



LUND UNIVERSITY

Sum rules and constraints on passive systems - a general approach and applications to electromagnetic scattering

Bernland, Anders

2010

[Link to publication](#)

Citation for published version (APA):

Bernland, A. (2010). *Sum rules and constraints on passive systems - a general approach and applications to electromagnetic scattering*. [Licentiate Thesis, Department of Electrical and Information Technology]. Department of Electrical and Information Technology, Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

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

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

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Licentiate dissertation

Sum Rules and Constraints on Passive Systems — a General Approach and Applications to Electromagnetic Scattering

Anders Bernland



Sum Rules and Constraints on Passive Systems — a General Approach and Applications to Electromagnetic Scattering

Anders Bernland

Licentiate Dissertation
Electromagnetic Theory

Lund University
Lund, Sweden
2010

Department of Electrical and Information Technology
Electromagnetic Theory
Lund University
P.O. Box 118, S-221 00 Lund, Sweden

Series of licentiate and doctoral theses
No. 24
ISSN 1654-790X

© 2010 by Anders Bernland, except where otherwise stated.
Typeset in Computer Modern 8 pt using $\text{\LaTeX} 2_{\epsilon}$ and $\text{BIB}\text{\TeX}$.
Printed in Sweden by Tryckeriet i E-huset, Lund University, Lund.
April 2010

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

Physical processes are often modeled as input-output systems. Many such systems obey passivity, which means that power is dissipated in the process. This thesis deals with the inevitable constraints imposed on physical systems due to passivity. A general approach to derive sum rules and physical limitations on passive systems is presented. The sum rules relate the dynamical behaviour of a system to its static and/or high-frequency properties. This is beneficial, since static properties are in general easier to determine. The physical limitations indicate what can, and what can not, be expected from certain passive systems. At the core of the general approach is a set of integral identities for Herglotz functions, a function class intimately related to the transfer functions of passive systems.

In this thesis, the general approach is also applied to a specific problem: the scattering and absorption of electromagnetic vector spherical waves by various objects. Physical limitations are derived, which limit the absorption of power from each individual spherical wave. They are particularly useful for electrically small scatterers. The results can be used in many fields where an understanding of the interaction between electromagnetic waves and matter is vital. One interesting application is within antenna theory, where the limitations are helpful from a designer's viewpoint; they can give an understanding as to what factors limit performance, and also indicate if there is room for improvement or not.

Sammanfattning (in Swedish)

Fysikaliska processer modelleras ofta som insignal-utsignalsystem. Många sådana system är passiva, vilket betyder att effekt går förlorad i processen. Den här avhandlingen behandlar de ofrånkomliga begränsningar som passivitet sätter på fysikaliska system. Ett allmänt angreppssätt för att härleda summeringsregler och fysikaliska begränsningar på passiva system presenteras. Summeringsreglerna relaterar ett systems dynamiska uppträdande till statiska egenskaper och/eller högfrekvensegenskaper. Detta är ofta fördelaktigt, eftersom statiska egenskaper i allmänhet är lättare att bestämma. De fysikaliska begränsningarna antyder vad som kan, och vad som inte kan, förväntas från vissa passiva system. Kärnan i det allmänna angreppssättet är en samling integralidentiteter för Herglotzfunktioner, en funktionsklass som är intimt förknippad med passiva systems överföringsfunktioner.

I den här avhandlingen tillämpas även det allmänna angreppssättet på ett specifikt problem: diverse objekts spridning och absorption av elektromagnetiska sfäriska vektorvågor. Fysikaliska begränsningar, vilka begränsar absorptionen av effekt från varje enskild sfärisk våg, härleds. De är särskilt användbara för elektriskt små spridare. Resultaten kan användas inom många fält där det är viktigt att förstå samspelet mellan elektromagnetiska vågor och materia. En intressant tillämpning är inom antennteorin, där begränsningarna är behjälpliga ur en utvecklades perspektiv; de kan ge en förståelse för vilka faktorer som begränsar prestanda, och även antyda om det finns möjligheter till förbättring eller inte.

List of included papers

This thesis consists of a General Introduction and the following scientific papers which are referred to in the text by their roman numerals:¹

- I. A. Bernland, A. Luger, and M. Gustafsson. Sum Rules and Constraints on Passive Systems. Technical Report LUTEDX/(TEAT-7193)/1-28/(2010), Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2010. <http://www.eit.lth.se>.
- II. A. Bernland, M. Gustafsson, and S. Nordebo. Physical Limitations on the Scattering of Electromagnetic Vector Spherical Waves. Technical Report LUTEDX/(TEAT-7194)/1-23/(2010), Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2010. <http://www.eit.lth.se>.

Other publications by the author

- III. M. Gustafsson, G. Kristensson, A. Bernland, D. Sjöberg, and B. L. G. Jonsson. Physical bounds on the partial realized gain. The 26th International Review of Progress in Applied Computational Electromagnetics, Tampere, Finland, pp. 1-6, April 25-29, 2010. (In press)
- IV. 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. European Conference on Antennas and Propagation, Berlin, Germany, pp. 1795-1798, March 23, 2009.
- V. A. Bernland, M. Gustafsson, and S. Nordebo. Summation rules and physical bounds for partial wave scattering in electromagnetics. The 9:th International Conference on Mathematical and Numerical Aspects of Waves Propagation, Pau, France, pp. 206-207, June 15-19, 2009.²
- VI. M. Gustafsson, G. Kristensson, S. Nordebo, C. Larsson, A. Bernland, and D. Sjöberg. Physical bounds and sum rules in scattering and antenna theory. International Conference on Electromagnetics in Advanced Applications (ICEAA), Torino, pp. 600-603, September 14-18, 2009.
- VII. C. Sohl, M. Gustafsson, and A. Bernland. Some paradoxes associated with a recent sum rule in scattering theory. URSI General Assembly, Chicago, U.S., August 7, 2008.³

¹The order of the authors names indicates their relative contributions to the publications.

²Presented by the author of this thesis at the symposium.

³Appointed Commission B's Best Student Paper Prize at the URSI General Assembly, Chicago, U.S., August 7-16, 2008

Summary of included papers

Paper I - Sum Rules and Constraints on Passive Systems

This paper presents a general approach to derive sum rules and physical limitations on passive systems. Passive systems are related to Herglotz functions, and Paper I uses properties of this function class to derive a set of integral identities. These identities are the foundation for the sum rules and physical limitations. The general approach is described in detail, and several examples are included.

The author of this thesis has carried out most of the analysis.

Paper II - Physical Limitations on the Scattering of Electromagnetic Vector Spherical Waves

This paper employs the general approach presented in Paper I in order to derive physical limitations on the scattering and absorption of electromagnetic vector spherical waves. The limitations state that the reflection coefficients cannot be arbitrarily small over a whole wavenumber interval; how small is determined by the size, shape, and static material properties of the scatterer. The bounds can be interpreted as limits on the absorption of power from the individual vector spherical waves. They are particularly useful for 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.

The author of this thesis has carried out most of the analysis.

Preface and acknowledgments

This is a thesis for the degree of Licentiate in Engineering in Electromagnetic Theory. It summarises the research I have carried out from the start of my doctoral studies on August 27, 2007 and (almost) up until the time of print. All of my work has been done at the Department of Electrical and Information Technology within Lund University, Lund, Sweden.

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 two and a half years as a doctoral student so far 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 PhD-student and giving me all the support, guidance and collaboration a supervisor should plus infinitely more. His door is always open for me, despite his hectic schedule. He has been a great role-model. In particular, I admire his positive spirit, uncanny 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. Without her, we would still be fumbling in the dark. Many thanks goes to my co-supervisor Gerhard Kristensson, for sharing his vast knowledge in physics, mathematics and electromagnetic theory in the best possible way, and for discussions and guidance on undergraduate teaching. I also thank my co-supervisor Fredrik Tufvesson and former colleague Andrés Alayon Glazunov for introducing me to communication channels.

I also thank Sven Nordebo for collaboration on vector spherical waves and matching, and for the week in Pau. Two especially warm thank you goes to former colleague Christian Sohl, for opening up a vast area of interesting research and guiding me through it during my first year, and Anders Melin, for sharing his excellent knowledge in mathematics and long experience in scientific research. I am grateful to Christian also for the elegant $\LaTeX 2_{\epsilon}$ template used to assemble this thesis.

I am grateful for 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, Andrés Alayon Glazunov, Marius Cişmaşu, Mats Gustafsson, Andreas Ioannidis, Anders Karlsson, Alireza Kazemzadeh, Gerhard Kristensson, Christer Larsson, Buon Kiong Lau, Richard Lundin, Anders Melin, Sven Nordebo, Kristin Persson, Vanja Plicanic, Daniel Sjöberg, Christian Sohl, Anders Sunesson, Elsbietta Szybicka, Ruiyuan Tian and Niklas Wellander, for letting me into your close-knit group where everyone looks after each other in a remarkable way.

This work is financed by the High Speed Wireless Communications Center of the Swedish Foundation for Strategic Research (SSF). This generous grant is gratefully acknowledged. A travel grant from Stiftelsen Sigfrid och Walborg Nordkvist for participation in "The 9:th International Conference on Mathematical and Numerical Aspects of Waves Propagation" in Pau, France, is also acknowledged.

Last, but not least, I thank my family and friends, for supporting me and giving me a happy life outside work. 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.

Borlänge, Easter Sunday, 2010

Anders Bernland

Contents

Abstract	v
Sammanfattning (in Swedish)	vi
List of included papers	vii
Other publications by the author	vii
Summary of included papers	viii
Preface and acknowledgments	ix
Contents	xi
General Introduction	1
1 Introduction	3
2 Dispersion relations, sum rules and physical limitations	4
2.1 Systems on convolution form	4
2.2 Causality and dispersion relations	6
2.3 Passive systems and Herglotz functions	11
3 Physical limitations in antenna theory	14
3.1 Chu (1948)	16
3.2 Fano (1950)	18
3.3 Sohl et al. (2007)	19
3.4 Spherical wave scattering and MIMO	20
4 Concluding remarks	23
I Sum Rules and Constraints on Passive Systems	29
1 Introduction	31
2 Herglotz functions and integral identities	32
3 Sum rules for passive systems	34
4 Proof of the integral identities	39
5 Examples	43
5.1 Elementary Herglotz functions	43
5.2 Lossless resonance circuit	44
5.3 Reflection coefficient (Fano's matching equations revisited)	45
5.4 Kramers-Kronig relations and ϵ near-zero materials	46
6 Conclusions	49
A Proofs	50
A.1 Calculation of the limits $\lim_{z \rightarrow \infty} h(z)/z$ and $\lim_{z \rightarrow 0} zh(z)$	50
A.2 Proof of Lemma 4.1	51
A.3 Proof of Corollary 4.1	53
A.4 Proof of Lemma 4.2	54
A.5 Proof of Theorem 4.1	55
A.6 Auxiliary theorems	56
II Physical Limitations on the Scattering of Electromagnetic Vector Spherical Waves	59
1 Introduction	61
2 A general approach to obtain sum rules and physical limitations on passive systems	62
2.1 Herglotz functions and integral identities	62

2.2	Sum rules for passive systems.	63
3	Vector spherical waves in the time and frequency domains	66
4	The scattering matrix $\tilde{\mathbf{S}}_S$	68
4.1	Implications of passivity on $\tilde{\mathbf{S}}_S$	69
4.2	Low-frequency asymptotic behaviour of $\tilde{\mathbf{S}}_S$	69
4.3	The polarizability dyadics and bounds on $\rho_{\nu,\nu}$	71
4.4	Sum rules and physical limitations on $\tilde{\mathbf{S}}_S$	72
5	Examples	75
5.1	Absorbing spherical nanoshells	75
5.2	Physical limitations on antennas	75
6	Conclusions	77
7	Acknowledgments	79
A	Definitions and derivations	79
A.1	Definition of vector spherical waves	79
A.2	Derivation of (4.6)	81
A.3	Derivation of (4.15)	82

General Introduction

Anders Bernland



1 Introduction

In physical sciences, there is an ambition to model various aspects of nature. Many physical processes are modeled 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. 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 ω (with unit Hertz, Hz).

Many physical systems obey passivity; that is, they cannot produce energy. If they do not consume energy either, they are called lossless. Passivity poses severe constraints on a system. As a result, dispersion relations may be derived, effectively describing realisable frequency dependencies of physical systems. This is a consequence of a classical result that states that the transfer functions of passive systems are related to the well studied class of Herglotz functions, see *e.g.*, [54, 56, 58]. In some cases, a set of sum rules follow the dispersion relations; in essence, they relate the dynamical behaviour of a system to its static and/or high-frequency properties. This is beneficial, since static properties are in general easier to determine. The sum rules can also be used to derive physical limitations, or constraints, indicating what can and cannot be expected from the system. In Section 2 of this General Introduction, the concept of input output systems in general, and those on convolution form in particular, are discussed. Furthermore, three different approaches to derive dispersion relations are considered. A general approach to derive sum rules and physical limitations on passive systems put forth in Paper I is also reviewed briefly.

In the second part of the thesis (Paper II), the results of Paper I are applied to a specific problem: the scattering and absorption of electromagnetic waves by various objects. Physical limitations for this interaction are derived in Paper II. They quantify the intuitively obvious statement that objects that are small compared to the wavelength can only absorb a limited amount of power. Understanding electromagnetic wave interaction with matter is vital in many applications, from classical optics to stealth technology. Recently, much attention have been devoted to so called metamaterials, synthetic materials designed to have extra-ordinary electromagnetic properties. The results of Paper II can potentially be used within all the mentioned fields.

Another interesting application is within antenna theory. This is discussed in Section 3 of this General Introduction, where also previous approaches to find physical limitations on antenna performance are reviewed. The physical limitations can be very helpful from a designer point of view, both to understand what factors limit performance, but also to determine if there is room for improvement.

2 Dispersion relations, sum rules and physical limitations

This section discusses general models of physical processes as input output systems. In particular, systems on convolution form, *i.e.*, systems that satisfy the basic assumptions of linearity, time-translational invariance and continuity, are considered. These systems are fully described by their impulse responses in the time domain, or equivalently by their transfer functions in the frequency domain.

One way to study these systems are to derive so called dispersion relations, quantifying their frequency dependence. This requires some extra assumptions on the system. Three approaches to derive dispersion relations, relying on somewhat different assumptions, are described briefly in Section 2.2. One of them relies on the assumption that the system is passive. From the dispersion relations, sum rules and physical limitations can sometimes be derived. This is described for passive systems in Paper I, and the procedure is outlined in Section 2.3. For a discussion on dispersion relations, see also the General Introduction in Sohl’s doctoral thesis [50] and references therein.

An early example of dispersion relations are the Kramer-Kronig relations, relating the real and imaginary part of the electric permittivity $\epsilon(\omega)$ to each other. They were derived independently by Kramers in [34] and Kronig in [11], see *e.g.*, [35] for a review of their results. The Kramers-Kronig relations can be used to derive sum rules as well. Classic examples of sum rules and physical limitations used within electromagnetic theory are Fano’s matching equations, presented in [14]. There are more recent examples as well, see *e.g.*, [6, 18, 20, 21, 23, 43, 48]. There are also sum rules within quantum mechanical scattering, see *e.g.*, [52].

2.1 Systems on convolution form

Systems on convolution form are discussed in this section, inspired by the book [58] by Zemanian. See also the books [45–47] by Schwartz. As mentioned in the introduction, many physical systems are modeled as 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 1. Here the signals are functions of time, t . It is desirable to allow u and v to be generalised functions, or distributions¹, *i.e.*, the domain $D(\mathcal{R})$ of the operator \mathcal{R} is some subset of \mathcal{D}' . This allows the modeling of functions having point support, *i.e.*, signals 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.

¹An introduction to distribution theory can be found in the books [47], [17] and [51]. More thorough treatises are *e.g.*, [12, 30, 45, 46, 58].

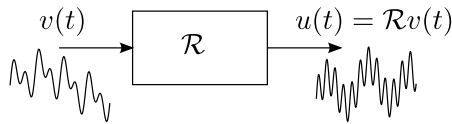


Figure 1: The physical system (2.1) relates the input signal $v(t)$ to the output signal $u(t)$.

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 “if you double the input, you double the output”.

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.

Continuity: An operator is continuous if

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

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 [58]. An interpretation of continuity is that a small change in the input signal only leads to a small change in the output signal.

It can be shown that a system satisfies these assumptions if and only if it is on convolution form, (*cf.*, Theorem 5.8-2 in [58] and pages 134–140 in [47]):

$$u(t) = w * v(t) = \int_{\mathbb{R}} w(t')v(t-t') dt', \quad (2.2)$$

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 [58]. The generalised function w is called the impulse response of the system, and it contains a complete description on the systems properties. It is clear now that the three assumptions of linearity, time-translational invariance and continuity can be replaced by one assumption: the assumption that the system is on convolution form.

In many applications, it is desirable to study physical systems in the frequency domain, *i.e.*, the Fourier transform² is applied to equation (2.2). In general, the Fourier transform $\tilde{f}(\omega) = (\mathcal{F}f)(\omega)$ measures the frequency dependence of a system. For example, if $f(t)$ is a representation of a sound wave, $\tilde{f}(\omega)$ states which frequencies (or tones) that are present in the sound.

The Fourier transform of (2.2) is

$$\tilde{u}(\omega) = \tilde{w}(\omega)\tilde{v}(\omega), \quad (2.3)$$

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.³ Equation (2.3) reveals one reason to study systems in the frequency domain; multiplication is in general preferable over convolution.

2.2 Causality and dispersion relations

As seen in the previous section, the impulse response $w(t)$ of a system contains a complete description of the systems behaviour. This is evidently also true for the transfer function $\tilde{w}(\omega)$. Scrutiny of one of these two functions is therefore a reasonable way to study a given physical system. A physical system with a frequency dependent transfer function $\tilde{w}(\omega)$ (as opposed to a constant transfer function $\tilde{w}(\omega) \equiv C$) is often called dispersive. Relations for this frequency dependency for systems satisfying certain assumptions are called dispersion relations. In this section, possible candidates for these assumptions are discussed. Three sets of assumptions are presented, which lead to three distinct approaches to derive dispersion relations.

One critical assumption on a physical system is:

Causality: The system (2.1) is causal if

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

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

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

²Here the Fourier transform of an integrable function $f(t)$ is defined as

$$\tilde{f}(\omega) = (\mathcal{F}f)(\omega) = \int_{\mathbb{R}} f(t)e^{i\omega t} dt.$$

For references concerning Fourier transforms of more general functions or distributions, see footnote 1.

³When both w and v are integrable functions, their Fourier transform are well defined and (2.3) applies. But it should be mentioned that some extra assumptions on w and v are needed in the case they are distributions. 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}' . If in addition w or v *e.g.*, have compact support, the system (2.2) is mapped to (2.3) under the Fourier transform. See the references of footnote 1.

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

For many physical systems it is obvious that causality holds; the action cannot precede the cause. However, it turns out that causality in itself is often not enough to obtain dispersion relations; more assumptions are required, and one candidate is

Rational transfer function: The transfer function $\tilde{w}(\omega)$ in (2.3) is a rational function if it is the quotient of two polynomials:

$$\tilde{w}(\omega) = \frac{c_n \omega^n + c_{n-1} \omega^{n-1} + \dots + c_1 \omega + c_0}{d_m \omega^m + d_{m-1} \omega^{m-1} + \dots + d_1 \omega + d_0}. \quad (2.4)$$

For example, impedance functions $Z(\omega)$ realisable with a finite number of lumped circuit elements are rational functions.

If a system is causal with a rational transfer function, then the transfer function $\tilde{w}(z)$ is holomorphic in the upper half-plane $\mathbb{C}^+ = \{z : \text{Im } z > 0\}$. Throughout this General introduction, z denotes a complex number, with $\omega = \text{Re } z$ and $y = \text{Im } z$. A holomorphic function (sometimes referred to as an analytic function) is a function of the complex variable $z \in \mathbb{C}$ that is complex-differentiable. As a result, very powerful tools from complex analysis can be employed to derive dispersion relations.⁴ However, even the electric engineer encounters non-rational transfer functions; one situation is if a time delay $e^{i\omega t_0}$ is introduced, for example by a transmission line. Also, the scattering of electromagnetic waves is in general modeled with non-rational transfer functions.

There is another candidate that can replace the assumption of a rational transfer function, namely:

Square-integrable transfer function: The transfer function in (2.3) is square-integrable ($\tilde{w} \in L^2$) if it is a regular function and

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

For square-integrable functions, the following theorem applies [32, 41]:

Theorem 2.1 (Titchmarsh's theorem). *Let $\tilde{f}(\omega)$ be a square-integrable function on the real line. If $\tilde{f}(\omega)$ satisfies one of the four conditions below, then it fulfills 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:

$$\text{Re } \tilde{f}(\omega) = -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0} \int_{|\omega - \xi| > \varepsilon} \frac{\text{Im } \tilde{f}(\xi)}{\omega - \xi} d\xi. \quad (2.5)$$

⁴For an introduction to complex analysis, see *e.g.*, [1, 16].

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

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

4. The function $\tilde{f}(z)$ is a holomorphic function in the open upper half-plane $\mathbb{C}^+ = \{z : \operatorname{Im} z > 0\}$. Furthermore, it holds that

$$\tilde{f}(\omega) = \lim_{y \rightarrow 0^+} \tilde{f}(\omega + iy), \quad \text{for almost all } \omega \in \mathbb{R},$$

and

$$\int_{\mathbb{R}} |\tilde{f}(\omega + iy)|^2 d\omega < \infty, \quad \text{for } y > 0.$$

If $\tilde{f} = \tilde{w}$ is the transfer function of a causal system, the two Plemelj formulae (2.5)–(2.6) directly give dispersion relations for the system by relating the real and imaginary parts to each other. More relations can be derived by exploiting the holomorphic properties of $\tilde{w}(z)$ in the open upper half-plane due to point 4 in the theorem.

The Plemelj formulae closely resembles the Hilbert transform, \mathcal{H} , which is defined for functions $F(x)$ on \mathbb{R} that are *e.g.*, square-integrable and locally integrable, as:

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

see *e.g.*, [32, 37]. Its inverse, under the above assumptions on F , is

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

The functions $F(x)$ and $\mathcal{H}F(x)$ in (2.7)–(2.8) constitute a Hilbert transform pair, and it is evident from Titchmarsh's theorem that the real and imaginary parts of a causal transform \tilde{w} constitute such a pair. Properties of the well-studied Hilbert Transform can thus be used to derive dispersion relations for physical systems. For a summary of integral relations connected to this approach, see *e.g.*, [32, 33, 36] and references therein.

It is in place here to discuss the meaning of the assumption that \tilde{w} is square-integrable (in L^2). A square-integrable function can be interpreted as a function of finite energy. Therefore, the input and output signals are frequently assumed to be in L^2 . The transfer function, on the other hand, need in general *not* be in L^2 .

The Titchmarsh/Hilbert transform approach to derive dispersion relations can be generalised to much larger classes of transfer functions, see [32, 33, 36, 41] and references therein. However, there is an alternative approach, that makes use of another fundamental property of many physical systems: passivity. 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 modeled, the definition of passivity comes in different forms. The two forms considered here have names borrowed from electric circuit theory:

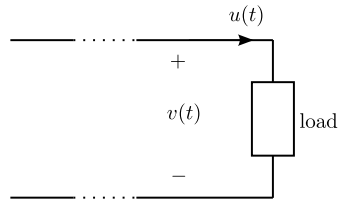


Figure 2: Here the input $v(t)$ is the electric voltage over the load, and the output $u(t)$ is the electric current running through it. They are related by the admittance operator of the load.

