

Note that there are only odd indices $m$ in the last summation above since the complex zeros appear in symmetric pairs $(z_n, -z_n^*)$.

With $p = 1, 2, \ldots, N$, the following relationships are now obtained from (2.3):

$$\frac{2}{\pi}\int_0^\infty \frac{1}{\omega^{2p}} \log|\rho(\omega)|^{-1}\,\mathrm{d}\omega = \left\{\delta_{p1}(t_0 - b_1) + a_{2p-1}^{(0)} + \frac{2}{2p-1}\sum_n \mathrm{Im}\left\{\frac{1}{z_n^{2p-1}}\right\}\right\}. \tag{2.7}$$

Here $\delta_{p1}$ denotes the Kronecker delta. Note that the term $b_1 \geqslant 0$ originates from the high-frequency asymptotics.

Denote $\rho_0 = \max_\omega |\rho(\omega)|$ where the maximum is taken over the angular frequency interval $\omega \in [\omega_0(1 - \frac{B}{2}), \omega_0(1 + \frac{B}{2})]$, $\omega_0$ is the center angular frequency and $B$ the relative bandwidth ($0 \leqslant B \leqslant 2$). The integral identities (2.7) then yield the following inequalities:

$$\begin{aligned}\log\rho_0^{-1}B \leqslant \log\rho_0^{-1}G_p(B) &\leqslant \omega_0^{2p-1}\int_0^\infty \frac{1}{\omega^{2p}}\log|\rho(\omega)|^{-1}\,\mathrm{d}\omega \\ &= \frac{\pi}{2}\left[\delta_{p1}t_0\omega_0^{2p-1} + a_{2p-1}^{(0)}\omega_0^{2p-1} + \frac{2}{2p-1}\sum_n \mathrm{Im}\left\{\left(\frac{\omega_0}{z_n}\right)^{2p-1}\right\}\right], \end{aligned} \tag{2.8}$$

where it has been used that $b_1 \geqslant 0$. The factor $G_p(B)$ is defined by

$$G_p(B) = \int_{1-B/2}^{1+B/2} \frac{1}{x^{2p}}\,\mathrm{d}x = \frac{1}{2p-1}\frac{(1+\frac{B}{2})^{2p-1} - (1-\frac{B}{2})^{2p-1}}{(1-\frac{B^2}{4})^{2p-1}}. \tag{2.9}$$

Note that $B \leqslant G_p(B)$ for all $0 \leqslant B \leqslant 2$, and $G_p(B) \approx B$ in the narrowband approximation when $B \ll 1$.

## 2.3 Fano broadband matching bounds for spherical waves

The classical broadband matching bounds for lossless networks by Fano [8] are revisited using the Herglotz function formulation and integral identities (2.3) and (2.7).

The Fano matching bounds are then used to formulate the problem of finding the broadband tuning limits of the wave impedance of the spherical waves as in [30, 32].

In circuit theory it is convenient to employ the Laplace variable $s = -\mathrm{i}z = \mathrm{j}\omega + \sigma$, with $\mathrm{j} = -\mathrm{i}$. The Herglotz function $h(z)$ then corresponds to a positive real (PR) function $g(s) = -\mathrm{i}h(z)$ with the property that $g(s)$ is analytic with $\operatorname{Re} g(s) \geqslant 0$ for $\operatorname{Re} s = \sigma > 0$, cf. [23, 38]. The symmetry (2.1) takes the form $g(s) = g^*(s^*)$. The low- and high-frequency asymptotics are given by

$$\begin{cases} g(s) = \displaystyle\sum_{m=0}^{N} A_{2m-1} s^{2m-1} + o(s^{2N-1}) \quad \text{as } \mathrm{i}s \hat{\to} 0 \\ g(s) = \displaystyle\sum_{m=0}^{M} B_{1-2m} s^{1-2m} + o(s^{1-2M}) \quad \text{as } \mathrm{i}s \hat{\to} \infty, \end{cases} \tag{2.10}$$

where all coefficients are real and the even indexed coefficients are zero. Furthermore, the PR function property implies that $A_{-1} \geqslant 0$ and $B_1 \geqslant 0$. Note also that the mapping $g(s) = -\mathrm{i}h(\mathrm{i}s)$ implies the relations $A_{2m-1} = (-1)^{m+1} a_{2m-1}$ and $B_{2m-1} = (-1)^{m+1} b_{2m-1}$ for the coefficients in (2.2) and (2.10). The following integral identity now follows directly from (2.3):

$$\frac{2}{\pi} \int_0^\infty \frac{\operatorname{Re} g(\mathrm{j}\omega)}{\omega^{2p}} \, \mathrm{d}\omega = (-1)^{p+1} \left( A_{2p-1} - B_{2p-1} \right), \quad \text{for } p = 1 - M, 2 - M, \ldots, N. \tag{2.11}$$

Let $\rho(s)$ denote the reflection coefficient corresponding to an arbitrary impedance function defined by a passive $RLC$ network. Such an impedance function can always be represented by a lossless two-port which is terminated in a pure resistance [8]. The appropriate PR function corresponding to (2.5) is given by

$$g(s) = -\log \left( \rho(s) \prod_n \frac{1 + s/s_n^*}{1 - s/s_n} \right), \tag{2.12}$$

where $s_n$ are the zeros of $\rho(s)$ with $\operatorname{Re} s_n > 0$. Note that the causality factor $\mathrm{e}^{\mathrm{i}zt_0}$ is not needed here, since the reflection coefficient corresponds to a causal convolution kernel.

Suppose that the low-frequency asymptotics of $-\log \rho(s)$ is given by $-\log \rho(s) = \sum_{m=0}^{M} A_{2m-1}^{(0)} s^{2m-1} + o(s^{2M-1})$, as $\mathrm{i}s \hat{\to} 0$. The low-frequency asymptotics of $g(s)$ is then given by

$$g(s) = \sum_{m=0}^{N} A_{2m-1}^{(0)} s^{2m-1} + o(s^{2N-1}) - \sum_{m=1,3,\ldots}^{\infty} \frac{2}{m} \sum_n \operatorname{Re} \left\{ \frac{1}{s_n^m} \right\} s^m, \quad \text{as } \mathrm{i}s \hat{\to} 0. \tag{2.13}$$

With $p = 1, 2, \ldots, N$, the following relationships are now obtained from (2.11) (cf. [8]):

$$\frac{2}{\pi} \int_0^\infty \frac{1}{\omega^{2p}} \log |\rho(\mathrm{j}\omega)|^{-1} \, \mathrm{d}\omega = (-1)^{p+1} \left\{ A_{2p-1}^{(0)} - \delta_{p1} B_1 - \frac{2}{2p-1} \sum_n \operatorname{Re} \left\{ \frac{1}{s_n^{2p-1}} \right\} \right\}. \tag{2.14}$$

Note that $B_1 = 0$ if the circuit consists of only lumped elements, since $\rho(s)$ is a rational function in this case. Furthermore, for rational functions $\rho(s)$ the asymptotic expansions (2.10) are valid as $s \to 0$ and $s \to \infty$, respectively.

Consider now the broadband matching problem as described in [8]. In Figure 2 is shown the cascade of two lossless and reciprocal two-ports $N'$ and $N''$ with a source at one side and a resistive termination at the other side. Let $N'$ be the fixed network and $N''$ the matching network. The reflection and transmission coefficients for the overall two-port are denoted by $\rho_1$, $\rho_2$ and $\varrho$. Since the overall two-port is lossless with $|\rho_1|^2 = 1 - |\varrho|^2 = |\rho_2|^2$, the optimal matching limitations for the input port of interest with coefficient $\rho_2$ may be conveniently analyzed by considering the opposite port with coefficient $\rho_1$, as depicted in Figure 2.

The reflection coefficient $\rho_1$ for the overall two-port is given by

$$\rho_1 = \rho_1' + \frac{\varrho'^2 \rho_1''}{1 - \rho_2' \rho_1''}, \tag{2.15}$$

where the primed scattering parameters corresponding to the two networks $N'$ and $N''$ have been defined as indicated in Figure 2. It is assumed that the network $N'$ is a high-pass $LC$ ladder with $\varrho'(0) = 0$ (where $s = 0$). Without loss of generality, it may then be assumed that $\rho_1'(0) = 1$. Furthermore, it is also assumed that $\rho_2'(0)\rho_1''(0) = -1$ so that there is no cancellation of zeros at $s = 0$ in (2.15). This condition is easily achieved by choosing the appropriate $LC$ ladder structure for $N''$ if $\rho_2'(0)$ is known, cf. also [8].

Suppose that the transmission coefficient $\varrho'$ has a zero of order $N$ at $s = 0$. This implies that $\rho_1'(s)\rho_1'(-s) = 1 - \varrho'(s)\varrho'(-s) = 1 + \mathcal{O}(s^{2N})$, where the big-O notation $\mathcal{O}$ is defined as in [22]. Suppose further that the low-frequency asymptotics of $\rho_1'$ is given by $-\log \rho_1'(s) = \sum_{m=0}^{\infty} A_m^{(0)\prime} s^m$ as $s \to 0$. Hence,

$$-\log \rho_1'(s) - \log \rho_1'(-s) = \sum_{m=0}^{\infty} A_m^{(0)\prime} s^m + \sum_{m=0}^{\infty} A_m^{(0)\prime} (-1)^m s^m = \mathcal{O}(s^{2N}), \quad \text{as } s \to 0, \tag{2.16}$$

implying that $A_m^{(0)} = A_m^{(0)\prime} = 0$ for $m = 0, 2, \ldots, 2N-2$. Furthermore, from (2.15) follows that $\frac{\partial^m}{\partial s^m} \log \rho_1|_{s=0} = \frac{\partial^m}{\partial s^m} \log \rho_1'|_{s=0}$ for $0 \leqslant m \leqslant 2N-1$, and hence the invariance of the Taylor coefficients $A_m^{(0)} = A_m^{(0)\prime}$ for $m = 1, 3, \ldots, 2N-1$. Thus, (2.14) can now be applied with $A_{2p-1}^{(0)} = A_{2p-1}^{(0)\prime}$ for $p = 1, 2, \ldots, N$. These are the original Fano matching equations formulated in [8].

Consider now the problem of finding the optimum broadband tuning limits of the wave impedance of the spherical waves, as described in e.g. [30, 32]. Hence, consider the matching problem of an outgoing $\mathrm{TE}_l$ ($\tau = 1$) or $\mathrm{TM}_l$ ($\tau = 2$) spherical wave of order $l$. As was shown by Chu [6], the wave impedance of the spherical waves as seen at a spherical boundary can be represented by a finite $LC$ high-pass ladder network terminated in a fixed resistance, cf. Figure 3. The impedance $Z_{\tau l}$ is the normalized wave impedance as seen at a spherical boundary of radius $a$, i.e. at the left (antenna) side of the equivalent circuit in Figure 3. The input impedance used in the Fano analysis is the impedance $Z_{1,\tau l}$ as seen from the opposite, right-hand

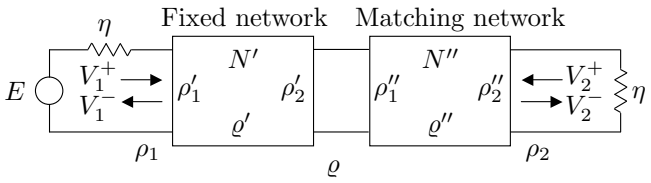



**Figure 2**: Cascade of two reciprocal two-ports $N'$ and $N''$. Here, $\rho_1$, $\rho_2$ and $\varrho$ denote the overall scattering parameters, and the corresponding primed scattering parameters refer to the two networks $N'$ and $N''$.

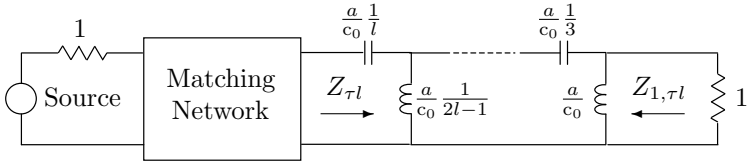



**Figure 3**: Matching network and equivalent circuit for the impedance of a $\mathrm{TM}_l$ wave at a spherical boundary of radius $a$. The circuit is drawn for a $\mathrm{TM}_l$ wave of odd order $l$.

side of the equivalent circuit when it is correctly terminated in a pure resistance. The corresponding reflection coefficient is given by $\rho_{1,\tau l} = (Z_{1,\tau l} - 1)/(Z_{1,\tau l} + 1)$.

It has been conjectured [30, 32] that the low-frequency asymptotics of $-\log \rho_{1,\tau l}$ is of the form

$$-\log(\pm\rho_{1,\tau l}) = A_1^{(0)} s + A_{2l+1}^{(0)} s^{2l+1} + \mathcal{O}(s^{2l+2}), \quad \text{as } s \to 0 \tag{2.17}$$

where

$$\begin{cases} A_1^{(0)} = 2\dfrac{a}{\mathrm{c}_0} \\ A_{2l+1}^{(0)} = 2(-1)^l (\dfrac{a}{\mathrm{c}_0})^{2l+1} c_{\tau l} \end{cases} \tag{2.18}$$

and where $c_{\tau l}$ is a constant determined from network analysis.

The conjecture (2.18) may be verified by using the equivalent circuits for a fixed order $l = 1, 2, \ldots$. However, from a network ($N'$) analysis point of view, it seem to be non-trivial to prove it for general order $l$. In the next two sections, it is shown that the conjecture (2.18) is true and an explicit expression for $A_{2l+1}^{(0)}$ is given by showing that the matching problem is equivalent to the problem of finding the optimal scattering limitations for a homogeneous sphere in its high-contrast limit, i.e. in the limit as the permittivity or the permeability tends to infinity.

# 3 Optimal limitations for scattering of spherical waves

Consider the scattering of vector spherical waves which is associated with an isotropic and homogeneous sphere of radius $a$, and with relative permeability and permittiv-

ity $\mu$ and $\epsilon$, respectively. The refractive index is $n = (\mu\epsilon)^{1/2}$ and the relative wave impedance $\eta = (\mu/\epsilon)^{1/2}$. The exterior of the sphere is free space, and $\mathrm{c}_0$ and $\eta_0$ are the speed of light and the wave impedance of free space, respectively. For convenience, introduce the angular wavenumber $\mathrm{Re}\{k\} = \omega/\mathrm{c}_0$. Allow $k$ to take values in the upper-half plane, so that $k$ corresponds to $z/\mathrm{c}_0$ in Section 2.2. Let $(r, \theta, \phi)$ denote the spherical coordinates and $\boldsymbol{r} = r\hat{\boldsymbol{r}}$ the corresponding radius vector.

## 3.1 Exterior of the sphere

The electric and magnetic fields outside the sphere, i.e. for $r \geqslant a$, are given by

$$\boldsymbol{E}(\boldsymbol{r}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \sum_{\tau=1}^{2} a_{\tau ml}^{(1)} \boldsymbol{u}_{\tau ml}^{(1)}(k\boldsymbol{r}) + a_{\tau ml}^{(2)} \boldsymbol{u}_{\tau ml}^{(2)}(k\boldsymbol{r}), \tag{3.1}$$

and

$$\boldsymbol{H}(\boldsymbol{r}) = \frac{1}{\mathrm{i}\eta_0} \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \sum_{\tau=1}^{2} a_{\tau ml}^{(1)} \boldsymbol{u}_{\bar{\tau} ml}^{(1)}(k\boldsymbol{r}) + a_{\tau ml}^{(2)} \boldsymbol{u}_{\bar{\tau} ml}^{(2)}(k\boldsymbol{r}) \tag{3.2}$$

where $\boldsymbol{u}_{\tau ml}^{(1)}(k\boldsymbol{r})$ and $\boldsymbol{u}_{\tau ml}^{(2)}(k\boldsymbol{r})$ are *outgoing* and *incoming* vector spherical waves, respectively, see e.g. [2, 17, 19], and $a_{\tau ml}^{(j)}$ the corresponding multipole coefficients. Here $\tau = 1$ corresponds to transverse electric (TE) waves, $\tau = 2$ corresponds to transverse magnetic (TM) waves, and $\bar{\tau} = 3 - \tau$ denotes the complementary index. The other indices are $l = 1, 2, \ldots$, and $m = -l, -l+1, \ldots, l$, where $l$ denotes the *order* of the spherical wave. The vector spherical waves are given by

$$\begin{aligned} \boldsymbol{u}_{1ml}^{(j)}(k\boldsymbol{r}) &= \mathrm{h}_l^{(j)}(kr)\boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}) \\ \boldsymbol{u}_{2ml}^{(j)}(k\boldsymbol{r}) &= \frac{(kr\mathrm{h}_l^{(j)}(kr))'}{kr}\boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}}) + \sqrt{l(l+1)}\frac{\mathrm{h}_l^{(j)}(kr)}{kr}\boldsymbol{A}_{3ml}(\hat{\boldsymbol{r}}) \end{aligned} \tag{3.3}$$

where $\boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}})$ are the *vector spherical harmonics* and $\mathrm{h}_l^{(j)}(x)$ the *spherical Hankel* functions of the $j$th kind, $j = 1, 2$, and order $l$, see e.g. [2, 17, 19]. Here, $(\cdot)'$ denotes differentiation with respect to the argument $kr$. The vector spherical harmonics $\boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}})$ are given by

$$\begin{cases} \boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}) = \dfrac{1}{\sqrt{l(l+1)}} \nabla \times (\boldsymbol{r}\mathrm{Y}_{ml}(\hat{\boldsymbol{r}})) \\ \boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}} \times \boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}) \\ \boldsymbol{A}_{3ml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}}\mathrm{Y}_{ml}(\hat{\boldsymbol{r}}) \end{cases} \tag{3.4}$$

where $\mathrm{Y}_{ml}(\hat{\boldsymbol{r}})$ are the scalar *spherical harmonics* given by

$$\mathrm{Y}_{ml}(\theta, \phi) = (-1)^m \sqrt{\frac{2l+1}{4\pi}} \sqrt{\frac{(l-m)!}{(l+m)!}} \mathrm{P}_l^m(\cos\theta) \mathrm{e}^{\mathrm{i}m\phi}, \tag{3.5}$$

and where $\mathrm{P}_l^m(x)$ are the *Associated Legendre functions*, see e.g. [2]. The vector spherical harmonics $\boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}})$ are orthonormal on the unit sphere and have the directional properties $\hat{\boldsymbol{r}} \cdot \boldsymbol{A}_{\tau ml}(\hat{\boldsymbol{r}}) = 0$ for $\tau = 1, 2$ and $\hat{\boldsymbol{r}} \times \boldsymbol{A}_{3ml}(\hat{\boldsymbol{r}}) = \boldsymbol{0}$.

## 3.2 Interior of the sphere and scattering coefficients

The electric and magnetic fields inside the sphere for $r \leqslant a$ are given by

$$\boldsymbol{E}(\boldsymbol{r}) = \sum_{l=1}^{\infty}\sum_{m=-l}^{l}\sum_{\tau=1}^{2} b_{\tau ml}\boldsymbol{v}_{\tau ml}(kn\boldsymbol{r}) \tag{3.6}$$

$$\boldsymbol{H}(\boldsymbol{r}) = \frac{1}{\mathrm{i}\eta_0\eta}\sum_{l=1}^{\infty}\sum_{m=-l}^{l}\sum_{\tau=1}^{2} b_{\tau ml}\boldsymbol{v}_{\bar{\tau} ml}(kn\boldsymbol{r}) \tag{3.7}$$

where $\boldsymbol{v}_{\tau ml}(kn\boldsymbol{r})$ are *regular* vector spherical waves, and $b_{\tau ml}$ the corresponding multipole coefficients. The regular vector spherical waves are defined by

$$\begin{aligned}\boldsymbol{v}_{1ml}(kn\boldsymbol{r}) &= \mathrm{j}_l(knr)\boldsymbol{A}_{1ml}(\hat{\boldsymbol{r}}) \\ \boldsymbol{v}_{2ml}(kn\boldsymbol{r}) &= \frac{(knr\mathrm{j}_l(knr))'}{knr}\boldsymbol{A}_{2ml}(\hat{\boldsymbol{r}}) + \sqrt{l(l+1)}\frac{\mathrm{j}_l(knr)}{knr}\boldsymbol{A}_{3ml}(\hat{\boldsymbol{r}})\end{aligned} \tag{3.8}$$

where $\mathrm{j}_l(x)$ are the *spherical Bessel* functions of order $l$, see e.g. [2, 17, 19]. Here, $(\cdot)'$ denotes differentiation with respect to the argument $knr$.

Continuity of the tangential fields $\boldsymbol{E}_{\mathrm{t}}$ and $\boldsymbol{H}_{\mathrm{t}}$ in (3.1), (3.2), (3.6) and (3.7) for $r = a$ yields the following solution for the reflection coefficient defined by $a^{(1)}_{\tau ml} = \rho_{\tau l}a^{(2)}_{\tau ml}$:

$$\rho_{\tau l}(k) = \frac{-\mathrm{h}^{(2)}_l(ka)\left(kna\mathrm{j}_l(kna)\right)' + \nu_\tau(k)\mathrm{j}_l(kna)\left(ka\mathrm{h}^{(2)}_l(ka)\right)'}{\mathrm{h}^{(1)}_l(ka)\left(kna\mathrm{j}_l(kna)\right)' - \nu_\tau(k)\mathrm{j}_l(kna)\left(ka\mathrm{h}^{(1)}_l(ka)\right)'} \tag{3.9}$$

where $\nu_1 = \mu$ and $\nu_2 = \epsilon$, cf. [28]. It is assumed that $\nu_\tau(k)$ can be represented by an asymptotic series at $k = 0$. It has been shown that $\rho_{\tau l} = \mathrm{e}^{-\mathrm{i}2ka}\rho^{\mathrm{c}}_{\tau l}$, where $\rho^{\mathrm{c}}_{\tau l}$ is the transform of a causal kernel, see [3]. It can be expected that $\rho^{\mathrm{c}}_{\tau l}(k) = o(k)$ as $k \hat{\to} \infty$, which means that $b_1 = 0$ for the Herglotz function corresponding to (2.5). A detailed study of the high-frequency asymptotics of the reflection coefficient has been performed in the Appendix A, including e.g. the Debye and Lorentz dispersion models, and it asserts this expectation for these material models. The low-frequency asymptotics is obtained from a Taylor series expansion yielding $\rho_{\tau l}(k) \sim 1 + \mathrm{i}2(ka)^{2l+1}c_{\tau l}$, or

$$-\mathrm{i}\log\rho_{\tau l}(k) \sim a^{(0)}_{2l+1}k^{2l+1} = 2(ka)^{2l+1}c_{\tau l}, \quad \text{as } k \to 0, \tag{3.10}$$

where

$$c_{\tau l} = \frac{2^{2l}(l+1)!l!}{(2l+1)!(2l)!}\frac{\nu_\tau(0) - 1}{l + 1 + \nu_\tau(0)l} \tag{3.11}$$

and $\nu_\tau(0)$ is the static response. The symbol $\sim$ denotes asymptotic equivalence and is defined in e.g. [22]. Note that the low-frequency asymptotics of the TM (TE) wave reflection is independent of $\mu(0)$ ($\epsilon(0)$).

