

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Learning-based physics-enhanced modeling and
inline material characterization for
electromagnetics

SIMON STENMARK

Department of Electrical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2026

Learning-based physics-enhanced modeling and inline material characterization for electromagnetics

SIMON STENMARK

ISBN 978-91-8103-464-6

Acknowledgements, dedications, and similar personal statements in this thesis, reflect the author's own views.

© SIMON STENMARK 2026 except where otherwise stated.

Selected material from the author's licentiate thesis: S. Stenmark, *Deep neural networks for electromagnetics applied to optimization and inverse problems*, Chalmers University of Technology, Gothenburg, Sweden, Aug. 2023, is republished in this Ph.D. thesis.

Doktorsavhandlingar vid Chalmers tekniska högskola

Ny serie nr 5921

ISSN 0346-718X

DOI: 10.63959/chalmers.dt/5921

Department of Electrical Engineering

Chalmers University of Technology

SE-412 96 Gothenburg, Sweden

Phone: +46 (0)31 772 1000

The work in Paper A, C and E was supported in part by the Swedish Governmental Agency for Innovation Systems (VINNOVA) under Grant 2016-00460. The work in Paper B was supported in part by the FFI MORFex project, which is partly funded by the Swedish Energy Agency.

Cover: Magnitude of the reflection coefficient $|S_{11}|$ for an optimized band pass filter under geometrical perturbations. See pp. 42 – 43 for details.

Printed by Chalmers Digital Printing

Gothenburg, Sweden, September 2026

Learning-based physics-enhanced modeling and inline material characterization for electromagnetics

SIMON STENMARK

Department of Electrical Engineering
Chalmers University of Technology

Abstract

This thesis explores data-driven modeling and inline material characterization methods for electromagnetic problems. In particular, it explores possibilities for combining physics-based and learning-based techniques to create efficient hybrid methods. The field of computational electromagnetics (CEM) offers many numerical methods to solve Maxwell's equations. However, they tend to be computationally demanding, especially in three spatial dimensions. In the first part of the thesis, we propose and evaluate Gradient-Informed Attentive Normalization Training (GIANT), a neural network training procedure for constructing computationally efficient surrogate models for CEM methods. GIANT consists of two parts: (i) Attentive Normalization (AttNorm), a reparameterization procedure that enables efficient training of very deep fully-connected neural networks; and (ii) gradient-informed training, which is used to minimize the number of samples that are needed to train the neural network. For the reciprocal microwave problems that we consider, we use continuum sensitivity analysis to compute these gradients at a very low computational cost. We demonstrate GIANT by constructing accurate surrogate models for a cylindrical resonator that is filled with an inhomogeneous dielectric and for an H-plane waveguide filter. In the second part of the thesis, we develop two inline material characterization techniques. The first is a method for determining unknown parameters that describe the materials of an electric motor. The method is based on a computationally efficient physics-based model of the motor, to which we apply a Bayesian estimation framework to exploit prior knowledge about the sought parameters. The second is an auto-calibration method that, given uncalibrated observations of scattering parameters, simultaneously determines the averaged permittivity of an inhomogeneous dielectric in the measurement domain and calibrates the measurement system. We demonstrate the methods by applying them to both synthetic and measured data.

Keywords: Surrogate models, neural networks, deep learning, material characterization, statistical signal processing, Bayesian estimation.

List of Publications

This thesis is based on the following publications:

[A] **Simon Stenmark**, Thomas Rylander, Tomas McKelvey, Andrei Ludvig-Osipov, “GIANT Networks: Very Deep Fully-Connected Neural Networks Applied to Microwave Problems”. Published in IET Microwaves, Antennas & Propagation, Vol. 20, 2026.

[B] **Simon Stenmark**, Thomas Rylander, Jan Carlsson, Mats Viberg, Matthys M. Botha, “Parameter Estimation for Electric Machines with Hairpin Conductors Utilizing Hybrid Modeling”. Submitted for publication in IEEE Transactions on Energy Conversion.

[C] Andrei Ludvig-Osipov, **Simon Stenmark**, Thomas Rylander, Tomas McKelvey, “Auto-calibration of near-field microwave measurements for complex permittivity estimation”. Published in Progress In Electromagnetics Research M, Vol. 136, 2025.

[D] **Simon Stenmark**, Thomas Rylander, Matthys M. Botha, Jan Carlsson, “Common-Mode Impedance of an Electric Motor and the Impact of Material and Geometry Uncertainties”. 2024 International Symposium on Electromagnetic Compatibility – EMC Europe.