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

$$e_{\text{adm}}(T) = \text{Re} \int_{-\infty}^T u^*(t)v(t) dt \quad (2.9)$$

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

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

$$e_{\text{scat}}(T) = \int_{-\infty}^T |v(t)|^2 - |u(t)|^2 dt \quad (2.10)$$

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

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 for other admissible input signals $v \in D(\mathcal{R})$. The above definition of scatter-passivity was introduced by Youla et al. in [56], while the definition of admittance-passivity was introduced by Zemanian in [57]. The connection between them is discussed by Wohlers and Beltrami in [54]. Both passivity concepts have been generalised to a Hilbert space setting, see [59] and references therein.

If the input signal $v(t)$ is the electric voltage over a load and the output signal $u(t)$ is the electric current running through it, then the operator \mathcal{R} is the so called admittance operator. See Figure 2. In this case, the electric energy absorbed by the load until time T is given by (2.9). Thus, the admittance operator of a passive circuit elements is an admittance-passive operator, as the name suggests. Note that admittance-passivity might as well have been called impedance-passivity, since the current could have been the input and the voltage the output in the example.

Consider now a transmission line ended in a load. Let $v(t)$ be the amplitude of the voltage wave traveling towards the load (measured by the load), and let the output $u(t)$ be the amplitude of the reflected wave, as in Figure 3. In this case the electric energy absorbed until time T is given by (2.10). Hence, passive reflection operators (or scatter operators) are scatter-passive.

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. Passivity has

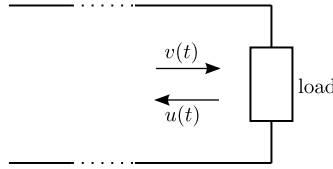


Figure 3: In this case, the input $v(t)$ and output $u(t)$ are the amplitudes (measured by the load) of the voltage waves traveling along the transmission line towards and from the load, respectively. They are related by the reflection operator, also called the scattering operator.

far-reaching implications on the physically realisable behaviour of a system. One consequence is that passive systems must also be causal. Another is that the impulse response w must be in \mathcal{S}' , and thus it is Fourier transformable in the distributional sense (see footnote 3). Combined, it guarantees that the transfer function $\tilde{w}(z)$ is well-defined also for z in the open upper half-plane $\mathbb{C}^+ = \{z : \text{Im } z > 0\}$, and that it is holomorphic there. Furthermore, the transfer function $\tilde{w}_{\text{adm}}(z)$ of an admittance-passive system satisfies $\text{Re } \tilde{w}_{\text{adm}}(z) \geq 0$ in \mathbb{C}^+ , while $|\tilde{w}_{\text{scat}}(z)| \leq 1$ in \mathbb{C}^+ when $\tilde{w}_{\text{scat}}(z)$ is the transfer function of a scatter-passive system. See *e.g.*, [54, 56, 58]. These properties of the transfer functions imply that they can be related to Herglotz functions, as described in the next section. Properties of the well-studied Herglotz functions is the starting point to derive dispersion relations for such systems. In Paper I, sum rules and physical limitations on passive systems are derived from there.

Summing up, the three approaches to derive dispersion relations for systems on convolution form discussed in this section are:

1. **The rational function approach:** This approach relies *e.g.*, on the assumption that the system is causal and that the transfer function is rational. It derives dispersion relations using straightforward complex analysis.
2. **The Titchmarsh's theorem/Hilbert transform approach:** It employs Titchmarsh's theorem or the Hilbert transform, and requires that the system is causal and that the transfer function is *e.g.*, square-integrable. It can be generalised to larger classes of transfer functions.
3. **The passive systems approach:** This approach assumes that the system is passive (and thereby causal). It relates the transfer functions to Herglotz functions, and derives dispersion relations from there.

Note here that the concept of causality is crucial to all three approaches. It should be stressed that since the approaches rely on different assumptions, they are complementary rather than in competition. For all the approaches, sum rules and physical limitations can sometimes be derived from the dispersion relations. This is described in more detail for the passive systems approach in the next section.

2.3 Passive systems and Herglotz functions

In this sections, Herglotz functions are presented, and their relation to the transfer functions of passive systems is clarified. Following a well-known representation theorem, Herglotz functions are represented by positive measures on the real line. This representation can be interpreted as a dispersion relation for passive systems. Furthermore, a set of integral identities for Herglotz functions are derived in Paper I from this representation. For physical systems, these are referred to as sum rules, relating dynamical behaviour to static and/or high frequency properties. One way to make use of the sum rules is to derive physical limitations by estimating the integrals.

The class of Herglotz functions is now introduced. Start with the definition:

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

In other words, they are complex differentiable mappings of the open upper half-plane to the closed upper half-plane. Herglotz functions are sometimes referred to as Nevanlinna [27], Pick [12], or R-functions [31]. They are closely related to positive harmonic functions and the Hardy space $H^\infty(\mathbb{C}^+)$ [13, 37], and appear within the theory of continued fractions and the problem of moments [2, 28], but also within functional analysis and spectral theory for self-adjoint operators [3, 27]. Because of this, they have been thoroughly studied. The aforementioned representation theorem is commonly attributed to Nevanlinna's paper [39], but it was presented in its final form by Cauey in [7]. See also [3] for a proof and discussion.

Theorem 2.2. *A necessary and sufficient condition for a function h to be a Herglotz function is that*

$$h(z) = \beta z + \alpha + \int_{\mathbb{R}} \left(\frac{1}{\xi - z} - \frac{\xi}{1 + \xi^2} \right) d\mu(\xi), \quad \text{Im } z > 0, \quad (2.11)$$

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

Note the resemblance of (2.11) to the Hilbert transform (2.7). The representation theorem follows from a similar representation theorem for positive harmonic functions on the unit disk due to Herglotz [29], hence the name Herglotz functions. A photograph of Gustav Herglotz is shown in Figure 4.

The Lebesgue integral over the measure μ in (2.11) is a generalisation of the Riemann integral. The Lebesgue integral is more complete in a sense, and thus often appears in representation theorems such as Theorem 2.2. The interested reader can find an introduction to measure and integration theory in the book [4] by Berezansky et al., and the book [44] by Rudin.

Since the transfer function $\tilde{w}(z)$ of an admittance-passive system is holomorphic and has a non-negative real part for $z \in \mathbb{C}^+$, a Herglotz function is given by

$$h(z) = i\tilde{w}(z). \quad (2.12)$$

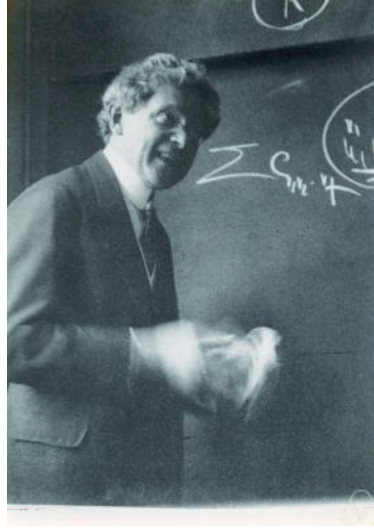


Figure 4: The German mathematician Gustav Herglotz (1881–1953) in the year 1930. The photograph is courtesy of Konrad Jacobs, reproduced here under license <http://creativecommons.org/licenses/by-sa/3.0/de/deed.en>.

For a scatter-passive system, a Herglotz function can be constructed by applying the inverse Cayley transform $z \mapsto (iz + i)/(1 - z)$ to $\tilde{w}(z)$. Alternatively, the complex logarithm may be used, see Paper I. For clarity, only admittance-passive systems are considered in this section. Following (2.12), a Herglotz function can be thought of as a generalised admittance, or impedance, function. The difference is that it is not necessarily rational, and thus can not in general be realised with a finite number of lumped circuit elements. The connection between the transfer functions of passive systems and Herglotz functions is well known, see *e.g.*, [54, 56, 58]. Note that some authors prefer the Laplace transform and the related function class of Positive Real (PR) functions over the Fourier transform and Herglotz functions.

The representation (2.11) is in a sense a dispersion relation. The reason is that the measure μ can be interpreted as the imaginary part of h , see Lemma 4.1 in Paper 1 and the discussion following the lemma. For example, when the measure is given by a continuous function $\mu'(\xi)$, *i.e.*, $d\mu(\xi) = \mu'(\xi) d\xi$, then $\lim_{y \rightarrow 0^+} h(x + iy) = \mu'(x)$ for almost all $x \in \mathbb{R}$. Additional dispersion relations can be derived by composing Herglotz functions with each other.

From the representation, it also follows that all Herglotz functions have low- and

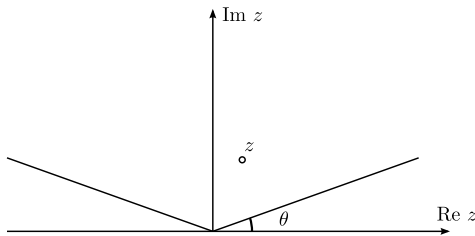


Figure 5: The Stoltz domain, $\{z : \theta \leq \arg z \leq \pi - \theta\}$ for some $\theta \in (0, \pi/2]$.

high-frequency asymptotic expansions of the forms:

$$h(z) = \sum_{n=-1}^{2N-1} a_n z^n + o(z^{2N-1}), \quad \text{as } z \hat{\rightarrow} 0, \quad (2.13)$$

$$h(z) = \sum_{m=1}^{1-2M} b_m z^m + o(z^{1-2M}), \quad \text{as } z \hat{\rightarrow} \infty, \quad (2.14)$$

where all $a_n, b_m \in \mathbb{R}$ and $N, M \geq 0$. Here $z \hat{\rightarrow} 0$ is a short-hand notation for $|z| \rightarrow 0$ in the Stoltz domain $\theta \leq \arg z \leq \pi - \theta$ for any $\theta \in (0, \pi/2]$, see Figure 5, and likewise for $z \hat{\rightarrow} \infty$. 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.

The main results of Paper I are the following integral identities for a Herglotz function h :

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{\varepsilon < |\omega| < \varepsilon^{-1}} \frac{\text{Im } h(\omega + iy)}{\omega^p} d\omega = a_{p-1} - b_{p-1}, \quad p = 2 - 2M, 3 - 2M, \dots, 2N. \quad (2.15)$$

The left-hand side of (2.15) is the integral of $\text{Im } h(\omega)/\omega^p$ in the distributional sense, *i.e.*, contributions from possible singularities in the interval $(0, \infty)$ are included, *cf.*, the discussion in Paper I. The derivation of the integral identities (2.15) for $p = 2, 3, \dots, 2N$ rely on two results; the first (Corollary 4.1 of Paper I) relates the left-hand sides to moments of the measure μ , while the other (Lemma 4.2) relates the convergence and explicit values of these moments to the expansion (2.13). A change of variables in the left-hand side of (2.15) enables a proof for $p = 2 - 2M, 3 - 2M, \dots, 1$. If the Herglotz function is a rational function, the identities (2.15) follow from the Cauchy integral formula. This derivation can be found in [49].

When the Herglotz function h is given by (2.12), the integral identities (2.15) constitute a set of sum rules for $\tilde{w}(\omega)$. A sum rule relates a sum of the dynamical behaviour of the system (the left-hand side in (2.15)) to its low- and/or high-frequency properties (the coefficients a_n and b_n in the right-hand side)⁵. As mentioned in the

⁵This is the meaning of the term “sum rule” adopted in this thesis. Elsewhere, the term can have a wider meaning, where the trademark of a sum rule is that one of its sides is a sum or integral (generalised sum).

introduction, this is very beneficial, since static properties are in general easier to determine than dynamical behaviour.

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

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

For many physical systems, this is taken for granted. Reality implies that the impulse response $w(t)$ is real, which in turn implies the symmetry $h(z) = -h^*(-z^*)$ when $h(z)$ is given by (2.12). This restricts the identities (2.15) to even powers and simplifies them to

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{y \rightarrow 0^+} \frac{2}{\pi} \int_{\varepsilon}^{\varepsilon^{-1}} \frac{\text{Im } h(\omega + iy)}{\omega^{2\hat{p}}} d\omega = a_{2\hat{p}-1} - b_{2\hat{p}-1}, \quad \hat{p} = 1 - M, \dots, N. \quad (2.16)$$

The identities (2.16) are the starting points to derive physical limitations on a system. Since the integrands are non-negative and the integrals over the whole real line are equal to the right-hand sides, the integrals over any subset of the real line must be bounded by the right-hand sides. Let the frequency band be $\mathcal{B} = [\omega_0(1 - B/2), \omega_0(1 + B/2)]$, with center frequency ω_0 and fractional bandwidth B . Then the following physical limitations may be derived from (2.16):

$$\frac{2\omega_0^{1-2\hat{p}} B}{\pi} \inf_{\mathcal{B}} \text{Im } h(\omega) \leq a_{2\hat{p}-1} - b_{2\hat{p}-1}, \quad \hat{p} = \max(1 - M, 0), 2, 3, \dots, N. \quad (2.17)$$

(For $p = 1 - M, 2 - M, \dots, -1$ the bounds have a slightly different appearance.) The bounds state that the imaginary part of h cannot be arbitrarily large over a frequency band. Often, the imaginary part of h models the losses of the physical system; for an impedance $\tilde{w}(\omega) = Z(\omega)$ in (2.12), the imaginary part of h is the real part of Z . The concept of sum rules and physical limitations is illustrated schematically in Figure 6.

3 Physical limitations in antenna theory

In the second part of this thesis, Paper II, the results of Paper I is employed in order to find physical limitations on the scattering of electromagnetic waves by various objects. Electromagnetic waves can be expanded in a sum of orthogonal vector spherical waves, also referred to as partial waves, (electric and magnetic) multipoles, or (TM and TE) modes. In Paper II, physical limitations on the scattering and absorption on the individual waves are derived. Other physical limitations on electromagnetic scattering have been derived recently by Sohl et al. [48]; instead of considering spherical waves, Sohl et al. derives limitations on the total scattering and absorption of an electromagnetic wave. The limitations on electromagnetic scattering presented in Paper II and [48] are particularly useful for electrically small scatterers, and are therefore well suited to small structures designed to be resonant in one or more frequency bands.

One example of a resonating structure is an antenna. The results of [48] have been applied to antenna theory in [20]. The implications of the limitations of Paper II

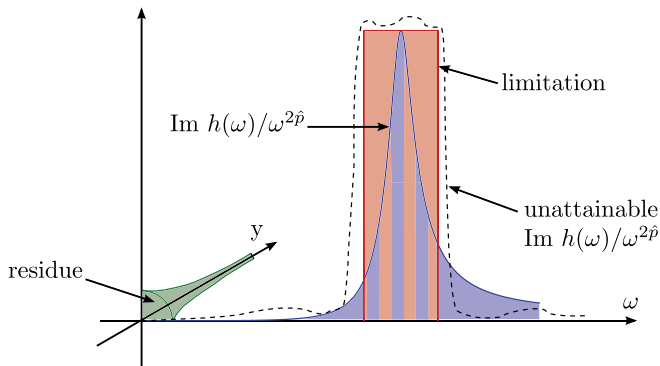


Figure 6: Illustration of the sum rules (2.16) and physical limitations (2.17). The integrals of $\text{Im } h(\omega)/\omega^{2\hat{p}}$ in the left-hand sides of (2.16) are determined by the low-frequency coefficients $a_{2\hat{p}-1}$ and/or high-frequency coefficients $b_{2\hat{p}-1}$. In an analogy to the calculus of residues, this can be interpreted as contributions from residues at the origin and infinity in the complex z -plane (green in the figure). Since the total integrals (blue) are equal to these residues, the curves have to intersect the boxes with the same area (red). This gives the physical limitations (2.17). Also shown in the figure is an unattainable curve of $\text{Im } h(\omega)/\omega^{2\hat{p}}$ (dashed).

on antenna performance is discussed briefly in Example 5.2 in Paper II. There are also numerous other publications addressing limitations on antenna performance; many of these fall into one of the following two categories: Either they are based on the pioneering paper [8] by Chu, or they use Fano's theory of optimal wideband matching presented in [14].

It is in general difficult to describe common properties of antennas; since different antennas are designed for completely different purposes, the behaviour of two antennas can differ radically. Furthermore, in many applications, the antennas are influenced by other objects nearby. For example, in a mobile phone the antenna must co-exist with the batteries, speaker, camera and so forth, see Figure 7. The physical limitations is one way to quantify some properties that all antennas satisfy by stating what can, and what can not, be achieved in terms of performance under certain constraints. One significant constraint that limits antenna performance is the size; this is intuitively reasonable, since objects that are small compared to the wavelength can only provide a limited interaction with electromagnetic waves. This section briefly summarizes the four mentioned approaches to find limitations on antenna performance: the approach based on Chu's paper in Section 3.1, limitations due to Fano's matching theory in Section 3.2, and the results due to Sohl et al. in Section 3.3. Finally, the spherical wave approach due to Paper II is covered in Section 3.4, and its connection to multiple-input multiple-output (MIMO) systems is discussed.

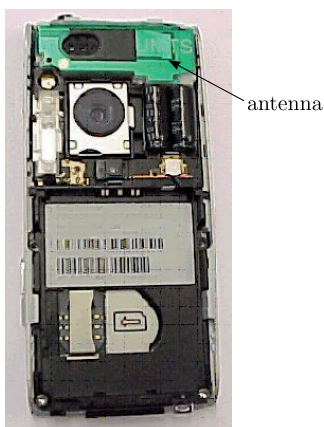


Figure 7: The back of a Sony Ericsson K800 mobile phone. The author thanks Anders Sunesson for the photograph.

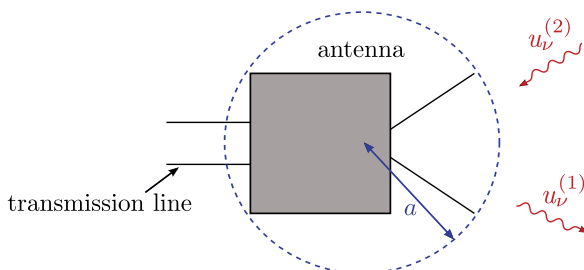


Figure 8: In Chu's paper [8] as well as in Paper II, the antenna is contained in a hypothetical sphere of radius a . Outside this sphere, the electric and magnetic fields are expanded in outgoing ($\mathbf{u}_\nu^{(1)}$) and incoming ($\mathbf{u}_\nu^{(2)}$) vector spherical waves, or modes, with index ν .

3.1 Chu (1948)

In his paper [8], Chu laid the foundation for much of the coming work on fundamental limitations on antenna performance. He circumscribed the antenna with a hypothetical sphere of radius a , and expanded the electric and magnetic fields into orthogonal vector spherical waves, or modes, see Figure 8. Furthermore, Chu derived lumped element circuit-equivalents of the respective modes; they take the form of ladders, where the length of the ladder (and hence the complexity of the circuit) is increased for higher order modes.

A radiating antenna is surrounded by electric and magnetic fields, the so called near-field. Chu considered the radiated power P_{rad} compared to the stored electric and magnetic energy, W_e and W_m , of the near field. He defined the quality factor,

Q , as [8, 26]

$$Q(\omega) = 2\omega \frac{\max(W_e(\omega), W_m(\omega))}{P_{\text{rad}}(\omega)}.$$

At the resonance frequency ω_0 of the antenna, there are equal amounts of stored electric and magnetic energy:

$$Q(\omega_0) = 2\omega_0 \frac{W_e(\omega_0)}{P_{\text{rad}}(\omega_0)}. \quad (3.1)$$

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 the half power bandwidth of the antenna impedance [26]. If it is low, it is an indication of a broadband antenna, *i.e.*, an antenna that can operate over a wide frequency band.

Chu considered linearly polarized omnidirectional antennas, and stated that an antenna radiating like an electric dipole (*i.e.*, only the lowest order TM-modes are present) yields a minimum Q . He determined this minimum $Q_{\text{TM,min}}$ as:

$$Q_{\text{TM,min}} = \frac{1}{k_0^3 a^3} + \frac{1}{k_0 a} \quad (3.2)$$

where a is the radius of the circumscribing sphere, $k_0 = \omega_0/c$ the resonant wavenumber, and c the speed of light in free space. This equation is not stated explicitly in [8], but a direct consequence of the results presented there, see *e.g.*, [38]. Collin and Rothschild derived closed form expressions for the minimum Q for all modes and circular polarisation in [9]. McLean re-derived the minimum Q of circularly polarized antennas:

$$Q_{\text{min}} = \frac{1}{2k_0^3 a^3} + \frac{1}{k_0 a}.$$

Yaghjian and Best [55] propose an alternative quality factor Q_Z expressed in the antenna impedance $Z(\omega)$:

$$Q_Z(\omega_0) = \frac{\omega_0}{2R} |Z'(\omega_0)|. \quad (3.3)$$

Here $R = Z(\omega_0)$ is the real-valued impedance at the resonance frequency, and a prime (') denotes differentiation with respect to the argument. For many antennas, $Q_Z(\omega_0) \approx Q(\omega_0)$, but this is not generally applicable [22]. One advantage of Q_Z over Q is that (3.3) is often easier to evaluate for real antennas than (3.1).

Harrington [26], as well as Geyi [15], discusses limitations on directivity and quality factor. Directivity is the quotient of the power radiated in the desired direction to the total power radiated. Notwithstanding the references cited above, there have been numerous other publications addressing limitations on antennas based on the original paper [8] by Chu. A summary of some important results can be found in Hansen's book [24].

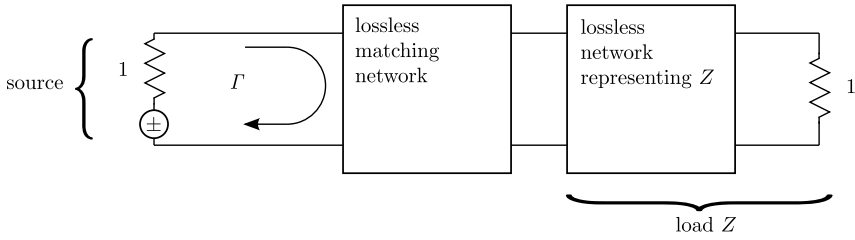


Figure 9: The matching problem as described by Fano in [14]. The internal resistance of the source as well as the resistance stemming from the representation of the load Z can be normalised to 1.

3.2 Fano (1950)

Fano is perhaps most known for his work within information theory, but his doctoral dissertation is concerned with electrical networks. This work has also been published in [14]. More specifically, he studied the problem of matching a source to a load over a frequency band. When a source is connected to a load, some, or all, of the power delivered by the source will be rejected by the load:

$$P_{\text{rejected}} = |\Gamma|^2 P_{\text{source}}.$$

Here Γ denotes the reflection coefficient. This is of course undesirable, since it diminishes efficiency. Furthermore, it may cause non-linearities and damage the source.

A given source may be matched perfectly to a load at one specific frequency. But if the matching network must be lossless (*i.e.*, neither producing nor consuming power), the source and load can not be matched over a whole frequency band. To investigate the limits, Fano used a representation presented by Darlington in [10], stating that a load may be represented as a lossless network terminated in a pure resistance, see Figure 9.