## 3.3 Optimization formulation

The following inequalities are obtained from (2.8) when applied to the reflection coefficient (3.9) using (3.10), where $p = 1, 2 \ldots, l+1$, $a_{2l+1}^{(0)} = 2a^{2l+1}c_{\tau l}/\mathrm{c}_0^{2l+1}$ and $t_0 = 2a/\mathrm{c}_0$:

$$\left\{\begin{array}{lcl} \frac{G_1}{\pi}\log|\rho_0|^{-1} & \leqslant & k_0 a + \sum_n \mathrm{Im}\left\{\left(\frac{k_0}{k_n}\right)\right\} \\ \frac{G_{l'+1}}{\pi}\log|\rho_0|^{-1} & \leqslant & \frac{1}{2l'+1}\sum_n \mathrm{Im}\left\{\left(\frac{k_0}{k_n}\right)^{2l'+1}\right\},\ l' = 1, 2, \ldots, l-1 \\ \frac{G_{l+1}}{\pi}\log|\rho_0|^{-1} & \leqslant & c_{\tau l}(k_0 a)^{2l+1} + \frac{1}{2l+1}\sum_n \mathrm{Im}\left\{\left(\frac{k_0}{k_n}\right)^{2l+1}\right\} \end{array}\right. \tag{3.12}$$

where $\rho_0$ is defined as in (2.8), $G_l$ is defined by (2.9), $k_0 = \omega_0/\mathrm{c}_0$ and $k_n = z_n/\mathrm{c}_0$ where $z_n$ are the zeros of the reflection coefficient $\rho$, defined as in (2.5) and (2.8). Note that the same relations are obtained by using (2.14) and (2.18) and the substitution $s = -\mathrm{i}k\mathrm{c}_0$.

The narrowband model is now assumed, i.e. let $G_p = B$ in (3.12). Note also that in general, $B \leqslant G_p$. Hence, the assumption $G_p = B$ will simplify the analysis below without loss of generality. Let $k_0/k_n = \alpha_n - \mathrm{i}\beta_n = r_n \mathrm{e}^{-\mathrm{i}\theta_n}$, where $\beta_n > 0$, $r_n > 0$ and $0 < \theta_n < \pi$, and let $f = \frac{B}{\pi}\log|\rho_0|^{-1}$. The optimum solution to the inequalities in (3.12) can then be formulated as the solution to the following constrained optimization problem:

$$\left\{\begin{array}{l} \max f \\ -\sum_n \mathrm{Im}\left\{\alpha_n - \mathrm{i}\beta_n\right\} + f \leqslant k_0 a \\ -\frac{1}{2l'+1}\sum_n \mathrm{Im}\left\{(\alpha_n - \mathrm{i}\beta_n)^{2l'+1}\right\} + f \leqslant 0,\ l' = 1, 2, \ldots, l-1 \\ -\frac{1}{2l+1}\sum_n \mathrm{Im}\left\{(\alpha_n - \mathrm{i}\beta_n)^{2l+1}\right\} + f \leqslant c_{\tau l}(k_0 a)^{2l+1} \\ f \geqslant 0,\ \beta_n \geqslant 0 \end{array}\right. \tag{3.13}$$

where the variables are $(f, \{\alpha_n\}, \{\beta_n\})$. The second constraint above is ignored when $l = 1$. Note that $\beta_n = 0$ is equivalent to removing the corresponding zeros from the summations above.

The solution to the optimization problem (3.13) defines the Fano limit[1] for the reflection coefficient of the spherical waves, i.e. $|\rho_0| \geqslant \rho_{\mathrm{Fano}} = \mathrm{e}^{-\pi f/B}$. When $l = 1$, it is sufficient to use one single zero, and the solution can be uniquely obtained from a $2 \times 2$ non-linear system of equations, see [8]. However, when $l > 1$ the numerical solution to the non-convex optimization problem (3.13) will in general require a global optimization routine and an exhaustive search. Furhermore, for $l > 1$ the optimal number of zeros is not known. A straightforward relaxation of the narrowband Fano equations (3.13) is considered in Section 5 below.

[1]The term Fano limit is used here even though the scattering problem is different from the matching problem. This is motivated by the equivalence of these problems as discussed in this paper.

# 4 Exact circuit analogy for the scattering of a homogeneous sphere

Recursive relationships for the spherical Hankel functions can be used to obtain an exact circuit analogy for the scattering of spherical waves as described below, cf. also [6, 29].

The spherical Hankel functions $\mathrm{h}_l^{(j)}(z)$ satisfy the following initial relations:

$$\begin{cases} \dfrac{(z\mathrm{h}_0^{(1)}(z))'}{z} = \mathrm{i}\mathrm{h}_0^{(1)}(z) \\ -\mathrm{h}_1^{(1)}(z) = \mathrm{i}\mathrm{h}_0^{(1)}(z)\left(1 - \dfrac{1}{\mathrm{i}z}\right), \end{cases} \quad \begin{cases} \dfrac{(z\mathrm{h}_0^{(2)}(z))'}{z} = -\mathrm{i}\mathrm{h}_0^{(2)}(z) \\ -\mathrm{h}_1^{(2)}(z) = -\mathrm{i}\mathrm{h}_0^{(2)}(z)\left(1 + \dfrac{1}{\mathrm{i}z}\right), \end{cases} \tag{4.1}$$

and the recursive relations

$$\begin{cases} \dfrac{(z\mathrm{h}_l^{(j)}(z))'}{z\mathrm{i}^l} = \dfrac{\mathrm{h}_{l-1}^{(j)}(z)}{\mathrm{i}^l} + \dfrac{l}{-\mathrm{i}z}\dfrac{-\mathrm{h}_l^{(j)}(z)}{\mathrm{i}^{l+1}} \\ \dfrac{\mathrm{h}_{l+1}^{(j)}(z)}{\mathrm{i}^{l+2}} = \dfrac{\mathrm{h}_{l-1}^{(j)}(z)}{\mathrm{i}^l} + \dfrac{2l+1}{-\mathrm{i}z}\dfrac{-\mathrm{h}_l^{(j)}(z)}{\mathrm{i}^{l+1}} \end{cases} \tag{4.2}$$

for $j = 1, 2$ and $l = 1, 2, \ldots$, see e.g. [2, 17, 19].

There are two possible dual circuits associated with the recursions in (4.1) and (4.2), cf. Figure 4. For circuit a), define

$$\frac{(z\mathrm{h}_l^{(j)}(z))'}{z\mathrm{i}^l} = \begin{cases} \eta I_l^{(j)}(z) & l = 0, 2, 4, \ldots \\ -V_l^{(j)}(z) & l = 1, 3, 5, \ldots \end{cases} \tag{4.3}$$

and

$$-\frac{\mathrm{h}_l^{(j)}(z)}{\mathrm{i}^{l+1}} = \begin{cases} V_l^{(j)}(z) & l = 0, 2, 4, \ldots \\ -\eta I_l^{(j)}(z) & l = 1, 3, 5, \ldots \end{cases} \tag{4.4}$$

where $V_l(z)$ and $I_l(z)$ represent voltages and currents, respectively.

Let $z = \kappa n$, where $\kappa = ka$ and $n = (\mu\epsilon)^{1/2}$. By introducing the normalized Laplace variable $S = -\mathrm{i}\kappa = sa/\mathrm{c}_0$ and employing the definitions in (4.3) and (4.4), the following initial relations for $V_l(z)$ and $I_l(z)$ corresponding to (4.1) are obtained:

$$\begin{cases} \eta I_0^{(1)}(z) = V_0^{(1)}(z) \\ I_1^{(1)}(z) = I_0^{(1)}(z) + V_0^{(1)}(z)\dfrac{1}{S\mu}, \end{cases} \quad \begin{cases} \eta I_0^{(2)}(z) = -V_0^{(2)}(z) \\ I_1^{(2)}(z) = I_0^{(2)}(z) + V_0^{(2)}(z)\dfrac{1}{S\mu}, \end{cases} \tag{4.5}$$

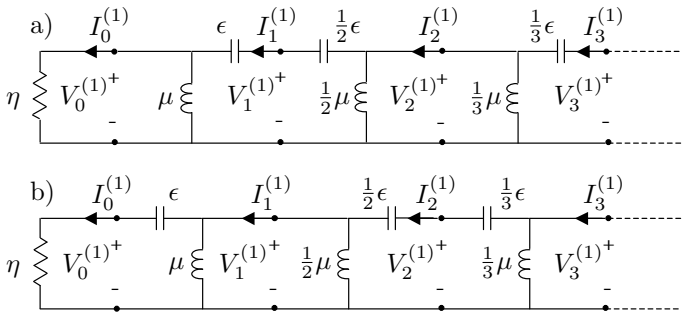



**Figure 4**: The two dual electric circuits with termination representing spherical Hankel functions of the first kind $\mathrm{h}_l^{(1)}(z)$, i.e. outgoing waves outside a sphere of radius $a$.

and the recursive relations corresponding to (4.2) are given by

$$\begin{cases} V_l^{(j)}(z) = V_{l-1}^{(j)}(z) + \dfrac{1}{S\epsilon\frac{1}{l}} I_l^{(j)}(z) & l = 1,3,5,\ldots \\ I_l^{(j)}(z) = I_{l-1}^{(j)}(z) + \dfrac{1}{S\mu\frac{1}{l}} V_l^{(j)}(z) & l = 2,4,6,\ldots \\ V_{l+1}^{(j)}(z) = V_{l-1}^{(j)}(z) + \dfrac{1}{S\epsilon\frac{1}{2l+1}} I_l^{(j)}(z) & l = 1,3,5,\ldots \\ I_{l+1}^{(j)}(z) = I_{l-1}^{(j)}(z) + \dfrac{1}{S\mu\frac{1}{2l+1}} V_l^{(j)}(z) & l = 2,4,6,\ldots \end{cases} \tag{4.6}$$

where $j = 1, 2$. The dual circuit b) is obtained by interchanging $V_l \leftrightarrow \eta I_l$, or equivalently, by simultaneously interchanging $V_l \leftrightarrow I_l$ and $\mu \leftrightarrow \epsilon$.

## 4.1 Exterior of the sphere

Consider now the free space exterior of the sphere where $r \geqslant a$ and $z = \kappa = ka$ ($\mu = \epsilon = \eta = n = 1$). In Figure 4 is shown the two dual electric circuits with termination representing spherical Hankel functions of the first kind $\mathrm{h}_l^{(1)}(z)$, corresponding to outgoing vector spherical waves. There are four different circuits representing the TM and TE waves of odd and even order, as depicted in Figure 5. In Figure 6 is shown the excitation with a Hankel function generator for the two dual electric circuits representing spherical Hankel functions of the second kind $\mathrm{h}_l^{(2)}(z)$, corresponding to incoming vector spherical waves.

From the field definition (3.1) and (3.2) and the circuit (and its dual) definition (4.3) and (4.4), the tangential fields $\boldsymbol{E}_{\mathrm{t},ml}$ and $\boldsymbol{H}_{\mathrm{t},ml}$ (spherical wave indices $m, l$ for $\tau = 1, 2$) outside the sphere are given by

$$\begin{cases} \boldsymbol{E}_{\mathrm{t},ml} = \mp \mathrm{i}^{l+1} \boldsymbol{A}_{1ml} \left( a_{1ml}^{(1)} V_l^{(1)} + a_{1ml}^{(2)} V_l^{(2)} \right) \pm \mathrm{i}^{l} \boldsymbol{A}_{2ml} \left( a_{2ml}^{(1)} V_l^{(1)} + a_{2ml}^{(2)} V_l^{(2)} \right) \\ \eta_0 \boldsymbol{H}_{\mathrm{t},ml} = \mp \mathrm{i}^{l} \boldsymbol{A}_{1ml} \left( a_{2ml}^{(1)} I_l^{(1)} + a_{2ml}^{(2)} I_l^{(2)} \right) \mp \mathrm{i}^{l+1} \boldsymbol{A}_{2ml} \left( a_{1ml}^{(1)} I_l^{(1)} + a_{1ml}^{(2)} I_l^{(2)} \right) \end{cases} \tag{4.7}$$

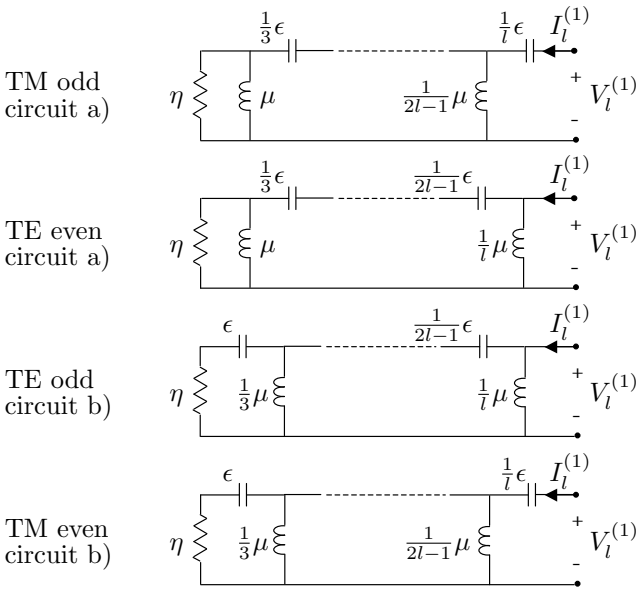



**Figure 5**: Electric circuit analogy for TM and TE waves of odd and even order, corresponding to spherical Hankel functions of the first kind $\mathrm{h}_l^{(1)}(z)$, i.e. outgoing waves outside a sphere of radius $a$.

where the arguments $\boldsymbol{r}$, $\hat{\boldsymbol{r}}$ and $z = \kappa$ have been suppressed for simplicity, and the upper and lower signs refer to even and odd orders, respectively. The normalized TE and TM wave impedances $Z_{\tau l}^{(j)}(z)$ are given by

$$\begin{cases} Z_{1l}^{(j)}(z) = \dfrac{V_l^{(j)}(z)}{I_l^{(j)}(z)} = \mathrm{i}\eta \dfrac{z\mathrm{h}_l^{(j)}(z)}{(z\mathrm{h}_l^{(j)}(z))'} \\ Z_{2l}^{(j)}(z) = \dfrac{V_l^{(j)}(z)}{I_l^{(j)}(z)} = -\mathrm{i}\eta \dfrac{(z\mathrm{h}_l^{(j)}(z))'}{z\mathrm{h}_l^{(j)}(z)} \end{cases} \tag{4.8}$$

where $j = 1, 2$ correspond to the outgoing and incoming waves, respectively.

## 4.2 Interior of the sphere

Next, consider the interior of the sphere where $r \leqslant a$, $z = \kappa n = kan$ and $n = (\mu\epsilon)^{1/2}$. In Figure 7 is shown the two dual electric circuits with termination representing spherical Hankel functions of the second kind $\mathrm{h}_l^{(2)}(z)$, corresponding to incoming vector spherical waves. The circuit definitions (4.3) and (4.4) and recursions (4.5) and (4.6) are the same, but the circuit interpretation is different with an opposite direction for $I_l^{(j)}(z)$ and a sign change of $\mu$ and $\epsilon$. These changes correspond precisely to the symmetry of the incoming and outgoing wave impedances

$$Z_{\tau l}^{(2)}(z) = -Z_{\tau l}^{(1)}(-z) \tag{4.9}$$

defined in (4.8). The four different circuits representing odd and even TM and TE waves in Figure 5 are changed accordingly. In Figure 8 is shown the excitation with

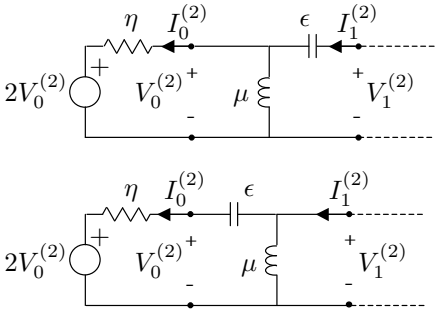



**Figure 6**: Excitation with a Hankel function generator for the two dual electric circuits representing spherical Hankel functions of the second kind $\mathrm{h}_l^{(2)}(z)$, i.e. incoming waves outside a sphere of radius $a$.

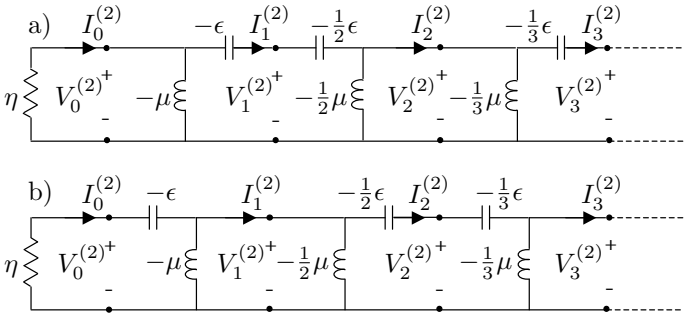



**Figure 7**: The two dual electric circuits with termination representing spherical Hankel functions of the second kind $\mathrm{h}_l^{(2)}(z)$, i.e. incoming waves inside a sphere of radius $a$.

a Hankel function generator for the two dual electric circuits representing spherical Hankel functions of the first kind $\mathrm{h}_l^{(1)}(z)$, corresponding to outgoing vector spherical waves. The circuit elements with impedances $S\mu$ and $1/S\epsilon$ are regarded as "generalized" inductors and capacitors in case the material is dispersive. However, these circuit elements behave asymptotically as "true" inductors and capacitors in the low-frequency limit. Hence, $S\mu \sim S\mu(0)$ and $1/S\epsilon \sim 1/S\epsilon(0)$ when $S \to 0$.

From the field definition (3.6) and (3.7) and the circuit (and its dual) definition (4.3) and (4.4), the tangential fields $\boldsymbol{E}_{\mathrm{t},ml}$ and $\boldsymbol{H}_{\mathrm{t},ml}$ (spherical wave indices $m, l$ for $\tau = 1, 2$) inside the sphere are given by

$$\begin{cases} \boldsymbol{E}_{\mathrm{t},ml} = \mp \mathrm{i}^{l+1} \boldsymbol{A}_{1ml} \dfrac{b_{1ml}}{2} \left( V_l^{(1)} + V_l^{(2)} \right) \pm \mathrm{i}^{l} \boldsymbol{A}_{2ml} \dfrac{b_{2ml}}{2} \left( V_l^{(1)} + V_l^{(2)} \right) \\ \eta_0 \boldsymbol{H}_{\mathrm{t},ml} = \mp \mathrm{i}^{l} \boldsymbol{A}_{1ml} \dfrac{b_{2ml}}{2} \left( I_l^{(1)} + I_l^{(2)} \right) \mp \mathrm{i}^{l+1} \boldsymbol{A}_{2ml} \dfrac{b_{1ml}}{2} \left( I_l^{(1)} + I_l^{(2)} \right) \end{cases} \tag{4.10}$$

where the arguments $\boldsymbol{r}$, $\hat{\boldsymbol{r}}$ and $z = \kappa n$ have been suppressed for simplicity, and the upper and lower signs refer to even and odd orders, respectively. The normalized TE and TM wave impedances $Z_{\tau l}^{(j)}$ are given by (4.8) with $z = \kappa n$.

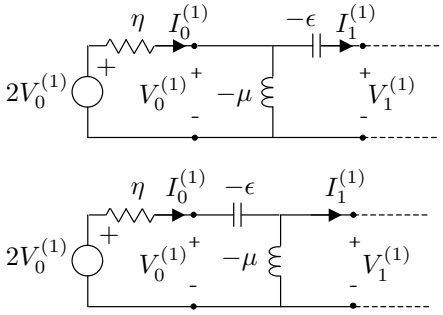



**Figure 8**: Excitation with a Hankel function generator for the two dual electric circuits representing spherical Hankel functions of the first kind $\mathrm{h}_l^{(1)}(z)$, i.e. outgoing waves inside a sphere of radius $a$.

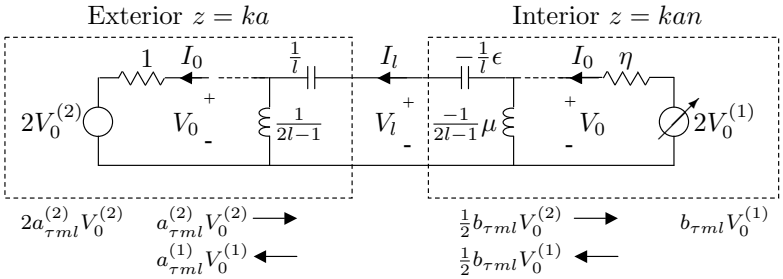



**Figure 9**: Scattering model with Hankel function generators and matching. The interior generator is dependent, creating Bessel functions corresponding to standing waves within the sphere. The circuits are drawn for $\mathrm{TM}_l$ waves. The $\mathrm{TE}_l$ waves are similar.

## 4.3 Exact circuit analogy for the scattering

The scattering problem in Section 3 can now be interpreted by using an exact (equivalent) circuit analogy where the exterior and the interior tangential fields (4.7) and (4.10) are perfectly matched as depicted in Figure 9. An independent exterior generator is used to generate the incoming waves, and a dependent interior generator is used to create the outgoing waves and hence the Bessel functions (obtained as the superposition of the two kinds of Hankel functions) within the sphere, see also [29]. The dependent interior generator and its internal resistance correspond to a reflection coefficient

$$\Gamma(\kappa n) = \frac{V_0^{(1)}(\kappa n)}{V_0^{(2)}(\kappa n)} = (-1)^{\tau+l}\mathrm{e}^{\mathrm{i}2\kappa n}. \tag{4.11}$$

Note that in the equivalent circuit analogy depicted in Figure 9, the voltage and current constituents $V_l^{(j)}$ and $I_l^{(j)}$ with $j = 1, 2$, correspond to a wave splitting with respect to the generator or termination impedance $\eta$, cf. also Figures 6 and 8.

The circuit problem, and hence the scattering problem, has a unique solution

through the scattering (S-matrix) relations

$$\begin{cases} a^{(1)}_{\tau ml} V^{(1)}_0(\kappa) = \rho^{\mathrm{c}}_{1,\tau l} a^{(2)}_{\tau ml} V^{(2)}_0(\kappa) + \varrho^{\mathrm{c}}_{2,\tau l} \frac{1}{2} b_{\tau ml} V^{(1)}_0(\kappa n) \\ \frac{1}{2} b_{\tau ml} V^{(2)}_0(\kappa n) = \varrho^{\mathrm{c}}_{1,\tau l} a^{(2)}_{\tau ml} V^{(2)}_0(\kappa) + \rho^{\mathrm{c}}_{2,\tau l} \frac{1}{2} b_{\tau ml} V^{(1)}_0(\kappa n) \end{cases} \tag{4.12}$$

where $(\rho^{\mathrm{c}}_{1,\tau l}, \varrho^{\mathrm{c}}_{1,\tau l}, \rho^{\mathrm{c}}_{2,\tau l}, \varrho^{\mathrm{c}}_{2,\tau l})$ are the scattering parameters of the equivalent circuit representing the exterior as well as the interior of the sphere. Here, $a^{(2)}_{\tau ml}$ is the amplitude of the incoming wave and (4.12) can be solved for the amplitudes of the outgoing wave $a^{(1)}_{\tau ml}$ and the Bessel function (standing wave) amplitude $b_{\tau ml}$. The overall reflection coefficient $\rho^{\mathrm{c}}_{\tau l}$ for the equivalent circuit is given by

$$\rho^{\mathrm{c}}_{\tau l} = \frac{V^{(1)}_0(\kappa) a^{(1)}_{\tau ml}}{V^{(2)}_0(\kappa) a^{(2)}_{\tau ml}} = (-1)^{\tau+l} \mathrm{e}^{\mathrm{i}2\kappa} \rho_{\tau l} \tag{4.13}$$

where $\rho_{\tau l}$ is the reflection coefficent given by (3.9).

Note that the presence of the negative circuit elements in Figure 9 is consistent with the fact that the wave impedance $Z^{(2)}_{\tau l}$ for incoming waves at $r = a$ is anticausal, cf. (4.8) and (4.9). However, note also that the overall equivalent circuit is causal due to the delay factor in (4.13) above.

The low-frequency asymptotics of the function $-\mathrm{i}\log\left\{(-1)^{\tau+l}\rho^{\mathrm{c}}_{\tau l}\right\}$ corresponding to (4.13) is given by

$$-\mathrm{i}\log\left\{(-1)^{\tau+l}\rho^{\mathrm{c}}_{\tau l}\right\} \sim 2\kappa + 2\kappa^{2l+1} c_{\tau l} \tag{4.14}$$

where $c_{\tau l}$ is given by (3.11). The high-frequency asymptotics is

$$-\mathrm{i}\log\left\{(-1)^{\tau+l}\rho^{\mathrm{c}}_{\tau l}(\kappa)\right\} = -\mathrm{i}\log\left\{\mathrm{e}^{\mathrm{i}2\kappa}\rho_{\tau l}(\kappa)\right\} = \kappa b_1 \mathrm{c}_0/a + o(\kappa), \quad \text{as } \kappa \to \infty,$$

where $b_1 \geqslant 0$. Furthermore, it is expected that $b_1 = 0$ for many material models as discussed in Section 3.2.