Fano only considered lumped circuit elements, and thus the impedance $Z(\omega)$ of the load, as well as the reflection coefficient $\Gamma(\omega)$, were rational. He could therefore derive sum rules for $\ln |\Gamma(\omega)|$ with the Cauchy integral formula. The sum rules are known under the name Fano's matching equations. They can also be obtained using the integral identities (2.15) for Herglotz functions, as done in Example 5.3 in Paper I. From the sum rules, Fano derived physical limitations on the reflection coefficient Γ . The limitations are sometimes referred to as the Bode-Fano limits, due to similar work by Bode in [5]. Fano also addressed the problem how the lossless matching network should be designed in order to obtain optimal matching.

To use Fano's equations for antenna matching, a model for the antenna impedance is required. The impedance $Z(\omega)$ of many antennas can be approximated by the resonance circuit in Figure 10 close to the resonance frequency ω_0 . Using Fano's limitations, it is straightforward to show that

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

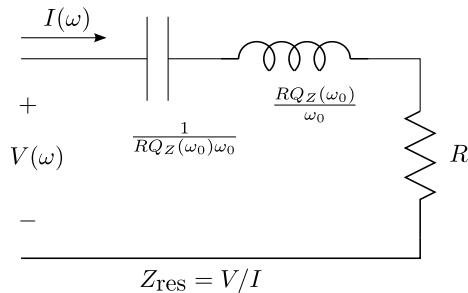


Figure 10: For many antennas, the impedance $Z_{\text{res}}(\omega)$ of the resonance circuit is a good approximation for the antenna impedance $Z(\omega)$ close to its resonance frequency ω_0 [22]. The quality factor $Q_Z(\omega_0)$ is given by (3.3).

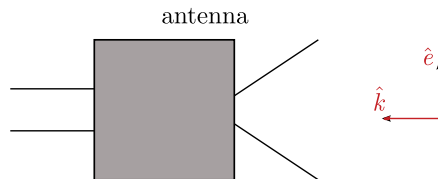


Figure 11: Sohl et al. consider a plane wave impinging in the $\hat{\mathbf{k}}$ direction on an arbitrary scatterer in [48], and on an antenna in [20].

where the frequency band is $\mathcal{B} = [\omega_0(1 - B/2), \omega_0(1 + B/2)]$ with center frequency ω_0 and fractional bandwidth B . This is treated in detail by Gustafsson and Nordebo in [22], where also the validity of the approximation by resonance circuits is discussed. Fano matching can be combined with the limits on Q in Section 3.1 to yield Chu-Fano limitations, see [20]. Fano matching of antennas is also considered in *e.g.*, [25, 53].

3.3 Sohl et al. (2007)

A different approach to find limitations on antennas is adopted by Sohl et al. In [48], they consider electromagnetic plane waves impinging on an antenna or other scatterer, see Figure 11. The absorption cross section σ_a is a measure on the total power absorbed from the wave. Some power will also be scattered, and the amount is measured by the scattering cross section σ_s . The sum of the absorption and scattering cross sections is the extinction cross section,

$$\sigma_e = \sigma_a + \sigma_s.$$

Using the optical theorem (see *e.g.*, [40]) and Cauchy integrals, they derive the following sum rule for the extinction cross section when the incoming wave is linearly

polarized:

$$\int_0^\infty \frac{\sigma_e(k, \hat{\mathbf{k}}, \hat{\mathbf{e}})}{k^2} dk = \frac{\pi}{2} \left(\hat{\mathbf{e}} \cdot \boldsymbol{\gamma}_e \cdot \hat{\mathbf{e}} + (\hat{\mathbf{k}} \times \hat{\mathbf{e}}) \cdot \boldsymbol{\gamma}_m \cdot (\hat{\mathbf{k}} \times \hat{\mathbf{e}}) \right), \quad (3.4)$$

where $k = \omega/c$ is the wavenumber. Here $\hat{\mathbf{e}}$ is the electric polarization of the incoming plane wave, and $\hat{\mathbf{k}}$ the direction of propagation for the wave, see Figure 11. The electric and magnetic polarizability dyadics, $\boldsymbol{\gamma}_e$ and $\boldsymbol{\gamma}_m$, quantify how much the scatterer responds to static electric and magnetic fields. Closed form expressions for the polarizability dyadics exists for homogeneous spheroidal scatterers, see [48]. For heterogeneous scatterers and other geometries, the right-hand side of (3.4) can be bounded. Thus, limitations of the form (2.17) can be derived where the right-hand side is only dependent on the geometry circumscribing the scatterer. If the scatterer is non-magnetic, the term $(\hat{\mathbf{k}} \times \hat{\mathbf{e}}) \cdot \boldsymbol{\gamma}_m \cdot (\hat{\mathbf{k}} \times \hat{\mathbf{e}})$ in the right-hand side of (3.4) is zero, which yields a sharper bound.

Many small scatterers absorb roughly the same amount of energy as they scatter, *i.e.*, $\sigma_a \approx \sigma_s$. This fact can be used to find sharp limits on the absorption cross section, since in that case $\sigma_a \approx \sigma_e/2$. This means that the results of [48] are well suited to limit antennas performance; a receiving antenna should ideally absorb the power of an incoming wave transmitted from some other antenna. By reciprocity, most antennas behave similarly whether they are acting as receivers or transmitters; therefore, limiting receiving antenna performance also limits transmitting antenna performance. The physical limitations on scattering derived by Sohl et al. are applied to antennas in [20]. They are comparable to the bounds based on Chu's paper for a circumscribing sphere, but sharper for non-spherical circumscribing geometries, see Figure 12. In [20], only linearly polarized antennas are considered. Elliptically polarized antennas are treated in [19]. Finally, it should be mentioned that the results in [48] of course can be used within other applications of electromagnetic scattering as well, and not only in antenna theory.

3.4 Spherical wave scattering and MIMO

In Paper II, the scattering of orthogonal vector spherical waves are considered. Therefore, the scatterer, or antenna, is inscribed in a sphere of radius a , see Figure 8. Outside this sphere, the electric and magnetic fields are expanded in incoming ($\mathbf{u}_\nu^{(2)}$) and outgoing ($\mathbf{u}_\nu^{(1)}$) vector spherical waves with index ν . The infinite dimensional scattering matrix $\tilde{\mathbf{S}}_S(\omega)$ of the scatterer relates the amplitudes of the outgoing waves ($\tilde{b}_\nu^{(1)}(\omega)$) to the amplitudes of the incoming waves ($\tilde{b}_\nu^{(2)}(\omega)$):

$$\tilde{b}_\nu^{(1)}(\omega) = \sum_{\nu'} \tilde{S}_{\nu, \nu'}(\omega) \tilde{b}_{\nu'}^{(2)}(\omega).$$

By considering the expressions for the vector spherical waves in the time domain, it is shown rigorously that the elements $S_{\nu, \nu'}(t - 2a/c)$ of the scattering matrix in the time domain are the impulse responses of scatter-passive systems. The time delay $-2a/c$ is due to the fact that the incoming waves $\mathbf{u}_{\nu'}^{(2)}$ does not have to reach the center of the sphere to interact with the scatterer and produce outgoing waves $\mathbf{u}_\nu^{(1)}$.

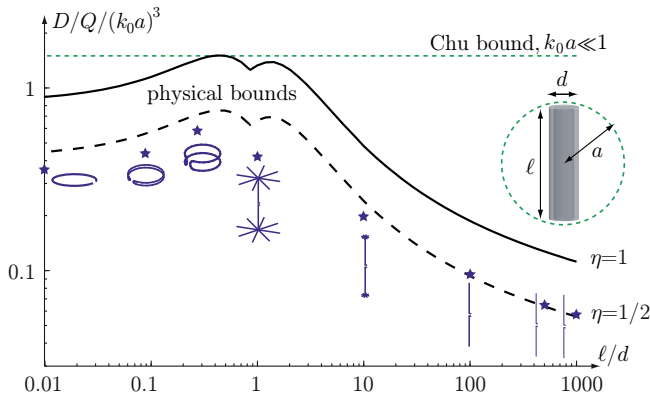


Figure 12: The physical limitations on directivity D over quality factor Q derived in [20]. All antennas inscribed in the cylinder are bounded by the curve labeled $\eta = 1$. If they scatter as much as they absorb, they are bounded by the curve labeled $\eta = 1/2$. For comparison, some sample antennas and the Chu-limit (3.2) with maximum directivity $D = 3/2$ are included.

The general approach to find sum rules and limitations on passive systems presented in Paper I can be used; two sum rules are derived, of which the following physical limitation is a consequence:

$$\frac{B \inf_{\mathcal{B}} \ln |\tilde{S}_{\nu,\nu'}(\omega)|^{-1}}{\pi} \leq k_0 a - \sqrt[3]{\iota + \zeta} + \sqrt[3]{\iota - \zeta} \quad (3.5)$$

$$= \left(\frac{1}{3} + \rho_{\nu,\nu} \right) (k_0^3 a^3 - k_0^5 a^5) + \mathcal{O}(k_0^7), \quad \text{as } k_0 \rightarrow 0,$$

where the frequency interval as before is defined as $\mathcal{B} = [(1 - B/2)\omega_0, (1 + B/2)\omega_0]$ with center frequency ω_0 , center wavenumber $k_0 = \omega_0/c$, and fractional bandwidth B . In the bound (3.5), the material and geometry of the scatterer is contained in $\zeta = 3k_0 a(1 - \rho_{\nu,\nu} k_0^2 a^2)/2$ and $\iota = \sqrt{1 + \zeta^2}$ where

$$\rho_{\nu,\nu} = \frac{1}{6\pi a^3} \begin{cases} \gamma_{e,nn}, & \text{if } \nu \text{ is the index of an electric dipole} \\ \gamma_{m,nn}, & \text{if } \nu \text{ is the index of a magnetic dipole} \\ 0, & \text{for higher order modes.} \end{cases} \quad (3.6)$$

Here $\gamma_{e,nn}$ is a diagonal element of the electric polarizability dyadic and $\gamma_{m,nn}$ is a diagonal element of the magnetic polarizability dyadic. From (3.6) it follows that the material parameter $\rho_{\nu,\nu}$ is bounded by $2/3$ for all modes if the scatterer is contained in sphere. If the scatterer is contained in a non-spherical geometry, ρ is bounded by a smaller constant. Furthermore, if the scatterer is non-magnetic, the elements of the magnetic polarizability dyadic γ_m are zero, which yields a sharper bound for the magnetic dipoles.

The power of the incoming mode ν that is rejected by the scatterer is

$$P_{\nu,\text{rejected}} \geq |\tilde{S}_{\nu,\nu}|^2 P_{\nu,\text{incoming}},$$

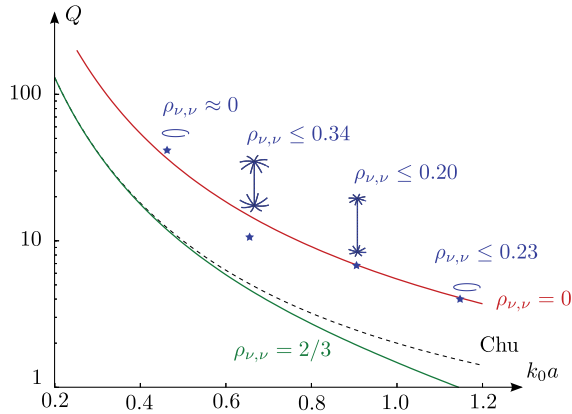


Figure 13: The bound (3.5) interpreted as a bound on the quality factor Q of an antenna. The bounds are plotted for $\rho_{\nu,\nu} = 2/3$ and $\rho_{\nu,\nu} = 0$, respectively. For comparison, the Chu-bound (3.2) is included, as well as the quality factor $Q_Z(\omega_0)$ of four sample antennas plotted at their respective resonance wavenumbers. Also, the bound on the material parameter $\rho_{\nu,\nu}$ for the four antennas are stated.

with approximate equality for small scatterers. Thus the bound (3.5) places a limit on the maximal power that an antenna can absorb from each individual incoming mode. It can be interpreted as a bound on the quality factor of an antenna for each mode (see Example 5.2 in Paper II), and it is compared to the Chu-bound (3.2) in Figure 13. It deserves mentioning that the results of Paper II can be applied within other applications of electromagnetic scattering as well, just like the results in [48].

The bounds derived in Paper II are not as sharp as those derived in [20], since the expansion in vector spherical waves requires the antenna to be inscribed in a hypothetical sphere. One reason to still consider spherical wave scattering is multiple-input multiple-output (MIMO) systems [42]. A MIMO system employs several antennas, and transmits and receives several signals at once. Each signal must be sent over an orthogonal communication channel, and, if the circumstances are correct, it follows that [42]

$$\text{Capacity} \propto N \ln(1 + \text{SNR}). \quad (3.7)$$

Here capacity is a measure on the amount of information that can be transmitted, N denotes the number of transmitted signals, and SNR is the signal to noise ratio. Transmitting and receiving several signals simultaneously is thus a good way of increasing capacity. But in order for (3.7) to apply, the N signals must be carried over an orthogonal set of spherical waves. In other words, there must be N orthogonal communication channels available. Otherwise, the individual signals can not be separated from each other. The bound (3.5) states that there are only six dominant modes available when the geometry circumscribing the antennas is small, and only three if the antennas are non-magnetic. Therefore, increasing the number of

transmitted signals N above three for a small geometry and non-magnetic antennas might not give the desired capacity gain. The bounds (3.5) on the non-dominant modes are only a factor of a third lower than the bounds on the dominant modes, however.

Intuitively, one might expect that there is in fact only two dominant modes if the non-magnetic antennas are circumscribed by a flat geometry, like a mobile phone. That is one open challenge for the future. Furthermore, it is likely that the bound for the non-dominant modes is sharper than (3.5) suggests. This supposition should be investigated in the future as well.

4 Concluding remarks

In the first paper of this thesis, a general approach to derive sum rules and physical limitations on passive physical systems is presented. The sum rules relate dynamical behaviour to static and/or high frequency properties. This is helpful, since static properties are often easier to determine. The physical limitations indicate what can, and what can not, be expected from the physical system. Since many physical systems obey passivity, the general approach of Paper I shows great potential; it may be applied to a wide range of problems, not only within electromagnetic theory.

Even though physical limitations on antenna performance have been discussed by researchers at least since Chu published his pioneering paper [8] in 1948, there is still much to be discovered in this area. The results of Paper I open up promising new ways to investigate this field of research, as indicated by the results of Paper II as well as by the recent work of Sohl et al. in [48] and [20].

References

- [1] L. V. Ahlfors. *Complex Analysis*. McGraw-Hill, New York, second edition, 1966.
- [2] N. I. Akhiezer. *The classical moment problem*. Oliver and Boyd, 1965.
- [3] N. I. Akhiezer and I. M. Glazman. *Theory of linear operators in Hilbert space*, volume 2. Frederick Ungar Publishing Co, New York, 1963.
- [4] Y. M. Berezansky, Z. G. Sheftel, and G. F. Us. *Functional Analysis*. Birkhäuser, Boston, 1996.
- [5] H. W. Bode. A method of impedance correction. *Bell System Technical Journal*, **9**, 794–835, October 1930.
- [6] C. R. Brewitt-Taylor. Limitation on the bandwidth of artificial perfect magnetic conductor surfaces. *Microwaves, Antennas & Propagation, IET*, **1**(1), 255–260, 2007.

-
- [7] W. Cauer. The Poisson integral for functions with positive real part. *Bulletin of the American Mathematical Society*, **38**(1919), 713–717, 1932.
- [8] L. J. Chu. Physical limitations of omni-directional antennas. *Appl. Phys.*, **19**, 1163–1175, 1948.
- [9] R. E. Collin and S. Rothschild. Evaluation of antenna Q. *IEEE Trans. Antennas Propagat.*, **12**, 23–27, January 1964.
- [10] S. Darlington. Synthesis of reactance 4-poles. *Journal of Mathematics and Physics*, **18**, 275–353, September 1939.
- [11] R. de L. Kronig. On the theory of dispersion of X-rays. *J. Opt. Soc. Am.*, **12**(6), 547–557, 1926.
- [12] W. F. Donoghue, Jr. *Distributions and Fourier Transforms*. Academic Press, New York, 1969.
- [13] P. L. Duren. *Theory of H^p Spaces*. Dover Publications, New York, 2000.
- [14] 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.
- [15] W. Geyi. Physical limitations of antenna. *IEEE Trans. Antennas Propagat.*, **51**(8), 2116–2123, August 2003.
- [16] R. E. Greene and S. G. Krantz. *Function theory of one complex variable*. John Wiley & Sons, New York, 1997.
- [17] D. H. Griffel. *Applied functional analysis*. Dover Publications, 2002.
- [18] M. Gustafsson and D. Sjöberg. Sum rules and physical bounds on passive metamaterials. *New Journal of Physics*, 2010. in press.
- [19] 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.
- [20] M. Gustafsson, C. Sohl, and G. Kristensson. Physical limitations on antennas of arbitrary shape. *Proc. R. Soc. A*, **463**, 2589–2607, 2007.
- [21] M. Gustafsson. Sum rule for the transmission cross section of apertures in thin opaque screens. *Opt. Lett.*, **34**(13), 2003–2005, 2009.
- [22] M. Gustafsson and S. Nordebo. Bandwidth, Q factor, and resonance models of antennas. *Progress in Electromagnetics Research*, **62**, 1–20, 2006.

- [23] 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.
- [24] R. C. Hansen. *Electrically small, superdirective, and superconductive antennas*. John Wiley & Sons, New Jersey, 2006.
- [25] R. C. Hansen. Bode and Fano impedance matching. *Microwave Opt. Techn. Lett.*, **50**(4), 875–877, 2008.
- [26] 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.
- [27] S. Hassi and A. Luger. Generalized zeros and poles of \mathcal{N}_κ -functions: On the underlying spectral structure. *Methods of Functional Analysis and Topology*, **12**(2), 131–150, 2006.
- [28] P. Henrici. *Applied and Computational Complex Analysis*, volume 2. John Wiley & Sons, New York, 1977.
- [29] G. Herglotz. Über Potenzreihen mit positivem, reellem Teil im Einheitskreis. *Leipziger Berichte*, **63**, 501–511, 1911.
- [30] L. Hörmander. *The Analysis of Linear Partial Differential Operators I*. Grundlehren der mathematischen Wissenschaften 256. Springer-Verlag, Berlin Heidelberg, second edition, 1990.
- [31] 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.
- [32] F. W. King. *Hilbert Transforms, Volume 1*. Cambridge University Press, 2009.
- [33] F. W. King. *Hilbert Transforms, Volume 2*. Cambridge University Press, 2009.
- [34] M. H. A. Kramers. La diffusion de la lumière par les atomes. *Atti. Congr. Int. Fis. Como*, **2**, 545–557, 1927.
- [35] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskĭĭ. *Electrodynamics of Continuous Media*. Pergamon, Oxford, second edition, 1984.
- [36] J. R. Macdonald and M. K. Brachman. Linear-system integral transform relations. *Reviews of modern physics*, **28**(4), 393–422, 1956.
- [37] J. Mashreghi. *Representation Theorems in Hardy Spaces*. Cambridge University Press, Cambridge, U.K., 2009.
- [38] 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.

- [39] R. H. Nevanlinna. *Asymptotische Entwicklungen beschränkter Funktionen und das Stieltjes'sche Momentenproblem*. Suomalaisen Tiedeakatemia Kustantama, 1922.
- [40] R. G. Newton. *Scattering Theory of Waves and Particles*. Dover Publications, New York, second edition, 2002.
- [41] H. M. Nussenzweig. *Causality and dispersion relations*. Academic Press, London, 1972.
- [42] A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space-Time Wireless Communications*. Cambridge University Press, Cambridge, U.K., 2003.
- [43] K. N. Rozanov. Ultimate thickness to bandwidth ratio of radar absorbers. *IEEE Trans. Antennas Propagat.*, **48**(8), 1230–1234, August 2000.
- [44] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 1987.
- [45] L. Schwartz. *Théorie des distributions*, volume 1. Hermann, Paris, 1957.
- [46] L. Schwartz. *Théorie des distributions*, volume 2. Hermann, Paris, 1959.
- [47] L. Schwartz. *Mathematics for the physical sciences*. Hermann, Paris, 1966.
- [48] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on broadband scattering by heterogeneous obstacles. *J. Phys. A: Math. Theor.*, **40**, 11165–11182, 2007.
- [49] 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*, 2008.
- [50] 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>.
- [51] R. S. Strichartz. *A guide to distribution theory and Fourier transforms*. World Scientific Publisher, Singapore, 2003.
- [52] J. R. Taylor. *Scattering theory: the quantum theory of nonrelativistic collisions*. Robert E. Krieger Publishing Company, Malabar, Florida, 1983.
- [53] A. Vassiliadis and R. L. Tanner. Evaluating the impedance broadbanding potential of antennas. *IRE Trans. on Antennas and Propagation*, **6**(3), 226–231, July 1958.
- [54] 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.

-
- [55] A. D. Yaghjian and S. R. Best. Impedance, bandwidth, and Q of antennas. *IEEE Trans. Antennas Propagat.*, **53**(4), 1298–1324, 2005.
 - [56] 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.
 - [57] A. H. Zemanian. An n -port realizability theory based on the theory of distributions. *IEEE Transactions on Circuit Theory*, **10**(2), 265–274, 1963.
 - [58] A. H. Zemanian. *Distribution theory and transform analysis: an introduction to generalized functions, with applications*. Dover Publications, New York, 1987.
 - [59] A. H. Zemanian. *Realizability Theory for Continuous Linear Systems*. Dover Publications, New York, 1995.

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. Technical Report LUTEDX/(TEAT-7193)/1-28/(2010), Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2010. <http://www.eit.lth.se>.

Abstract

A passive system is one that cannot produce energy, a property that naturally poses constraints on the system. A system on 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; one revisits Fano's matching equations, while another makes a link to the Kramers-Kronig dispersion relations and discusses physical limitations on metamaterials.

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 physical limitations is presented.

A system on 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 [20] (sometimes referred to as a Nevanlinna [14], Pick [6], or R-function [16]). The Laplace transform and the related function class of positive real (PR) functions are commonly preferred in system theory [9, 27].

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 [7, 19]. Herglotz functions appear in literature concerning continued fractions and the problem of moments [1, 15], but also within functional analysis and spectral theory for self-adjoint operators [2, 14]. 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.

In physical applications the integral identities are often called sum rules, effectively relating dynamic behaviour to static and/or high-frequency properties. This is very beneficial, since static properties are often easier to determine than dynamical behaviour in various applications. The representation in itself can also provide information on a system in the form of dispersion relations; consider *e.g.*, the Kramers-Kronig relations [18] discussed in Example 5.4. One way to take advantage of the sum rules is to derive constraints, or physical limitations, by considering finite frequency intervals. In essence, the physical limitations indicate what can and cannot be expected from a system. Some examples of applications in electromagnetic theory are in the analysis of bandwidth versus mismatch for matching networks [8], temporal dispersion for metamaterials [10], broadband electromagnetic interaction with objects [24], bandwidth and directivity for antennas of certain sizes [11], extra ordinary transmission through sub-wavelength apertures [12], thickness influence on performance of radar absorbers [21] and high-impedance surfaces [4], and impact of inter element coupling on frequency selective surfaces [13]. 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 on convolution form. They rely on somewhat different assumptions, and therefore does not apply to all the same problems. One approach, presented in [8, 25], relies on the Cauchy integral formula and is valid *e.g.*, if the transfer function is rational. Another assumes *e.g.*, that the transfer function is square-integrable, and derives sum rules from the Hilbert transform [17]. For square-integrable transfer functions, a theorem by Titchmarsh can also be used to find dispersion relations [20], which in some cases yield sum rules. As mentioned above, the crucial assumptions for the approach presented in this paper are that the system is causal and passive, and there does not seem to be a previous account on an approach to derive sum rules for causal and passive systems together with the rigorous proofs required.

This paper is divided into a number of distinct parts: 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, which constitute the core of the paper. After this section there is a discussion about passive systems and the possibility to constrain these. 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 main results of the paper, namely the integral identities. They are presented in the end of the section. Start with the definition of a Herglotz function:

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

There is a powerful representation theorem for the set of Herglotz functions \mathcal{H} due to Nevanlinna [2]:

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

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

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

Note the resemblance of (2.1) to the Hilbert transform [17, 19]. The proof of the representation theorem is not included here (it can be found in [2]), but in order to make it believable, (2.1) is cast into the slightly different form

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

where $d\nu(\xi) = d\mu(\xi)/(1 + \xi^2)$ is a positive and finite measure. The function $F(\xi, z) = (1 + \xi z)/(\xi - z)$ is a Herglotz function in z for all $\xi \in \mathbb{R} \cup \{\infty\}$, and sums of Herglotz functions are Herglotz functions. The constant β may be interpreted as $\nu(\{\infty\})$ (the point mass of ν at the point ∞ of the extended real line $\mathbb{R} \cup \{\infty\}$), since $F(\xi, z) \rightarrow z$ as $|\xi| \rightarrow \infty$. A real constant α may also be added to a Herglotz function, so the function given by (2.2) is a Herglotz function. That (2.1) exhausts the set \mathcal{H} follows *e.g.*, from a representation theorem for positive harmonic functions on the unit disk due to Herglotz. This representation theorem relies on the Riesz representation theorem for continuous, linear functionals on a compact metric space. 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 $h(z)/z \rightarrow \beta$, as $z \hat{\rightarrow} \infty$, where $z \hat{\rightarrow} \infty$ is a short-hand notation for $|z| \rightarrow \infty$ in the Stoltz domain $\theta \leq \arg z \leq \pi - \theta$ for any $\theta \in (0, \pi/2]$ (see Appendix A.1). Hence it makes sense to consider Herglotz functions with the asymptotic expansion

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

where $b_m \in \mathbb{R}$. Since $b_1 = \beta$, this expansion is always possible for some integer $M \geq 0$. It will simplify notation to define $b_m = 0$ for $m > 1$. The representation

¹The following notation is adopted throughout this paper (*cf.*, [3, 22]): If μ is a positive measure on the Borel subsets \mathcal{E} of \mathbb{R} and $E \in \mathcal{E}$, denote $\mu(E) = \int_E d\mu(\xi)$. The measure is referred to as μ or $d\mu$. The Lebesgue integral of f with respect to μ is denoted $\int_{\mathbb{R}} f(\xi) d\mu(\xi)$ whenever f is a complex-valued measurable function on \mathbb{R} . The measure that maps E to $\int_E u(\xi) d\mu(\xi)$ for some non-negative measurable function u on \mathbb{R} is denoted $u d\mu$.

also implies that $zh(z) \rightarrow -\mu(\{0\})$, as $z \hat{\rightarrow} 0$ (once more, see Appendix A.1), and so an asymptotic expansion

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

where $a_{-1} = -\mu(\{0\})$ and all a_n are real, is available for some integer $N \geq 0$. 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).

The main results of this paper are the identities:

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{\varepsilon < |x| < \varepsilon^{-1}} \frac{\text{Im } h(x + iy)}{x^p} dx = a_{p-1} - b_{p-1}, \quad p = 2 - 2M, 3 - 2M, \dots, 2N. \quad (2.5)$$

Throughout this paper i denotes the imaginary unit ($i^2 = -1$), and $x = \text{Re } z$ and $y = \text{Im } z$ 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 the examples 5.1 and 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 generalised function $h(x) = \lim_{y \rightarrow 0^+} h(x + iy)$, where the right hand side is interpreted as a limit of distributions, is a distribution of slow growth. In a discussion following Lemma 4.1 it is shown that, for almost all $x \in \mathbb{R}$, the limit $\lim_{y \rightarrow 0^+} \text{Im } h(x + iy)$ 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), Example 5.1 and Example 5.2.

In some special cases the integral identities follow directly from the Cauchy integral formula [8, 25]. This requires some extra assumptions, *e.g.*, that the Herglotz function is the restriction to \mathbb{C}^+ of a rational function. An alternative approach to obtain integral identities from the Hilbert transform is adopted by King [17].

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.

In [26] and [27], Zemanian shows that Herglotz functions appear in the context of linear, time translational invariant, continuous, causal and passive systems. These treatises are in the context of distributions, while a study in a more general setting is given in [28]. A short summary of some important results are given in this section. See also the book [20] by Nussenzveig.

Let \mathcal{D}' denote the space of distributions of one variable, and let \mathcal{D}'_0 denote distributions with compact support [27]. Consider an operator $\mathcal{R} : D(\mathcal{R}) \subseteq \mathcal{D}' \rightarrow \mathcal{D}'$. It is a convolution operator if and only if it is linear, time translational invariant, and continuous [27, Theorem 5.8-2]:

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

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 [27]. The output signal u is given by (3.1) at least for all input signals $v \in \mathcal{D}'$ so that convolution with w is defined in the sense of Theorems 5.4-1 and 5.7-1 in [27]. Since $w \in \mathcal{D}'$, $u = w * v$ at least for all $v \in \mathcal{D}'_0$. If, for example, w is in \mathcal{S}' , then $u = w * v$ for all $v \in \mathcal{S}$. Here \mathcal{S}' denotes distributions of slow growth and \mathcal{S} denotes smooth functions of rapid descent [27].

The operator is causal if w is not supported in $t < 0$, *i.e.*, $\text{supp } w \subseteq [0, \infty)$. 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 said port. Assume that the voltage and current are almost time-harmonic with an amplitude varying over a timescale much larger than the dominating frequency, so that u and v are complex valued distributions. The power absorbed by the system at the time t is $\text{Re } u^*(t)v(t)$ (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 [27, 28]:

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

$$e_{\text{adm}}(T) = \text{Re} \int_{-\infty}^T u^*(t)v(t) dt$$

and

$$e_{\text{scat}}(T) = \int_{-\infty}^T |v(t)|^2 - |u(t)|^2 dt.$$

The operator is admittance-passive (scatter-passive) if $e_{\text{adm}}(T)$ ($e_{\text{scat}}(T)$) 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. Reference [26] and [27] only treat admittance-passivity, and furthermore assume that w is real. The results referenced below are formulated only for real operators, but the proofs do not make use of this. Reference [28] deals with both admittance- and scatter-passive operators, but in a general Hilbert space setting. The results for scatter-passive operators may also be proved in a distributional context in a similar manner as the results for admittance-passive operators, see also [20].

An operator which is admittance-passive or scatter-passive is called passive in this paper. As it turns out, passivity implies causality for operators on convolution form. Furthermore, in this case the impulse response w must be a distribution of slow growth, *i.e.*, $w \in \mathcal{S}'$, and thus (3.1) is defined for smooth input signals of rapid descent, $v \in \mathcal{S}$. Note that (3.1) is also defined for all input signals v with support bounded on the left, since $\text{supp } w \subseteq [0, \infty)$ [27, 28].

Since the impulse response is in \mathcal{S}' , its Fourier transform may be defined as

$$\langle \mathcal{F}w, \varphi \rangle = \langle w, \mathcal{F}\varphi \rangle, \quad \text{for all } \varphi \in \mathcal{S},$$

where $\langle f, \varphi \rangle$ is the value in \mathbb{C} that $f \in \mathcal{S}'$ assigns to $\varphi \in \mathcal{S}$ [27]. The Fourier transform of φ is defined as

$$\mathcal{F}\varphi(\omega) = \int_{\mathbb{R}} \varphi(t)e^{i\omega t} dt.$$

The Fourier transform of w is the transfer function \tilde{w} of the system, *viz.*,

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

The convolution in (3.1) is mapped to multiplication if *e.g.*, $v \in \mathcal{D}'_0$ or $v \in \mathcal{S}$. In that case the frequency domain system is modeled 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 ω , 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 $\tilde{w}_{\text{Laplace}}(s) = \tilde{w}(is)$. Scrutinising the transfer function, the following theorem is proved (*cf.*, Theorem 10.4-1 in [27] and Theorems 7.4-3 and 8.12-1 in [28]):

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 $\text{Re } \tilde{w}(\omega) \geq 0$ for all $\omega \in \mathbb{C}^+$. If \mathcal{R} is scatter-passive, then $|\tilde{w}(\omega)| \leq 1$ 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 [27, Theorem 10.6-1], [28, Theorems

7.5-1 and 8.12-1]. Therefore the generalised function $h(x) = \lim_{y \rightarrow 0^+} h(x + iy)$ is a distribution of slow growth for all Herglotz functions h .

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

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

is a Blaschke product [7, 19]. Here the zeros ω_n of \tilde{w} are repeated according to their multiplicity and $k \geq 0$ is the order of the possible zero at $\omega = i$. 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 [7, 19]. Moving on, the function H may be represented as $H(\omega) = e^{ih(\omega)}$ 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

$$\tilde{w}(\omega) = B(\omega)e^{ih(\omega)}, \quad (3.4)$$

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

The formula (3.4) may be inverted:

$$h(\omega) = -i \log \left(\frac{\tilde{w}(\omega)}{B(\omega)} \right),$$

if the logarithm is defined as

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

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

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \lim_{y \rightarrow 0^+} \int_{\varepsilon < |x| < \varepsilon^{-1}} \frac{\operatorname{Im} h(x + iy)}{x^p} dx \\ = \lim_{\varepsilon \rightarrow 0^+} \lim_{y \rightarrow 0^+} \int_{\varepsilon < |x| < \varepsilon^{-1}} \frac{-\ln |\tilde{w}(x + iy)/B(x + iy)|}{x^p} dx. \end{aligned}$$

The modulus $|B(z)|$ tends to 1 as $z \rightarrow x$ for almost all $x \in \mathbb{R}$ (the exceptions are the x which are accumulation points of the zeros of \tilde{w} [19]). If the origin is not an

accumulation point of the zeros of \tilde{w} , the low-frequency asymptotic expansion of h is

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

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, see Example 5.3.

For operators \mathcal{R} mapping real input to real output, the impulse response w has to be real. This implies the symmetry

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

which is transferred to the Herglotz function as

$$h(\omega) = -h^*(-\omega^*) \quad (3.8)$$

if it is defined by $h = i\tilde{w}(\omega)$ (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 $e^{ih(\omega)}$ for purely imaginary ω . The symmetry restricts the identities (2.5) to even powers and simplifies them to

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{y \rightarrow 0^+} \frac{2}{\pi} \int_{\varepsilon}^{\varepsilon^{-1}} \frac{\text{Im } h(x + iy)}{x^{2\hat{p}}} dx = a_{2\hat{p}-1} - b_{2\hat{p}-1}, \quad \hat{p} = 1 - M, \dots, N. \quad (3.9)$$

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 three essentially equivalent ways to evaluate if a system can be modeled with a Herglotz function and potentially be constrained according to (2.5): First, just based on a priori knowledge of linearity, time-translational invariance, continuity, and passivity. Secondly, the time-domain characterization given by the convolution form (3.1) together with passivity. These approaches can often be applied directly to various physical systems. The third, frequency domain case is often more involved and requires direct verification that $h(\omega)$ is holomorphic and $\text{Im } h(\omega) \geq 0$ for $\text{Im } \omega > 0$. In many fields of physics it is common to consider time-harmonic signals as approximations of physically realisable signals. It becomes clear here that this could have inspired assumptions too severe; one could easily have been lead to the assumption that the transfer function needs to be defined pointwise for real frequencies.

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

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

The second term in the right hand side goes to zero as $\omega \hat{\rightarrow} \infty$ (but not as $|\omega| \rightarrow \infty$ on the real line) for any $\varepsilon > 0$. This verifies the statement for $w \in L^1$. For a general $w \in \mathcal{S}'$, consider the equivalent definition of $\tilde{w}(\omega)$ for $\text{Im } \omega > 0$ [27]:

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

Here $\lambda(t)$ is a smooth function with support bounded on the left, and such that $\lambda(t) \equiv 1$ for $t \geq 0$. It is decomposed into two non-negative smooth functions, $\lambda = \lambda_1 + \lambda_2$, where $\lambda_2 \equiv 0$ for $t \leq \varepsilon$ for some $\varepsilon > 0$. The second term in the right hand side vanishes as $\omega \hat{\rightarrow} \infty$. 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 $p = 2, 3, \dots, 2N$ they rely on two results: The first (Corollary 4.1) states that the left-hand side of (2.5) is equal to moments of the measure $d\mu(\xi)$. The second (Lemma 4.2) relates the convergence and explicit value of these moments to the expansion (2.4). A change of variables in the left-hand side of (2.5) enables a proof for $p = 2 - 2M, 3 - 2M, \dots, 1$.

A Herglotz function $h(z)$ is in general not defined pointwise for $\text{Im } z = 0$, but integrals of the type $\lim_{y \rightarrow 0^+} \int_{\mathbb{R}} \varphi(x) \text{Im } h(x + iy) dx$ are well defined under certain conditions on φ . The following lemma gives such sufficient conditions. They are stronger than needed, but weak enough to lead to the needed Corollary 4.1. This is a well known result, see *e.g.*, Theorem 11.9 in [19] and Lemma S1.2.1 in [16]. The lemma and proof are included here for clarity.

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

$$\lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{\mathbb{R}} \varphi(x) \text{Im } h(x + iy) dx = \int_{\mathbb{R}} \check{\varphi}(\xi) d\mu(\xi), \quad (4.1)$$

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

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

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

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 $\lim_{u \hat{\rightarrow} 0} \int_{\mathbb{R}} \varphi(x) \text{Im } h(x + u) dx$.

Note that the lemma is in some sense an inversion formula; whereas the representation (2.1) gives the Herglotz function h from the measure μ , (4.1) makes possible

the retrieval of μ when h is known. The inversion is clarified by decomposing the measure as $\mu = \mu_a + \mu_s$, where μ_a is absolutely continuous with respect to the Lebesgue measure $d\xi$ and μ_s is singular in the same sense [3]. Recall that \mathcal{E} denotes the set of Borel subsets of \mathbb{R} . Then

$$\mu_a(E) = \int_E \mu'_a(\xi) d\xi, \quad \text{for all } E \in \mathcal{E},$$

where the Radon-Nikodym derivative μ'_a of μ_a with respect to dx is a finite, integrable function and for almost all $x \in \mathbb{R}$ uniquely defined as [3]

$$\mu'_a(x) = \lim_{s \rightarrow 0} \frac{\mu_a([x-s, x+s])}{2s}.$$

“Almost all” is with respect to dx . Furthermore,

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

Hence Lemma 4.1 implies that

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

See also [19].

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 $\mu = \mu_a$. Otherwise set $h = h_a + h_s$, where h_a (h_s) is represented by μ_a (μ_s), to get

$$\begin{aligned} \lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{\mathbb{R}} \varphi(x) \operatorname{Im} h(x+iy) dx \\ = \frac{1}{\pi} \int_{\mathbb{R}} \varphi(x) \operatorname{Im} h_a(x) dx + \lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{\mathbb{R}} \varphi(x) \operatorname{Im} h_s(x+iy) dx, \end{aligned} \quad (4.3)$$

where $\operatorname{Im} h_a(x) = \lim_{y \rightarrow 0^+} \operatorname{Im} h(x+iy)$ whenever the limit exists finitely. Equivalently, the left-hand side of (4.1) may be interpreted as an integral over the real line in the distributional sense. Recall that the generalised function $h(x)$ is a distribution of slow growth.

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*

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \lim_{\xi \rightarrow 0^+} \lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{-\xi-1}^{-\varepsilon} \frac{\operatorname{Im} h(x+iy)}{x^p} dx + \lim_{\varepsilon \rightarrow 0^+} \lim_{\xi \rightarrow 0^+} \lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{\varepsilon}^{\xi-1} \frac{\operatorname{Im} h(x+iy)}{x^p} dx \\ = \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^p}, \quad p = 0, \pm 1, \pm 2, \dots \end{aligned}$$

Here $\mu_0 = \mu - \mu(\{0\})\delta_0$, i.e., the measure in the representation (2.1) with the point mass in 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

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

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

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

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

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

for z in the Stoltz domain $\theta \leq \arg z \leq \pi - \theta$ 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*

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

The expansion coefficients in (2.4) equal:

$$a_0 = \alpha + \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi(1 + \xi^2)}, \quad (4.5)$$

$$a_{p-1} = \delta_{p,2}\beta + \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^p}, \quad p = 2, 3, \dots, 2N, \quad (4.6)$$

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]. See also Lemma 6.1 in [14]. Note that the case $N = 0$ is trivial, since then all three statements are true for all Herglotz functions. The proof for $N \geq 1$ can be found in Appendix A.4. The convergence of $\int_{\mathbb{R}} d\mu_0(\xi)/(|\xi^{2N+1}|(1 + \xi^2))$ does guarantee an expansion with real coefficients up to $o(z^{2N})$, but the converse is not true. A counterexample for $N = 0$ is given by the measure $d\mu_0(\xi) = \mu'_0(\xi) d\xi$ where $\mu'_0(\xi) = -(\ln |\xi|)^{-1}$ when $\xi < 1$ and $\mu'_0(\xi) = 0$ otherwise.

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

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

$$\check{h}(z) = \frac{-\beta}{z} + \alpha + \mu(\{0\})z + \int_{\mathbb{R}} \frac{1 + \xi z^{-1}}{1 + \xi^2} d\nu_0(\xi), \quad \text{Im } z > 0,$$

where $d\nu_0(\xi) = d\mu_0(\xi)/(1 + \xi^2)$. It would be desirable to make a change of variables $\xi \mapsto -1/\xi$ in the integral. Therefore, consider the continuous bijection $j : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R} \setminus \{0\}$ 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 $j : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R} \setminus \{0\}$ be the mapping that takes ξ to $-1/\xi$. Let $\mathcal{E}(\mathbb{R} \setminus \{0\})$ be the Borel sets of $\mathbb{R} \setminus \{0\}$ and $M(\mathbb{R} \setminus \{0\})$ be the set of measures on $\mathcal{E}(\mathbb{R} \setminus \{0\})$. Define the mapping $J : M(\mathbb{R} \setminus \{0\}) \rightarrow M(\mathbb{R} \setminus \{0\})$ through*

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

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

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

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

for all measurable functions f on $\mathbb{R} \setminus \{0\}$, since it holds if f is a simple measurable function [22]. The representation of \check{h} can now be rewritten:

$$\check{h}(z) = \frac{-\beta}{z} + \alpha + \mu(\{0\})z + \int_{\mathbb{R}} \frac{1 - \xi z}{1 + \xi^2} d(J\nu_0)(\xi), \quad \text{Im } z > 0.$$

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

$$\begin{aligned} \lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{\varepsilon}^{\varepsilon^{-1}} \frac{\text{Im } h(x + iy)}{x^p} dx &= \int_{\mathbb{R}} \check{\varphi}_{p,\varepsilon,\varepsilon}(\xi) d\mu_0(\xi) = \int_{\mathbb{R}} \check{\varphi}_{p,\varepsilon,\varepsilon}(-1/\xi) \frac{d\check{\mu}_0(\xi)}{\xi^2} \\ &= \lim_{y \rightarrow 0^+} (-1)^p \frac{1}{\pi} \int_{-\varepsilon^{-1}}^{-\varepsilon} \frac{\text{Im } \check{h}(x + iy)}{x^{2-p}} dx, \quad \text{for } p = 0, \pm 1, \pm 2, \dots \text{ and } 0 < \varepsilon < \varepsilon^{-1}, \end{aligned} \quad (4.7)$$

and likewise for the corresponding integral over $(-\varepsilon^{-1}, -\varepsilon)$. Here $\check{\varphi}_{p,\varepsilon,\varepsilon}$ is given by (A.2) and (4.2). The proof of the integral identities (2.5) for $p = 2 - 2M, 3 - 2M, \dots, 0$ have now been returned to the case $p = 2, 3, \dots, 2N$. 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 proof can be found in Appendix A.5. The integrals in the left-hand side of (2.5) may be taken over the set $\{x : \varepsilon < |x| < \infty\}$ when $p = 2, 3, \dots, 2N$ and $\{x : 0 < |x| < \varepsilon^{-1}\}$ when $p = 2 - 2M, 3 - 2M, \dots, 0$, 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.

5 Examples

5.1 Elementary Herglotz functions

Examples of elementary Herglotz functions are

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

with $\beta \geq 0$, $\text{Im } C \geq 0$, and appropriate branch cuts for $\sqrt{\cdot}$ and \log .

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

$$h_e(z) = \frac{ie^{iz} + i}{1 - e^{iz}}$$

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

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

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 (3.9) for $\hat{p} = 1, 2, \dots$:

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{y \rightarrow 0} \frac{2}{\pi} \int_{\varepsilon}^{\infty} \frac{\text{Im } \tan(x + iy)}{x^{2\hat{p}}} dx = \begin{cases} 1 & \text{for } \hat{p} = 1 \\ 1/3 & \text{for } \hat{p} = 2 \\ 2/15 & \text{for } \hat{p} = 3. \\ \vdots & \end{cases}$$

On the real axis except for $x = n\pi$, where $n = 0, \pm 1, \pm 2, \dots$, $\tan(x)$ is C^∞ and $\text{Im } \tan(x) = 0$. It is not locally integrable around $x = n\pi$, where $\tan z$ has simple poles. There is an essential singularity at ∞ , and the limit as $x \rightarrow \infty$ 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

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

yielding an identity of the type (3.9):

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{y \rightarrow 0^+} \frac{2}{\pi} \int_{\varepsilon}^{\infty} \frac{\ln |1 - i \tan(x + iy)|}{x^2} dx = 1.$$

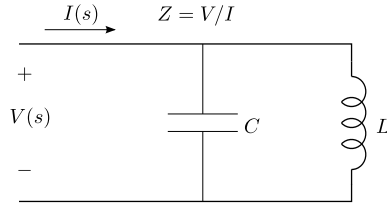


Figure 1: The lossless resonance circuit of Example 5.2.

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

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

This gives the (2.5)-identities

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

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 $\sqrt{h_1 h_2}$.

5.2 Lossless resonance circuit

Consider a parallel resonance circuit consisting of a lumped inductance, L , and a lumped capacitance, C , see Figure 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. Therefore the transfer function given by (3.2) multiplied by i is a Herglotz function:

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

where $\omega_0 = 1/\sqrt{LC}$ is the resonance frequency of the LC circuit. In general, the imaginary part of $h(\omega) = iZ(-i\omega)$ corresponds to the power absorbed by the impedance Z .