Note that the circuit elements corresponding to the interior in Figure 9 behave as $S\mu \sim S\mu(0)$ and $1/S\epsilon \sim 1/S\epsilon(0)$ when $S = -\mathrm{i}\kappa \to 0$. Note also that the low-frequency asymptotics of the TM (TE) reflection coefficient $\rho_{\tau l}$, i.e. the coefficient $c_{\tau l}$, is independent of $\mu(0)$ ($\epsilon(0)$). Hence, when considering the high-contrast limit of the low-frequency asymptotics (4.14) in the TM (TE) case, the limit $\epsilon(0) \to \infty$ ($\mu(0) \to \infty$) may be carried out using $\epsilon(0) = \mu(0) \to \infty$. In this limit, the circuit elements with impedances $S\mu(0)$ and $1/S\epsilon(0)$ behave as open and short circuits, respectively. Further, the low-frequency asymptotics of (4.11) is $\mathit{\Gamma}(\kappa n) \sim (-1)^{\tau+l}$ as $\kappa \to 0$. Hence, the high-contrast limit of the low-frequency asymptotics in (4.14) may be obtained equivalently by using the exterior circuit with open or short termination as depicted in Figure 10. This means that the low-frequency asymptotics of $\rho_{1,\tau l}$ according to the conjecture (2.17) and (2.18) is identical to (4.14) with $-\mathrm{i}\kappa = s\frac{a}{\mathrm{c}_0}$, and hence

$$-\log\left\{(-1)^{\tau+l}\rho_{1,\tau l}\right\} \sim 2\frac{a}{\mathrm{c}_0}s + 2(-1)^l(\frac{a}{\mathrm{c}_0})^{2l+1} c_{\tau l} s^{2l+1} \tag{4.15}$$

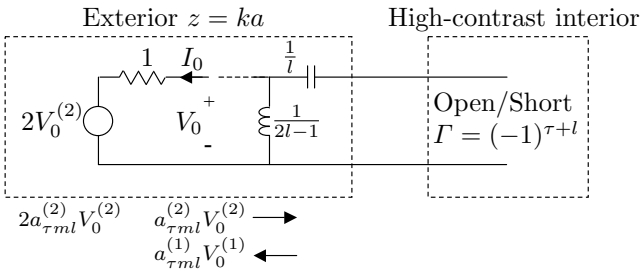



**Figure 10**: Interpretation of the scattering model in the high-contrast limit, $\epsilon(0) = \mu(0) \to \infty$. The exterior circuit is drawn for $\mathrm{TM}_l$ waves. The $\mathrm{TE}_l$ waves are similar.

where

$$c_{\tau l} = \frac{2^{2l}(l+1)!(l-1)!}{(2l+1)!(2l)!} \tag{4.16}$$

is the high-contrast limit of (3.11) when $\nu_\tau(0) \to \infty$. Note that the exterior circuit has a transmission zero of order $l+1$ at $S = 0$ and the term $2\kappa^{2l+1}c_{\tau l}$ of the reflection coefficient $\rho_{1,\tau l}$ is therefore invariant to whether the circuit is terminated with a short, open or match, cf. (2.15). Note also the interesting distinguishing feature that the integral identity (2.14) contains no causality term $t_0$ as in (2.7), instead this term $t_0 = 2a/c_0$ appears in the low-frequency asymptotics of $-\log\{\rho_{1,\tau l}\}$ as $A_1^{(0)}$ in (2.17) and (4.15).

In conclusion, the optimal Fano matching problem for the exterior circuit as described in Section 2.3 is equivalent to the problem of determining the optimal limitations for scattering of spherical waves in the high-contrast limit as described in Section 3.3. An exact expression for the low-frequency asymptotics of $-\log\{\rho_{1,\tau l}\}$ is given by (4.15) and (4.16). The exact expression agree perfectly with the numerical results given in [30].

# 5 Relaxation of the Fano equations

As a numerical example, a relaxation of the narrowband Fano equations (3.13) is considered below. To solve (3.13) for $l \geqslant 2$, one has to resort to global optimization and computationally expensive numerical experiments. Hence, a straightforward relaxation yielding an upper bound on the objective function $f$ may be useful.

In order to relax the constraints in (3.13), consider the minimization of the expression $-\frac{1}{2l+1}\sum_n \mathrm{Im}\left\{(\alpha_n - \mathrm{i}\beta_n)^{2l+1}\right\} = \frac{1}{2l+1}\sum_n \beta_n^{2l+1}\frac{\sin(\theta_n(2l+1))}{\sin^{2l+1}\theta_n}$ when $\beta_n$ is fixed. This implies the stationarity condition $\frac{\partial}{\partial\alpha_n}\mathrm{Im}\left\{(\alpha_n - \mathrm{i}\beta_n)^{2l+1}\right\} = 0$ yielding the solutions, $(r_n, \theta_n) = \left(\frac{\beta_n}{\sin(m\frac{\pi}{2l})}, m\frac{\pi}{2l}\right)$ where $m = 1, \ldots, 2l-1$. Hence, by choosing

$$-d_l = \min_{1 \leqslant m \leqslant 2l-1} \frac{1}{2l+1}\frac{\sin(m\frac{\pi}{2l}(2l+1))}{\sin^{2l+1}(m\frac{\pi}{2l})} \tag{5.1}$$

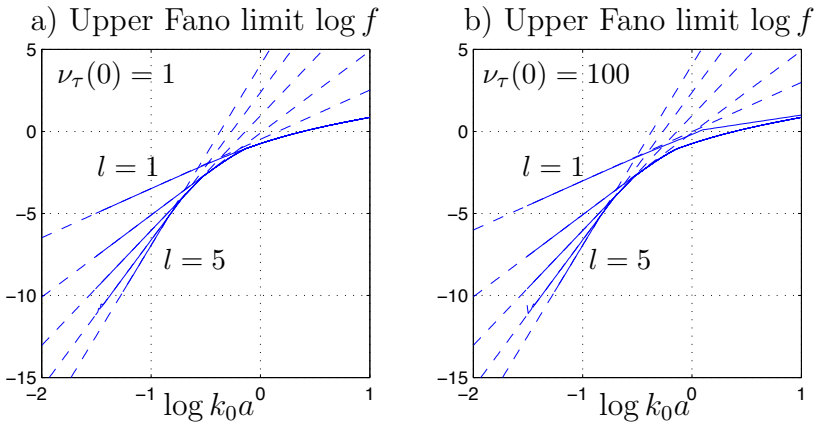



**Figure 11**: Upper Fano limit $f$ as a function of $k_0a$ for $l = 1, 2, \ldots, 5$. Graphs a and b show $\log f$ for $\nu_\tau(0) = 1$ and $\nu_\tau(0) = 100$, respectively. The dashed lines show the asymptotic upper bounds $(d_l + c_{\tau l})(k_0a)^{2l+1}$ in the narrowband approximation where $G_l = B$.

where $d_l > 0$, and by employing $\left(\sum_{n=1}^{N} \beta_n\right)^{2l+1} \geqslant \sum_{n=1}^{N} \beta_n^{2l+1}$, a relaxation of (3.13) valid for all $N$ is given by

$$\begin{cases} \max f \\ \beta + f \leqslant k_0a \\ -d_{l'}\beta^{2l'+1} + f \leqslant 0, \ l' = 1, \ldots, l-1 \\ -d_l\beta^{2l+1} + f \leqslant c_{\tau l}(k_0a)^{2l+1} \\ f \geqslant 0, \ \beta \geqslant 0 \end{cases} \tag{5.2}$$

where there are two variables $(f, \beta)$. The solution to (5.2) yields an upper bound for the corresponding Fano limit in the variable $f$. Hence, $|\rho_0| \geqslant \rho_{\text{Fano}} \geqslant \mathrm{e}^{-\pi f/B}$. When $l = 1$, the relaxation becomes tight and the solution to (5.2) is identical to the Fano limit ($l = 1 \Rightarrow \theta_n = \pi/2$). Furthermore, for $l = 1$ there is a transition point where the second constraint becomes inactive and hence $f = k_0a$ for $k_0a \geqslant \sqrt{1/c_{\tau 1}}$. To solve (5.2) for $l \geqslant 2$, it is noted that the first (linear) constraint is always active. Since the polynomial constraints are monotonic in $\beta$ for $\beta \geqslant 0$, the optimum solution is found as the minimum of $f$ over the $l$ constraint subsets corresponding to a $2 \times 2$ non-linear system of equations containing the first linear constraint $\beta + f = k_0a$. Note that each such constraint subset has a unique solution for $\beta \geqslant 0$. The asymptotic solution to (5.2) when $k_0a \to 0$ is given by

$$f = (d_l + c_{\tau l})(k_0a)^{2l+1} + \mathcal{O}\left((k_0a)^{2l+3}\right). \tag{5.3}$$

For $l \geqslant 2$, the asymptotic solution to (5.2) when $k_0a \to \infty$ is governed by the lowest index $l' = 1$ and is hence given by the solution to the first two constraints, i.e. the real-valued root of $f = (k_0a - f)^3/3$. In Figure 11 is shown the upper Fano limit $f$ as a function of $k_0a$ for $l = 1, 2, \ldots, 5$ and $\nu_\tau(0) = 1, 100$, respectively.

# 6 Summary

Optimal limitations for the scattering of vector spherical waves are considered where the geometry of the object is known but the temporal dispersion is unknown. Using integral relations similar to the derivation of Fano's broadband matching bounds, the optimal scattering limitations are determined by the static response as well as the high-frequency asymptotics of the reflection coefficient. Using an exact circuit analogy for the scattering of spherical waves, it is shown how the problem of determining the optimal scattering bounds for a homogeneous sphere in its high-contrast limit becomes identical to the closely related, and yet very different problem of finding the broadband tuning limits of the spherical waves. Furthermore, the scattering view of the matching problem yields explicitly the necessary low frequency asymptotics of the reflection coefficient that is used with Fano's broadband matching bounds for spherical waves, something that appears to be non-trivial to derive from the classical network point of view.

As with the Fano approach, the integral relations yield a non-convex global optimization problem which in general is quite difficult to handle. As a numerical example, a relaxation of the Fano equations is considered which is easily solved and which is especially useful in the regime of Rayleigh scattering.

# Acknowledgments

The financial support by the High Speed Wireless Communications Center of the Swedish Foundation for Strategic Research (SSF) is gratefully acknowledged.

# Appendix A High-frequency asymptotics of scattering coefficients

To find the dominant behavior of the reflection coefficients $\rho_{\tau l}$ in (3.9) for high frequencies, the asymptotic behavior of the spherical Bessel and Hankel functions are needed. For large arguments the spherical Hankel functions behave as [1]

$$\begin{cases} \mathrm{h}_l^{(1)}(z) = (-\mathrm{i})^{l+1}\dfrac{\mathrm{e}^{\mathrm{i}z}}{z}\left(1+\mathrm{i}\dfrac{a_l}{z}-\dfrac{b_l}{z^2}+\mathcal{O}(z^{-3})\right) \\ \mathrm{h}_l^{(2)}(z) = \mathrm{i}^{l+1}\dfrac{\mathrm{e}^{-\mathrm{i}z}}{z}\left(1-\mathrm{i}\dfrac{a_l}{z}-\dfrac{b_l}{z^2}+\mathcal{O}(z^{-3})\right) \\ \left(z\mathrm{h}_l^{(1)}(z)\right)' = (-\mathrm{i})^{l+1}\mathrm{e}^{\mathrm{i}z}\left(\mathrm{i}-\dfrac{a_l}{z}-\mathrm{i}\dfrac{a_l+b_l}{z^2}+\mathcal{O}(z^{-3})\right) \\ \left(z\mathrm{h}_l^{(2)}(z)\right)' = \mathrm{i}^{l+1}\mathrm{e}^{-\mathrm{i}z}\left(-\mathrm{i}-\dfrac{a_l}{z}+\mathrm{i}\dfrac{a_l+b_l}{z^2}+\mathcal{O}(z^{-3})\right) \end{cases} \tag{A.1}$$

as $z \to \infty$, where $z$ is complex, $a_l = (l+1)l/2$ and $b_l = (l+2)(l+1)l(l-1)/8$, and the big-O notation $\mathcal{O}$ is defined as in [22]. Moreover, as $z \to \infty$ the spherical Bessel

functions behave as [1]

$$\begin{cases} \mathrm{j}_l(z) = \dfrac{1}{z}\left\{\left(1 - \dfrac{b_l}{z^2} + \mathcal{O}(z^{-4})\right)\sin\left(z - \dfrac{l\pi}{2}\right) + \left(\dfrac{a_l}{z} + \mathcal{O}(z^{-3})\right)\cos\left(z - \dfrac{l\pi}{2}\right)\right\} \\ (z\mathrm{j}_l(z))' = \left(1 - \dfrac{a_l + b_l}{z^2} + \mathcal{O}(z^{-4})\right)\cos\left(z - \dfrac{l\pi}{2}\right) - \left(\dfrac{a_l}{z} + \mathcal{O}(z^{-3})\right)\sin\left(z - \dfrac{l\pi}{2}\right). \end{cases} \tag{A.2}$$

To find the high-frequency behavior of (3.9), special care must be taken to separate the exponential behavior of $ka$ and the algebraic behavior of $ka$. To this end, expand the material parameters as a power series at infinity, i.e.

$$\begin{cases} \nu_\tau = \alpha_0 + \dfrac{\mathrm{i}\alpha_1}{\kappa} + \dfrac{\alpha_2}{\kappa^2} + \mathcal{O}\left(\kappa^{-3}\right) \\ \kappa n = \kappa\left(\beta_0 + \frac{\mathrm{i}\beta_1}{\kappa} + \frac{\beta_2}{\kappa^2} + \frac{\mathrm{i}\beta_3}{\kappa^3} + \mathcal{O}\left(\kappa^{-4}\right)\right) = \beta_0\kappa + \mathrm{i}\beta_1 + \frac{\beta_2}{\kappa} + \frac{\mathrm{i}\beta_3}{\kappa^2} + \mathcal{O}\left(\kappa^{-3}\right) \end{cases} \tag{A.3}$$

where $\kappa = ka$, $\kappa \to \infty$, and where $\alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}$ and $\beta_0, \beta_1, \beta_2, \beta_3 \in \mathbb{R}$. The last power series includes the Debye and the Lorentz dispersion models [19]. In particular, the Debye dispersion model (with real and positive parameters $\epsilon_\infty$, $\epsilon_\mathrm{s}$ and $\tau$) is given by

$$\epsilon(\kappa) = \epsilon_\infty + \frac{\epsilon_\mathrm{s} - \epsilon_\infty}{1 - \mathrm{i}\kappa\tau} = \epsilon_\infty + \mathrm{i}\frac{\epsilon_\mathrm{s} - \epsilon_\infty}{\kappa\tau} + \frac{\epsilon_\mathrm{s} - \epsilon_\infty}{\kappa^2\tau^2} + \mathcal{O}(\kappa^{-3}) \tag{A.4}$$

and the Lorentz dispersion model (with real and positive parameters $\epsilon_\infty$, $\kappa_\mathrm{p}$, $\kappa_0$ and $\varsigma$)

$$\epsilon(\kappa) = \epsilon_\infty - \frac{\kappa_\mathrm{p}^2}{\kappa^2 - \kappa_0^2 + \mathrm{i}\kappa\varsigma} = \epsilon_\infty - \frac{\kappa_\mathrm{p}^2}{\kappa^2} + \mathrm{i}\frac{\kappa_\mathrm{p}^2\varsigma}{\kappa^3} + \mathcal{O}(\kappa^{-4}) \tag{A.5}$$

as $\kappa \to \infty$, which also motivates the assumption of real coefficients in the expansion. If $\alpha_1$ or $\beta_1$ is non-zero, then $\nu_\tau$ corresponds effectively to a Debye model or a conductivity model. If both are zero, the model is of Lorentz' type. These expansions imply

$$\begin{cases} \sin\left(\kappa n - \dfrac{l\pi}{2}\right) &= \sin\left(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1 + \frac{\beta_2}{\kappa} + \frac{\mathrm{i}\beta_3}{\kappa^2} + \mathcal{O}(\kappa^{-3})\right) \\ &= A\sin\left(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1\right) + B\cos\left(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1\right) \\ \cos\left(\kappa n - \dfrac{l\pi}{2}\right) &= \cos\left(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1 + \frac{\beta_2}{\kappa} + \frac{\mathrm{i}\beta_3}{\kappa^2} + \mathcal{O}(\kappa^{-3})\right) \\ &= A\cos\left(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1\right) - B\sin\left(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1\right) \end{cases} \tag{A.6}$$

where

$$\begin{cases} A = \cos\left(\frac{\beta_2}{\kappa} + \frac{\mathrm{i}\beta_3}{\kappa^2} + \mathcal{O}(\kappa^{-3})\right) = 1 - \frac{\beta_2^2}{2\kappa^2} + \mathcal{O}(\kappa^{-3}) \\ B = \sin\left(\frac{\beta_2}{\kappa} + \frac{\mathrm{i}\beta_3}{\kappa^2} + \mathcal{O}(\kappa^{-3})\right) = \frac{\beta_2}{\kappa} + \frac{\mathrm{i}\beta_3}{\kappa^2} + \mathcal{O}(\kappa^{-3}). \end{cases} \tag{A.7}$$

The quantities $\rho_{\tau l}$ are now studied. Introduce the appropriate numerator $N_l$ and denominator $D_l$ such that

$$\rho_{\tau l} = -\mathrm{e}^{-2\mathrm{i}\kappa}(-1)^{l+1}\rho_{\tau l}^\mathrm{c} = \mathrm{e}^{-2\mathrm{i}(\kappa - l\pi/2)}\rho_{\tau l}^\mathrm{c} = \mathrm{e}^{-2\mathrm{i}(\kappa - l\pi/2)}\frac{N_l}{D_l} \tag{A.8}$$

where the numerator $N_l$ is

$$N_l = \left(1 - \mathrm{i}\frac{a_l}{\kappa} - \frac{b_l}{\kappa^2} + \mathcal{O}(\kappa^{-3})\right)(\kappa n \mathrm{j}_l(\kappa n))' - \nu_\tau \kappa \left(-\mathrm{i} - \frac{a_l}{\kappa} + \mathrm{i}\frac{a_l + b_l}{\kappa^2} + \mathcal{O}(\kappa^{-3})\right)\mathrm{j}_l(\kappa n) \quad \text{(A.9)}$$

and the denominator is

$$D_l = \left(1 + \mathrm{i}\frac{a_l}{\kappa} - \frac{b_l}{\kappa^2} + \mathcal{O}(\kappa^{-3})\right)(\kappa n \mathrm{j}_l(\kappa n))' - \nu_\tau \kappa \left(\mathrm{i} - \frac{a_l}{\kappa} - \mathrm{i}\frac{a_l + b_l}{\kappa^2} + \mathcal{O}(\kappa^{-3})\right)\mathrm{j}_l(\kappa n). \quad \text{(A.10)}$$

Moreover, as $ka = \kappa \to \infty$ the power series expansions defined above yield after some algebra

$$\begin{cases} N_l &= \left(1 + \mathrm{i}\frac{\alpha_0(a_l+\beta_0\beta_2)-a_l\beta_0^2}{\beta_0^2\kappa} + \mathcal{O}(\kappa^{-2})\right)\cos(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1) \\ &\quad + \mathrm{i}\left(\frac{\alpha_0}{\beta_0} - \mathrm{i}\frac{(\alpha_0-1)a_l\beta_0+\alpha_0\beta_1-\beta_0(\alpha_1+\beta_0\beta_2)}{\beta_0^2\kappa} + \mathcal{O}(\kappa^{-2})\right)\sin(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1) \\ D_l &= \left(1 - \mathrm{i}\frac{\alpha_0(a_l+\beta_0\beta_2)-a_l\beta_0^2}{\beta_0^2\kappa} + \mathcal{O}(\kappa^{-2})\right)\cos(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1) \\ &\quad - \mathrm{i}\left(\frac{\alpha_0}{\beta_0} + \mathrm{i}\frac{(\alpha_0-1)a_l\beta_0-\alpha_0\beta_1+\beta_0(\alpha_1-\beta_0\beta_2)}{\beta_0^2\kappa} + \mathcal{O}(\kappa^{-2})\right)\sin(\beta_0\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1). \end{cases} \quad \text{(A.11)}$$

For simplicity, assume that there is no optical response i.e. $\alpha_0 = \beta_0 = 1$. Then (A.11) implies as $\kappa \to \infty$

$$\begin{cases} N_l &= \mathrm{e}^{\mathrm{i}(\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1)}\left(1 + \mathrm{i}\beta_2\kappa^{-1} + \mathcal{O}(\kappa^{-2})\right) \\ &\quad + \left((\beta_1 - \alpha_1)\kappa^{-1} + \mathcal{O}(\kappa^{-2})\right)\sin(\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1) \\ D_l &= \mathrm{e}^{-\mathrm{i}(\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1)}\left(1 - \mathrm{i}\beta_2\kappa^{-1} + \mathcal{O}(\kappa^{-2})\right) \\ &\quad - \left((\beta_1 - \alpha_1)\kappa^{-1} + \mathcal{O}(\kappa^{-2})\right)\sin(\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1). \end{cases} \quad \text{(A.12)}$$

Along the real axis all the exponential terms contribute, and the quotient is

$$\rho^{\mathrm{c}}_{\tau l} = \mathrm{e}^{\mathrm{i}2(\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1)}\left(1 + \mathrm{i}2\beta_2\kappa^{-1} + (\beta_1 - \alpha_1)\kappa^{-1}\sin(2\kappa - l\pi + \mathrm{i}2\beta_1) + \mathcal{O}(\kappa^{-2})\right). \quad \text{(A.13)}$$

In the upper half-plane as $\kappa \hat{\to} \infty$, the term $\mathrm{e}^{\mathrm{i}2\kappa}$ is exponentially small and the main contribution comes from terms of the form $\mathrm{e}^{-\mathrm{i}2\kappa}$. Therefore, the dominant contribution is given by

$$\begin{aligned} \rho^{\mathrm{c}}_{\tau l} = \mathrm{e}^{\mathrm{i}2(\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1)}\big[&1 + \mathrm{i}2\beta_2\kappa^{-1} + (\beta_1 - \alpha_1)\kappa^{-1}\sin(2\kappa - l\pi + \mathrm{i}2\beta_1) \\ &+ \mathcal{O}(\kappa^{-2})\sin(\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1)\mathrm{e}^{-\mathrm{i}(\kappa - \frac{l\pi}{2} + \mathrm{i}\beta_1)} + \mathcal{O}(\kappa^{-2})\big] \\ &= \mathrm{i}\frac{\beta_1 - \alpha_1}{2ka} + \mathcal{O}((ka)^{-2}) \end{aligned} \quad \text{(A.14)}$$

where $\kappa = ka$ has been inserted.

# References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Applied Mathematics Series No. 55. National Bureau of Standards, Washington D.C., 1970.

[2] G. B. Arfken and H. J. Weber. *Mathematical Methods for Physicists*. Academic Press, New York, fifth edition, 2001.

[3] A. Bernland, M. Gustafsson, and S. Nordebo. Physical limitations on the scattering of electromagnetic vector spherical waves. *J. Phys. A: Math. Theor.*, **44**(14), 145401, 2011.

[4] A. Bernland, A. Luger, and M. Gustafsson. Sum rules and constraints on passive systems. *J. Phys. A: Math. Theor.*, **44**(14), 145205, 2011.