Use of the identities (3.9) gives the sum rules

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{\omega' \rightarrow 0^+} \frac{2}{\pi} \int_{\varepsilon}^{\varepsilon^{-1}} \frac{\text{Im } h(\omega' + i\omega'')}{\omega'^{2p}} d\omega' = \sqrt{\frac{L}{C}} \omega_0^{-2p+1}, \quad \text{for } p = 0, \pm 1, \pm 2, \dots \quad (5.1)$$

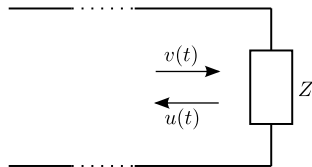


Figure 2: The voltage waves traveling along the transmission line has the amplitudes $v(t)$ and $u(t)$, respectively, measured by the load.

Note that on the real axis $\text{Im} h(\omega') = 0$ 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 Reflection coefficient (Fano’s matching equations revisited)

Consider a transmission line ended in a load impedance. The transmission line is assumed to be distortionless, *i.e.*, its characteristic impedance is not a function of frequency. Normalise so that the characteristic impedance of the transmission line is 1 and the lumped impedance is $Z(s)$, where $s = -i\omega$ denotes the Laplace parameter. The load impedance is assumed to be realisable with a finite number of linear passive elements (but otherwise arbitrary), so Z is a rational function.

The reflection coefficient $\rho(s) = (Z(s) - 1)/(Z(s) + 1)$ is of interest, since it determines the power rejected by the load. It is the Laplace-transfer function of the system where the input v and output u are the amplitudes of the voltage waves travelling along the transmission line toward or from the load, respectively. See Figure 2. The Fourier transfer function is $\tilde{w}(\omega) = \rho(-i\omega)$, satisfying (3.7). This is clearly a scatter-passive system, so $\tilde{w}(\omega)$ is holomorphic and bounded in magnitude by one in \mathbb{C}^+ .

Assume the asymptotic expansion

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

where $\arg \tilde{w}(0) = \lim_{\omega \rightarrow 0} \arg \tilde{w}(\omega)$ and all c_i are real. This is the case *e.g.*, if the impedance Z can be represented as a lossless network terminated in another impedance, Z_2 (*cf.*, Figure 3), and the network has a transmission zero of order N at $\omega = 0$ [8]. The low-frequency asymptotic expansion of the Herglotz-function in (3.4) is

$$h(\omega) = \arg \tilde{w}(0) + c_1\omega + c_3\omega^3 + \dots + c_{2N-1}\omega^{2N-1} + o(\omega^{2N-1}) \\ - \arg B(0) - 2 \sum_{m=1,3,\dots}^{\infty} \frac{\omega^m}{m} \sum_{\omega_n} \text{Im} \omega_n^{-m}, \quad \text{as } \omega \rightarrow 0,$$

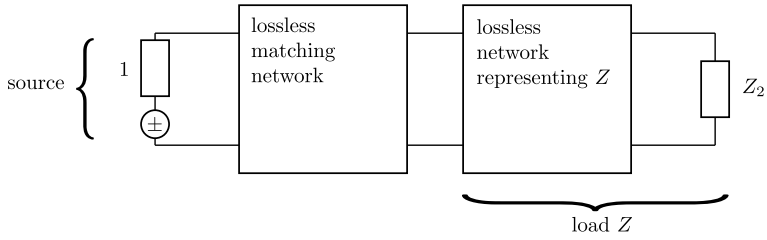


Figure 3: The matching problem as described in [8].

according to (3.6). In this case only odd terms appear in the sum originating from the Blaschke product due to the symmetry (3.7). The high-frequency asymptotic expansion of h is $o(\omega)$ since \tilde{w} is a rational function. This implies the (3.9)-identities

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \lim_{\omega'' \rightarrow 0^+} \frac{2}{\pi} \int_{\varepsilon}^{\infty} \frac{-\ln |\tilde{w}(\omega' + i\omega'')|}{\omega'^{2\hat{p}}} d\omega' \\ = c_{2\hat{p}-1} - \frac{2}{2\hat{p}-1} \sum_{\omega_n} \text{Im} \omega_n^{1-2\hat{p}}, \quad \text{for } \hat{p} = 1, 2, \dots, N. \end{aligned}$$

If ρ has no zeros at the imaginary axis, the limit as $\omega'' \rightarrow 0^+$ may be moved inside the integral. These are the original Fano matching equations, derived with the Cauchy integral formula in [8]. In said paper they are used to derive the best possible match of a source to a load over an open frequency interval, and how the lossless matching network should be constructed to obtain this best match. See Figure 3. When ρ 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 as long as asymptotic expansions of the type (5.2) are valid as $\omega \hat{\rightarrow} 0$ and/or $\omega \hat{\rightarrow} \infty$, respectively. It should be mentioned that Fano's results have been treated more generally in *e.g.*, [5], which however only covers rational reflection coefficients ρ .

5.4 Kramers-Kronig relations and ϵ near-zero materials

Suppose there is an isotropic constitutive relation on convolution form relating the electric field $\mathbf{E} = E\hat{\mathbf{e}}$ to the electric displacement $\mathbf{D} = D\hat{\mathbf{e}}$ [18]:

$$D(t) = \epsilon_0 \chi * E(t).$$

The permittivity of free space is denoted ϵ_0 , and a possible instantaneous response is included in $\chi(t)$ as a term $\epsilon_{\infty}\delta(t)$, where $\epsilon_{\infty} \geq 0$. 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 [18]

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

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

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

$$\omega\epsilon(\omega) = \omega\epsilon_\infty + \lim_{\psi \rightarrow 0^+} \frac{1}{\pi} \int_{\mathbb{R}} \left[\frac{1}{\xi - \omega} - \frac{\xi}{1 + \xi^2} \right] \left[\psi \operatorname{Re} \epsilon(\xi + i\psi) + \xi \operatorname{Im} \epsilon(\xi + i\psi) \right] d\xi, \quad (5.3)$$

for $\operatorname{Im} \omega > 0$. This is one of the two Kramers-Kronig relations [17, 18] 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 $\epsilon(\omega') = \lim_{\omega'' \rightarrow 0^+} \epsilon(\omega' + i\omega'')$ is sufficiently well-behaved. Here the notation $\omega' = \operatorname{Re} \omega$ and $\omega'' = \operatorname{Im} \omega$ has been used. If for instance $\epsilon(\omega')$ is a continuous and bounded function, the limit may be moved inside the integral in (5.3):

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

Employing the fact that $\operatorname{Im} \epsilon(\xi)$ is odd gives (after division with ω)

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

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

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

The real part of this equation is the Kramers-Kronig relation (5.3) as presented in *e.g.*, [18]:

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

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 $\omega \hat{\rightarrow} 0$, there is a sum rule of the type (3.9) for $\hat{p} = 1$ (also presented in *e.g.*, [18]):

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

It shows that the losses are related to the difference between the static and instantaneous responses of the medium.

In applications such as high-impedance surfaces and waveguides, it is desirable to have so called ϵ near-zero materials [23], *i.e.*, materials with $\epsilon(\omega') \approx 0$ in a frequency interval around some center frequency ω_0 . Define the Herglotz function

$$h_1(\omega) = \frac{\omega}{\omega_0} \epsilon(\omega) = \begin{cases} o(\omega^{-1}), & \text{as } \omega \hat{\rightarrow} 0 \\ \frac{\omega}{\omega_0} \epsilon_\infty + o(\omega), & \text{as } \omega \hat{\rightarrow} \infty. \end{cases} \quad (5.4)$$

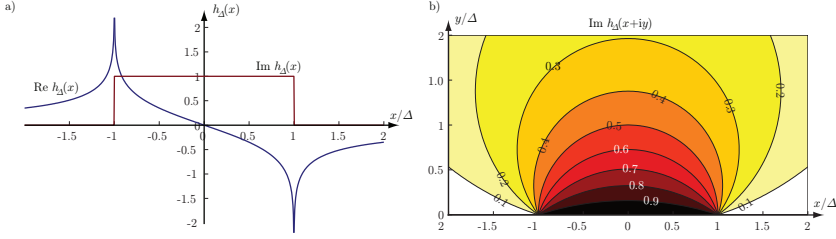


Figure 4: The function $h_\Delta(x + iy)$ given by (5.5) illustrated by its limit as $y \rightarrow 0$ to the left and the contours of $\text{Im } h_\Delta(x + iy)$ to the right.

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

$$h_\Delta(z) = \frac{1}{\pi} \int_{-\Delta}^{\Delta} \frac{1}{\xi - z} d\xi = \frac{1}{\pi} \ln \frac{z - \Delta}{z + \Delta} = \begin{cases} i + o(1), & \text{as } \omega \rightarrow 0 \\ \frac{-2\Delta}{\pi z} + o(z^{-1}), & \text{as } \omega \rightarrow \infty, \end{cases} \quad (5.5)$$

designed such that $\text{Im } h_\Delta(z) \approx 1$ for $\text{Im } z \approx 0$ and $|\text{Re } z| \leq \Delta$, see Figure 4. Here the logarithm has its branch cut along the negative imaginary axis. The asymptotic expansions of the composition are

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

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

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0^+} \lim_{\omega'' \rightarrow 0^+} \int_0^{\epsilon^{-1}} \text{Im } h_\Delta(h_1(\omega' + i\omega'')) d\omega' \\ &= \lim_{\epsilon \rightarrow 0^+} \lim_{\omega'' \rightarrow 0^+} \int_0^{\epsilon^{-1}} \arg \left(\frac{(\omega' + i\omega'')\epsilon_\infty - \Delta\omega_0}{(\omega' + i\omega'')\epsilon_\infty - \Delta\omega_0} \right) d\omega' = \frac{\omega_0\Delta}{\epsilon_\infty}. \end{aligned} \quad (5.6)$$

An illustration of $\lim_{\omega'' \rightarrow 0} \text{Im } h_\Delta(h_1(\omega' + i\omega''))$ for a permittivity function ϵ described by a Drude model can be found in Figure 5.

Let the frequency interval be $\mathcal{B} = [\omega_0(1 - B_F/2), \omega_0(1 + B_F/2)]$, where B_F denotes the fractional bandwidth. Assume that $h_1(\omega') = \lim_{\omega'' \rightarrow 0^+} h_1(\omega' + i\omega'')$ exists finitely in this interval and let $\Delta = \sup_{\omega' \in \mathcal{B}} |h_1(\omega')|$. Then $\inf_{\omega' \in \mathcal{B}} \lim_{\omega'' \rightarrow 0^+} \text{Im } h_\Delta(h_1(\omega' + i\omega'')) \geq 1/2$ which yields the bound

$$\sup_{\omega' \in \mathcal{B}} |h_1(\omega')| \geq \frac{B_F}{2} \epsilon_\infty$$

or

$$\sup_{\omega' \in \mathcal{B}} |\epsilon(\omega')| \geq \frac{B_F}{2 + B_F} \epsilon_\infty.$$

This shows that ϵ near-zero materials are dispersive and that the deviation from zero is proportional to the fractional bandwidth when $B_F \ll 1$.

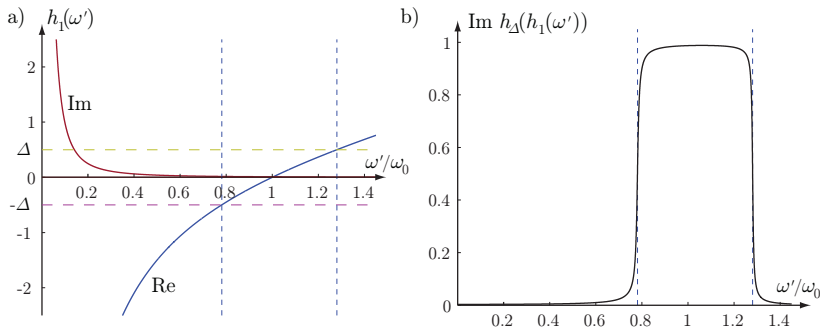


Figure 5: The left figure depicts the real and imaginary part of $h_1(\omega') = \lim_{\omega'' \rightarrow 0} h_1(\omega' + i\omega'')$, where h_1 is given by (5.4) and the permittivity is described by the Drude model $\epsilon(\omega) = 1 - (\omega/\omega_0 (\omega/\omega_0 - 0.01i))^{-1}$. The right figure depicts the integrand $\text{Im } h_\Delta(h_1(\omega')) = \lim_{\omega'' \rightarrow 0} \text{Im } h_\Delta(h_1(\omega' + i\omega''))$ in (5.6) for this choice of $\epsilon(\omega)$ and $\Delta = 1/2$.

6 Conclusions

Many physical systems are modeled 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 generalised functions such as the delta function should be allowed. Under the general assumptions of linearity, continuity and time-translational invariance, such a system is on convolution form, and thus fully described by its impulse response. The assumption of passivity (and thereby causality, as described in Section 3), imply that the transfer function is related to a Herglotz function, h [20, 27, 28]. In many areas it is convenient to analyse systems in the frequency domain, where the transfer function plays the role of the impulse response.

The main result of this paper is a set of integral identities for Herglotz functions, showing that weighted integrals of the function h over infinite intervals are determined by its high- and low-frequency asymptotic expansions. The identities rely on a well-known representation theorem for Herglotz functions [2], and furthermore makes use of results from the classical problem of moments [1].

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 \tilde{w} is related to a Herglotz function. No additional assumptions, like \tilde{w} being square-integrable, are needed. 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. One way to make use of the sum rules is to derive physical limitations, which 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.*, electromagnetics. Two famous examples are the Kramers-Kronig relations for the frequency dependence of the electric permittivity [18], discussed in Example 5.4, and Fano’s matching equations [8], considered in Example 5.3. There are more recent examples as well, see *e.g.*, [4, 10–13, 21, 24]. For example, the physical limitations may answer questions like what the maximal broadband electromagnetic interaction for an object is [24], how “well” an antenna can perform over a prescribed frequency band [11], or how close the electric permittivity can come to a specific value over a frequency interval [10]. This can be very helpful, both as a means to understand what factors that constrain performance, but also to determine if there is room for improvement.

The present paper seems to be the first to describe and rigorously prove a general approach to obtain sum rules for systems under the assumptions of convolution form and passivity. There are alternative approaches to derive sum rules for systems on convolution form, relying on somewhat different assumptions. Many previous papers use the Cauchy integral formula, see *e.g.*, [8, 25]. This approach demands *e.g.*, that the transfer function \tilde{w} is rational. Another approach assumes *e.g.*, that \tilde{w} is square-integrable, and derives sum rules from the Hilbert transform [17]. Alternatively, Titchmarsh’s theorem may be used to find dispersion relations and in some cases sum-rules when \tilde{w} is square-integrable [20]. It should be stressed that since the three approaches listed here works under different assumptions, they are complementary rather than in competition. It should also be pointed out that both the rational function and square-integrable function approach can be generalised to larger classes of impulse responses. 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 Example 5.4.

Acknowledgments

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

Appendix A Proofs

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

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

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

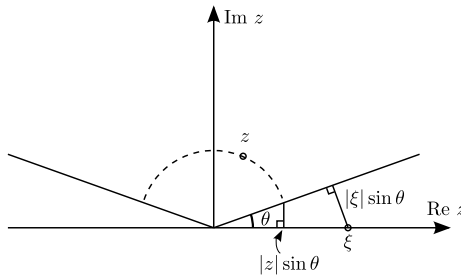


Figure 6: The Stoltz domain, $\{z : \theta \leq \arg z \leq \pi - \theta\}$ for some $\theta \in (0, \pi/2]$.

and (2.2) implies that

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

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

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

A.2 Proof of Lemma 4.1

The left-hand side of (4.1) is

$$\begin{aligned} \lim_{y \rightarrow 0^+} \int_{\mathbb{R}} \varphi(x) \left(\beta y + \int_{\mathbb{R}} \frac{y}{(x - \xi)^2 + y^2} d\mu(\xi) \right) dx \\ = \lim_{y \rightarrow 0^+} \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x) \frac{y}{(x - \xi)^2 + y^2} dx d\mu(\xi). \end{aligned}$$

Here Fubini's Theorem [22, pp. 164–165] has been used to change the order of integration.

Theorem A.2 is used to show that the order of the limit and the integrals may be interchanged. First set

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

To find an integrable majorant $g \in L^1(\mu)$ such that $|f_y(\xi)| \leq g(\xi)$ for all $\xi \in \mathbb{R}$ and $y \geq 0$, handle the cases $|\xi| < 2$ and $|\xi| \geq 2$ separately. For $|\xi| < 2$, the boundedness of φ guarantees that

$$|f_y(\xi)| \leq \int_{\mathbb{R}} D \frac{y}{(x - \xi)^2 + y^2} dx = D\pi.$$

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

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

and

$$\begin{aligned} \left| \int_{|x-\xi|\geq 1} \varphi(x) \frac{y}{(x-\xi)^2 + y^2} dx \right| &\leq \int_{|x-\xi|\geq 1} \frac{D}{1+x^2} \frac{y}{(x-\xi)^2} dx \\ &= Dy \left[\frac{\xi}{(\xi^2+1)^2} \ln \left| \frac{(\xi-1)^2+1}{(\xi+1)^2+1} \right| + \frac{2}{1+\xi^2} + \frac{\xi^2-1}{(\xi^2+1)^2} \frac{\pi}{2} \right] \leq \frac{D_1 y}{\xi^2+1}. \end{aligned}$$

Summing up, for all y less than some arbitrary constant there is a constant $D_2 \geq 0$ such that

$$|f_y(\xi)| \leq g(\xi) = \frac{D_2}{\xi^2+1},$$

which is an integrable majorant. Since $\lim_{y \rightarrow 0^+} f_y(\xi)$ 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 ξ 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 φ is continuously differentiable in said interval, except possibly at the point $x = \xi$. For $x \in [\xi - K, \xi + K]$,

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

for some constant $D_3 \geq 0$. Here it has been used that $|\varphi'(x)|$ is bounded in $[\xi - K, \xi + K]$, and that $|sy/(s^2 + y^2)|$ is bounded. An integrable majorant for $f_{y,\xi}(x)$ is

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

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

$$\lim_{y \rightarrow 0^+} \int_{\mathbb{R}} (\varphi(x) - \varphi(\xi)) \frac{y}{(x-\xi)^2 + y^2} dx = 0,$$

which is equivalent to

$$\lim_{y \rightarrow 0^+} \int_{\mathbb{R}} \varphi(x) \frac{y}{(x-\xi)^2 + y^2} dx = \pi \varphi(\xi).$$

This proves the lemma for continuous φ .

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

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

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

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

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

$$\lim_{y \rightarrow 0^+} \int_{\mathbb{R}} \varphi_{\text{even}}(x) \frac{y}{(x - \xi)^2 + y^2} dx = \pi \check{\varphi}(\xi).$$

Together with (A.1) this concludes the proof of the lemma for φ that are not continuous everywhere.

A.3 Proof of Corollary 4.1

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

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

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

$$\lim_{y \rightarrow 0^+} \frac{1}{\pi} \int_{\varepsilon}^{\check{\varepsilon}^{-1}} \frac{\text{Im } h(x + iy)}{x^p} dx = \int_{\mathbb{R}} \check{\varphi}_{p,\varepsilon,\check{\varepsilon}}(\xi) d\mu(\xi),$$

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

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

Implement Theorem A.1 to get

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{\check{\varepsilon} \rightarrow 0^+} \int_{\mathbb{R}} \check{\varphi}_{p,\varepsilon,\check{\varepsilon}}(\xi) d\mu(\xi) = \int_{\xi > 0} \frac{d\mu(\xi)}{\xi^p}, \quad p = 0, \pm 1, \pm 2, \dots$$

The integral over $(-\check{\varepsilon}^{-1}, -\varepsilon)$ is treated in the same manner. This proves the lemma, seeing that

$$\int_{\xi < 0} \frac{d\mu(\xi)}{\xi^p} + \int_{\xi > 0} \frac{d\mu(\xi)}{\xi^p} = \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^p},$$

unless the left-hand side is $-\infty + \infty$. In this case the right-hand side is not defined.

For $p = 2, 3, \dots$, the order of the limits $\varepsilon \rightarrow 0^+$ and $y \rightarrow 0^+$ may be interchanged. Likewise, for $p = 0, -1, -2, \dots$ the order of the limits $\varepsilon \rightarrow 0^+$ and $y \rightarrow 0^+$ 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 $\lim_{\varepsilon \rightarrow 0^+} \varphi_{p,\varepsilon,\varepsilon}(x)$ and $\lim_{\varepsilon \rightarrow 0^+} \varphi_{p,\varepsilon,\varepsilon}(x)$, respectively.

A.4 Proof of Lemma 4.2

Evidently, statement 1 always implies 2. Here it will be shown that 2 implies 3 and that 3 implies 1. Start with the case $N = 1$ and assume that 3 holds. Consider the Herglotz function $h_0(z) = h(z) + \mu\{0\}/z$, represented by the measure μ_0 . Set

$$a_0 = \lim_{z \rightarrow 0} h_0(z) = \alpha + \lim_{z \rightarrow 0} \int_{\mathbb{R}} \frac{1 + \xi z}{(\xi - z)(1 + \xi^2)} d\mu_0(\xi) = \alpha + \int_{\mathbb{R}} \frac{1}{\xi(1 + \xi^2)} d\mu_0(\xi).$$

Here Theorem A.2 could be used to move the limit under the integral sign, since for z restricted to the Stoltz domain $\theta \leq \arg z \leq \pi - \theta$ it holds that $|\xi - z| \geq |\xi| \sin \theta$ (see Appendix A.1) and $\int_{\mathbb{R}} \xi^{-2} d\mu_0(\xi)$ is finite by assumption. Use this expression for a_0 :

$$\lim_{z \rightarrow 0} \frac{h_0(z) - a_0}{z} = \beta + \lim_{z \rightarrow 0} \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{(\xi - z)\xi} = \beta + \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^2} = a_1,$$

where Theorem A.2 was used once more. Summing up, statement 1 is true.

Now assume that statement 2 is valid (still $N = 1$), *i.e.*,

$$h_0(iy) = a_0 + a_1 iy + o(y), \quad \text{as } y \rightarrow 0^+,$$

where $a_0, a_1 \in \mathbb{R}$. From this condition it follows that

$$\lim_{y \rightarrow 0^+} \frac{h_0(iy) - h_0^*(iy)}{2iy} = \lim_{y \rightarrow 0^+} \left(a_1 + \frac{o(y)}{iy} \right) = a_1.$$

But on the other hand,

$$\lim_{y \rightarrow 0^+} \frac{h_0(iy) - h_0^*(iy)}{2iy} = \beta + \lim_{y \rightarrow 0^+} \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^2 + y^2} = \beta + \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^2}.$$

The exchange of the limit and integral is motivated by Theorem A.1. Ergo,

$$\int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^2} = a_1 - \beta < \infty,$$

and thus statement 3 is true.

The equivalence of the statements for all $N = 0, 1, 2, \dots$ is proved by induction. For this reason, suppose that the equivalence has been proven for some $N \geq 1$, and that statement 3 holds for $N + 1$:

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

Consider the function

$$h_1(z) = \frac{h_0(z) - a_0 - a_1 z}{z^2}.$$

This function may be expressed as:

$$\begin{aligned} h_1(z) = \frac{1}{z^2} & \left[\beta z + \alpha + \int_{\mathbb{R}} \left(\frac{1}{\xi - z} - \frac{\xi}{1 + \xi^2} \right) d\mu_0(\xi) \right. \\ & \left. - \left(\alpha + \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi(1 + \xi^2)} \right) - z \left(\beta + \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^2} \right) \right] = \int_{\mathbb{R}} \frac{d\mu_1(\xi)}{\xi - z}, \end{aligned}$$

where $d\mu_1(\xi) = d\mu_0(\xi)/\xi^2$. Hence h_1 is a Herglotz function, and furthermore

$$\int_{\mathbb{R}} \frac{d\mu_1(\xi)}{\xi^{2N}} < \infty,$$

so h_1 has the asymptotic expansion

$$h_1(z) = \sum_{n=0}^{2N-1} a_{n+2} z^n + o(z^{2N-1}) \quad \text{as } z \rightarrow 0,$$

where all a_n are real. This proves statement 1 for $N + 1$.