[5] C. R. Brewitt-Taylor. Limitation on the bandwidth of artificial perfect magnetic conductor surfaces. *Microwaves, Antennas & Propagation, IET*, **1**(1), 255–260, 2007.

[6] L. J. Chu. Physical limitations of omni-directional antennas. *J. Appl. Phys.*, **19**, 1163–1175, 1948.

[7] P. L. Duren. *Theory of $H^p$ Spaces*. Dover Publications, New York, 2000.

[8] R. M. Fano. Theoretical limitations on the broadband matching of arbitrary impedances. *Journal of the Franklin Institute*, **249**(1,2), 57–83 and 139–154, 1950.

[9] W. Geyi. Physical limitations of antenna. *IEEE Trans. Antennas Propagat.*, **51**(8), 2116–2123, August 2003.

[10] M. Gustafsson and D. Sjöberg. Sum rules and physical bounds on passive metamaterials. *New Journal of Physics*, **12**, 043046, 2010.

[11] M. Gustafsson, C. Sohl, and G. Kristensson. Physical limitations on antennas of arbitrary shape. *Proc. R. Soc. A*, **463**, 2589–2607, 2007.

[12] M. Gustafsson, C. Sohl, and G. Kristensson. Illustrations of new physical bounds on linearly polarized antennas. *IEEE Trans. Antennas Propagat.*, **57**(5), 1319–1327, May 2009.

[13] M. Gustafsson. Sum rule for the transmission cross section of apertures in thin opaque screens. *Opt. Lett.*, **34**(13), 2003–2005, 2009.

[14] M. Gustafsson and S. Nordebo. Bandwidth, Q factor, and resonance models of antennas. *Progress in Electromagnetics Research*, **62**, 1–20, 2006.

[15] M. Gustafsson and S. Nordebo. Characterization of MIMO antennas using spherical vector waves. *IEEE Trans. Antennas Propagat.*, **54**(9), 2679–2682, 2006.

[16] M. Gustafsson, C. Sohl, C. Larsson, and D. Sjöberg. Physical bounds on the all-spectrum transmission through periodic arrays. *EPL Europhysics Letters*, **87**(3), 34002 (6pp), 2009.

[17] J. E. Hansen, editor. *Spherical Near-Field Antenna Measurements*. Number 26 in IEE electromagnetic waves series. Peter Peregrinus Ltd., Stevenage, UK, 1988. ISBN: 0-86341-110-X.

[18] A. Hujanen, J. Holmberg, and J. C.-E. Sten. Bandwidth limitations of impedance matched ideal dipoles. *IEEE Trans. Antennas Propagat.*, **53**(10), 3236–3239, 2005.

[19] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, New York, third edition, 1999.

[20] B. Kogan. Comments on "Broadband matching limitations for higher order spherical modes". *IEEE Trans. Antennas Propagat.*, **58**(5), 1826, 2010.

[21] H. M. Nussenzveig. *Causality and dispersion relations*. Academic Press, London, 1972.

[22] F. W. J. Olver. *Asymptotics and special functions*. A K Peters, Ltd, Natick, Massachusetts, 1997.

[23] A. Papoulis. *The Fourier integral and its applications*. McGraw-Hill, Inc., New York, 1962.

[24] K. N. Rozanov. Ultimate thickness to bandwidth ratio of radar absorbers. *IEEE Trans. Antennas Propagat.*, **48**(8), 1230–1234, August 2000.

[25] C. Sohl and M. Gustafsson. A priori estimates on the partial realized gain of Ultra-Wideband (UWB) antennas. *Quart. J. Mech. Appl. Math.*, **61**(3), 415–430, 2008.

[26] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on broadband scattering by heterogeneous obstacles. *J. Phys. A: Math. Theor.*, **40**, 11165–11182, 2007.

[27] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on metamaterials: Restrictions on scattering and absorption over a frequency interval. *J. Phys. D: Applied Phys.*, **40**, 7146–7151, 2007.

[28] J. A. Stratton. *Electromagnetic Theory*. McGraw-Hill, New York, 1941.

[29] H. L. Thal. Exact circuit analysis of spherical waves. *IEEE Trans. Antennas Propagat.*, **26**(2), 282–287, March 1978.

[30] M. C. Villalobos, H. D. Foltz, and J. S. McLean. Broadband matching limitations for higher order spherical modes. *IEEE Trans. Antennas Propagat.*, **57**(4), 1018–1026, 2009.

[31] M. C. Villalobos, H. D. Foltz, and J. S. McLean. Reply to comments on "Broadband matching limitations for higher order spherical modes". *IEEE Trans. Antennas Propagat.*, **58**(5), 1827–1827, 2010.

[32] M. C. Villalobos, H. D. Foltz, J. S. McLean, and I. S. Gupta. Broadband tuning limits on UWB antennas based on Fano's formulation. In *Proceedings of Antennas and Propagation Society International Symposium 2006*, pages 171–174, 2006.

[33] J. Volakis, C. C. Chen, and K. Fujimoto. *Small Antennas: Miniaturization Techniques & Applications*. McGraw-Hill, New York, 2010.

[34] H. A. Wheeler. The wide-band matching area for a small antenna. *IEEE Trans. Antennas Propagat.*, **31**, 364–367, 1983.

[35] M. Wohlers and E. Beltrami. Distribution theory as the basis of generalized passive-network analysis. *IEEE Transactions on Circuit Theory*, **12**(2), 164–170, 1965.

[36] A. D. Yaghjian and S. R. Best. Impedance, bandwidth, and $Q$ of antennas. *IEEE Trans. Antennas Propagat.*, **53**(4), 1298–1324, 2005.

[37] D. Youla, L. Castriota, and H. Carlin. Bounded real scattering matrices and the foundations of linear passive network theory. *IRE Transactions on Circuit Theory*, **6**(1), 102–124, 1959.

[38] A. H. Zemanian. *Distribution theory and transform analysis: an introduction to generalized functions, with applications*. McGraw-Hill, New York, 1965.

# Bandwidth Limitations for Scattering of Higher Order Electromagnetic Spherical Waves with Implications for the Antenna Scattering Matrix

**Paper IV**

Anders Bernland

**Based on:** A. Bernland. Bandwidth limitations for scattering of higher order electromagnetic spherical waves with implications for the antenna scattering matrix. *IEEE Transactions on Antennas and Propagation*, accepted for publication, April 2012.

**Abstract**

Various physical limitations in electromagnetic theory and antenna theory have received considerable attention recently. However, there are no previous limitations on the scattering of higher order electromagnetic vector spherical waves, despite the widespread use of spherical wave decompositions. In the present paper, bandwidth limitations on the scattering matrix are derived for a wide class of heterogeneous objects, in terms of their electrical size, shape and static material properties. In particular, it is seen that the order of the dominating term in the Rayleigh limit increases with the order of the spherical wave. Furthermore, it is shown how the limitations place bounds on the antenna scattering matrix, thus introducing a new approach to physical limitations on antennas. Comparisons to other types of antenna limitations are given, and numerical simulations for two folded spherical helix antennas and a directive Yagi-Uda antenna are included to illuminate and validate the theory. The results in this paper are derived using a general approach to derive limitations for passive systems: First, the low-frequency asymptotic expansion of the scattering matrix of a general scatterer is derived. This gives a set of sum rules, from which the limitations follow.

# 1 Introduction

Scattering of electromagnetic waves is essential to a wide range of applications, from classical optics to wireless communication and radar. In many cases it is beneficial to decompose the fields in electromagnetic vector spherical waves [28] (also referred to as partial waves, TM- and TE-modes or electric and magnetic multipoles). For instance, spherical waves are used for analysing scattering by spherical particles (Mie theory) [6], in Waterman's $T$-matrix method [27], in antenna measurements [19], and recently also for modelling wireless communication channels [12].

In the last few years, there has been an interest in physical limitations for electromagnetic scattering; several interesting attempts have been made to quantify the intuitively obvious statement that objects which are small compared to the wavelength can only provide limited interaction with electromagnetic waves [35]. Specific issues addressed are e.g. radar absorbers [32], high-impedance surfaces [8, 18] and metamaterials [13]. Various antenna limitations have received considerable attention recently (a review can be found in the book by Volakis et al. [39]). Despite the widespread use of spherical wave decompositions, however, there are no previous limitations on higher order spherical wave scattering.

The main results of the present paper are improved limitations for scattering of higher order electromagnetic vector spherical waves (quadrupoles, octopoles and so forth), originally derived for the dipole case in [3]. The limitations imply that the diagonal elements of the scattering matrix, which relate the coefficients of the incoming and outgoing waves, cannot be arbitrarily small over a whole wavelength interval; the bounds depend on the fractional bandwidth as well as the size, shape and static material properties of the scatterer.

The results of this paper pave the way for a new approach to physical limitations on antennas. In form, the sum rules and limitations derived here are similar to those from optimal broadband matching of the spherical waves. Matching of ideal dipoles, the lowest order spherical waves, was considered by Hujanen et al. [21], while higher order waves were treated by Villalobos et al. [38], Nordebo et al. [29], as well as Kogan (see [24] and references therein). One advantage that follows from the approach adopted in the present paper is that the shape and static material properties of the antenna are highlighted. Many previous publications on antenna limitations were concerned with the quality factors ($Q$-factors) of the spherical waves, but it is in general not straightforward to relate the $Q$-factor to the operating bandwidth of the antenna [16, 36]. Recently, a different method, based on sum rules for the extinction cross section, has been proposed by Gustafsson et al. [14, 15, 34]. Unfortunately, these results cannot handle antennas placed in a dielectric background or spherical wave decompositions. Spherical waves are a useful tool e.g. for analysing multiple-input multiple-output (MIMO) antenna systems [12, 17, 26].

The derivations in this paper follow a general approach to achieve sum rules and physical limitations for passive systems presented in [4], cf. also [3, 29]. It relies on the well-known connection between passive systems and Herglotz (or positive real) functions [43, 44] in conjunction with a set of integral identities for that class of functions. To use the approach, an intermediate result needs to be derived in the paper: the low-frequency asymptotic expansion of the scattering matrix of a general scatterer.

The outline of the paper is the following: Section 2 introduces the scattering and transition matrices as well as the electromagnetic vector spherical waves. Their low-frequency asymptotic expansions and static counterparts are also covered. The sum rules and limitations for the scattering matrix are derived in Section 3. Implications for the antenna scattering matrix are given in Section 4, and the results are compared to other types of antenna limitations. Simulation results for two folded spherical helix antennas and a directive Yagi-Uda antenna are also presented. Section 5 concludes the paper.

# 2 The scattering and transition matrices

This section presents the scattering problem considered in this paper. It is described in terms of vector spherical waves and the scattering and transition matrices, which are introduced in Section 2.1. Time-harmonic fields and sources are considered throughout this paper, and the time convention $\mathrm{e}^{-\mathrm{i}\omega t}$, where i is the imaginary unit and $\omega$ is the angular frequency, is used. The low-frequency and static cases, which are essential to the further analysis in later sections, are treated in Section 2.2.

## 2.1 Scattering geometry

Consider an uncharged scatterer in free space. Let the scatterer be contained in a hypothetical sphere of radius $a$, centered at the origin, as in Figure 1. The electric

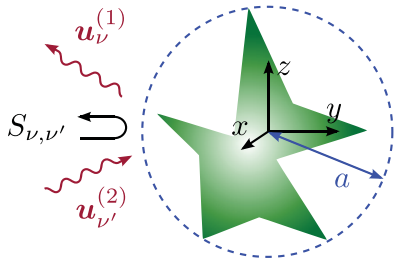



**Figure 1**: The scatterer is placed in free space and circumscribed by a hypothetical sphere of radius $a$, centered at the origin. The electric and magnetic fields are written as sums of incoming ($\boldsymbol{u}_\nu^{(2)}$) and outgoing ($\boldsymbol{u}_\nu^{(1)}$) electromagnetic vector spherical waves, with index $\nu$, outside the sphere. The scattering matrix $\mathbf{S}$, with elements $S_{\nu,\nu'}$, relates the coefficients of the incoming and outgoing waves.

and magnetic fields can be written as sums of incoming ($\boldsymbol{u}^{(2)}$) and outgoing ($\boldsymbol{u}^{(1)}$) vector spherical waves outside the circumscribing sphere [28]. The scattering matrix $\mathbf{S}$ relates the coefficients of the incoming and outgoing waves, and is thus a measure on the incoming power that is rejected by the scatterer.

The spherical wave decomposition of the electric and magnetic fields outside the circumscribing sphere is [28]

$$\boldsymbol{E}(\boldsymbol{r},k) = k\sqrt{2\eta_0}\sum_\nu b_\nu^{(1)}\boldsymbol{u}_\nu^{(1)}(k\boldsymbol{r}) + b_\nu^{(2)}\boldsymbol{u}_\nu^{(2)}(k\boldsymbol{r}). \tag{2.1}$$

Here the free space parameters are wavenumber $k = \omega/\mathrm{c}$, speed of light c and impedance $\eta_0$. The spatial coordinate is denoted $\boldsymbol{r}$, with $r = |\boldsymbol{r}|$ and $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$. The vector spherical waves are defined as in [7], see Appendix A. The multi-index $\nu = 2(l^2 + l - 1 + (-1)^s m) + \tau$ is introduced in place of the indices $\{\tau, s, m, l\}$ to simplify the notation. It is defined so that $\tau = 1$ (odd $\nu$) corresponds to a magnetic $2^l$-pole (TE$_l$-mode), while $\tau = 2$ (even $\nu$) identifies an electric $2^l$-pole (TM$_l$-mode). Hence, $l = 1$ denotes dipoles, $l = 2$ quadrupoles, and so on. The corresponding magnetic field is $\boldsymbol{H}(\boldsymbol{r},k) = \frac{1}{\mathrm{i}k\eta_0}\nabla \times \boldsymbol{E}(\boldsymbol{r},k)$. With this normalization, the time-average of the power passing out through a sphere of radius $r > a$ is

$$\langle P(t)\rangle = \int_{\Omega_{\hat{\boldsymbol{r}}}} \hat{\boldsymbol{r}} \cdot \mathrm{Re}\left(\frac{1}{2}\boldsymbol{E}(\boldsymbol{r},k) \times \boldsymbol{H}^*(\boldsymbol{r},k)\right) r^2 \,\mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \sum_\nu |b_\nu^{(1)}|^2 - |b_\nu^{(2)}|^2, \tag{2.2}$$

where $\Omega_{\hat{\boldsymbol{r}}} = \{(\theta,\phi) : 0 \leqslant \theta < \pi, 0 \leqslant \phi < 2\pi\}$ is the unit sphere and $\mathrm{d}\Omega_{\hat{\boldsymbol{r}}} = \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi$.

Alternatively, the fields can be decomposed into outgoing and regular waves $\boldsymbol{v}_\nu$:

$$\boldsymbol{E}(\boldsymbol{r},k) = k\sqrt{2\eta_0}\sum_\nu d_\nu^{(1)}\boldsymbol{u}_\nu^{(1)}(k\boldsymbol{r}) + d_\nu^{(2)}\boldsymbol{v}_\nu(k\boldsymbol{r}). \tag{2.3}$$

An incident field is regular at the origin, and so constitutes the sum over the regular waves, while the scattered field makes up the sum over the outgoing waves in (2.3).

The infinite dimensional scattering matrix $\mathbf{S}$ relates the coefficients in (2.1): $b_\nu^{(1)} = \sum_{\nu'} S_{\nu,\nu'} b_{\nu'}^{(2)}$. The counterpart for (2.3) is the transition matrix $\mathbf{T}$: $d_\nu^{(1)} = \sum_{\nu'} T_{\nu,\nu'} d_{\nu'}^{(2)}$. The scattering and transition matrices are related as $\mathbf{S} = 2\mathbf{T} + \mathbf{I}$, where $\mathbf{I}$ is the infinite dimensional identity matrix. Note that it has now been implicitly assumed that the constitutive relations of the scatterer are in convolution form in the time domain [3]. The convolution form assumption is closely related to the assumptions of linearity and time-translational invariance [44], and is commonly used.

The main results of this paper are limitations for the diagonal elements $S_{\nu,\nu}$ of the scattering matrix. A general approach to derive sum rules and physical limitations for passive systems presented in [4] is used, cf. also [3, 29]. In order to use it, expressions for the low-frequency asymptotic expansions of the scattering and transition matrix elements are required, and this is the topic of Section 2.2.

## 2.2 Low-frequency asymptotics and statics

The low-frequency and static transition matrices have been considered by a number of authors. Peterson [31] introduced the transition matrices of the static field problem, and noted that the problem is the low-frequency limit of the dynamic scattering problem. Waterman showed how the electric and magnetic components decouple in the static limit [42], and Olsson treated the elastodynamic case similarly [30]. Recently, Waterman has derived expressions for the low-frequency electromagnetic transition matrix in two dimensions [40] and the acoustic counterpart in three dimensions [41]. A review of results on low-frequency approximations until 2006 can be found in the book by Martin [25]. However, none of these previous publications provides the necessary expressions for the scattering problem considered here.

To be able to derive the required low-frequency asymptotic expansions of the scattering and transition matrix elements, consider a static electric field $\boldsymbol{E}$. The electric field is given by the electrostatic potential, $\boldsymbol{E} = -\nabla\phi$, and the potential can be expanded in scalar spherical harmonics $Y_{sml}$ (defined in (A.3)) [28, 42]:

$$\phi(\boldsymbol{r}) = \sum_{l=0}^{\infty} \sum_{m=0}^{l} \sum_{s=1}^{2} f_{sml}^{(1)} r^{-l-1} Y_{sml}(\hat{\boldsymbol{r}}) - f_{sml}^{(2)} r^{l} Y_{sml}(\hat{\boldsymbol{r}}). \tag{2.4}$$

In this case, the electrostatic transition matrix $\mathbf{T}^{[2]}$ relates the coefficients: $f_{sml}^{(1)} = \sum_{s'm'l'} T_{sml,s'm'l'}^{[2]} f_{s'm'l'}^{(2)}$, cf. [31, 42]. The magnetostatic transition matrix $\mathbf{T}^{[1]}$ is defined analogously; since $\nabla \times \boldsymbol{H} = \boldsymbol{0}$ outside the circumscribing sphere, a magnetostatic scalar potential $\phi^{\mathrm{MS}}$ can be defined there such that $\boldsymbol{H} = -\nabla\phi^{\mathrm{MS}}$. Note that an applied external potential is regular at the origin and constitutes the sum over the terms $f_{sml}^{(2)} r^l Y_{sml}(\hat{\boldsymbol{r}})$, while the scattered potential decays at infinity and thus is given by the terms $f_{sml}^{(1)} r^{-l-1} Y_{sml}(\hat{\boldsymbol{r}})$.

To obtain expressions for the low-frequency expansions of the electrodynamic transition matrix, consider the asymptotic expansions of the spherical waves (which follow from the asymptotic expansions for the spherical Bessel and Hankel functions

[2] appearing in their definitions, and some algebra, see Appendix A.1):

$$\begin{aligned}\boldsymbol{v}_{2sml} &= \frac{2^l(l+1)!}{(2l+1)!}\frac{\nabla(r^l Y_{sml}(\hat{\boldsymbol{r}}))}{\sqrt{l(l+1)}}k^{l-1} + \mathcal{O}(k^{l+1}) \\ \boldsymbol{u}^{(1)}_{2sml} &= \frac{\mathrm{i}(2l)!}{2^l(l-1)!}\frac{\nabla(r^{-l-1} Y_{sml}(\hat{\boldsymbol{r}}))}{\sqrt{l(l+1)}}k^{-l-2} + \mathcal{O}(k^{-l}),\end{aligned} \tag{2.5}$$

while $\boldsymbol{v}_{1sml} = \mathcal{O}(k^l)$ and $\boldsymbol{u}^{(1)}_{1sml} = \mathcal{O}(k^{-l-1})$, as $k \to 0$. For the electric case ($\tau = 2$), combining (2.5) with (2.3) and (2.4) readily yields

$$T_{\nu,\nu'} = \frac{\mathrm{i}2^{l+l'}(l-1)!(l'+1)!\sqrt{l(l+1)}}{(2l)!(2l'+1)!\sqrt{l'(l'+1)}}\delta_{\tau,\tau'}T^{[\tau]}_{sml,s'm'l'}k^{l+l'+1} + \mathcal{O}(k^{l+l'+3}) \quad \text{as } k \to 0. \tag{2.6}$$

Here $\delta_{\tau,\tau'}$ is the Kronecker delta. The same equation holds also for the magnetic case ($\tau = 1$), which is seen by also making use of $\boldsymbol{H} = \frac{1}{\mathrm{i}k\eta_0}\nabla \times \boldsymbol{E}$ and (A.1). Recall that the multi-index $\nu$ represents the indices $\{\tau, s, m, l\}$.

Equation (2.6), which is needed in the following section in order to derive the limitations for the scattering matrix, cannot be found in any previous publication. The equation explicitly shows how the electrostatic and magnetostatic transition matrices are the low-frequency limits of the electrodynamic counterpart, cf. [42]. Consequently, the static transition matrices are crucial to the limitations. For dipoles ($l = l' = 1$), the elements of $\mathbf{T}^{[2]}$ and $\mathbf{T}^{[1]}$ for an uncharged body are (apart from normalization) equal to the elements of the well-studied static electric and magnetic polarizability dyadics, defined in [23]. The elements of $\mathbf{T}^{[2]}$ and $\mathbf{T}^{[1]}$ for higher order spherical waves can be seen as generalizations of the polarizability dyadics in spherical coordinates, see Appendix B for details.

# 3 Sum rules and limitations for the scattering matrix

The limitations on the scattering matrix, which are the main results of the paper, are derived in this section. First, in Section 3.1, it is shown that the low-frequency expansion (2.6) implies that a set of sum rules, or integral identities, apply. The sum rules, in turn, are used to obtain the limitations, or inequalities. Similarly as in the case of optimal broadband matching [29, 38], the limitations presented in this paper make up an optimization problem. Its solution is discussed briefly. After that, physical interpretations are given in Section 3.2. Further discussion on interpretations for antennas is given later in Section 4.

## 3.1 Results

As mentioned in the introduction, the derivations in the present paper rely on a general approach presented in [4] for deriving sum rules and limitations on passive

systems. The approach relies on the connection between passive systems and Herglotz (or positive real) functions [43, 44], and it can be used here since $\mathrm{e}^{\mathrm{i}2ka}S_{\nu,\nu'}(k)$ is a passive reflection coefficient corresponding to a real-valued and causal convolution kernel, under the assumption that the material of the scatterer is passive [3]. The limitations for scattering of dipoles were derived previously in [3], whereas [29] derives matching limitations by relating the matching problem to scattering of spherical waves by a high-contrast sphere (with infinite static relative permeability and permittivity). Both these references contain more mathematical background, and the interested reader is referred there.

The following low-frequency expansion of the diagonal elements of the scattering matrix is required:

$$-\mathrm{i}\log(\mathrm{e}^{\mathrm{i}2ka}S_{\nu,\nu}) = 2ka + 2k^{2l+1}c_l T^{[\tau]}_{sml,sml} + \mathcal{O}(k^{2l+3}), \quad \text{as } k \to 0, \tag{3.1}$$

where $c_l = [2^{2l}(l+1)!(l-1)!]/[(2l+1)!(2l)!]$ is a constant. The equation (3.1) is a straightforward consequence of the low-frequency expansion (2.6) for the transition matrix $\mathbf{T}$, the relation $\mathbf{S} = 2\mathbf{T} + \mathbf{I}$ and the asymptotic expansion $\log(1+z) = z + \mathcal{O}(z^2)$ as $z \to 0$. The off-diagonal elements of $\mathbf{S}$ tend to zero as $k \to 0$, and so the logarithms of them are not well-behaved in the low-frequency limit. For this reason, only the diagonal elements are considered from now on.

Following (3.1), $l+1$ sum rules can be derived [3]:

$$\begin{cases} \frac{1}{\pi}\int_0^\infty \frac{1}{k^2}\ln\frac{1}{|S_{\nu,\nu}(k)|}\,\mathrm{d}k & = a - \frac{\beta_{\nu,\nu}}{2} + \sum_n \mathrm{Im}\,\frac{1}{k_n} \\ \frac{1}{\pi}\int_0^\infty \frac{1}{k^{2p}}\ln\frac{1}{|S_{\nu,\nu}(k)|}\,\mathrm{d}k & = \frac{1}{2p-1}\sum_n \mathrm{Im}\,\frac{1}{k_n^{2p-1}}, \quad \text{for } p = 2,3,\ldots,l \\ \frac{1}{\pi}\int_0^\infty \frac{1}{k^{2l+2}}\ln\frac{1}{|S_{\nu,\nu}(k)|}\,\mathrm{d}k & = c_l T^{[\tau]}_{sml,sml} + \frac{1}{2l+1}\sum_n \mathrm{Im}\,\frac{1}{k_n^{2l+1}}, \end{cases} \tag{3.2}$$

where $k_n$ are the zeros of $S_{\nu,\nu}(k)$ in the open upper half of the complex plane ($\mathrm{Im}\,k > 0$). The parameter $\beta_{\nu,\nu} \geqslant 0$ is expected to be zero if the circumscribing sphere is chosen as small as possible [29]. Note the close likeness to Fano's matching equations [11]. In [3], the asymptotic expansion (3.1) was only derived to order $k^3$, and hence only 2 sum rules were available in (3.2). Ref. [29] used the expansion (3.1) to order $k^{2l+1}$, but only for the simple case of an isotropic sphere.

To derive limitations, consider a finite wavenumber interval $[k_0(1-B/2), k_0(1+B/2)]$, where $k_0$ is the center wavenumber and $B$ the relative bandwidth. Denote $S_{0,\nu} = \max_{[k_0(1-B/2),k_0(1+B/2)]}|S_{\nu,\nu}(k)|$. The sum rules then give $l+1$ limitations:

$$\begin{cases} \frac{G_1(B)\ln S_{0,\nu}^{-1}}{\pi} \leqslant k_0 a + \sum_n \mathrm{Im}\,\frac{k_0}{k_n} \\ \frac{G_p(B)\ln S_{0,\nu}^{-1}}{\pi} \leqslant \frac{1}{2p-1}\sum_n \mathrm{Im}\left(\frac{k_0}{k_n}\right)^{2p-1}, \text{ for } p = 2,3,\ldots,l \\ \frac{G_{l+1}(B)\ln S_{0,\nu}^{-1}}{\pi} \leqslant k_0^{2l+1} c_l T^{[\tau]}_{sml,sml} + \frac{1}{2l+1}\sum_n \mathrm{Im}\left(\frac{k_0}{k_n}\right)^{2l+1}, \end{cases} \tag{3.3}$$

where the bandwidth factor $G_p(B)$ for $p = 1, 2, \ldots$ is defined by

$$G_p(B) = \int_{1-B/2}^{1+B/2} \frac{1}{x^{2p}}\,\mathrm{d}x = \frac{1}{2p-1}\frac{(1+B/2)^{2p-1} - (1-B/2)^{2p-1}}{(1-B^2/4)^{2p-1}}.$$

Note that $G_p(B) \approx B$ in the narrowband approximation where $B \ll 1$. Furthermore, seeing that $B \leqslant G_p(B)$ for all $0 \leqslant B \leqslant 2$, the inequalities in (3.3) are valid with $G_p(B)$ replaced by $B$.

Although they place bounds on the scattering matrix rather than the mismatch, the limitations (3.3) are in form similar to the limitations on optimal wideband matching presented in [29, 38]. Note, however, that the last right-hand side differs: It includes an element of a static transition matrix $T^{[\tau]}_{sml,sml}$, which describes the shape and static material properties of the scatterer. The limitations (3.3) coincide with the corresponding limitations in [29, 38] in the simple case when the scatterer is a high-contrast sphere, since then $T^{[\tau]}_{sml,sml} = a^{2l+1}$ [37]. This fact was also noted in [29].

The system of inequalities suffer from a drawback: they incorporate the unknown zeros $k_n$ of $S_{\nu,\nu}(k)$. However, limitations not containing the zeros can be derived by solving the constrained optimization problem given by (3.3), so that

$$\frac{B \ln S_{0,\nu}^{-1}}{\pi} \leqslant f_\nu(T^{[\tau]}_{sml,sml}; k_0 a), \tag{3.4}$$

where $f_\nu(T^{[\tau]}_{sml,sml}; k_0 a)$ is the solution to (3.3). For the dipole case ($l = 1$), it is sufficient to consider a single complex zero $k_n$, which gives the closed form solution [3]:

$$f_\nu(T^{[\tau]}_{sm1,sm1}; k_0 a) = \begin{cases} k_0 a - \sqrt[3]{\iota + \xi} + \sqrt[3]{\iota - \xi} & \text{for } k_0 a \leqslant \sqrt{\frac{a^3}{c_1 T^{[\tau]}_{sm1,sm1}}} \\ k_0 a & \text{otherwise,} \end{cases}$$

where $\xi = 3(k_0 a - k_0^3 c_1 T^{[\tau]}_{sm1,sm1})/2$ and $\iota = \sqrt{1 + \xi^2}$. For higher order waves, it has been conjectured that $l$ complex zeros are sufficient to obtain an optimal solution [38]. The computationally expensive numerical problem is solved by Villalobos et al. in [38]. Alternatively, Kogan has shown that the solution can be found by solving a polynomial equation of order $2l + 1$, see [24]. However, upper bounds on $f_\nu(T^{[\tau]}_{sml,sml}; k_0 a)$ can be derived by considering a single complex zero also for higher order waves, which gives a problem that is straightforward to solve numerically [29]. For this reason, this procedure is chosen in this paper, and the results can be found in Figure 2. The dominating term for small $k_0 a$ (Rayleigh scattering) is [29]:

$$f_\nu(T^{[\tau]}_{sml,sml}; k_0 a) = \left( g_l + \frac{c_l T^{[\tau]}_{sml,sml}}{a^{2l+1}} \right) k_0^{2l+1} a^{2l+1} + \mathcal{O}(k_0^{2l+3} a^{2l+3}), \tag{3.5}$$

as $k_0 a \to 0$, where the term $g_l > 0$ is given by

$$g_l = - \min_{1 \leqslant m \leqslant 2l-1} \frac{1}{2l+1} \frac{\sin\left(m\pi \frac{2l+1}{2l}\right)}{\sin^{2l+1}\left(m\pi \frac{1}{2l}\right)}.$$

Equation (3.5) shows that the order of the dominating term in the Rayleigh regime increases with the order of the spherical wave, something that is also evident from the tangentials of the curves in Figure 2.

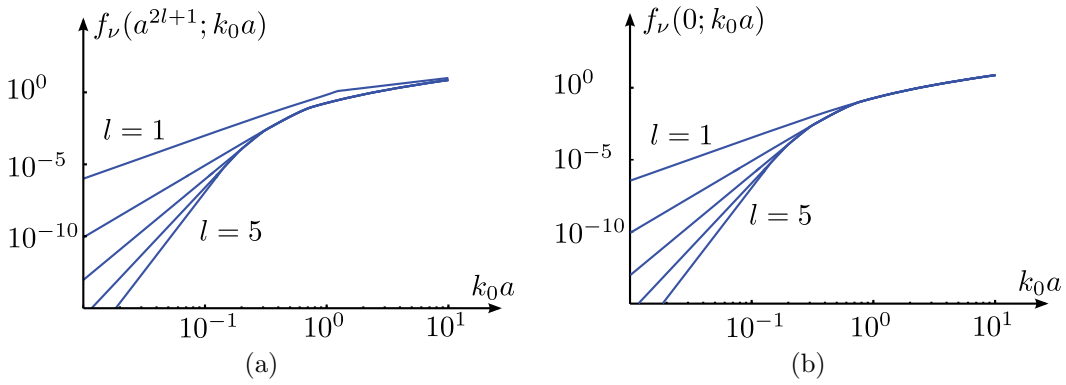



**Figure 2**: Upper bounds, $B \ln S_{0,\nu}^{-1}/\pi \leqslant f_\nu(T^{[\tau]}_{sml,sml}; k_0 a)$, for $l = 1 \ldots 5$. The bounds are for (a): $T^{[\tau]}_{sml,sml} = a^{2l+1}$ (high-contrast sphere) and (b): $T^{[\tau]}_{sml,sml} = 0$.

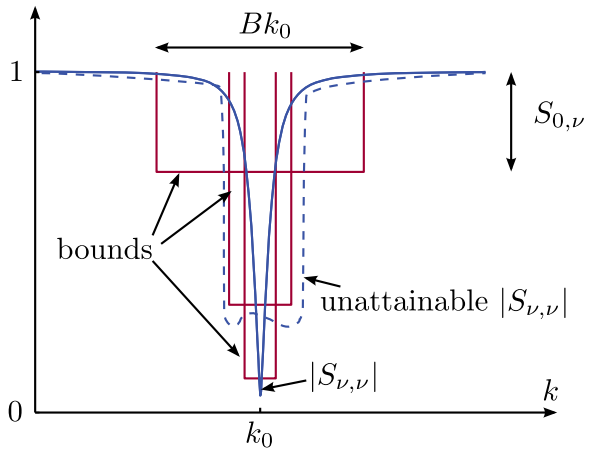



**Figure 3**: Interpretation of the limitations (3.4). Bounds for a given wave index $\nu$ and center wavenumber $k_0$ are shown for three different values of $S_{0,\nu}$ (and thus three different values of $B$). The limitations state that $|S_{\nu,\nu}(k)|$ have to intersect the boxes. The figure also shows one attainable and one unattainable element $S_{\nu,\nu}$.

## 3.2 Physical interpretations

The limitations (3.4) imply that the moduli of the scattering matrix elements $S_{\nu,\nu}$ cannot be arbitrarily small over a whole wavelength interval, see Figure 3. How small they can be is determined by the relative bandwidth $B$, as well as the electrical size of the scatterer (center wavenumber $k_0$ times radius $a$ of the circumscribing sphere) and its shape and static material properties (described by the static transition matrix elements $T^{[\tau]}$). Alternatively, any chosen value of $S_{0,\nu} \in [0, 1]$ determines how large the fractional bandwidth $B$ may be.

The absorption efficiency

$$\eta_\nu(k) = 1 - \sum_{\nu'} |S_{\nu',\nu}(k)|^2 \leqslant 1 - |S_{\nu,\nu}(k)|^2 \tag{3.6}$$

is the relative power of the incoming spherical wave with index $\nu$ that is absorbed

by the scatterer [3]. Recall that the off-diagonal terms $S_{\nu',\nu}(k)$ tend to zero as $k \to 0$. The limitations (3.3) imply that $\eta_\nu(k)$ cannot be arbitrarily high over a whole wavelength interval:

$$\min_{[k_0(1-B/2),k_0(1+B/2)]} \eta_\nu(k) \leqslant 1 - \mathrm{e}^{-2\pi f_\nu(T^{[\tau]}_{sml,sml};k_0a)/B}. \tag{3.7}$$

Many applications concerned with electromagnetic scattering can make use of the limitations (3.4) and (3.7). An example of the limitations (3.7) for the dipole case applied to nanoshells can be found in Section 5.1 in [3].

The static transition matrix elements $T^{[\tau]}_{sml,sml}$ are well understood for dipoles ($l = 1$), see [3] and references therein. The higher order static transition matrix elements are not as well-studied; there are, however, a few previous publications, see [25, 31, 40, 42]. A couple of general remarks can also be made: Firstly, note that the magnetostatic transition matrix $\mathbf{T}^{[1]}$ vanishes when the scatterer is non-magnetic. This gives the upper bounds $f_\nu(0; k_0a)$ in Figure 2b. Secondly, variational principles put forth by Sjöberg in [33] show that $T^{[\tau]}_{sm1,sm1}$ ($l = 1$) is bounded from above by its value for the high-contrast sphere, i.e. $T^{[\tau]}_{sm1,sm1} \leqslant a^3$. If the same holds also for higher order modes, namely that $T^{[\tau]}_{sml,sml} \leqslant a^{2l+1}$, it means that the upper bounds $f_\nu(a^{2l+1}; k_0a)$ in Figure 2a are absolute upper bounds. Also recall that for a high contrast sphere, the scattering matrix limitations (3.3) are identical to the broadband matching limitations in [29, 38].

# 4 Interpretations for antennas

Since the limitations (3.4) can be interpreted as bounds on the absorption of power from each spherical wave, they are well suited to study antennas. More precisely, the limitations have implications for the *antenna scattering matrix*, defined in [19]. This is explained in Section 4.1. Furthermore, it was also mentioned above that the limitations are similar in form to the broadband matching limitations presented in [29, 38]. The interpretations, however, are different, as discussed in Section 4.2. Comparisons to $Q$-factor and gain-bandwidth limitations are given in Section 4.3. Finally, simulation results for two folded spherical helix antennas (one linearly polarized and the other elliptically polarized) and a directive Yagi-Uda antenna are presented in Section 4.4.

## 4.1 Limitations on the antenna scattering matrix

Consider an antenna as in Figure 4, connected to a local port through a matching network. The antenna scattering matrix $\mathbf{S}^{\mathrm{A}}$ completely describes the antenna properties:

$$\underbrace{\begin{pmatrix} \Gamma & \mathbf{R}^{\mathrm{A}} \\ \mathbf{T}^{\mathrm{A}} & \mathbf{S} \end{pmatrix}}_{\mathbf{S}^{\mathrm{A}}} \begin{pmatrix} w^{(2)} \\ \boldsymbol{b}^{(2)} \end{pmatrix} = \begin{pmatrix} w^{(1)} \\ \boldsymbol{b}^{(1)} \end{pmatrix}, \tag{4.1}$$

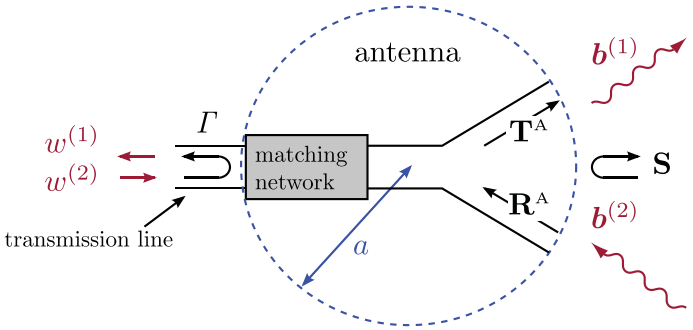



**Figure 4**: The antenna scattering matrix $\mathbf{S}^{\mathrm{A}}$ in (4.1) completely describes the antenna properties.

Here $\boldsymbol{b}^{(1)} = (b_1^{(1)}\ b_2^{(1)}\ \ldots)^{\mathrm{T}}$ and $w^{(1)}$ are the coefficients of the outgoing waves and received signal, respectively, whereas $\boldsymbol{b}^{(2)} = (b_1^{(2)}\ b_2^{(2)}\ \ldots)^{\mathrm{T}}$ and $w^{(2)}$ are the coefficients of the incoming waves and transmitted signal. The signals are normalized so that their power content is $|w^{(2)}|^2$ and $|w^{(1)}|^2$, respectively. Recall that the spherical waves are normalized similarly, see (2.2). Apart from the scattering matrix $\mathbf{S}$, which describes the scattering properties of the antenna, the antenna scattering matrix $\mathbf{S}^{\mathrm{A}}$ also incorporates the antenna reflection coefficient $\Gamma$ as well as the transmitting coefficients $T_\nu^{\mathrm{A}}$ in $\mathbf{T}^{\mathrm{A}}$ and the receiving coefficients $R_\nu^{\mathrm{A}}$ in $\mathbf{R}^{\mathrm{A}}$ [19]. If the alternative decomposition in (2.3) is used instead of (2.1), equation (4.1) becomes [19]:

$$\begin{pmatrix} \Gamma & \frac{1}{2}\mathbf{R}^{\mathrm{A}} \\ \mathbf{T}^{\mathrm{A}} & \frac{1}{2}(\mathbf{S}-\mathbf{I}) \end{pmatrix} \begin{pmatrix} w^{(2)} \\ \boldsymbol{d}^{(2)} \end{pmatrix} = \begin{pmatrix} w^{(1)} \\ \boldsymbol{d}^{(1)} \end{pmatrix}. \tag{4.2}$$

This is beneficial for use in numerical simulations, see Section 4.4. The antenna scattering matrix can also be generalized for multi-port antennas [17].

The limitations (3.4) place bounds on the antenna scattering matrix. The receiving coefficients $R_\nu^{\mathrm{A}}$ are evidently bounded by the absorption efficiency $\eta_\nu$, defined in (3.6):

$$|R_\nu^{\mathrm{A}}| \leqslant \eta_\nu \leqslant 1 - |S_{\nu,\nu}|^2. \tag{4.3}$$

The first inequality is an equality for lossless antennas. Consequently, from (3.7) it follows that

$$\min_{[k_0(1-B/2),k_0(1+B/2)]} |R_\nu^{\mathrm{A}}(k)| \leqslant 1 - \mathrm{e}^{-2\pi f_\nu(T^{[\tau]}_{sml,sml};k_0a)/B}. \tag{4.4}$$

For reciprocal antennas, the transmitting and receiving coefficients are related as $R_\nu^{\mathrm{A}} = (-1)^s T_\nu^{\mathrm{A}}$ [17] (recall the indices $\{\tau, s, m, l\}$, see Appendix A), and therefore (4.4) applies also with $R_\nu^{\mathrm{A}}$ replaced by $T_\nu^{\mathrm{A}}$ in this case.

Consequently, there is an upper bound on the maximum achievable bandwidth of an antenna when it is receiving (or transmitting) a certain spherical wave. As discussed in Section 3.2, the bound depends on the electrical size of the antenna as well as its shape and static material properties. Furthermore, due to (3.5) it is clear that it is increasingly harder to take advantage of the higher order spherical waves for an electrically small antenna. This is also known previously due to the high reactive energies associated with higher order spherical waves [10].

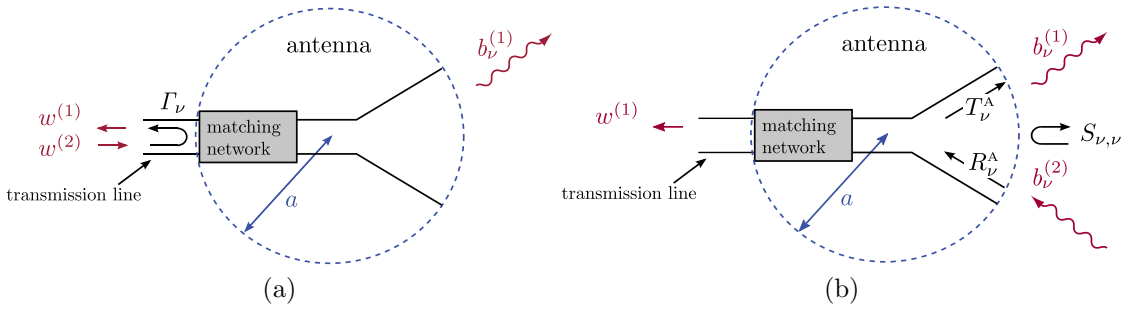



**Figure 5**: Comparison of the broadband matching limitations presented in [29, 38] and the scattering limitations presented in this paper. (a): The matching limitations place a lower bound on the antenna reflection coefficient $\Gamma_\nu$ when the antenna is transmitting a certain spherical wave with index $\nu$. (b): The scattering limitations place lower bounds on the scattering coefficients $S_{\nu,\nu}$ and upper bounds on the antenna receiving and transmitting coefficients $R_\nu^{\mathrm{A}}$ and $T_\nu^{\mathrm{A}}$ due to (4.3).

## 4.2 Comparison to broadband matching limitations

The matching limitations in [29, 38] also place upper bounds on the maximum achievable bandwidth of an antenna receiving or transmitting a certain spherical wave. However, they are not directly comparable to the limitations presented in this paper: The matching limitations place a lower bound on the antenna reflection coefficient $\Gamma_\nu$ when the antenna is transmitting a certain spherical wave with index $\nu$, see Figure 5a. Recall that the limitations (3.4) in this paper instead place lower bounds on the scattering coefficients $S_{\nu,\nu}$ and upper bounds on the antenna receiving and transmitting coefficients $R_\nu^{\mathrm{A}}$ and $T_\nu^{\mathrm{A}}$ due to (4.3), see Figure 5b.

One advantage thanks to the approach chosen in the present paper is that the derived limitations highlight the shape and static material properties of the antenna, and not just its electrical size as in [29, 38]. This can lead to sharper bounds in some cases. If, for instance, the antenna is non-magnetic, the bounds for the magnetic spherical waves (TE-modes) are sharpened since the magnetostatic transition matrix $\mathbf{T}^{[1]}$ vanishes. Recall, though, that the limitations (3.4) coincide with the corresponding limitations in [29, 38] for the simple case of a high contrast sphere. Another advantage of the scattering approach to antenna limitations is that it is directly applicable to other areas concerned with electromagnetic scattering, as discussed in Section 3.2 and [3].

## 4.3 Comparisons to $Q$-factor and gain-bandwidth limitations

It is hard to make a direct comparison between the scattering (or matching) limitations for spherical waves with other bounds on antennas, but it is still worthwhile to make a consistency check. The various approaches reach different conclusions, and should therefore be considered as complementary rather than in competition; there is not one approach that reaches the best result for every case.

Lower bounds on the $Q$-factor were first presented by Chu in [9], and closed form expressions for higher order waves were derived by Collin and Rothschild, see equation (10) in [10]. Even though there is no general relationship between $Q$-factor and bandwidth, for many antennas it can be argued that $B \ln \Gamma_0^{-1}/\pi \leqslant 1/Q$, where $\Gamma_0 = \max_{[k_0(1-B/2),k_0(1+B/2)]} |\Gamma(k)|$ [16]. Hence the upper bounds on $1/Q$ in [10] are on equal footing to the upper bounds on $f_\nu$ in (3.4) and Figure 2. A closer comparison reveals that the numerical values are comparable for a high-contrast sphere, and the $Q$-bounds also show the same asymptotic behaviour for small $k_0 a$. For $l = 1$, the results are almost identical, whereas the $Q$-bounds are better for $l > 1$; this is probably due to the simplified optimization procedure adopted in this paper, see Section 3.1. However, the results in [10] do not take shape and material properties into account, as do (3.4).

Bounds on gain and bandwidth were derived by Gustafsson et al. using sum rules for the extinction cross section, see equation (3.4) in [14]. Inserting the directivity of a spherical wave ($D = 1.5$ for $l = 1$, $D = 2.5$ for $l = 2$, and so fort) yields an upper bound on $B(1 - \Gamma_0^2)$, which can be compared to (3.4). The numerical values are comparable for electric dipoles ($\tau = 2, l = 1$) and non-magnetic materials; the bounds in (3.4) are slightly sharper for narrow bandwidths, and the other way around for wider bandwidths. The results in [14] *do* take shape and material properties into account (in terms of the static polarizability dyadics), and provide sharper bounds for non-spherical circumscribing geometries. However, the results presented in this paper provide sharper bounds for the case of electric dipoles *with* magnetic materials, magnetic dipoles *without* magnetic materials, as well as for higher order waves ($l > 1$).