On the other hand, assume that statement 2 holds for $N + 1$, where $N \geq 1$. Consider the function h_1 once more. The induction assumption ensures that

$$\int_{\mathbb{R}} \frac{d\mu_1(\xi)}{\xi^{2N}} = \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^{2N+2}} < \infty,$$

which proves that statement 3 is true for $N + 1$.

Finally, note that from the representation of h_1 it is clear that

$$a_3 = \int_{\mathbb{R}} \frac{d\mu_1(\xi)}{\xi^2} = \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^4}.$$

Furthermore,

$$a_2 = \lim_{z \rightarrow 0} h_1(z) = \int_{\mathbb{R}} \frac{d\mu_1(\xi)}{\xi} = \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi^3}.$$

This procedure may be continued for $a_4, a_5, \dots, a_{2N-1}$ to prove (4.6), concluding the proof of the lemma.

A.5 Proof of Theorem 4.1

The theorem for $p = 2, 3, \dots, 2N$ follows directly from Corollary 4.1 and Lemma 4.2. For $p = 2 - 2M, 3 - 2M, \dots, 0$ it also requires (4.7) and the relation between the asymptotic expansions of h and \check{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 \check{h} :

$$\begin{aligned} a_0 - b_0 &= (a_0 - \alpha) - (\check{a}_0 - \check{\alpha}) = \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi(1+\xi^2)} - \int_{\mathbb{R}} \frac{d\check{\mu}_0(\xi)}{\xi(1+\xi^2)} \\ &= \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi(1+\xi^2)} - \int_{\mathbb{R}} \frac{\xi d\mu_0(\xi)}{1+\xi^2} = \int_{\mathbb{R}} \frac{d\mu_0(\xi)}{\xi} \\ &= \lim_{\varepsilon \rightarrow 0^+} \lim_{\xi \rightarrow 0^+} \lim_{y \rightarrow 0^+} \int_{\varepsilon < |x| < \xi^{-1}} \frac{\operatorname{Im} h(x + iy)}{x} dx. \end{aligned}$$

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.

A.6 Auxiliary theorems

The following theorem can be found in *e.g.*, [22], 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*

$$0 \leq f_1(x) \leq f_2(x) \leq \dots \leq \infty, \quad \text{for all } x \in X$$

and

$$f_n(x) \rightarrow f(x), \quad \text{as } n \rightarrow \infty \text{ for all } x \in X.$$

Then f is measurable, and

$$\lim_{n \rightarrow \infty} \int_X f_n(x) d\mu(x) = \int_X f(x) d\mu(x).$$

The next theorem is also available in *e.g.*, [22], page 26:

Theorem A.2 (Lebesgue's Dominated Convergence Theorem). *Suppose $\{f_n\}$ is a sequence of complex-valued measurable functions on X such that*

$$f(x) = \lim_{n \rightarrow \infty} f_n(x)$$

exists for every $x \in X$. If there is a function $g \in L^1(\mu)$ such that

$$|f_n(x)| \leq g(x), \quad \text{for all } n = 1, 2, \dots \text{ and } x \in X,$$

then $f \in L^1(\mu)$,

$$\lim_{n \rightarrow \infty} \int_X |f_n(x) - f(x)| d\mu(x) = 0$$

and

$$\lim_{n \rightarrow \infty} \int_X f_n(x) d\mu(x) = \int_X f(x) d\mu(x).$$

References

- [1] N. I. Akhiezer. *The classical moment problem*. Oliver and Boyd, 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. F. Donoghue, Jr. *Distributions and Fourier Transforms*. Academic Press, New York, 1969.
- [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] E. A. Guillemin. *The Mathematics of Circuit Analysis*. MIT Press, 1949.
- [10] M. Gustafsson and D. Sjöberg. Sum rules and physical bounds on passive metamaterials. *New Journal of Physics*, 2010. in press.
- [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. Sum rule for the transmission cross section of apertures in thin opaque screens. *Opt. Lett.*, **34**(13), 2003–2005, 2009.
- [13] 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.
- [14] S. Hassi and A. Luger. Generalized zeros and poles of \mathcal{N}_κ -functions: On the underlying spectral structure. *Methods of Functional Analysis and Topology*, **12**(2), 131–150, 2006.
- [15] P. Henrici. *Applied and Computational Complex Analysis*, volume 2. John Wiley & Sons, New York, 1977.
- [16] 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.

-
- [17] F. W. King. *Hilbert Transforms, Volume 2*. Cambridge University Press, 2009.
- [18] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii. *Electrodynamics of Continuous Media*. Pergamon, Oxford, second edition, 1984.
- [19] J. Mashreghi. *Representation Theorems in Hardy Spaces*. Cambridge University Press, Cambridge, U.K., 2009.
- [20] H. M. Nussenzveig. *Causality and dispersion relations*. Academic Press, London, 1972.
- [21] K. N. Rozanov. Ultimate thickness to bandwidth ratio of radar absorbers. *IEEE Trans. Antennas Propagat.*, **48**(8), 1230–1234, August 2000.
- [22] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 1987.
- [23] M. G. Silveirinha and N. Engheta. Theory of supercoupling, squeezing wave energy, and field confinement in narrow channels and tight bends using ϵ near-zero metamaterials. *Phys. Rev. B*, **76**(24), 245109, 2007.
- [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] 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*, 2008.
- [26] A. H. Zemanian. An n-port realizability theory based on the theory of distributions. *IEEE Transactions on Circuit Theory*, **10**(2), 265–274, 1963.
- [27] A. H. Zemanian. *Distribution theory and transform analysis: an introduction to generalized functions, with applications*. Dover Publications, New York, 1987.
- [28] A. H. Zemanian. *Realizability Theory for Continuous Linear Systems*. Dover Publications, New York, 1995.

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. Technical Report LUTEDX/(TEAT-7194)/1-23/(2010), Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2010. <http://www.eit.lth.se>.

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 on 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 on 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 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 [16], discussing radiation emission and absorption by interstellar dust. Results similar to Purcell's can also be found in [3]. Limitations on antenna performance were introduced by Chu in [5]. Sohl et al. derives limitations on the extinction cross sections of arbitrary heterogeneous, anisotropic objects in [20], results that are directly applicable to antenna theory [7].

Electromagnetic fields can be decomposed into orthogonal vector spherical waves [10], 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.

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 on 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 on convolution form, and anisotropic in the static limit. The bounds 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 bounds 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, like antennas and metamaterials.

This paper is divided into sections as follows: 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 obtain the bounds in this paper, expressions for the vector spherical waves in the time domain 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 metallic nano-shells with dielectric cores, and another which 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 on convolution form presented in [2]. This section summarises this general approach in order to put the derivations of the limitations on scattering in the right context.

There are three major steps in the approach [2] to obtain sum rules 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.

The general approach is described more thoroughly in [2], where all the necessary proofs can be found. For a discussion on passive and causal systems, see also the books [22] and [23] by Zemanian and [14] by Nussenzveig and references therein.

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 $\mathbb{C}^+ = \{z, \text{Im } z > 0\}$, satisfying $\text{Im } h(z) \geq 0$ there. Furthermore, many Herglotz functions appearing in various applications are

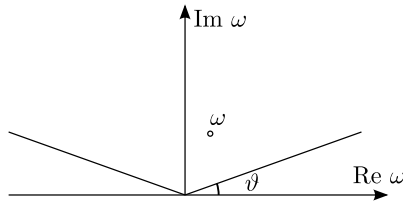


Figure 1: The cone $\{\omega : \vartheta \leq \arg \omega \leq \pi - \vartheta\}$ for some $\vartheta \in (0, \pi/2]$.

of the form $h(\omega) = \alpha + h_1(\omega)$, where h_1 exhibits the symmetry $h_1(\omega) = -h_1^*(-\omega^*)$ and $\alpha \in \mathbb{R}$ [2]. Such a function h is called symmetric in this paper, and it satisfies the low-frequency expansion

$$h(\omega) = \alpha + \sum_{n=0}^N A_{2n-1} \omega^{2n-1} + o(\omega^{2N-1}), \quad \text{as } \omega \hat{\rightarrow} 0, \quad (2.1)$$

for some integer $N \geq 0$. Here $A_{-1} \leq 0$ and all A_n are real. The limit $\omega \hat{\rightarrow} 0$ is a short-hand notation for $|\omega| \rightarrow 0$ for ω in the cone $\vartheta \leq \arg \omega \leq \pi - \vartheta$ for any $\vartheta \in (0, \pi/2]$, see Figure 1. The asymptotic expansion (2.1) is clearly valid as $\omega \rightarrow 0$ 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 :

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{\omega'' \rightarrow 0^+} \frac{2}{\pi} \int_{\varepsilon}^{\infty} \frac{\text{Im } h(\omega' + i\omega'')}{\omega'^{2p}} d\omega' = A_{2p-1} - \delta_{p,1}\beta, \quad p = 1, 2, \dots, N. \quad (2.2)$$

Here $\delta_{p,q}$ denotes the Kronecker delta and $\beta = \lim_{\omega \hat{\rightarrow} \infty} h(\omega)/\omega \geq 0$, which always exists finitely. The Herglotz function h is not necessarily holomorphic in a neighbourhood of the real line, but the distributional limit $\lim_{\omega'' \rightarrow 0^+} h(\omega' + i\omega'') \stackrel{\text{def}}{=} h(\omega')$ exists. The notation $\omega' = \text{Re } \omega$ and $\omega'' = \text{Im } \omega$ is used throughout this paper. The left-hand side of (2.2) is the integral of $\text{Im } h(\omega')/\omega'^{2p}$ in the distributional sense, *i.e.*, contributions from possible singularities in the interval $(0, \infty)$ are included [2].

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 on convolution form, *i.e.*, the first step of the general approach. This is done here. How the integral identities (2.2) can be used to derive sum rules for such systems once the low-frequency asymptotic behaviour of the transfer function has been determined is also explained here.

Consider a general mathematical model of a physical system in the time domain, $u(t) = \mathcal{R}v(t)$, where v and u are the input and output signals, respectively, related to each other by the operator \mathcal{R} . The context of distributions is natural, since generalised functions such as the delta function should be allowed; hence, the domain $D(\mathcal{R})$ of the operator \mathcal{R} is assumed to be some subset of the space of distributions

\mathcal{D}' . The assumptions of linearity, continuity, and time-translational invariance imply that the operator is on convolution form [22], *i.e.*,

$$u(t) = \mathcal{R}v(t) = w * v(t). \quad (2.3)$$

Such a system is fully described by its impulse response w . Many physical systems obey causality, which intuitively means that the output cannot precede the input. For the mathematical model (2.3) it means that $\text{supp } w \subseteq [0, \infty)$.

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

$$e(T) = \int_{-\infty}^T |v(t)|^2 - |u(t)|^2 dt \quad (2.4)$$

is non-negative for all $T \in \mathbb{R}$ and $v \in \mathcal{D}$, where \mathcal{D} denotes smooth functions of compact support [23].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 $v \in D(\mathcal{R})$.

One might expect that passive systems must be causal, and it turns out that this expectation is correct for operators on convolution form [23]. Also, passivity implies that the impulse response is a distribution of slow growth, $w \in \mathcal{S}'$ and hence Fourier transformable in the distributional sense [23]. 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 $\varphi \in \mathcal{S}$ [22], and the Fourier transform of φ is defined as $\mathcal{F}\varphi(\omega) = \int_{\mathbb{R}} \varphi(t)e^{i\omega t} dt$. The frequency domain version of (2.3) is

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

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} contains \mathbb{C}^+ and \tilde{w} is holomorphic there. Furthermore, the transfer-function $\tilde{w}(\omega)$ is bounded in magnitude by one for $\omega \in \mathbb{C}^+$ [23]. The transfer function \tilde{w} is not necessarily holomorphic in a neighbourhood of the real axis, but $\tilde{w}(\omega') = \lim_{\omega'' \rightarrow 0} \tilde{w}(\omega' + i\omega'')$ is well-defined for almost all $\omega' \in \mathbb{R}$ and bounded in magnitude by one [13].

One more assumption on the physical system is convenient (but not necessary): It is assumed that it maps real-valued input signals to real-valued output, which means that w is real. This implies the symmetry

$$\tilde{w}(\omega) = \tilde{w}^*(-\omega^*), \quad \text{Im } \omega > 0, \quad (2.5)$$

¹This is not the only way to classify passive systems, see [2, 23].

where the superscript $*$ is used to denote the complex conjugate.

The assumption of passivity assures that the transfer function \tilde{w} is holomorphic and bounded in magnitude by one in \mathbb{C}^+ . 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

$$h(\omega) = -i \log \left(\frac{\tilde{w}(\omega)}{B(\omega)} \right), \quad (2.6)$$

where

$$B(\omega) = \prod_{\omega_n} \frac{1 - \omega/\omega_n}{1 - \omega/\omega_n^*} \quad (2.7)$$

is a Blaschke product [13], repeating the possible zeros ω_n of \tilde{w} in \mathbb{C}^+ according to their multiplicity. The logarithm is defined in [2]. The symmetry (2.5) implies that $h(\omega)$ is symmetric in the sense discussed in Section 2.1, with $\alpha = \arg \tilde{w}(i\omega'')$.

The integral identities (2.2) applied to the function in (2.6) yield

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{\omega'' \rightarrow 0^+} \frac{2}{\pi} \int_{\varepsilon}^{\infty} \frac{1}{\omega'^{2p}} \ln \frac{1}{|\tilde{w}(\omega' + i\omega'')|} d\omega' = A_{2p-1} - \delta_{p,1}\beta, \quad p = 1, 2, \dots, N, \quad (2.8)$$

where it has been used that $|B(\omega' + i\omega'')| \rightarrow 1$ as $\omega'' \rightarrow 0$ for almost all $\omega' \in \mathbb{R}$ [13].² The low-frequency asymptotic expansion in (2.1) may be related to the behaviour of $\tilde{w}(\omega)$ as $\omega \hat{\rightarrow} 0$, where as before $\omega \hat{\rightarrow} 0$ is short-hand notation for $|\omega| \rightarrow 0$ for ω in the cone $\vartheta \leq \arg \omega \leq \pi - \vartheta$ for any $\vartheta \in (0, \pi/2]$. The cone assures that the low-frequency limit is only dependent on the behaviour of $w(t)$ for arbitrarily large times t [2]. If, however, $\tilde{w}(\omega)$ is holomorphic in a neighbourhood of the origin, the low frequency limit is identical whatever the argument of ω . The asymptotic behaviour of $\tilde{w}(\omega)$ as $\omega \hat{\rightarrow} 0$ 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 (2.8) is non-negative. In some cases, variational principles are used to bound the expansion coefficients A_p of h when they are unknown.

²It might be that $\tilde{w}(\omega)$ has an accumulation point of zeros for one or more $\omega_j \in \mathbb{R}$, in which case $B(\omega_j + i\omega'')$ does not tend to 1 as $\omega'' \rightarrow 0$. Then the Blaschke product must be included for the left-hand side to make sense, *i.e.*, (2.8) reads [2]:

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{\omega'' \rightarrow 0^+} \frac{2}{\pi} \int_{\varepsilon}^{\infty} \frac{1}{\omega'^{2p}} \ln \frac{|B(\omega' + i\omega'')|}{|\tilde{w}(\omega' + i\omega'')|} d\omega' = A_{2p-1} - \delta_{p,1}\beta, \quad p = 1, 2, \dots, N. \quad (2.9)$$

Equation (2.8) is understood to be replaced by (2.9) whenever necessary throughout this paper.

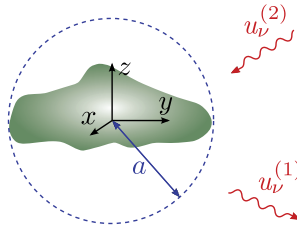


Figure 2: The scatterer is contained in a sphere of radius a centered at the origin. Outside this sphere, the electric and magnetic fields are expanded in outgoing and incoming vector spherical waves, $\mathbf{u}_\nu^{(1)}$ and $\mathbf{u}_\nu^{(2)}$, with index ν .

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.*, [10]. Their counterparts in the time domain have been treated by Shlivinski and Heyman [17, 18]. Both the frequency and time domain vector spherical waves are considered in this section, to be able to later derive sum rules and physical limitations on the scattering of these by following the approach presented in Section 2. A tilde ($\tilde{\cdot}$) is used in the remainder of this paper to denote functions in the frequency domain, and it is also convenient to employ the wavenumber $k = \omega/c$, so that $\tilde{f}(k) = \mathcal{F}f(\omega)$. Here c is the speed of light in free space.

Consider an 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 $\mathbf{u}_\nu^{(1)}$ and $\mathbf{u}_\nu^{(2)}$, respectively:

$$\tilde{\mathbf{E}}(\mathbf{r}, k) = k\sqrt{\eta_0} \sum_{\nu} i^{l+2-\tau} \tilde{b}_\nu^{(1)}(k) \mathbf{u}_\nu^{(1)}(k\mathbf{r}) + i^{l+2-\tau} \tilde{b}_\nu^{(2)}(k) \mathbf{u}_\nu^{(2)}(k\mathbf{r}). \quad (3.1)$$

Here η_0 is the wave impedance in free space. The spatial coordinate is denoted \mathbf{r} , and in the rest of the paper the notation $r = |\mathbf{r}|$ and $\hat{\mathbf{r}} = \mathbf{r}/r$ 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 $k\sqrt{\eta_0} i^{l+2-\tau}$ are included for consistency with the time domain expansion described below. The corresponding magnetic field is

$$\tilde{\mathbf{H}}(\mathbf{r}, k) = \frac{k}{\sqrt{\eta_0}} \sum_{\nu} i^{l+1-\tau} \tilde{b}_\nu^{(1)}(k) \mathbf{u}_\nu^{(1)}(k\mathbf{r}) + i^{l+1-\tau} \tilde{b}_\nu^{(2)}(k) \mathbf{u}_\nu^{(2)}(k\mathbf{r}), \quad (3.2)$$

where the dual multi-index $\bar{\nu} = \{\bar{\tau}, s, m, l\}$ with $\bar{\tau} = 3 - \tau$ has been introduced.

The time domain versions of (3.1) and (3.2) are helpful in order to make the assumptions on the scatterer of convolution form and passivity. Outgoing vector

spherical waves in the time domain are described thoroughly in [17, 18]. A short description, also covering incoming waves, is included here for clarity. Assuming that the fields vanish as $t \rightarrow -\infty$, the inverse Laplace transform may be applied to (3.1)–(3.2) with $k = is/c$ 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 $\tilde{\mathbf{E}}_{\text{T}} = \tilde{\mathbf{E}} - \hat{\mathbf{r}}(\hat{\mathbf{r}} \cdot \tilde{\mathbf{E}})$:

$$\tilde{\mathbf{E}}_{\text{T}}(\mathbf{r}, k) = \frac{\sqrt{\eta_0}}{r} \sum_{\nu} \left[\tilde{b}_{\nu}^{(1)}(k) e^{-sr/c} R_{\tau,l}^{(1)}(sr/c) + \tilde{b}_{\nu}^{(2)}(k) e^{sr/c} R_{\tau,l}^{(2)}(sr/c) \right] \mathbf{A}_{\nu}(\hat{\mathbf{r}}), \quad (3.3)$$

where

$$\begin{aligned} R_{1,l}^{(1)}(s) &= \sum_{n=0}^l D_{n,l} s^{-n} \\ R_{1,l}^{(2)}(s) &= (-1)^{l-1} R_{1,l}^{(1)}(-s) \\ R_{2,l}^{(j)}(s) &= R_{1,l-1}^{(j)}(s) + \frac{l}{s} R_{1,l}^{(j)}(s), \quad j = 1, 2, \end{aligned}$$

and $D_{n,l} = (l+n)!/(2^n n!(l-n)!)$ according to (A.8)–(A.9). The vector spherical harmonics \mathbf{A}_{ν} are defined in Appendix A.1. Applying the inverse Laplace transform yields

$$\mathbf{E}_{\text{T}}(\mathbf{r}, t) = \frac{\sqrt{\eta_0}}{r} \sum_{\nu} \left[\mathcal{R}_{\tau,l}^{(1)} b_{\nu}^{(1)}(t - r/c) + \mathcal{R}_{\tau,l}^{(2)} b_{\nu}^{(2)}(t + r/c) \right] \mathbf{A}_{\nu}(\hat{\mathbf{r}}).$$

Here the operators $\mathcal{R}_{\tau,l}^{(j)} : \mathcal{D} \rightarrow \mathcal{D}$ in the time domain are defined by

$$\begin{aligned} \mathcal{R}_{1,l}^{(j)} f(t) &= (\pm 1)^{l-1} \sum_{n=0}^l D_{n,l} \left(\pm \frac{c}{r} d_t^{-1} \right)^n f(t) \\ \mathcal{R}_{2,l}^{(j)} f(t) &= \mathcal{R}_{1,l-1}^{(j)} f(t) \pm l \frac{c}{r} d_t^{-1} \mathcal{R}_{1,l}^{(j)} f(t), \end{aligned}$$

where the upper (lower) signs are for $j = 1$ ($j = 2$). The inverse to differentiation d_t^{-1} is chosen so that $d_t^{-1} f(t)$ is the distributional primitive to f that vanishes at $t = -\infty$, *i.e.*, $d_t^{-1} f(t) = \int_{-\infty}^t f(t') dt'$ for regular functions f . A similar representation is used for the magnetic field, giving

$$\mathbf{H}_{\text{T}}(\mathbf{r}, t) = \frac{1}{r\sqrt{\eta_0}} \sum_{\nu} \left[\mathcal{R}_{\tau,l}^{(1)} b_{\nu}^{(1)}(t - r/c) + \mathcal{R}_{\tau,l}^{(2)} b_{\nu}^{(2)}(t + r/c) \right] (-1)^{\tau-1} \mathbf{A}_{\nu}(\hat{\mathbf{r}}).$$

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

$$\begin{aligned} P(r, t) &= \int_{\Omega_{\hat{\mathbf{r}}}} r^2 (-\hat{\mathbf{r}}) \cdot [\mathbf{E}_{\text{T}}(\mathbf{r}, t) \times \mathbf{H}_{\text{T}}(\mathbf{r}, t)] d\Omega_{\hat{\mathbf{r}}} = - \int_{\Omega_{\hat{\mathbf{r}}}} r^2 \mathbf{E}_{\text{T}}(\mathbf{r}, t) \cdot [\mathbf{H}_{\text{T}}(\mathbf{r}, t) \times \hat{\mathbf{r}}] d\Omega_{\hat{\mathbf{r}}} \\ &= - \int_{\Omega_{\hat{\mathbf{r}}}} \left[\sum_{\nu} \sum_{j=1}^2 \mathcal{R}_{\tau,l}^{(j)} b_{\nu}^{(j)}(t \mp r/c) \mathbf{A}_{\nu}(\hat{\mathbf{r}}) \right] \cdot \left[\sum_{\nu} \sum_{j=1}^2 \mathcal{R}_{\tau,l}^{(j)} b_{\nu}^{(j)}(t \mp r/c) \mathbf{A}_{\nu}(\hat{\mathbf{r}}) \right] d\Omega_{\hat{\mathbf{r}}}, \end{aligned}$$

where $\Omega_{\hat{r}} = \{(\theta, \phi) : 0 \leq \theta \leq \pi, 0 \leq \phi \leq 2\pi\}$ is the unit sphere and $d\Omega_{\hat{r}} = \sin\theta d\theta d\phi$. 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 ν may be replaced by one. Also, all cross-terms in j cancel each other:

$$P(r, t) = - \sum_{\nu} \sum_{j=1}^2 \left[\mathcal{R}_{\tau, i}^{(j)} b_{\nu}^{(j)}(t \mp r/c) \right] \left[\mathcal{R}_{\bar{\tau}, i}^{(j)} b_{\nu}^{(j)}(t \mp r/c) \right].$$

The power $P(r, t)$ may be divided into one radiating part, and another part appertaining to the reactive near-field:

$$P(r, t) = P_{\text{rad}}(r, t) + P_{\text{react}}(r, t), \quad (3.4)$$

where

$$P_{\text{rad}}(r, t) = \sum_{\nu} |b_{\nu}^{(2)}(t + r/c)|^2 - |b_{\nu}^{(1)}(t - r/c)|^2 \quad (3.5)$$

is only dependent on r via $t \mp r/c$. The reactive power $P_{\text{react}}(r, t)$ tends to zero as $r \rightarrow \infty$, and furthermore it has a zero mean for all $r \geq a$, *i.e.*,

$$\int_{-\infty}^{\infty} P_{\text{react}}(r, t) dt = 0. \quad (3.6)$$

This result is derived in [18], where also P_{rad} and P_{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 [18]. It is made clear there that the reactive power becomes larger for higher order modes if the radiative power is the same.