## 4.4 Numerical examples

To illustrate the limitations (3.4) and (4.4), two folded spherical helix antenna designs proposed by Best [5] have been considered. These designs were chosen since their quality factors are close to the Chu-bound [5]. Both antennas fit into a sphere of radius $a = 4.18$ cm. The first design is linearly polarized, and it turns out that the spherical wave with multi-index $\nu = 4$ (i.e. $\{\tau, s, m, l\} = \{2, 2, 0, 1\}$) is dominant. This corresponds to an electric dipole ($\mathrm{TM}_1$-mode) in the $z-$direction. The antenna geometry, scattering matrix element $S_{4,4}$, reflection coefficient $\Gamma$, and transmitting and receiving coefficients $T_4^{\mathrm{A}}$ and $R_4^{\mathrm{A}}$ are depicted in Figure 6 along with the limitations. The second design is elliptically polarized, and it radiates two spherical waves: the electric dipole with multi-index $\nu = 4$, and the magnetic dipole ($\mathrm{TE}_1$-mode) in the $z-$direction with multi-index $\nu = 3$ (i.e. $\{\tau, s, m, l\} = \{1, 2, 0, 1\}$). The results can be found in Figure 7. Note that the limitations are sharper for the magnetic dipole, since the magnetostatic transition matrix $T^{[1]}$ vanishes for a non-magnetic antenna. It can be seen that both spherical helices approach the limitations (3.4) and (4.4).

An antenna with directivity greater than 3 must have a radiation pattern that includes spherical waves of orders higher than dipoles, since an antenna radiating only dipole modes must have directivity $D \leqslant 3$ [20]. As a consequence of the

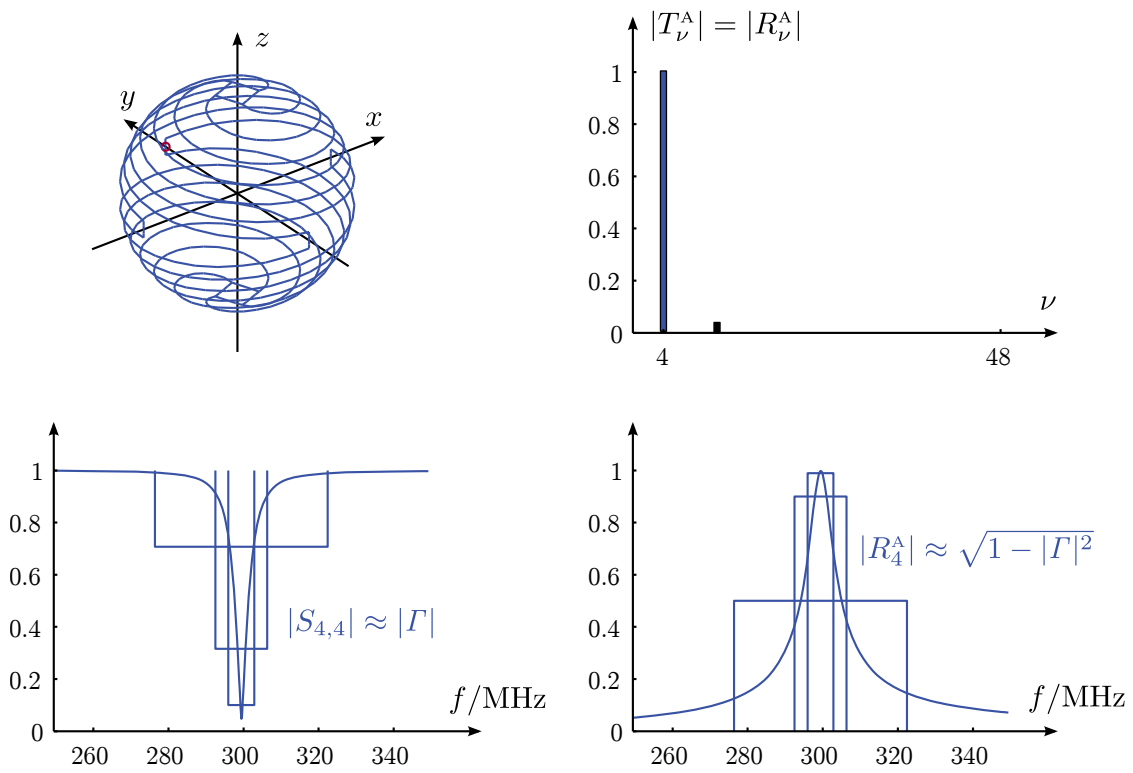



**Figure 6**: Upper left: The geometry of the linearly polarized folded spherical helix (the red circle marks the location of the feed). Upper right: The moduli of the transmitting and receiving coefficients, $|T_\nu^{\mathrm{A}}| = |R_\nu^{\mathrm{A}}|$, at the resonance frequency $f = 299\,\mathrm{MHz}$. Lower left: Reflection coefficient $\Gamma$ and scattering matrix element $S_{4,4}$ with the bound (3.4) for three different values of $S_{0,\nu}$. Lower right: Square root of the mismatch $(1 - |\Gamma|^2)$, and receiving coefficient $R_4^{\mathrm{A}}$ with the bound (4.4).

limitations (3.4), such an antenna must be narrowband and/or electrically large. A design of a directive Yagi-Uda antenna recently proposed by Arceo and Balanis in [1] has been simulated here (the specific dimensions labelled "C" in Table I in [1] was used). It has a maximum directivity of $D = 5.7$, and the results for the antenna scattering matrix can be found in Figure 8. It is seen that there are three dominating modes: $\nu = 1$ (magnetic dipole), $\nu = 4$ (electric dipole), and $\nu = 14$ (electric quadrupole). This antenna is electrically large, (it has $k_0 a = 1.42$), and therefore the numerical values of the limitations (3.4) do not give much useful information and are not included in the figure. The design of an electrically small antenna that approaches the limitations (3.4) for higher order waves is an open problem.

It should be noted that the simulation results in this paper do not perfectly match those from the references [5] and [1]. One reason is that the exact dimensions of the antennas were not clear. Another is that the wires were modelled as perfectly conducting in this paper (the wire diameter is 2.6 mm for the spherical helices and 3.0 mm for the Yagi-Uda), whereas the simulations and measurements in the references are for realistic material parameters. Lastly, the Yagi-Uda antenna was modelled as a dipole in this paper, rather than as a monopole over a ground plane.

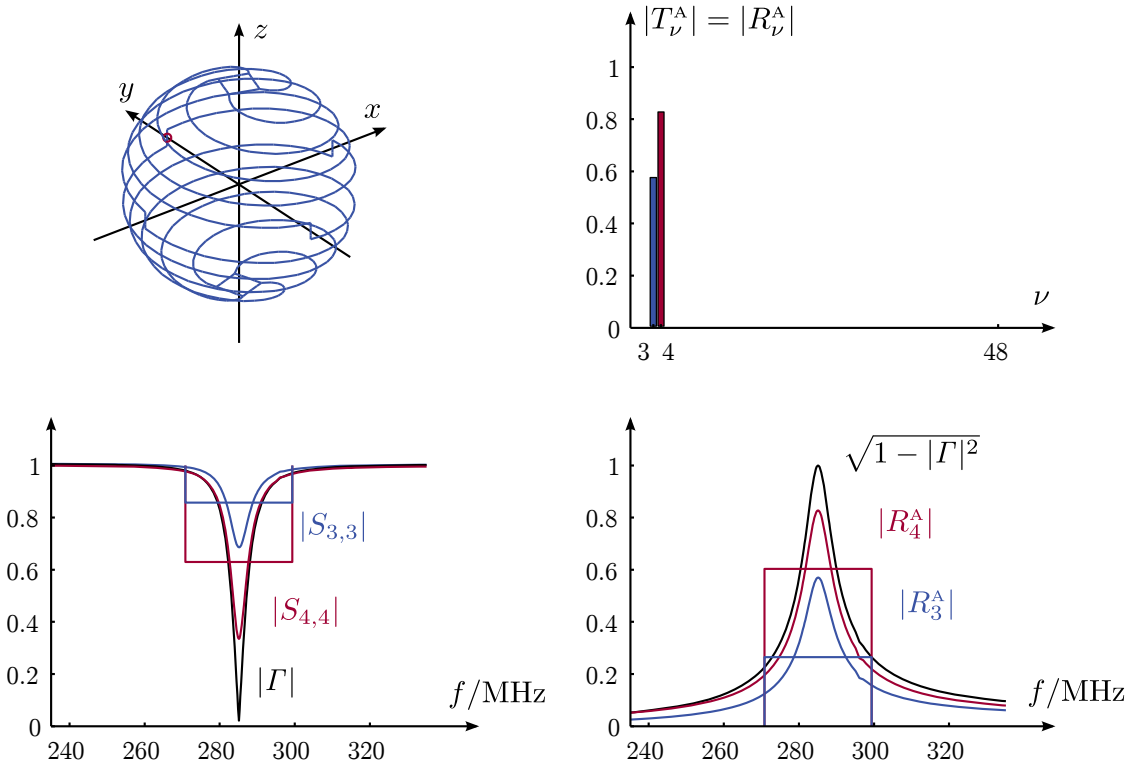



**Figure 7**: Upper left: The geometry of the elliptically polarized folded spherical helix (the red circle marks the location of the feed). Upper right: The moduli of the transmitting and receiving coefficients, $|T_\nu^{\mathrm{A}}| = |R_\nu^{\mathrm{A}}|$, at the resonance frequency $f = 285\,\mathrm{MHz}$. Lower left: Reflection coefficient $\Gamma$ and scattering matrix elements $S_{3,3}$ and $S_{4,4}$ with the bound (3.4). Note that the bound on $S_{3,3}$ is tighter since $\nu = 3$ corresponds to a magnetic spherical wave (TE-mode), and the antenna is non-magnetic. Lower right: Square root of the mismatch $(1 - |\Gamma|^2)$, and receiving coefficients $R_3^{\mathrm{A}}$ and $R_4^{\mathrm{A}}$ with the bound (4.4).

However, the task was not to verify the results of the references, but to pick clever antenna designs to illustrate the theoretical results of this paper.

All simulations have been done in the commercial software Efield (`http://www.efieldsolutions.com`). For all antennas, two separate simulations had to be carried out: In the first the antenna is transmitting, excited by a voltage source. This allows calculations of the antenna reflection coefficient $\Gamma$ and the far-field $\boldsymbol{F}$. With the far-field, the spherical wave coefficients $d_\nu^{(1)}$ of the outgoing waves and the transmitting coefficients $T_\nu^{\mathrm{A}}$ in (4.2) can be calculated, see Appendix A.2. The integral in (A.5) is solved numerically in Matlab. In the second simulation, the antenna is receiving: The voltage source is replaced by a load, and the antenna is excited with one regular spherical wave $\boldsymbol{v}_\nu$ at the time. The scattered far-field is calculated, and this in turn allows the coefficients of the outgoing waves $d_\nu^{(1)}$ and hence the scattering matrix elements $S_{\nu,\nu'}$ in (4.2) to be determined. The receiving coefficients $R_\nu^{\mathrm{A}}$ are determined by calculating the power in the load. Recall that $R_\nu^{\mathrm{A}} = (-1)^s T_\nu^{\mathrm{A}}$ holds for a reciprocal antenna; this is a good error-check. With the procedure described

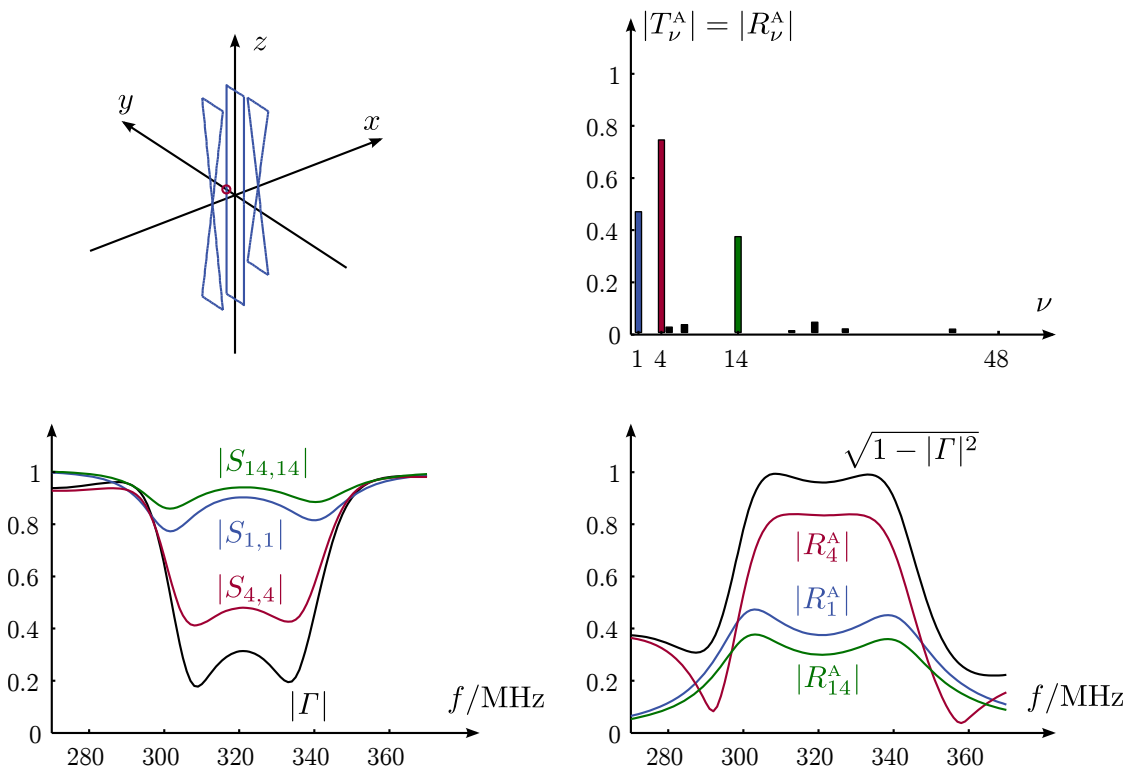



**Figure 8**: Upper left: The geometry of the directive Yagi-Uda antenna (the red circle marks the location of the feed). Upper right: The moduli of the transmitting and receiving coefficients, $|T_\nu^A| = |R_\nu^A|$, at the first resonance frequency $f = 304\,\mathrm{MHz}$. Lower left: Reflection coefficient $\Gamma$ and scattering matrix elements $S_{1,1}$, $S_{4,4}$ and $S_{14,14}$. Lower right: Square root of the mismatch $(1 - |\Gamma|^2)$, and receiving coefficients $R_1^A$, $R_4^A$ and $R_{14}^A$.

here, the complete antenna scattering matrix $\mathbf{S}^A$ in (4.1) can be determined.

To see the influence of the complex zeros $k_n$ of $S_{\nu,\nu}$ in the sum rules (3.2), the integrals in the left-hand sides as well as the static transition matrix elements appearing in the right-hand sides have also been determined, see Table 1. The integrals were determined numerically over the finite frequency intervals in Figure 6 and Figure 7, respectively; this gives estimates from below, since all integrands are positive. The electrostatic transition matrix elements were calculated with an in-house Method of Moments code. Although only the results for the folded spherical helix antennas are included here, the static transition matrix elements of other antennas and higher order spherical waves can be determined in the same way. Note that the magnetostatic transition matrices vanish since the antennas are non-magnetic. The difference between the columns in Table 1 are due to the zeros $k_n$.

| | | | |
|---|---|---|---|
| LP | $\frac{1}{\pi}\int_0^\infty \frac{1}{k^2}\ln\frac{1}{\lvert S_{4,4}(k)\rvert}\,\mathrm{d}k \geqslant 1.58\,\mathrm{mm}$ | $a$ | $= 41.8\,\mathrm{mm}$ |
| | $\frac{1}{\pi}\int_0^\infty \frac{1}{k^4}\ln\frac{1}{\lvert S_{4,4}(k)\rvert}\,\mathrm{d}k \geqslant 40.2\cdot 10^3\,\mathrm{mm}^3$ | $c_1 T^{[2]}_{201,201}$ | $= 42.7\cdot 10^3\,\mathrm{mm}^3$ |
| EP | $\frac{1}{\pi}\int_0^\infty \frac{1}{k^2}\ln\frac{1}{\lvert S_{3,3}(k)\rvert}\,\mathrm{d}k \geqslant 0.658\,\mathrm{mm}$ | $a$ | $= 41.8\,\mathrm{mm}$ |
| | $\frac{1}{\pi}\int_0^\infty \frac{1}{k^4}\ln\frac{1}{\lvert S_{3,3}(k)\rvert}\,\mathrm{d}k \geqslant 18.3\cdot 10^3\,\mathrm{mm}^3$ | $c_1 T^{[1]}_{201,201}$ | $= 0$ |
| | $\frac{1}{\pi}\int_0^\infty \frac{1}{k^2}\ln\frac{1}{\lvert S_{4,4}(k)\rvert}\,\mathrm{d}k \geqslant 1.37\,\mathrm{mm}$ | $a$ | $= 41.8\,\mathrm{mm}$ |
| | $\frac{1}{\pi}\int_0^\infty \frac{1}{k^4}\ln\frac{1}{\lvert S_{4,4}(k)\rvert}\,\mathrm{d}k \geqslant 38.4\cdot 10^3\,\mathrm{mm}^3$ | $c_1 T^{[2]}_{201,201}$ | $= 41.4\cdot 10^3\,\mathrm{mm}^3$ |

**Table 1**: The middle column presents the left-hand sides of the applicable sum rules in (3.2) for the linearly and elliptically polarized folded spherical helix antennas in Figure 6 and Figure 7, respectively. The right column presents the respective right-hand sides without the complex zeros $k_n$. The differences between the columns are due to the zeros.

# 5 Conclusions

The limitations (3.4) on the diagonal elements $S_{\nu,\nu}$ of the scattering matrix, which relate the coefficients of the incoming and outgoing vector spherical waves, were derived in this paper. The heterogeneous scatterer was assumed to be passive, with constitutive relations in convolution form in the time domain. The limitations state that the scattering matrix elements cannot be arbitrarily small over a whole wavenumber interval; the bounds depend on the fractional bandwidth $B$, as well as the electrical size of the scatterer (wavenumber $k$ times radius $a$ of the circumscribing sphere) and its shape and static material properties (given by the electrostatic and magnetostatic transition matrix elements $T^{[\tau]}_{sml,sml}$). Specifically, it was seen that the order of the dominating term in the bandwidth bounds for electrically small scatterers (Rayleigh scattering) increases with the order of the spherical wave, due to (3.5). A physical interpretation of the limitations (3.4) is that the absorption of power from each spherical wave is limited, as discussed in Section 3.2.

The derivations relied on a general approach for deriving sum rules and physical limitations for passive systems presented in [4], cf. also [3, 29]. A crucial intermediate result was the low-frequency asymptotic expansion (3.1) of the scattering matrix elements, which implied a set of sum rules, given by (3.2), from which the limitations (3.3) and (3.4) followed.

The limitations place bounds on the antenna scattering matrix $\mathbf{S}^{\mathrm{A}}$, given by (4.1). The limitations derived in the present paper are in form similar to the limitations on optimal broadband matching derived in [29, 38], although the interpretations are different, as discussed in Section 4.2. One advantage of the approach presented in this paper is that the limitations (3.4) incorporate the shape and static material properties of the antenna, and not just its electrical size as in [29, 38].

Finally, the antenna scattering matrix $\mathbf{S}^{\mathrm{A}}$ was calculated numerically for two folded spherical helix antennas and a directive Yagi-Uda antenna in Section 4.4. It was seen that the folded spherical helix antennas, which radiate dipole-patterns, performed close to the limitations. The electrically large Yagi-Uda antenna, with

directivity $D = 5.7$, had a quadrupole contribution in the far-field. Due to the limitations, such an antenna must be narrowband and/or electrically large.


# Acknowledgments

The author would like to thank Mats Gustafsson and Henrik Gyllstad for truly helpful discussions during the work on this paper, and Mats Gustafsson also for his advice regarding some of the numerical computations. The financial support by the High Speed Wireless Communications Center of the Swedish Foundation for Strategic Research (SSF) is gratefully acknowledged.


# Appendix A Details on vector spherical waves

The definitions of the incoming ($j = 2$) and outgoing ($j = 1$) vector spherical waves are those of Boström et al. [7], which only differs in normalization from those employed by Morse and Feshbach [28]:

$$\begin{cases} \boldsymbol{u}^{(j)}_{1sml}(k\boldsymbol{r}) = \mathrm{h}^{(j)}_l(kr)\boldsymbol{A}_{1sml}(\hat{\boldsymbol{r}}) = \dfrac{\nabla \times \boldsymbol{u}^{(j)}_{2sml}(k\boldsymbol{r})}{k} \\ \boldsymbol{u}^{(j)}_{2sml}(k\boldsymbol{r}) = \dfrac{(kr\mathrm{h}^{(j)}_l(kr))'}{kr}\boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) + \sqrt{l(l+1)}\dfrac{\mathrm{h}^{(j)}_l(kr)}{kr}\boldsymbol{A}_{3sml}(\hat{\boldsymbol{r}}) \\ \qquad\qquad = \dfrac{\nabla \times \boldsymbol{u}^{(j)}_{1sml}(k\boldsymbol{r})}{k}. \end{cases} \tag{A.1}$$

The same definitions are also used in [3], where more details can be found. Here $\mathrm{h}^{(j)}_l$ denotes the spherical Hankel function [2] of the $j$:th kind and order $l$, and a prime denotes differentiation with respect to the argument $kr$. The regular vector spherical waves $\boldsymbol{v}_\nu$ contain spherical Bessel functions $\mathrm{j}_l$ instead. The vector spherical harmonics $\boldsymbol{A}_{\tau sml}$ are defined by

$$\begin{cases} \boldsymbol{A}_{1sml}(\hat{\boldsymbol{r}}) = \dfrac{1}{\sqrt{l(l+1)}}\nabla \times (\boldsymbol{r}Y_{sml}(\hat{\boldsymbol{r}})) \\ \boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) = \dfrac{1}{\sqrt{l(l+1)}}r\nabla Y_{sml}(\hat{\boldsymbol{r}}) \\ \boldsymbol{A}_{3sml}(\hat{\boldsymbol{r}}) = \hat{\boldsymbol{r}}Y_{sml}(\hat{\boldsymbol{r}}). \end{cases} \tag{A.2}$$

Here $Y_{sml}$ are the (scalar) spherical harmonics

$$Y_{sml}(\theta,\phi) = \sqrt{\frac{2-\delta_{m0}}{2\pi}}\sqrt{\frac{2l+1}{2}\frac{(l-m)!}{(l+m)!}}P^m_l(\cos\theta)\left\{\begin{array}{c}\sin m\phi\\ \cos m\phi\end{array}\right\}, \tag{A.3}$$

where $\delta_{mm'}$ denotes the Kronecker delta and $P^m_l$ are associated Legendre polynomials [28]. The polar angle is denoted $\theta$ while $\phi$ is the azimuth angle. The upper (lower) expression is for $s = 1$ ($s = 2$), and the range of the indices are $l = 1, 2, \ldots$, $m = 0, 1, \ldots, l$, $s = 2$ when $m = 0$ and $s = 1, 2$ otherwise.

## A.1 Low-frequency asymptotic expansion

To derive the low-frequency asymptotic expansion (2.5), consider the following low-frequency asymptotic expansion of the spherical Bessel and Hankel functions [2]:

$$\begin{cases} \mathrm{j}_l(z) = \dfrac{2^l l! z^l}{(2l+1)!} + \mathcal{O}(z^{l+2}) \\ \mathrm{h}_l^{(1)}(z) = -\mathrm{i}\dfrac{2l!}{2^l l! z^{l+1}} + \mathcal{O}(z^{-l+1}) \end{cases} \quad \text{as } z \to 0.$$

Inserting these into (A.1) gives

$$\begin{cases} \boldsymbol{v}_{2sml} = \dfrac{2^l(l+1)!}{(2l+1)!} k^{l-1} r^{l-1} \left[ \boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) + \sqrt{\dfrac{l}{l+1}} \boldsymbol{A}_{3sml}(\hat{\boldsymbol{r}}) \right] + \mathcal{O}(k^{l+1}) \\ \boldsymbol{u}_{2sml}^{(1)} = \dfrac{\mathrm{i}(2l)!}{2^l(l-1)!} k^{-l-2} r^{-l-2} \left[ \boldsymbol{A}_{2sml}(\hat{\boldsymbol{r}}) - \sqrt{\dfrac{l+1}{l}} \boldsymbol{A}_{3sml}(\hat{\boldsymbol{r}}) \right] + \mathcal{O}(k^{-l}), \end{cases}$$

as $k \to 0$. Due to (A.2), this is equal to (2.5). In the same manner, it is straightforward to show that $\boldsymbol{v}_{1sml} = \mathcal{O}(k^l)$ and $\boldsymbol{u}_{1sml}^{(1)} = \mathcal{O}(k^{-l-1})$ as $k \to 0$.