4 The scattering matrix $\tilde{\mathbf{S}}_{\text{S}}$

This section introduces the scattering matrix $\tilde{\mathbf{S}}_{\text{S}}$, 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. Herglotz functions corresponding to (2.6) 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 bounds on the diagonal elements of $\tilde{\mathbf{S}}_{\text{S}}$.

Assume that the scatterer is linear, continuous, and time-translational invariant, *i.e.*, that the constitutive relations relating the electric and magnetic flux densities $\mathbf{D}(t)$ and $\mathbf{B}(t)$ to the electric and magnetic fields $\mathbf{E}(t)$ and $\mathbf{H}(t)$ are on convolution form, as discussed in Section 2.2. In this case the relation between the outgoing and incoming amplitudes $b_{\nu}^{(1)}(t)$ and $b_{\nu}^{(2)}(t)$ must also be on convolution form, $b_{\nu}^{(1)}(t) = \sum_{\nu'} S_{\nu, \nu'} * b_{\nu'}^{(2)}(t)$. With matrix notation,

$$\mathbf{b}^{(1)}(t) = \mathbf{S}_{\text{S}} * \mathbf{b}^{(2)}(t), \quad (4.1)$$

where $\mathbf{b}^{(1)} = [b_1^{(1)} b_2^{(1)} \dots]^T$ and $\mathbf{b}^{(2)}$ is defined analogously. The order of the multi-index is specified in Appendix A.1. In the frequency domain, (4.1) reads

$$\tilde{\mathbf{b}}^{(1)}(k) = \tilde{\mathbf{S}}_S(k) \tilde{\mathbf{b}}^{(2)}(k), \quad (4.2)$$

where $\tilde{\mathbf{S}}_S(k)$ is the infinite dimensional scattering matrix.

4.1 Implications of passivity on $\tilde{\mathbf{S}}_S$

It is now shown that the elements of $\mathbf{S}_S(t-2a/c)$ 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 $r \geq a$ before the time T must be non-negative. This means

$$\int_{-\infty}^T P_{\text{rad}}(r, t) dt = \int_{-\infty}^T \sum_{\nu} |b_{\nu}^{(2)}(t+r/c)|^2 - |b_{\nu}^{(1)}(t-r/c)|^2 dt \geq 0,$$

for all $T \in \mathbb{R}$ and $r \geq a$, 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

$$\int_{-\infty}^T P_{\text{rad}, \nu'}(r, t) dt = \int_{-\infty}^T |b_{\nu'}^{(2)}(t+r/c) - \sum_{\nu} |S_{\nu, \nu'} * b_{\nu}^{(2)}(t-r/c)|^2 dt \geq 0,$$

for all $T \in \mathbb{R}$, $r \geq a$ and ν' . Note that the above energy expression closely resembles that in (2.4), except for the time shifts $-2r/c$ in the outgoing waves. Hence $S_{\nu, \nu'}(t-2a/c)$ is the impulse response of a passive operator for all ν, ν' , and so its Fourier-transform $e^{i2ka} \tilde{S}_{\nu, \nu'}(k)$ is holomorphic and bounded in magnitude by one for $k \in \mathbb{C}^+$, see Section 2.2 and [2, 23]. Furthermore, $e^{i2ka} \tilde{S}_{\nu, \nu'}(k)$ satisfies the symmetry (2.5).

The time shift $-2a/c$ 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 [14].

4.2 Low-frequency asymptotic behaviour of $\tilde{\mathbf{S}}_S$

To derive equalities of the type (2.8), the low-frequency asymptotic expansion of the $\tilde{\mathbf{S}}_S$ -matrix is required. For this reason, consider the alternative decomposition of the electric field in outgoing and regular vector spherical waves:

$$\tilde{\mathbf{E}}(\mathbf{r}, k) = k\sqrt{\eta_0} \sum_{\nu} i^{l+2-\tau} \tilde{d}_{\nu}^{(1)}(k) \mathbf{u}_{\nu}^{(1)}(k\mathbf{r}) + i^{l+2-\tau} \tilde{d}_{\nu}^{(2)}(k) \mathbf{v}_{\nu}(k\mathbf{r}). \quad (4.3)$$

Here $\mathbf{v}_{\nu}(k\mathbf{r})$ denotes regular vector spherical waves, defined as $\mathbf{v}_{\nu}(k\mathbf{r}) = (\mathbf{u}_{\nu}^{(1)}(k\mathbf{r}) + \mathbf{u}_{\nu}^{(2)}(k\mathbf{r}))/2$ (see Appendix A.1). The relation corresponding to (4.2) is $\tilde{\mathbf{d}}^{(1)}(k) =$

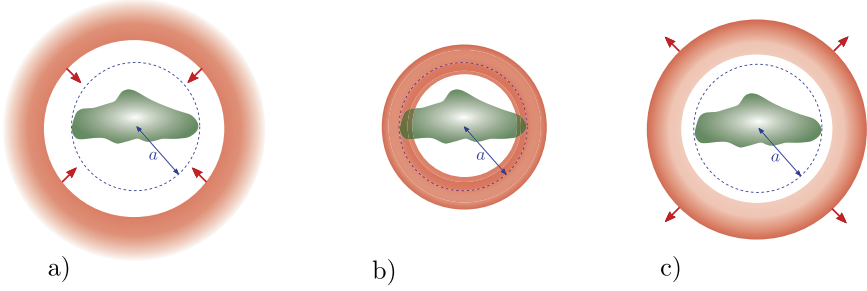


Figure 3: An incoming spherical wave $b_{\nu'}^{(2)}(t + r/c)$: a) impinges on the scatterer, b) interacts with the scatterer, and c) creates outgoing waves $\sum_{\nu} S_{\nu, \nu'} * b_{\nu'}^{(2)}(t - r/c)$. Note that the picture is over-simplified, but it makes it believable that $S_{\nu, \nu'}(t - 2a/c)$ is the impulse response of a passive operator for all ν and ν' .

$\tilde{\mathbf{T}}(k) \tilde{\mathbf{d}}^{(2)}(k)$, where $\tilde{\mathbf{T}}$ is the so called transition, or T -, matrix. Evidently, $\tilde{\mathbf{S}}_s = 2\tilde{\mathbf{T}} + \mathbf{I}$, where \mathbf{I} is the infinite dimensional identity matrix.

The advantage of a decomposition in regular and outgoing waves is that a plane wave $\tilde{\mathbf{E}}_i$ impinging on the scatterer is regular everywhere, while the produced scattered field $\tilde{\mathbf{E}}_s$ has to satisfy the radiation condition. Accordingly, in this situation $\tilde{\mathbf{E}}_i$ equals the sum over \mathbf{v}_{ν} , while $\tilde{\mathbf{E}}_s$ is the sum over $\mathbf{u}_{\nu}^{(1)}$. Consider a plane wave $\mathbf{E}_0(t - \mathbf{r} \cdot \hat{\mathbf{k}}/c)$ propagating in the $\hat{\mathbf{k}}$ -direction, corresponding to $e^{i\mathbf{r} \cdot \mathbf{k}} \tilde{\mathbf{E}}_0(k)$ in the frequency domain. Here $\mathbf{k} = k\hat{\mathbf{k}}$ and as usual $\tilde{\mathbf{E}}_0(k) = (\mathcal{F}\mathbf{E}_0)(\omega)$ with $k = \omega/c$.

The radiating part of the scattered field is described by the far-field amplitude \mathbf{F} , viz.,

$$\mathbf{E}_s(t, \mathbf{r}) = \frac{\mathbf{F}(t - r/c, \hat{\mathbf{r}})}{r} + \mathcal{O}(r^{-2}), \quad \tilde{\mathbf{E}}_s(k, \mathbf{r}) = \frac{e^{ikr} \tilde{\mathbf{F}}(k, \hat{\mathbf{r}})}{r} + \mathcal{O}(r^{-2}), \quad \text{as } r \rightarrow \infty. \quad (4.4)$$

Due to the assumption of convolution form for the constitutive relations, a scattering dyadic $\tilde{\mathbf{S}}$ may be defined:

$$\mathbf{F}(t, \hat{\mathbf{r}}) = \mathbf{S}(\cdot, \hat{\mathbf{r}}, \hat{\mathbf{k}}) * \mathbf{E}_0(t), \quad \tilde{\mathbf{F}}(k, \hat{\mathbf{r}}) = \tilde{\mathbf{S}}(k, \hat{\mathbf{r}}, \hat{\mathbf{k}}) \cdot \tilde{\mathbf{E}}_0(k). \quad (4.5)$$

The elements of the T -matrix can be deduced from the scattering dyadic:

$$\tilde{T}_{\nu, \nu'}(k) = \frac{ik}{4\pi} \int \int \mathbf{A}_{\nu}(\hat{\mathbf{r}}) \cdot \tilde{\mathbf{S}}(k; \hat{\mathbf{r}}, \hat{\mathbf{k}}) \cdot \mathbf{A}_{\nu'}(\hat{\mathbf{k}}) d\Omega_{\hat{\mathbf{r}}} d\Omega_{\hat{\mathbf{k}}}. \quad (4.6)$$

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

$$\begin{aligned} \tilde{\mathbf{D}}(0, \mathbf{r}) &= \epsilon_0 \boldsymbol{\epsilon}(0, \mathbf{r}) \cdot \tilde{\mathbf{E}}(0, \mathbf{r}) \\ \tilde{\mathbf{B}}(0, \mathbf{r}) &= \mu_0 \boldsymbol{\mu}(0, \mathbf{r}) \cdot \tilde{\mathbf{H}}(0, \mathbf{r}). \end{aligned}$$

Here $\tilde{\mathbf{D}}(k, \mathbf{r})$ denotes the electric flux density and $\tilde{\mathbf{B}}(k, \mathbf{r})$ the magnetic flux density at the point \mathbf{r} and wavenumber k . The relative permittivity and permeability dyadics are denoted $\boldsymbol{\epsilon}(k, \mathbf{r})$ and $\boldsymbol{\mu}(k, \mathbf{r})$, respectively, and ϵ_0 and μ_0 are the permittivity and permeability of free space, respectively. The low frequency expansion of $\tilde{\mathbf{S}}(k, \hat{\mathbf{r}}, \hat{\mathbf{k}})$ is then

$$\tilde{\mathbf{S}}(k, \hat{\mathbf{r}}, \hat{\mathbf{k}}) \cdot \mathbf{E} = \frac{k^2}{4\pi} \left[\hat{\mathbf{r}} \times \left((\boldsymbol{\gamma}_e \cdot \mathbf{E}) \times \hat{\mathbf{r}} \right) + \left(\boldsymbol{\gamma}_m \cdot (\hat{\mathbf{k}} \times \mathbf{E}) \right) \times \hat{\mathbf{r}} \right] + \mathcal{O}(k^3), \quad \text{as } k \rightarrow 0, \quad (4.7)$$

where \mathbf{E} is a constant vector. The electric polarizability dyadic $\boldsymbol{\gamma}_e$ relates the electric dipole moment induced in the scatterer to an applied static homogeneous electric field $\tilde{\mathbf{E}}(0)$, *viz.*, $\mathbf{p} = \epsilon_0 \boldsymbol{\gamma}_e \cdot \tilde{\mathbf{E}}(0)$. Similarly, the magnetic dipole moment induced by an applied static homogeneous magnetic field $\tilde{\mathbf{H}}(0)$ is given by $\mathbf{m} = \boldsymbol{\gamma}_e \cdot \tilde{\mathbf{H}}(0)$. The polarizability dyadics are thoroughly discussed in [12] and [20]. Now let $\mathbf{E} = \mathbf{A}_{\nu'}(\hat{\mathbf{k}})$. From (4.6) and (4.7) it follows that

$$\tilde{S}_{\nu, \nu'}(k) = \delta_{\nu, \nu'} + i2\rho_{\nu, \nu'} k^3 a^3 + \mathcal{O}(k^4), \quad \text{as } k \rightarrow 0, \quad (4.8)$$

where

$$\rho_{\nu, \nu'} = \frac{1}{16\pi^2 a^3} \iint \mathbf{A}_{\nu}(\hat{\mathbf{r}}) \cdot \boldsymbol{\gamma}_e \cdot \mathbf{A}_{\nu'}(\hat{\mathbf{k}}) + (-1)^{\tau + \tau'} \mathbf{A}_{\bar{\nu}}(\hat{\mathbf{r}}) \cdot \boldsymbol{\gamma}_m \cdot \mathbf{A}_{\bar{\nu}'}(\hat{\mathbf{k}}) d\Omega_{\hat{\mathbf{r}}} d\Omega_{\hat{\mathbf{k}}}.$$

Here (A.2) was used, and the dual multi-index is still defined as $\bar{\nu} = \{\bar{\tau}, s, m, l\}$ with $\bar{\tau} = 3 - \tau$.

Let $\gamma_{e,xx} = \hat{\mathbf{x}} \cdot \boldsymbol{\gamma}_e \cdot \hat{\mathbf{x}}$, $\gamma_{e,xy} = \hat{\mathbf{x}} \cdot \boldsymbol{\gamma}_e \cdot \hat{\mathbf{y}}$ and so on, and use the identities (A.5)–(A.7). This gives explicit expressions for $\rho_{\nu, \nu'}$:

$$\rho_{\nu, \nu'} = \frac{1}{6\pi a^3} \delta_{l,1} \delta_{l',1} \delta_{\tau, \tau'} \gamma_{\{m/e\}, mn'}, \quad (4.9)$$

where m (e) should be chosen for $\tau = 1$ ($\tau = 2$) and

$$n = \begin{cases} x, & \text{for } s = 1, m = 1 \\ y, & \text{for } s = 2, m = 1 \\ z, & \text{for } s = 1, m = 0, \end{cases}$$

and similarly for n' . Note that $\rho_{\nu, \nu'} = 0$ for non-dipole modes ($l \geq 2$ or $l' \geq 2$), 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 on 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 \mathbf{r} within the scatterer. It is common to assume that the permittivity and permeability dyadics are symmetric in the static limit, *i.e.*, $\boldsymbol{\epsilon}(0, \mathbf{r}) = \boldsymbol{\epsilon}(0, \mathbf{r})^T$ and

$\boldsymbol{\mu}(0, \mathbf{r}) = \boldsymbol{\mu}(0, \mathbf{r})^T$. This implies that the polarizability dyadics are also symmetric [20], and hence diagonal for a suitable choice of coordinates. Closed form expressions for the polarizability dyadics exists for anisotropic homogeneous spheroidal scatterers, see [20] and references therein. For the simple case of an isotropic sphere of radius a , they are

$$\begin{aligned}\gamma_e &= 4\pi a^3 \frac{\epsilon(0) - 1}{\epsilon(0) + 2} \mathbf{I} \\ \gamma_m &= 4\pi a^3 \frac{\mu(0) - 1}{\mu(0) + 2} \mathbf{I},\end{aligned}$$

where \mathbf{I} is the identity dyadic.

Furthermore, under the assumption of symmetry it can be shown that γ_e and γ_m are nondecreasing as functions of $\epsilon(0, \mathbf{r})$ and $\boldsymbol{\mu}(0, \mathbf{r})$ [19]. More specifically, consider two objects with permittivity $\epsilon(0, \mathbf{r})$ and $\epsilon'(0, \mathbf{r})$, respectively. If $\epsilon'(0, \mathbf{r}) - \epsilon(0, \mathbf{r})$ is a positive semidefinite dyadic for all \mathbf{r} in the object, then $\gamma'_e - \gamma_e$ is positive semidefinite as well. The same holds for γ_m , with $\epsilon(0, \mathbf{r})$ replaced by $\boldsymbol{\mu}(0, \mathbf{r})$. The diagonal elements of γ_e and γ_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 nondecreasing as functions of $\epsilon(0, \mathbf{r})$ and $\boldsymbol{\mu}(0, \mathbf{r})$, 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 γ_e and γ_m are bounded by the largest eigenvalue $\gamma_1 \leq 4\pi a^3$ of the high-contrast polarizability dyadic γ_∞ of that geometry. Therefore a sharper bound on $\rho_{\nu, \nu'}$, given by $\rho_{\nu, \nu'} \leq \gamma_1 / (6\pi a^3) \leq 2/3$, can be determined. The high-contrast polarizability dyadics γ_∞ of many geometries can be calculated numerically, see [9] for some examples.

A widely used material model is the perfect electric conductor (PEC). For a PEC inclusion, $\epsilon(0) = \infty$ and $\mu(0) = 0$. Consequently, γ_e (γ_m) is nondecreasing (nonincreasing) as the volume of the PEC inclusion increases [19].

4.4 Sum rules and physical limitations on $\tilde{\mathbf{S}}_s$

Now it has been shown that $e^{i2ka} \tilde{S}_{\nu, \nu'}(k)$ is a holomorphic function bounded in magnitude by one in \mathbb{C}^+ for all ν and ν' , 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 (2.8). The Herglotz function corresponding to (2.6) is

$$h_{\nu, \nu'}(k) = -i \log \left(\frac{e^{i2ka} \tilde{S}_{\nu, \nu'}(k)}{B_{\nu, \nu'}(k)} \right).$$

Here $B_{\nu, \nu'}$ is a Blaschke products of the form (2.7) for each pair (ν, ν') . Since $e^{i2ka} \tilde{S}_{\nu, \nu'}(k) \rightarrow 1$ as $k \rightarrow 0$ 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]):

$$h_{\nu,\nu}(k) = 2ka + 2\rho_{\nu,\nu}k^3a^3 + 2 \sum_n \sum_{q=1,3,\dots}^{\infty} \frac{k^q}{q} \operatorname{Im} \frac{1}{k_n^q} + \mathcal{O}(k^4), \quad \text{as } k \rightarrow 0. \quad (4.10)$$

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 $k \rightarrow 0$. Only terms with odd q appear in (4.10) due to the symmetry (2.5).

Note that the low-frequency asymptotic expansions (4.7) and (4.10) are valid as $k \rightarrow 0$ for all arguments of k , and especially as $k \rightarrow 0$. With the notation of Section 2.1, $N = 2$ and hence two sum rules of the type (2.8) (using $p = 1, 2$) can be deduced:

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{k'' \rightarrow 0^+} \frac{1}{\pi} \int_{\varepsilon}^{\infty} \frac{1}{k'^2} \ln \frac{1}{|\tilde{S}_{\nu,\nu}(k' + ik'')|} dk' = a - \frac{\beta_{\nu,\nu}}{2} + \sum_n \operatorname{Im} \frac{1}{k_n} \quad (4.11)$$

and

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{k'' \rightarrow 0^+} \frac{1}{\pi} \int_{\varepsilon}^{\infty} \frac{1}{k'^4} \ln \frac{1}{|\tilde{S}_{\nu,\nu}(k' + ik'')|} dk' = a^3 \rho_{\nu,\nu} + \frac{1}{3} \sum_n \operatorname{Im} \frac{1}{k_n^3}. \quad (4.12)$$

Here $k' = \operatorname{Re} k$ and $k'' = \operatorname{Im} k$, in consistency with previous notation. As discussed in Section 2.1, the left-hand sides may be interpreted as integrals of $-\ln |\tilde{S}_{\nu,\nu}(k')|/k'^{2p}$ in the distributional sense, *i.e.*, contributions from possible singularities in the interval $(0, \infty)$ should be included.

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 $\beta_{\nu,\nu} = \lim_{k \rightarrow \infty} h_{\nu,\nu}(k)/k$ 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 $S_0 = \sup_{k' \in \mathcal{K}} |\tilde{S}_{\nu,\nu}(k')|$, it follows that³

$$\frac{B_K \ln S_0^{-1}}{\pi} \leq k_0 a + \sum_n \operatorname{Im} \frac{k_0}{k_n} \quad (4.13)$$

and

$$\frac{B_K \ln S_0^{-1}}{\pi} \leq k_0^3 a^3 \rho_{\nu,\nu} + \frac{1}{3} \sum_n \operatorname{Im} \frac{k_0^3}{k_n^3}. \quad (4.14)$$

³Here $S_0 = \sup_{k' \in \mathcal{K}} |\tilde{S}_{\nu,\nu}(k')|$ should be interpreted as the supremum over those $k' \in \mathcal{K}$ such that $\tilde{S}_{\nu,\nu}(k')$ is well-defined (recall that it is well-defined for almost all $k' \in \mathbb{R}$). Also note here that the inequalities (4.13)–(4.14) are valid even if (4.11)–(4.12) must be interpreted as (2.9).

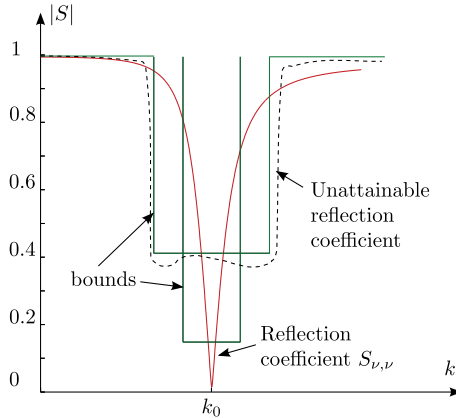


Figure 4: Interpretation of the bound (4.15). In the figure, bounds for a given center wavenumber k_0 are depicted for two different values of S_0 (and thus two different values of B_K). The bound states that the magnitude of all reflection coefficients $S_{\nu,\nu}$ have to intersect the boxes, when the scatterer satisfies the aforementioned assumptions; also shown in the figure is one attainable and one unattainable reflection coefficient.