## A.2 Farfield to spherical wave coefficients

In the numerical simulations, there is a need to extract the coefficients of the outgoing spherical waves from a calculated far-field. The electric field of a transmitting antenna or the scattered field of a receiving antenna consists of only the outgoing spherical waves $\boldsymbol{u}^{(1)}$ in (2.3). In the far-field zone, this becomes

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{\mathrm{e}^{\mathrm{i}kr}}{r} \left(1 + \mathcal{O}((kr)^{-1})\right) \underbrace{\sqrt{2\eta_0} \sum_{\nu} \mathrm{i}^{-l-2+\tau} d_\nu^{(1)} \boldsymbol{A}_\nu(\hat{\boldsymbol{r}})}_{\boldsymbol{F}(\hat{\boldsymbol{r}})} \quad \text{as } kr \to \infty, \tag{A.4}$$

where $\boldsymbol{F}$ is the far-field. Here the following expression was used [2]:

$$h_l^{(1)}(z) = \frac{\mathrm{e}^{\mathrm{i}z}}{\mathrm{i}^{l+1} z} \sum_{n=0}^{l} \frac{(l+n)!}{n!(l-n)!} (-2\mathrm{i}z)^{-n}.$$

The coefficients are given by [3]

$$d_\nu^{(1)} = \frac{\mathrm{i}^{l+2-\tau}}{\sqrt{2\eta_0}} \int_{\Omega_{\hat{\boldsymbol{r}}}} \boldsymbol{F}(\hat{\boldsymbol{r}}) \cdot \boldsymbol{A}_\nu(\hat{\boldsymbol{r}}) \, \mathrm{d}\Omega_{\hat{\boldsymbol{r}}}. \tag{A.5}$$

# Appendix B More details on the static transition matrices

It was mentioned in Section 2.2 that the electrostatic and magnetostatic transition matrices for an uncharged scatterer can be seen as generalizations of the electric

and magnetic polarizability dyadics, defined in [23]. This statement is clarified here. The electric polarizability dyadic $\boldsymbol{\gamma}_{\mathrm{e}}$ relates the induced Cartesian electric dipole moment $\boldsymbol{p} = \int \boldsymbol{r}\rho(\boldsymbol{r})\,\mathrm{d}v$ in the scatterer to the applied electrostatic field $\boldsymbol{E}$ as $\boldsymbol{p} = \epsilon_0\boldsymbol{\gamma}_{\mathrm{e}} \cdot \boldsymbol{E}$, where $\epsilon_0$ denotes the permittivity of free space. Similarly, the magnetic polarizability dyadic $\boldsymbol{\gamma}_{\mathrm{m}}$ gives the induced Cartesian magnetic dipole moment $\boldsymbol{m} = \frac{1}{2}\int \boldsymbol{r} \times \boldsymbol{J}(\boldsymbol{r})\,\mathrm{d}v$ in the scatterer caused by an applied static magnetic field $\boldsymbol{H}$: $\boldsymbol{m} = \boldsymbol{\gamma}_{\mathrm{m}} \cdot \boldsymbol{H}$. Here the induced charge and current densities in the scatterer are denoted $\rho$ and $\boldsymbol{J}$, respectively.

To see in what way the electrostatic transition matrix is a generalization of the polarizability dyadic, use the static free space Green's function to describe the scattered electrostatic potential [22]:

$$\phi(\boldsymbol{r}) = \frac{1}{\epsilon_0}\int \frac{\rho(\boldsymbol{r}')}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}v'.$$

The Green's function can be expanded into a sum of spherical harmonics; outside the sphere circumscribing the scatterer (where $r > r'$) it is [28, 31]

$$\frac{1}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|} = \sum_{l=0}^{\infty}\sum_{m=0}^{l}\sum_{s=1}^{2}\frac{1}{2l+1}r'^{l}Y_{sml}(\hat{\boldsymbol{r}}')r^{-l-1}Y_{sml}(\hat{\boldsymbol{r}}).$$

Multiply $\phi(\boldsymbol{r})$ with $Y_{sml}(\hat{\boldsymbol{r}})$ and integrate over the unit sphere. This results in the following expressions for the coefficients $f^{(1)}_{sml}$ in (2.4):

$$f^{(1)}_{sml} = \frac{1}{\epsilon_0}\frac{1}{(2l+1)}p_{sml},$$

where the electric multipole moment is [28]

$$p_{sml} = \int r^l Y_{sml}(\hat{\boldsymbol{r}})\rho(\boldsymbol{r})\,\mathrm{d}v.$$

Consequently, a scatterer subject to the external potential

$$\phi(\boldsymbol{r}) = -r^{l'}Y_{s'm'l'}(\hat{\boldsymbol{r}})$$

gets an induced multipole moment given by [28]

$$p_{sml} = \epsilon_0(2l+1)T^{[2]}_{sml,s'm'l'}.$$

For the dipole case ($l = l' = 1$), this reduces to the relation [3]

$$T^{[2]}_{sm1,s'm'1} = \frac{1}{4\pi}\hat{\boldsymbol{n}}_{sm} \cdot \boldsymbol{\gamma}_{\mathrm{e}} \cdot \hat{\boldsymbol{n}}_{s'm'}.$$

where

$$\hat{\boldsymbol{n}}_{sm} = \begin{cases} \hat{\boldsymbol{x}}, & \text{for s=2, m=1} \\ \hat{\boldsymbol{y}}, & \text{for s=1, m=1} \\ \hat{\boldsymbol{z}}, & \text{for s=2, m=0} \end{cases}$$

and $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$ are the Cartesian unit vectors.

Likewise, an applied magnetostatic potential

$$\phi^{\mathrm{MS}}(\boldsymbol{r}) = -r^{l'} Y_{s'm'l'}(\hat{\boldsymbol{r}})$$

induces a magnetic multipole moment

$$m_{sml} = (2l+1) T^{[1]}_{sml,s'm'l'},$$

where [28]

$$m_{sml} = \frac{1}{l+1} \int (\boldsymbol{r} \times \boldsymbol{J}(\boldsymbol{r})) \cdot \nabla \left( r^l Y_{sml}(\boldsymbol{r}) \right) \mathrm{d}v.$$

The dipole case is [3]

$$T^{[1]}_{sm1,s'm'1} = \frac{1}{4\pi} \hat{\boldsymbol{n}}_{sm} \cdot \boldsymbol{\gamma}_{\mathrm{m}} \cdot \hat{\boldsymbol{n}}_{s'm'}.$$

# References


[1] D. Arceo and C. A. Balanis. A compact Yagi-Uda antenna with enhanced bandwidth. *IEEE Antennas and Wireless Propagation Letters*, **10**, 442–445, 2011.

[2] G. B. Arfken and H. J. Weber. *Mathematical Methods for Physicists*. Academic Press, New York, fifth edition, 2001.

[3] A. Bernland, M. Gustafsson, and S. Nordebo. Physical limitations on the scattering of electromagnetic vector spherical waves. *J. Phys. A: Math. Theor.*, **44**(14), 145401, 2011.

[4] A. Bernland, A. Luger, and M. Gustafsson. Sum rules and constraints on passive systems. *J. Phys. A: Math. Theor.*, **44**(14), 145205, 2011.

[5] S. R. Best. Low Q electrically small linear and elliptical polarized spherical dipole antennas. *IEEE Trans. Antennas Propagat.*, **53**(3), 1047–1053, 2005.

[6] C. F. Bohren and D. R. Huffman. *Absorption and Scattering of Light by Small Particles*. John Wiley & Sons, New York, 1983.

[7] A. Boström, G. Kristensson, and S. Ström. Transformation properties of plane, spherical and cylindrical scalar and vector wave functions. In V. V. Varadan, A. Lakhtakia, and V. K. Varadan, editors, *Field Representations and Introduction to Scattering*, Acoustic, Electromagnetic and Elastic Wave Scattering, chapter 4, pages 165–210. Elsevier Science Publishers, Amsterdam, 1991.

[8] C. R. Brewitt-Taylor. Limitation on the bandwidth of artificial perfect magnetic conductor surfaces. *Microwaves, Antennas & Propagation, IET*, **1**(1), 255–260, 2007.

[9] L. J. Chu. Physical limitations of omni-directional antennas. *J. Appl. Phys.*, **19**, 1163–1175, 1948.

[10] R. E. Collin and S. Rothschild. Evaluation of antenna Q. *IEEE Trans. Antennas Propagat.*, **12**, 23–27, January 1964.

[11] R. M. Fano. Theoretical limitations on the broadband matching of arbitrary impedances. *Journal of the Franklin Institute*, **249**(1,2), 57–83 and 139–154, 1950.

[12] A. A. Glazunov, M. Gustafsson, A. Molisch, and F. Tufvesson. Physical modeling of multiple-input multiple-output antennas and channels by means of the spherical vector wave expansion. *IET Microwaves, Antennas & Propagation*, **4**(6), 778–791, 2010.

[13] M. Gustafsson and D. Sjöberg. Sum rules and physical bounds on passive metamaterials. *New Journal of Physics*, **12**, 043046, 2010.

[14] M. Gustafsson, C. Sohl, and G. Kristensson. Physical limitations on antennas of arbitrary shape. *Proc. R. Soc. A*, **463**, 2589–2607, 2007.

[15] M. Gustafsson, C. Sohl, and G. Kristensson. Illustrations of new physical bounds on linearly polarized antennas. *IEEE Trans. Antennas Propagat.*, **57**(5), 1319–1327, May 2009.

[16] M. Gustafsson and S. Nordebo. Bandwidth, Q factor, and resonance models of antennas. *Progress in Electromagnetics Research*, **62**, 1–20, 2006.

[17] M. Gustafsson and S. Nordebo. Characterization of MIMO antennas using spherical vector waves. *IEEE Trans. Antennas Propagat.*, **54**(9), 2679–2682, 2006.

[18] M. Gustafsson and D. Sjöberg. Physical bounds and sum rules for high-impedance surfaces. *IEEE Trans. Antennas Propagat.*, **59**(6), 2196–2204, 2011.

[19] J. E. Hansen, editor. *Spherical Near-Field Antenna Measurements*. Number 26 in IEE electromagnetic waves series. Peter Peregrinus Ltd., Stevenage, UK, 1988. ISBN: 0-86341-110-X.

[20] R. F. Harrington. *Time Harmonic Electromagnetic Fields*. McGraw-Hill, New York, 1961.

[21] A. Hujanen, J. Holmberg, and J. C.-E. Sten. Bandwidth limitations of impedance matched ideal dipoles. *IEEE Trans. Antennas Propagat.*, **53**(10), 3236–3239, 2005.

[22] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, New York, third edition, 1999.

[23] R. E. Kleinman and T. B. A. Senior. Rayleigh scattering. In V. V. Varadan and V. K. Varadan, editors, *Low and high frequency asymptotics*, volume 2 of *Handbook on Acoustic, Electromagnetic and Elastic Wave Scattering*, chapter 1, pages 1–70. Elsevier Science Publishers, Amsterdam, 1986.

[24] B. Kogan. Comments on "Broadband matching limitations for higher order spherical modes". *IEEE Trans. Antennas Propagat.*, **58**(5), 1826, 2010.

[25] P. A. Martin. *Multiple Scattering: Interaction of Time-Harmonic Waves with N Obstacles*, volume 107 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, U.K., 2006.

[26] M. Migliore. On electromagnetics and information theory. *IEEE Trans. Antennas Propagat.*, **56**(10), 3188–3200, October 2008.

[27] M. I. Mishchenko, L. D. Travis, and D. W. Mackowski. T-matrix method and its applications to electromagnetic scattering by particles: A current perspective. *J. Quant. Spectrosc. Radiat. Transfer*, **111**(11), 1700–1703, 2010.

[28] P. M. Morse and H. Feshbach. *Methods of Theoretical Physics*, volume 2. McGraw-Hill, New York, 1953.

[29] S. Nordebo, A. Bernland, M. Gustafsson, C. Sohl, and G. Kristensson. On the relation between optimal wideband matching and scattering of spherical waves. *IEEE Trans. Antennas Propagat.*, **59**(9), 3358–3369, 2011.

[30] P. Olsson. Elastostatics as a limit of elastodynamics — a matrix formulation. *Applied Scientific Research*, **41**, 125–151, 1984.

[31] B. Peterson. Matrix formulation of static field problems involving an arbitrary number of bodies. *Inst. Theoretical Physics, S-412 96 Göteborg, Sweden*, **Report 75–12**, 1975.

[32] K. N. Rozanov. Ultimate thickness to bandwidth ratio of radar absorbers. *IEEE Trans. Antennas Propagat.*, **48**(8), 1230–1234, August 2000.

[33] D. Sjöberg. Variational principles for the static electric and magnetic polarizabilities of anisotropic media with perfect electric conductor inclusions. *J. Phys. A: Math. Theor.*, **42**, 335403, 2009.

[34] C. Sohl and M. Gustafsson. A priori estimates on the partial realized gain of Ultra-Wideband (UWB) antennas. *Quart. J. Mech. Appl. Math.*, **61**(3), 415–430, 2008.

[35] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on broadband scattering by heterogeneous obstacles. *J. Phys. A: Math. Theor.*, **40**, 11165–11182, 2007.

[36] J. Sten and A. Hujanen. Notes on the quality factor and bandwidth of radiating systems. *Electrical Engineering (Archiv fur Elektrotechnik)*, **84**(4), 189–195, 2002.

[37] J. A. Stratton. *Electromagnetic Theory*. McGraw-Hill, New York, 1941.

[38] M. C. Villalobos, H. D. Foltz, and J. S. McLean. Broadband matching limitations for higher order spherical modes. *IEEE Trans. Antennas Propagat.*, **57**(4), 1018–1026, 2009.

[39] J. Volakis, C. C. Chen, and K. Fujimoto. *Small Antennas: Miniaturization Techniques & Applications*. McGraw-Hill, New York, 2010.

[40] P. C. Waterman. The T-matrix revisited. *J. Opt. Soc. Am. A*, **24**(8), 2257–2267, Aug 2007.

[41] P. C. Waterman. T-matrix methods in acoustic scattering. *J. Acoust. Soc. Am.*, **125**(1), 42–51, 2009.

[42] P. C. Waterman. Matrix methods in potential theory and electromagnetic scattering. *J. Appl. Phys.*, **50**(7), 4550–4566, 1979.

[43] M. Wohlers and E. Beltrami. Distribution theory as the basis of generalized passive-network analysis. *IEEE Transactions on Circuit Theory*, **12**(2), 164–170, 1965.

[44] A. H. Zemanian. *Realizability theory for continuous linear systems*. Academic Press, New York, 1972.

# Estimation of Spherical Wave Coefficients from 3D Positioner Channel Measurements

Anders Bernland, Mats Gustafsson, Carl Gustafson, and Fredrik Tufvesson.

Paper V

**Based on:** A. Bernland, M. Gustafsson, C. Gustafson, and F. Tufvesson. Estimation of spherical wave coefficients from 3D positioner channel measurements. Technical Report LUTEDX/TEAT-7215, Lund University, Sweden, pp. 1–11, March 2012. http://www.eit.lth.se. Submitted to *IEEE Antennas and Wireless Propagation Letters*.

**Abstract**

Electromagnetic vector spherical waves have been used recently to model antenna-channel interaction and the available degrees of freedom in MIMO systems. However, there are no previous accounts of a method to estimate spherical wave coefficients from channel measurements. One approach, using a 3D positioner, is presented in this letter, both in theory and practice. Measurement results are presented and discussed. One conclusion is that using randomly positioned measurements within a volume is less sensitive to noise than using only measurements on the surface.

# 1 Introduction

Real-world measurements and theoretical modelling of antennas and propagation channels are crucial to wireless communication, and have been the focus of extensive research for many years [10]. One way to the increase the capacity in wireless systems is to use multiple-input multiple-output (MIMO) technology. MIMO requires several degrees of freedom, and the degrees of freedom depend both on the mutual coupling between antenna elements and the richness of the channel. It is therefore desirable to separate the antenna and channel influence. One approach to do this is the double-directional channel model, which describes the channel in terms of plane waves, or multi-path components [7, 10].

Electromagnetic vector spherical waves provide a compact description of a single- or multi-port antenna in terms of the antenna scattering matrix, which describes the antenna receiving, transmitting and scattering properties [4]. One benefit is that only a few terms are needed for a small antenna. Furthermore, spherical waves are used within spherical near-field antenna measurements, where they enable the necessary probe corrections and near-field to far-field transforms [4, 8].

Spherical vector wave approaches to theoretically model antenna-channel interaction and the available degrees of freedom have been given in [2, 3, 9], separating the antenna from the channel in a compact and intuitive way. It is well known that a small antenna only can excite a limited number of spherical waves [1], which restricts the available degrees of freedom for a small multi-port antenna. It is not, however well-known how many degrees of freedom a given propagation channel can support. Furthermore, to the authors' best knowledge there are no previous publications where spherical waves are estimated from channel measurements, although some preliminary studies have been done [6].

The main objective of this letter is to present a method to estimate spherical wave coefficients from channel measurements. For this, a 3D positioner is used to move the receiving antenna to different positions and orientations within a cube, and probe correction [4] is used to separate the influence of the receiving antenna. The whole volume of the cube as well as different subsets are used in separate estimations to determine how the measurements points should be positioned.

## 2 Preliminaries

In a source-free region enclosed by spherical surfaces, the electric field can be written as a sum of regular ($\boldsymbol{v}$) and outgoing ($\boldsymbol{u}^{(1)}$) vector spherical waves (time convention $\mathrm{e}^{-\mathrm{i}\omega t}$):

$$\boldsymbol{E}(\boldsymbol{r},k) = k\sqrt{\eta_0}\sum_{\nu} d_\nu^{(1)}\boldsymbol{u}_\nu^{(1)}(k\boldsymbol{r}) + d_\nu^{(2)}\boldsymbol{v}_\nu(k\boldsymbol{r}). \tag{2.1}$$

Here the free space parameters are wavenumber $k = \omega/\mathrm{c}$, speed of light c and impedance $\eta_0$. The spatial coordinate is denoted $\boldsymbol{r}$, with $r = |\boldsymbol{r}|$ and $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$. The spherical waves are defined as in the book [4] by Hansen, but with slightly different notation (see Appendix A). The multi-index $\nu = 2(l^2+l-1+m)+\tau$ is introduced in place of the indices $\{\tau, m, l\}$, where $\tau = 1$ (odd $\nu$) corresponds to a magnetic $2^l$-pole ($\mathrm{TE}_l$-mode), while $\tau = 2$ (even $\nu$) identifies an electric $2^l$-pole ($\mathrm{TM}_l$-mode). The basis function in the azimuth angle $\phi$ is $\mathrm{e}^{\mathrm{i}m\phi}$.

The antenna source scattering matrix completely describes the antenna properties:

$$\begin{pmatrix} \mathit{\Gamma} & \mathbf{R}' \\ \mathbf{T}' & \mathbf{S}' \end{pmatrix} \begin{pmatrix} w^{(2)} \\ \boldsymbol{d}^{(2)} \end{pmatrix} = \begin{pmatrix} w^{(1)} \\ \boldsymbol{d}^{(1)} \end{pmatrix}. \tag{2.2}$$

Here $\boldsymbol{d}^{(2)} = (d_1^{(2)}\ d_2^{(2)}\ \dots\ )^{\mathrm{T}}$ and $w^{(2)}$ are the coefficients of the incident regular waves and transmitted signal, whereas $\boldsymbol{d}^{(1)}$ and $w^{(1)}$ are the coefficients of the scattered or transmitted outgoing waves and received signal. The transmitted and received signals are vectors in the case of a multi-port antenna. The primes are included here to indicate that the source scattering matrix formulation in (2.163) in [4] is used. The transmitting coefficients $T'_\nu$ and receiving coefficients $R'_\nu$ are included in $\mathbf{T}'$ and $\mathbf{R}'$, respectively [4].

The main purpose of this letter is to determine the spherical wave coefficients $d_\nu^{(2)}$ from channel measurements. More precisely, consider a transmitting antenna in a propagation channel, as in Figure 1a. Within any sphere containing no scatterers, only the regular waves contribute to the sum in (2.1). The coefficients $d_\nu^{(2)}$ will be estimated from measurements with the receiving antenna placed in a number of different positions and orientations. It is assumed that the scattered field that is in turn scattered back from nearby objects is negligible.

## 3 Method and measurement setup

In order to estimate the spherical wave coefficients, the receiving antenna is placed in a number of different positions and orientations. When the antenna is placed at the origin in its original orientation, it receives the signal $w^{(1)}$ given by (2.2). When it is moved and/or rotated, expressions for $w^{(1)}$ are derived by expressing the spherical waves in the original coordinate system $(x, y, z)$ as sums over the spherical waves in the translated and rotated coordinate system $(x_i, y_i, z_i)$:

$$\boldsymbol{v}_\nu(k\boldsymbol{r}) = \sum_{\nu_i} B_{\nu,\nu_i}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)\boldsymbol{v}_{\nu_i}(k\boldsymbol{r}_i). \tag{3.1}$$

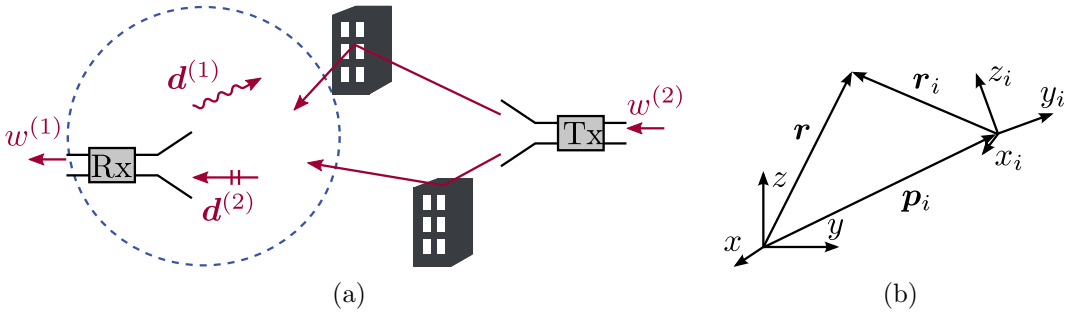



**Figure 1**: (a): A receiving antenna is used to estimate the coefficients of the incident field $\boldsymbol{d}^{(2)}$. It is assumed that the scattered field ($\boldsymbol{d}^{(1)}$) that is in turn scattered back from nearby objects is negligible. (b): The receiving antenna is placed at a number of different positions and orientations, described by the translated and rotated coordinate systems $(x_i, y_i, z_i)$. The original coordinate system $(x, y, z)$ is centered in the sampling region.