Here it has been used that $k_0^{2p-1} \int_{\mathcal{K}} 1/k^{2p} dk \geq B_K$ for $p = 0, 1, \dots$. Note also that $k_0^{2p-1} \int_{\mathcal{K}} 1/k^{2p} dk \approx B_K$ when $B_K \ll 1$.

The sum in the right-hand side of (4.13) is non-positive (since $\text{Im } k_n > 0$ for all k_n), and so

$$\frac{B_K \ln S_0^{-1}}{\pi} \leq k_0 a.$$

An alternative bound not containing the sum over all zeros can also be derived (see Appendix A.3):

$$\begin{aligned} \frac{B_K \ln S_0^{-1}}{\pi} &\leq k_0 a - \sqrt[3]{\iota + \zeta} + \sqrt[3]{\iota - \zeta} \\ &= \left(\frac{1}{3} + \rho_{\nu,\nu} \right) (k_0^3 a^3 - k_0^5 a^5) + \mathcal{O}(k_0^7), \quad \text{as } k_0 \rightarrow 0. \end{aligned} \quad (4.15)$$

Here the material and geometry of the scatterer are contained in $\rho_{\nu,\nu}$ via $\zeta = 3k_0 a(1 - \rho_{\nu,\nu} k_0^2 a^2)/2$ and $\iota = \sqrt{1 + \zeta^2}$. The term $k_0^3 a^3/3$ in the bound stems from the circumscribing sphere. The bound (4.15) states that, somewhere on the wavenumber interval \mathcal{K} , the reflection coefficient $\tilde{S}_{\nu,\nu}(k')$ for mode ν 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.

An interpretation of the bound is as a limitation on the absorption of a vector spherical wave over a bandwidth. To see this, consider the total energy $e(\infty)$ absorbed by the scatterer when the incoming field consists of only the mode ν , and

$b_\nu^{(2)}(t)$ is assumed to be in L^2 . By (3.4)–(3.6), it is (with $r \geq a$)

$$e(\infty) = \int_{-\infty}^{\infty} P(r, t) dt = \int_{-\infty}^{\infty} P_{\text{rad}}(r, t) dt = \int_{-\infty}^{\infty} |b_\nu^{(2)}(t)|^2 - \sum_{\nu'} |S_{\nu', \nu} * b_\nu^{(2)}(t)|^2 dt.$$

The expression for $e(\infty)$ may be rewritten with Parseval's equation:

$$e(\infty) = \frac{1}{2\pi c} \int_{-\infty}^{\infty} |\tilde{b}_\nu^{(2)}(k')|^2 \left(1 - \sum_{\nu'} |\tilde{S}_{\nu', \nu}(k')|^2 \right) dk'.$$

Hence $1 - \sum_{\nu'} |\tilde{S}_{\nu', \nu}(k')|^2$ is the normalised energy of the incoming mode ν that is absorbed by the scatterer at wavenumber k' ; all of the incoming energy is absorbed if $\sum_{\nu'} |\tilde{S}_{\nu', \nu}(k')|^2 = 0$, while no energy is absorbed in the case $\sum_{\nu'} |\tilde{S}_{\nu', \nu}(k')|^2 = 1$. The absorbed normalised energy is obviously less than or equal to $1 - |\tilde{S}_{\nu, \nu}(k')|^2$. Also recall that $\tilde{S}_{\nu', \nu}(k') \rightarrow 0$, as $k' \rightarrow 0$, when $\nu' \neq \nu$, due to (4.8).

5 Examples

5.1 Absorbing spherical 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 [4]. Hereafter they are illuminated by laser light, causing most of the cancer cells to die, see Figure 1 in [4]. It is thus desirable to design nanoshells that absorb large parts of the laser energy. In [4, 15], 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 (4.15) is well suited to study this problem, since the normalised absorbed energy from mode ν is bounded by $1 - |\tilde{S}_{\nu, \nu}|^2$ as discussed in Section 4.4. An illustration can be found in Figure 5.

5.2 Physical limitations on antennas

As discussed in Section 4.4, (4.15) 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. It is unusual to compute the $\tilde{\mathbf{S}}_S$ -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 to minimise 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

$$P_{\text{rad}}(k) \leq (1 - |\Gamma(k)|^2) P_{\text{in}}(k),$$

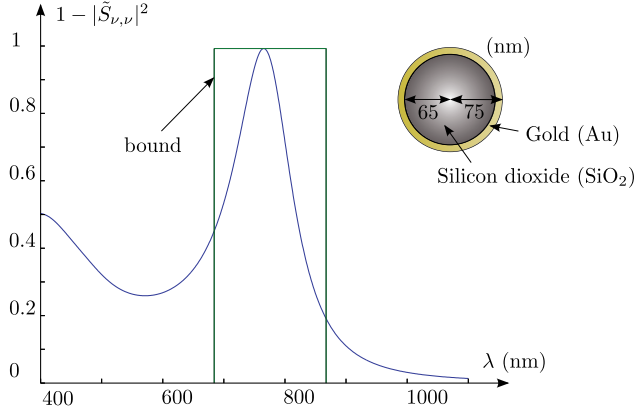


Figure 5: The reflection coefficient $\tilde{S}_{\nu,\nu}$ of the electric dipole modes ($\tau = 2, l = 1$) for a spherical silicon dioxide core of radius 65 nm covered by a layer of gold of thickness 10 nm. Here λ denotes the wavelength. The bound is (4.15) with $\rho_{\nu,\nu} = 2/3$, and it states that the curve has to intersect the box. The reflection coefficient $\tilde{S}_{\nu,\nu}$ was calculated from the closed form expression, using a Matlab-script for a Lorentz-Drude model for gold by Ung et al. [21]. Silicon dioxide has negligible losses and refractive index $n \approx 1.5$ for wavelengths 400–1100 nm [11].

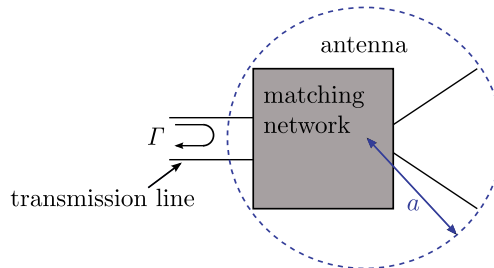


Figure 6: The antenna and matching network considered in Example 5.2.

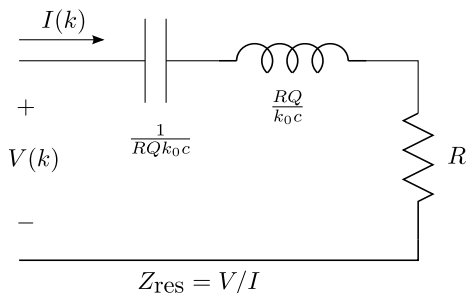


Figure 7: For many antennas, the impedance $Z_{\text{res}}(k)$ of the resonance circuit is a good approximation for the antenna input impedance $Z(k)$ close to its resonance wavenumber k_0 [8]. The quality factor Q is given by (5.1).

with equality if there are no ohmic losses in the antenna.

Many antennas can be modeled by the resonance circuit in Figure 7 in a frequency interval close to their respective resonance frequencies [8]. Here the quality factor is

$$Q = k_0 c \frac{Z'(k_0)}{2R}, \quad (5.1)$$

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 [6], it is straightforward to show that [8]

$$\frac{B_K \ln \Gamma_0^{-1}}{\pi} \leq \frac{1}{Q}, \quad (5.2)$$

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 .

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 (4.15) to the quality factor of an antenna; since $1 - |T|^2$ places a bound on the radiated power in terms of the input power, and $1 - |S_{\nu,\nu}|$ limits the absorbed power from a single mode ν , Γ and $S_{\nu,\nu}$ are on equal footing. In Figure 8, the bound in (4.15) 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 between electromagnetic waves and matter is vital in many applications, from classical optics to antenna theory. One

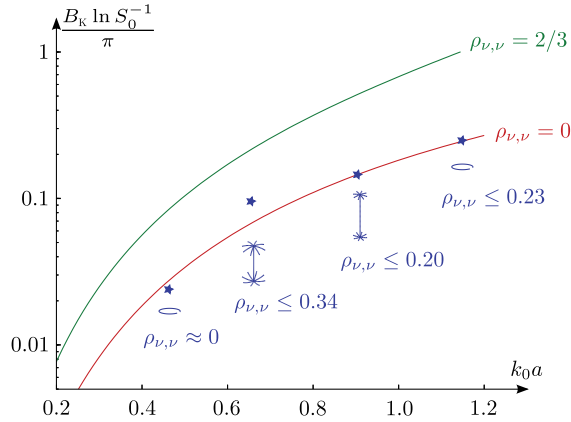


Figure 8: The lines are the bound (4.15) for $\rho = 2/3$ and $\rho = 0$, respectively. Four wire antennas were used in the example, with wires modeled as perfect electric conductors of diameter 2 mm. The radii of the loops are 60 mm (giving $a = 62$ mm), and the heights of the umbrellas are 100 mm (so that $a = 52$ 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 γ_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.

way to analyse the interaction is to apply physical limitations to it; in essence, the physical limitations state what cannot be expected from a certain physical system.

There are several publications addressing physical limitations in scattering and antenna theory, see *e.g.*, [3, 5, 7, 16, 20]. 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 [10]. In wireless communication, they are intimately linked to the orthogonal communication channels of multiple-input multiple-output (MIMO) systems.

The derivation makes use of a general approach to obtain sum rules and physical limitations on passive physical systems on convolution form presented in [2]. The limitations in this paper are valid for all heterogeneous passive scatterers with constitutive relations on 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 bounds can be interpreted as limits 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.

7 Acknowledgments

The financial support by the High Speed Wireless Communications Center of the Swedish Foundation for Strategic Research (SSF) is gratefully acknowledged.

Appendix A Definitions and derivations

A.1 Definition of vector spherical waves

The incoming ($j = 2$) and outgoing ($j = 1$) vector spherical waves [10] are defined as

$$\begin{cases} \mathbf{u}_{1sml}^{(j)}(k\mathbf{r}) = h_l^{(j)}(kr) \mathbf{A}_{1sml}(\hat{\mathbf{r}}) \\ \mathbf{u}_{2sml}^{(j)}(k\mathbf{r}) = \frac{(kr h_l^{(j)}(kr))'}{kr} \mathbf{A}_{2sml}(\hat{\mathbf{r}}) + \sqrt{l(l+1)} \frac{h_l^{(j)}(kr)}{kr} \mathbf{A}_{3sml}(\hat{\mathbf{r}}). \end{cases} \quad (\text{A.1})$$

Here $h_l^{(j)}$ 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 \mathbf{v}_ν are almost identical; for them the spherical Hankel functions have been replaced by spherical Bessel functions $j_l = (h_l^{(1)} + h_l^{(2)})/2$. The vector spherical

harmonics $\mathbf{A}_{\tau sml}$ are defined by

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

Here Y_{sml} are the (scalar) spherical harmonics

$$Y_{sml}(\theta, \phi) = -\sqrt{\frac{2 - \delta_{m0}}{2\pi}} \sqrt{\frac{2l+1}{2} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) \begin{Bmatrix} \cos m\phi \\ \sin m\phi \end{Bmatrix}$$

and P_l^m are associated Legendre polynomials [1]. The polar angle is denoted θ while ϕ is the azimuth angle. The upper (lower) expression is for $s = 1$ ($s = 2$), and the range of the indices are $l = 1, 2, \dots$, $m = 0, 1, \dots, l$, $\tau = 1, 2$, $s = 1$ 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

$$\begin{cases} \hat{\mathbf{r}} \cdot \mathbf{A}_{1sml}(\hat{\mathbf{r}}) = \hat{\mathbf{r}} \cdot \mathbf{A}_{2sml}(\hat{\mathbf{r}}) = 0 \\ \hat{\mathbf{r}} \times \mathbf{A}_{3sml}(\hat{\mathbf{r}}) = \mathbf{0}, \end{cases}$$

for which reason $\tau = 1$ (odd ν) identifies a TE mode (magnetic 2^l -pole) while $\tau = 2$ (even ν) identifies a TM mode (electric 2^l -pole) when the electric and magnetic fields are defined by (3.1) and (3.2), respectively. Furthermore,

$$\begin{cases} \mathbf{A}_{1sml}(\hat{\mathbf{r}}) = \mathbf{A}_{2sml}(\hat{\mathbf{r}}) \times \hat{\mathbf{r}} \\ \mathbf{A}_{2sml}(\hat{\mathbf{r}}) = \hat{\mathbf{r}} \times \mathbf{A}_{1sml}(\hat{\mathbf{r}}). \end{cases} \quad (\text{A.2})$$

The vector spherical harmonics are orthonormal on the unit sphere. More specifically, they satisfy

$$\int_{\Omega_{\hat{\mathbf{r}}}} \mathbf{A}_{\nu}(\hat{\mathbf{r}}) \cdot \mathbf{A}_{\nu'}(\hat{\mathbf{r}}) d\Omega_{\hat{\mathbf{r}}} = \delta_{\nu, \nu'}, \quad (\text{A.3})$$

where $\Omega_{\hat{\mathbf{r}}} = \{(\theta, \phi) : 0 \leq \theta \leq \pi, 0 \leq \phi \leq 2\pi\}$ is the unit sphere and $d\Omega_{\hat{\mathbf{r}}} = \sin \theta d\theta d\phi$. Define the L^2 -norm $\|\cdot\|$ for vector-valued functions on $\Omega_{\hat{\mathbf{r}}}$:

$$\|\mathbf{G}\|^2 = \int_{\Omega_{\hat{\mathbf{r}}}} \mathbf{G}(\hat{\mathbf{r}}) \cdot \mathbf{G}^*(\hat{\mathbf{r}}) d\Omega_{\hat{\mathbf{r}}}.$$

If the norm of \mathbf{G} is finite, it may be expanded in vector spherical harmonics:

$$\mathbf{G}(\hat{\mathbf{r}}) = \sum_{\nu} c_{\nu} \mathbf{A}_{\nu}(\hat{\mathbf{r}}), \quad (\text{A.4})$$

where the coefficients c_{ν} are given by

$$c_{\nu} = \int_{\Omega_{\hat{\mathbf{r}}}} \mathbf{G}(\hat{\mathbf{r}}) \cdot \mathbf{A}_{\nu}(\hat{\mathbf{r}}) d\Omega_{\hat{\mathbf{r}}},$$

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):

$$\hat{\mathbf{x}} = \sqrt{\frac{4\pi}{3}} \mathbf{A}_{3e11}(\hat{\mathbf{r}}) + \sqrt{\frac{8\pi}{3}} \mathbf{A}_{2e11}(\hat{\mathbf{r}}) \quad (\text{A.5})$$

$$\hat{\mathbf{y}} = \sqrt{\frac{4\pi}{3}} \mathbf{A}_{3o11}(\hat{\mathbf{r}}) + \sqrt{\frac{8\pi}{3}} \mathbf{A}_{2o11}(\hat{\mathbf{r}}) \quad (\text{A.6})$$

$$\hat{\mathbf{z}} = \sqrt{\frac{4\pi}{3}} \mathbf{A}_{3e01}(\hat{\mathbf{r}}) + \sqrt{\frac{8\pi}{3}} \mathbf{A}_{2e01}(\hat{\mathbf{r}}). \quad (\text{A.7})$$

There are expansions for the Hankel functions, used to determine the polynomials $R_{\tau,l}^{(j)}$ in (3.3):

$$h_l^{(1)}(z) = \frac{e^{iz}}{i^{l+1}z} \sum_{n=1}^l \frac{(l+n)!}{n!(l-n)!} (-2iz)^{-k} \quad (\text{A.8})$$

$$h_l^{(2)}(z) = \frac{i^{l+1}e^{-iz}}{z} \sum_{n=1}^l \frac{(l+n)!}{n!(l-n)!} (2iz)^{-k}. \quad (\text{A.9})$$

A.2 Derivation of (4.6)

The scattered field $\tilde{\mathbf{E}}_s$ is the sum over $\mathbf{u}_{\nu}^{(1)}$ in (4.3), *viz.*,

$$\tilde{\mathbf{E}}_s(k, \mathbf{r}) = \sqrt{\eta_0} \sum_{\nu} \tilde{d}_{\nu}^{(1)}(k) \mathbf{A}_{\nu}(\hat{\mathbf{r}}) \frac{e^{i\mathbf{k}\mathbf{r}}}{r} (1 + \mathcal{O}(r^{-1})), \quad \text{as } r \rightarrow \infty,$$

where (A.1) and (A.8) have been used. From the above equation it is clear that the far-field amplitude $\tilde{\mathbf{F}}(k, \hat{\mathbf{r}})$ in (4.4) is given by

$$\tilde{\mathbf{F}}(k, \hat{\mathbf{r}}) = \sqrt{\eta_0} \sum_{\nu} \tilde{d}_{\nu}^{(1)}(k) \mathbf{A}_{\nu}(\hat{\mathbf{r}}).$$

Using (4.5), multiplying with $\mathbf{A}_{\nu'}(\hat{\mathbf{r}})$ and integrating over the unit sphere yield

$$\int \mathbf{A}_{\nu'}(\hat{\mathbf{r}}) \cdot \tilde{\mathbf{S}}(k, \hat{\mathbf{r}}, \hat{\mathbf{k}}) \cdot \tilde{\mathbf{E}}_0(k) d\Omega_{\hat{\mathbf{r}}} = \sqrt{\eta_0} \tilde{d}_{\nu'}^{(1)}(k), \quad (\text{A.10})$$

due to (A.3).

The coefficients $\tilde{d}_{\nu'}^{(1)}(k)$ are given by

$$\sqrt{\eta_0} \tilde{d}_{\nu'}^{(1)}(k) = \sqrt{\eta_0} \sum_{\nu''} \tilde{T}_{\nu',\nu''}(k) \tilde{d}_{\nu''}^{(2)}(k) = \frac{4\pi}{ik} \sum_{\nu''} \tilde{T}_{\nu',\nu''}(k) \tilde{\mathbf{E}}_0(k) \cdot \mathbf{A}_{\nu''}(\hat{\mathbf{k}}),$$

where the expansion coefficients $\tilde{d}_{\nu''}^{(2)}(k)$ of a plane wave $e^{i\mathbf{r}\cdot\mathbf{k}} \tilde{\mathbf{E}}_0(k)$ have been used. Inserting this into (A.10) gives

$$\int \mathbf{A}_{\nu'}(\hat{\mathbf{r}}) \cdot \tilde{\mathbf{S}}(k, \hat{\mathbf{r}}, \hat{\mathbf{k}}) \cdot \tilde{\mathbf{E}}_0(k) d\Omega_{\hat{\mathbf{r}}} = \frac{4\pi}{ik} \sum_{\nu''} \tilde{T}_{\nu',\nu''}(k) \tilde{\mathbf{E}}_0(k) \cdot \mathbf{A}_{\nu''}(\hat{\mathbf{k}}),$$

which must be valid for all $\hat{\mathbf{k}}$ and $\tilde{\mathbf{E}}_0$. Letting $\tilde{\mathbf{E}}_0(k, \hat{\mathbf{k}}) = \mathbf{A}_{\nu'''}(\hat{\mathbf{k}})\varphi(k)$ for some $\varphi \in \mathcal{S}$ and integrating once more over the unit sphere leads to

$$\int \int \mathbf{A}_{\nu'}(\hat{\mathbf{r}}) \cdot \tilde{\mathbf{S}}(k, \hat{\mathbf{r}}, \hat{\mathbf{k}}) \cdot \mathbf{A}_{\nu'''}(\hat{\mathbf{k}})\varphi(k) d\Omega_{\hat{\mathbf{r}}} d\Omega_{\hat{\mathbf{k}}} = \frac{4\pi}{ik} \tilde{T}_{\nu', \nu'''}(k)\varphi(k),$$

and (4.6) is proven.

A.3 Derivation of (4.15)

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

$$\frac{B_K \ln S_0^{-1}}{\pi} \leq k_0 a - \theta_0. \quad (\text{A.11})$$

Furthermore, it follows that $\sum_n \text{Im } k_0^3/k_n^3 \leq \sum_n \theta''_n{}^3 \leq \theta_0^3$, since

$$\text{Im } \frac{k_0^3}{k_n^3} = \frac{k_0^3}{|k_n|^6} [(\text{Im } k_n)^3 - (\text{Re } k_n)^2 \text{Im } k_n] \leq \frac{k_0^3}{|k_n|^6} (\text{Im } k_n)^3 = \theta''_n{}^3.$$

Hence (4.14) becomes

$$\frac{B_K \ln S_0^{-1}}{\pi} \leq k_0^3 a^3 \rho_{\nu, \nu} + \frac{\theta_0^3}{3}. \quad (\text{A.12})$$

Combining (A.11) and (A.12) yields

$$\frac{B_K \ln S_0^{-1}}{\pi} \leq k_0^3 a^3 \rho_{\nu, \nu} + \frac{1}{3} \left(k_0 a - \frac{B_K \ln S_0^{-1}}{\pi} \right)^3,$$

with solution (4.15).

References

- [1] G. Arfken. *Mathematical Methods for Physicists*. Academic Press, Orlando, third edition, 1985.
- [2] A. Bernland, A. Luger, and M. Gustafsson. Sum rules and constraints on passive systems. Technical Report LUTEDX/(TEAT-7193)/1-28/(2010), Lund University, Department of Electrical and Information Technology, P.O. Box 118, S-221 00 Lund, Sweden, 2010. <http://www.eit.lth.se>.
- [3] C. F. Bohren and D. R. Huffman. *Absorption and Scattering of Light by Small Particles*. John Wiley & Sons, New York, 1983.
- [4] 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.

- [5] L. J. Chu. Physical limitations of omni-directional antennas. *Appl. Phys.*, **19**, 1163–1175, 1948.
- [6] 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.
- [7] M. Gustafsson, C. Sohl, and G. Kristensson. Physical limitations on antennas of arbitrary shape. *Proc. R. Soc. A*, **463**, 2589–2607, 2007.
- [8] M. Gustafsson and S. Nordebo. Bandwidth, Q factor, and resonance models of antennas. *Progress in Electromagnetics Research*, **62**, 1–20, 2006.
- [9] 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>.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] J. Mashreghi. *Representation Theorems in Hardy Spaces*. Cambridge University Press, Cambridge, U.K., 2009.
- [14] H. M. Nussenzveig. *Causality and dispersion relations*. Academic Press, London, 1972.
- [15] 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.
- [16] E. M. Purcell. On the absorption and emission of light by interstellar grains. *J. Astrophys.*, **158**, 433–440, 1969.
- [17] 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.

- [18] 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.
- [19] 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.
- [20] C. Sohl, M. Gustafsson, and G. Kristensson. Physical limitations on broadband scattering by heterogeneous obstacles. *J. Phys. A: Math. Theor.*, **40**, 11165–11182, 2007.
- [21] B. Ung and Y. Sheng. Interference of surface waves in a metallic nanoslit. *Optics Express*, **15**(3), 1182–1190, 2007.
- [22] A. H. Zemanian. *Distribution theory and transform analysis: an introduction to generalized functions, with applications*. Dover Publications, New York, 1987.
- [23] A. H. Zemanian. *Realizability Theory for Continuous Linear Systems*. Dover Publications, New York, 1995.



LUND
UNIVERSITY

Series of licentiate and doctoral theses
Department of Electrical and Information Technology
1654-790X
No. 24

<http://www.eit.lth.se>