Here $\boldsymbol{p}_i$ is the position of the translated origin, and $\hat{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{z}}_i$ are the orientations of the $y_i$- and $z_i$-axes, see Figure 1b. Explicit expressions for $B_{\nu,\nu_i}$ can be found in [4]. From (2.2) and (3.1) it follows that the antenna receives the signal

$$w_i^{(1)} = \mathbf{R}'\ \mathbf{B}^{\mathrm{T}}(\boldsymbol{p}_i, \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)\ \boldsymbol{d}^{(2)} \tag{3.2}$$

when positioned at $\boldsymbol{p}_i$ and oriented according to $(\hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)$, where $\mathbf{B}$ is the infinite-dimensional matrix with elements $B_{\nu,\nu_i}$, and $^{\mathrm{T}}$ denotes transpose. If the receiving antenna had been an ideal dipole (i.e. $\mathbf{R}' = [0\ 0\ 0\ 1/2\ 0\ 0 \ldots]$), the expression for the received signal in (3.2) simplifies to [4]

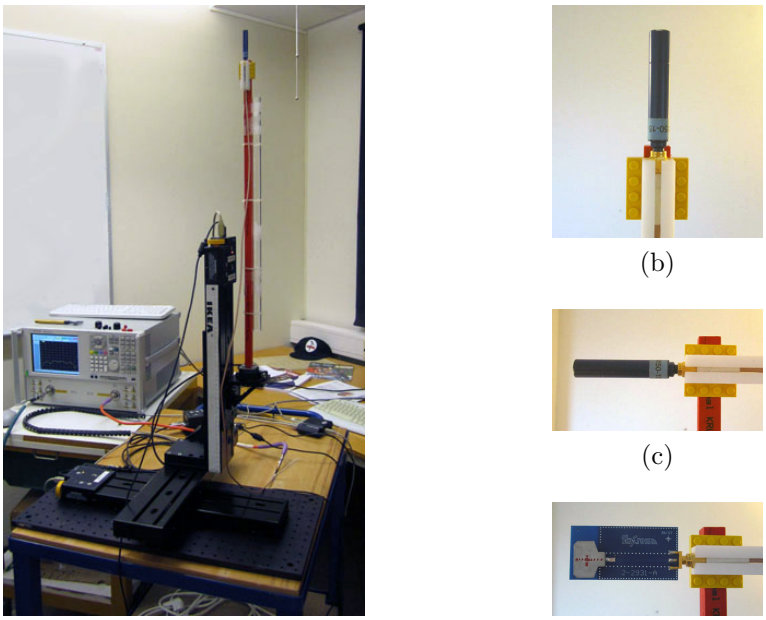
$$w_i^{(1)} = \frac{\sqrt{3\pi}}{2k\sqrt{\eta_0}} \hat{\boldsymbol{z}}_i \cdot \boldsymbol{E}(\boldsymbol{r}_i),$$

which is a good error-check for numerical implementations. The expansion in (2.1) is truncated at $\nu = N = 2L(L+2)$ by choosing a maximum order $L = \max l$ [4]. With $M$ measurements, this leads to:

$$\underbrace{\begin{bmatrix} w_1^{(1)} \\ w_2^{(1)} \\ \vdots \\ w_M^{(1)} \end{bmatrix}}_{\boldsymbol{w}^{(1)}} = \underbrace{\begin{bmatrix} \mathbf{R}'\ \mathbf{B}(\boldsymbol{p}_1, \hat{\boldsymbol{y}}_1, \hat{\boldsymbol{z}}_1)^{\mathrm{T}} \\ \mathbf{R}'\ \mathbf{B}(\boldsymbol{p}_2, \hat{\boldsymbol{y}}_2, \hat{\boldsymbol{z}}_2)^{\mathrm{T}} \\ \vdots \\ \mathbf{R}'\ \mathbf{B}(\boldsymbol{p}_M, \hat{\boldsymbol{y}}_M, \hat{\boldsymbol{z}}_M)^{\mathrm{T}} \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} d_1^{(2)} \\ d_2^{(2)} \\ \vdots \\ d_N^{(2)} \end{bmatrix}}_{\boldsymbol{d}^{(2)}} + \boldsymbol{e}, \tag{3.3}$$

where $\boldsymbol{e}$ is error due to the truncation. In practice, also noise is present in the measurement.

The measurements were carried out with the 3D positioner in Figure 2a. An in-house patch antenna, kept at a fixed position, is used as the transmit antenna at 5.15 GHz. A Satimo 5.15 GHz sleeve dipole (SD5150) was chosen as the receiving antenna; it is placed on the 3D positioner as depicted in Figure 2b–2c and moved



(a) (b) (c) (d)

**Figure 2**: (a): The 3D positioner used in the measurements. (b)–(c): The Satimo 5.15 GHz sleeve dipole used as receiving antenna to estimate coefficients, mounted in (b): $z$-polarization and (c): $x$-polarization. The 3D positioner rotates the antenna in $x-$polarization 90° to measure $y-$polarization. (d): The Skycross UWB antenna used for validation.

in a $10 \times 10 \times 10$ cubical grid with stepsize 15mm ($\approx 0.26\lambda$), measuring $x$, $y$, and $z$ polarization at each point for a total of 3000 measurements. Here the coordinates given to the 3D positioner must first be corrected for the offset in phase center as the antenna is rotated. The transmitting and receiving antennas are connected to port 1 and 2 of an Agilent E8361A vector network analyzer, which was calibrated and used to measure the transfer function $S_{21} = w_{\mathrm{Rx}}^{(1)}/w_{\mathrm{Tx}}^{(2)}$. An amplifier was used at the transmitter side. For later use as validation of the estimated coefficients, the Skycross UWB antenna (SMT-2TO6MB-A) in Figure 2d (frequency range 2.3-5.9 GHz) is used as the receiving antenna in place of the sleeve dipole in otherwise identical measurements; it is moved along a subset of the points in the cubical grid for a total of 90 measurements.

For verification purposes, data has also been simulated: 100 random plane waves with independent polarization, complex Gaussian amplitude and angles of arrival uniformly distributed over the sphere, distorted by zero mean white Gaussian noise. In this case, closed form expressions of the coefficients in $d^{(2)}$ are known [4], which makes it possible to check the accuracy of the method as a function of SNR.

The matrix $\mathbf{G}$ in (3.3) is determined with in-house Matlab-scripts, using the positions $\boldsymbol{p}_i$ and orientations $(\hat{\boldsymbol{y}}_i, \hat{\boldsymbol{z}}_i)$ from the 3D positioner and the receiving coefficients $R'_\nu$ of the antennas. For this reason, both the sleeve dipole and UWB antenna have been characterized in a Satimo Stargate-24 chamber where the antenna transmitting coefficients $T'_\nu$ are given as output. The receiving coefficients

$R'_\nu$ are given by $R'_{\{\tau,m,l\}} = (-1)^m T'_{\{\tau,-m,l\}}/2$ [4]. The sleeve dipole is very close to a Hertzian dipole, and the higher orders contribute little to the results of the 3D positioner measurements. It is observed that the smallest errors are obtained when the receiving coefficients $R'_\nu$ are truncated to contain only the dipole term.

An estimate $\bar{\boldsymbol{d}}^{(2)}$ of the unknown coefficients in $\boldsymbol{d}^{(2)}$ can be computed from the system of equations in (3.3). A first approach is the least-squares solution

$$\bar{\boldsymbol{d}}^{(2)}_{\text{LSQ}} = \arg\min_{\boldsymbol{d}^{(2)}} ||\mathbf{G}\boldsymbol{d}^{(2)} - \boldsymbol{w}^{(1)}||^2,$$

but the singular values of $\mathbf{G}$ suggest that this is an ill-posed problem. Furthermore, when computing least-squares solutions from simulated data, it is seen that large errors are introduced for the coefficients of high orders $l$. Therefore, a more elaborate method should be used. Here, a regularized solution by means of Tikhonov's method is chosen [5]:

$$\bar{\boldsymbol{d}}^{(2)} = \arg\min_{\boldsymbol{d}^{(2)}} \left( ||\mathbf{G}\boldsymbol{d}^{(2)} - \boldsymbol{w}^{(1)}||^2 + ||\lambda_{\text{reg}}\boldsymbol{d}^{(2)}||^2 \right).$$

The regularization parameter $\lambda_{\text{reg}}$ is determined with the L-curve criterion [5]. The regularization works well when tested on simulated data, and, based on the experiments carried out, the regularization gives plausible results also for measured data. It is also seen that the estimated coefficients $\bar{\boldsymbol{d}}^{(2)}$ are independent of the truncation order $L$, as long as it is chosen large enough. A rule of thumb is $L > kr_{\text{circ}} \approx 12.6$, where $r_{\text{circ}}$ is the radius of the smallest sphere circumscribing the cube.

The measurement problem considered here shows some similarities with near-field antenna measurements, see e.g. [4, 8, 11]. However, none of these methods are directly applicable here.

# 4 Results and discussion

The measurement scenario is a small room with many scatterers and obstructed-line-of-sight (OLOS), see Figure 3. It is chosen to get a rich channel, and a challenging problem to estimate the spherical wave coefficients. Measurements were also carried out in a large, empty room under line-of-sight conditions, with similar results.

The spherical wave coefficients $d^{(2)}_\nu$ are estimated from the measured data as described above. For a first validation, 30 randomly chosen measurements out of the 3000 measurements are excluded from the estimation, and the estimated coefficients $\bar{d}^{(2)}_\nu$ are used to predict those transfer functions $S_{21,i} = w^{(1)}_{\text{Rx},i}/w^{(2)}_{\text{Tx},i}$. The results can be found in Figure 4a. In the second validation, the same estimates $\bar{d}^{(2)}_\nu$ are used to estimate the transfer function for the UWB antenna, see Figure 4b. The errors are slightly larger here; at 5.15 GHz the UWB antenna has a complicated radiation pattern, and small errors in positioning and influence from nearby objects give large errors for the received signal. The results in Figure 4 give an indication of the magnitude of the errors in the estimation of the spherical wave coefficients.

To investigate if, and how, the number of measurements can be reduced, the spherical wave coefficients are estimated using three different subsets of the 3000

**Figure 3**: The measurement scenario, a small room with many scatterers and obstructed-line-of-sight (OLOS). The transmitting patch antenna is mounted on the stand to the right, and the receiving sleeve dipole is mounted on the 3D positioner to the left.

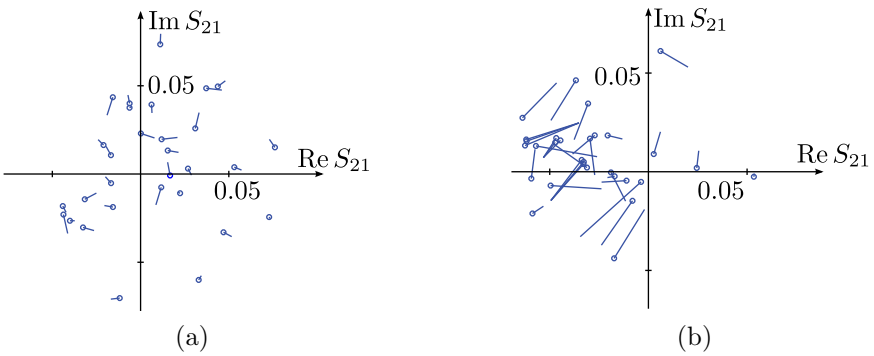



**Figure 4**: (a): 30 randomly chosen measurements out of the 3000 measurements are excluded, and the estimated coefficients $\bar{d}_\nu^{(2)}$ are used to predict those transfer functions $S_{21,i} = w_{\mathrm{Rx},i}^{(1)}/w_{\mathrm{Tx},i}^{(2)}$. The circle marks the measured transfer function, and a line is drawn to the estimated value. (b): Same as (a), but the estimated coefficients $\bar{d}_\nu^{(2)}$ are used to estimate the transfer function when the UWB antenna is receiving.

measurements: I) 1464 measurements on the surface of the cube, II) 1536 measurements in the $8 \times 8 \times 8$ inner cube, and III) 1500 measurements chosen at random. The estimated coefficients $\bar{d}_{\nu,\mathrm{sub}}^{(2)}$ for all three cases are compared to the estimated coefficients $\bar{d}_{\nu,\mathrm{cube}}^{(2)}$ when all the measurements are used, see Figure 5. It is seen that I) introduces errors for all the coefficients, II) introduces errors for high order coefficients (large multi-index $\nu$), and III) works well for all the coefficients.

Some observations: I) In theory, it would suffice to measure only on the sur-

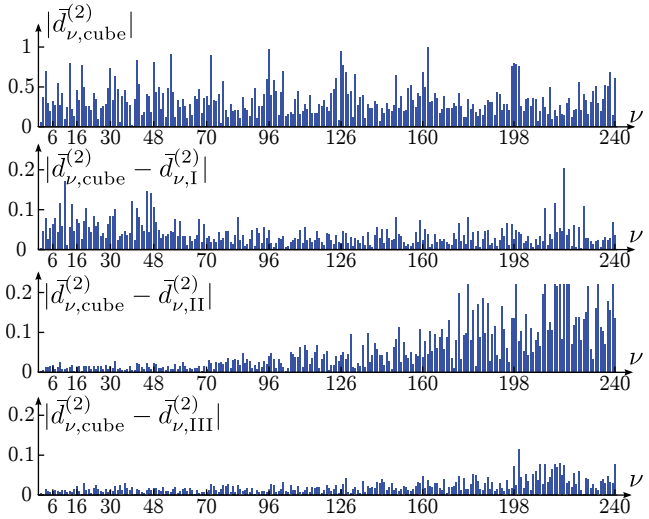



**Figure 5**: Top: Estimated coefficients $\bar{d}^{(2)}_{\nu,\mathrm{cube}}$ using all the 3000 measurements in the cube. The coefficients are normalized so that $\max_\nu |\bar{d}^{(2)}_{\nu,\mathrm{cube}}| = 1$. The other graphs depict the difference for the estimated coefficients, $|\bar{d}^{(2)}_{\nu,\mathrm{cube}} - \bar{d}^{(2)}_{\nu,\mathrm{sub}}|$ when the subsets I)-III) are used.

face [4], but it fails when noise is introduced. II) As expected, using the inner points works equally well for the low order waves, but the high order waves are not detectable since they vanish close to the origin in the middle of the cube. A rule of thumb is that $N = 2L(L+1)$ coefficients can be estimated (cf. [4]), where $L = kr_{\mathrm{insc}}$ and $r_{\mathrm{insc}}$ is the radius of the largest sphere inscribed in the cube. This gives $N = 126$ for the $10 \times 10 \times 10$ cube and $N = 70$ for the $8 \times 8 \times 8$ cube. III) The randomly chosen points cover the same volume as the whole cube, but the errors go up slightly since fewer measurements are used.

To check the accuracy of the method as a function of SNR, simulated data with SNR = 15 dB and SNR = 30 dB, respectively, is also used. For the lower SNR, where also small errors have been introduced in the characterization of the UWB antenna, similar results as those from the measurements are obtained, see Figure 6–7. It is therefore expected that this simulation represents the measurements well, and it is seen that the coefficients up to $\nu = 126$ (which corresponds to $L = 7$) are estimated with less than 5% error, and that the coefficients up to $\nu = 198$ ($L = 9$) are estimated with less than 10% error. For the high SNR (Figure 8), it is seen that using either the surface of the cube or randomly chosen measurements works well, and therefore it is expected that fewer measurement points are sufficient for high SNR. However, using only the inner $8 \times 8 \times 8$ cube fails for the high order waves as expected.

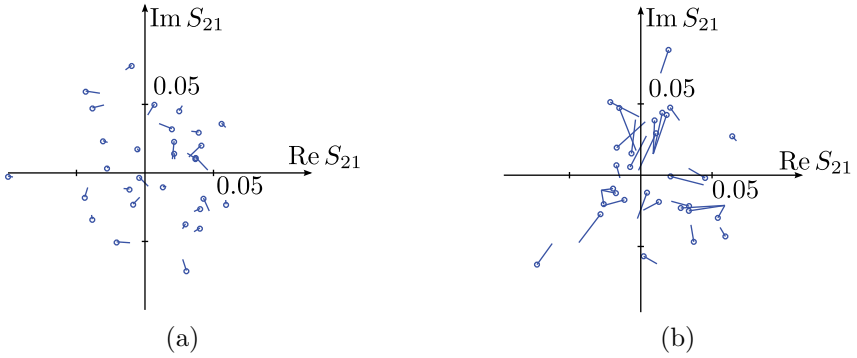



**Figure 6**: Same as Figure 4, but using simulated data with SNR=15 dB and small errors in the characterization of the UWB antenna.

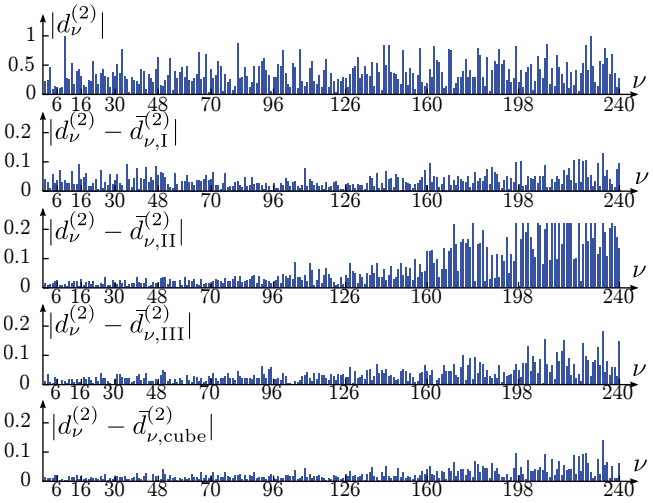



**Figure 7**: Simulated data, SNR = 15 dB. Top: Analytic solution for the coefficients $d_\nu^{(2)}$, normalized so that $\max_\nu |d_\nu^{(2)}| = 1$. The other graphs depict the difference $|d_\nu^{(2)} - \bar{d}_\nu^{(2)}|$ when the coefficients have been estimated using the subsets I)-III) and the whole cube.

# 5 Conclusions

A method to estimate spherical wave coefficients from channel measurements was presented in this letter. A 3D positioner was used to move the receiving sleeve dipole antenna within a $10 \times 10 \times 10$ cubical grid, measuring $x-$, $y-$ and $z-$polarization at each point. The receiving antenna was characterized, and expressions for translations and rotations of spherical waves were then used to obtain a system of equations for the unknown coefficients $d_\nu^{(2)}$, which was solved numerically with Tikhonov regularization. The results were validated by using the estimated coefficients to estimate the received signal, both in the sleeve dipole and in a UWB antenna. Simulated data was also used to check the accuracy of the method as a function of SNR.

Furthermore, different subsets of the measurements were used to estimate the coefficients. It was seen that using only the surface gave large errors, using only inner

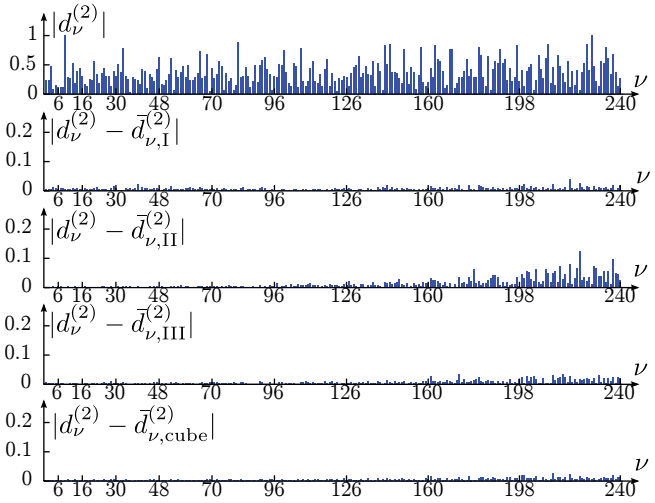



**Figure 8**: Same as Figure 7, but with SNR = 30 dB.

points failed for high order waves, whereas a randomly chosen subset worked well. With a comparison to simulated data, it can be expected that fewer measurement points are sufficient for higher SNR, and in this case they can be placed uniformly, randomly, or on the surface.

In future measurements, it would be desirable to use something else than the rather slow 3D positioner. A real array that measures on a surface does not seem to be feasible for low SNR. However, the good results for estimation using the randomly chosen points indicate that it is not necessary to use a device that controls the positions precisely, as long as they are measured correctly. Hopefully, this fact can be taken advantage of in order to simplify and speed up the measurements.

# Acknowledgments

The authors would like to thank Anders Sunesson for his help with the antenna characterizations. The financial support by the High Speed Wireless Communications Center of the Swedish Foundation for Strategic Research (SSF) is gratefully acknowledged.

# Appendix A Definitions of vector spherical waves

The vector spherical waves are defined as in [4], but with different notation. The regular waves are defined as

$$\begin{cases} \boldsymbol{v}_{1sml}(k\boldsymbol{r}) = \mathrm{j}_l(kr)\dfrac{\nabla \times (\boldsymbol{r} Y_{sml}(\hat{\boldsymbol{r}}))}{\sqrt{l(l+1)}} \\ \boldsymbol{v}_{2sml}(k\boldsymbol{r}) = \dfrac{\nabla \times \boldsymbol{v}_{1sml}(k\boldsymbol{r})}{k}. \end{cases}$$

Here $\mathrm{j}_l$ denotes the spherical Bessel function of order $l$ [4]. The Bessel function is replaced with a spherical Hankel function of the first kind to get outgoing vector spherical waves $\boldsymbol{u}^{(1)}_{\tau sml}$. The spherical harmonics $Y_{sml}$ are given by

$$Y_{sml}(\theta,\phi) = (-1)^m \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta)\mathrm{e}^{\mathrm{i}m\phi},$$

where $P_l^m$ are associated Legendre polynomials [4]. The polar angle is denoted $\theta$ while $\phi$ is the azimuth angle. The range of the indices are $l = 1, 2, \ldots$ and $m = -l, -l+1, \ldots l$.

# References

[1] A. Bernland. Bandwidth limitations for scattering of higher order electromagnetic spherical waves with implications for the antenna scattering matrix. *IEEE Trans. Antennas Propagat.*, 2012. Accepted for publication.

[2] A. A. Glazunov, M. Gustafsson, A. Molisch, and F. Tufvesson. Physical modeling of multiple-input multiple-output antennas and channels by means of the spherical vector wave expansion. *IET Microwaves, Antennas & Propagation*, **4**(6), 778–791, 2010.

[3] A. A. Glazunov, M. Gustafsson, A. Molisch, F. Tufvesson, and G. Kristensson. Spherical vector wave expansion of gaussian electromagnetic fields for antenna-channel interaction analysis. *IEEE Trans. Antennas Propagat.*, **3**(2), 214–227, 2009.

[4] J. E. Hansen, editor. *Spherical Near-Field Antenna Measurements*. Number 26 in IEE electromagnetic waves series. Peter Peregrinus Ltd., Stevenage, UK, 1988. ISBN: 0-86341-110-X.

[5] P. C. Hansen. Regularization tools version 4.0 for Matlab 7.3. *Numerical Algorithms*, **46**, 189–194, 2007.

[6] A. Khatun, T. Laitinen, and P. Vainikainen. Noise sensitivity analysis of spherical wave modelling of radio channels using linear scanners. In *Microwave Conference Proceedings (APMC), 2010 Asia-Pacific*, pages 2119–2122, Yokohama, December 2010.

[7] V.-M. Kolmonen, P. Almers, J. Salmi, J. Koivunen, K. Haneda, A. Richter, F. Tufvesson, A. Molisch, and P. Vainikainen. A dynamic dual-link wideband MIMO channel sounder for 5.3 GHz. *IEEE Trans. Instrumentation and Measurement*, **59**(4), 873–883, April 2010.

[8] T. Laitinen, S. Pivnenko, J. M. Nielsen, and O. Breinbjerg. Theory and practice of the FFT/matrix inversion technique for probe-corrected spherical near-field antenna measurements with high-order probes. *IEEE Trans. Antennas Propagat.*, **58**(8), 2623–2631, August 2010.

[9] M. D. Migliore. An intuitive electromagnetic approach to MIMO communication systems. *IEEE Trans. Antennas Propagat.*, **48**(3), 128–137, June 2006.

[10] A. F. Molisch. *Wireless Communications*. John Wiley & Sons, New York, second edition, 2011.

[11] C. H. Schmidt, M. M. Leibfritz, and T. F. Eibert. Fully probe-corrected near-field far-field transformation employing plane wave expansion and diagonal translation operators. *IEEE Trans. Antennas Propagat.*, **56**(3), 737–746, March 2008.

LUND
UNIVERSITY

Series of licentiate and doctoral theses
Department of Electrical and Information Technology
ISSN 1654-790X
No. 40
ISBN 978-91-7473-302-0

http://www.eit.lth.se