[E] **Simon Stenmark**, Thomas Rylander, Tomas McKelvey, “Neural Networks for the Estimation of Low-Order Statistical Moments of a Stochastic Dielectric”. 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC).

Contents

Abstract	i
List of Papers	iii
Acknowledgements	xi
Acronyms	xi
I Overview	1
1 Introduction	3
1.1 Computational electromagnetics	3
Learning-based surrogate models for Maxwell's equations . . .	6
In between physics-based and learning-based models	7
1.2 Aim of thesis	7
1.3 Canonical electromagnetic problem	8
Scattering parameters	9
Solution by finite element method	10
Sensitivity analysis for the scattering parameters	10
1.4 Abstract formulation of the canonical electromagnetic problem	12

2	Learning-Based Models for Electromagnetic Problems	13
2.1	Supervised learning workflow	14
2.2	Foundations of fully-connected neural networks	14
2.3	The GIANT neural network training framework	17
	Attentive Normalization	17
	Gradient-informed training of neural networks	19
2.4	Surrogate models based on Padé approximants	21
2.5	Towards global optimization	23
	Sampling strategies and zeroth-order methods	24
	Iterative methods for local optimization	26
3	Methods for Inline Material Characterization	29
3.1	Formulating the material characterization problem	30
	Estimators	31
	Cramér-Rao lower bounds	31
3.2	Bayesian estimation	32
	Choice of prior distribution	33
	Bayesian Cramér-Rao lower bounds	33
	Maximum a posteriori estimator	34
3.3	Choice of model parameterization	35
4	Applications and Results	37
4.1	Monitoring of inhomogeneous media in a metal pipe	37
	Estimation of statistical properties of dielectric medium	38
	High-dimensional histograms for detection problems	38
	Inline calibration and permittivity estimation	40
4.2	Robust geometry optimization for microwave waveguide filter	42
4.3	Electromagnetic compatibility for electric drivelines	43
	Perturbation analysis for model of hairpin electric motor	44
	Parameter estimation for model of hairpin electric motor	47
5	Author Contributions to Included Papers	49
5.1	Paper A	49
5.2	Paper B	50
5.3	Paper C	50
5.4	Paper D	51
5.5	Paper E	51

6 Summary and Concluding Remarks	53
References	57

II Papers 67

A GIANT Networks	A1
1 Introduction	A3
2 Method	A6
2.1 AttNorm Network Setup	A6
2.2 Training, Validation and Test Sets	A8
2.3 AttNorm: Updating Scheme	A9
2.4 Low-Variance Preserving Scaling Function	A12
2.5 Sobolev Training	A14
2.6 The GIANT Network	A15
3 Modelling of Passive Microwave Circuits	A15
3.1 Boundary Value Problem	A16
3.2 Sensitivity Analysis	A16
3.3 Normalization	A17
4 Results	A17
4.1 Computer Implementation	A18
4.2 Test 1 – Stochastic Inhomogeneous Medium in Circular Cavity	A18
4.3 Test 2 – Optimized H-Plane Filter Under Uncertainty .	A32
5 Conclusion	A38
References	A41
B Parameter Estimation for Electric Machines	B1
1 Introduction	B3
2 Method	B7
2.1 Model of Electric Motor	B7
2.2 Measurement Fixture	B11
2.3 Parameter Estimation	B13
3 Results	B16
3.1 Determination of Prior Distribution	B17
3.2 Numerical Studies	B20

3.3	Experimental Measurement Data	B24
4	Conclusion	B28
	References	B29
C	Auto-Calibration of Near-Field Microwave Measurements	C1
1	Introduction	C3
2	Measurement system model and auto-calibration problem . . .	C5
2.1	Scattering-parameter model	C6
2.2	Noise model	C7
2.3	Measurement-domain parametrization	C8
2.4	Problem formulation	C8
3	Method	C10
3.1	Averaging operators	C11
3.2	Solving the auto-calibration problem	C11
3.3	Local approximation functions	C14
4	Numerical examples	C15
4.1	Geometry	C15
4.2	Signal-to-noise ratio	C17
4.3	Amplification matrices	C18
4.4	Error assessment	C18
4.5	Numerical tests	C18
5	Discussion	C20
5.1	Computational complexity	C26
5.2	Comparison with other methods	C27
5.3	Extensions and future work	C27
6	Conclusion	C28
	References	C29
D	Common-Mode Impedance of an Electric Motor	D1
1	Introduction	D3
2	Motor geometry	D4
3	Method	D6
3.1	Differential equations	D6
3.2	Transmission line equations	D9
4	Results	D11
4.1	Test #1 – simplified cross-section of electric motor . . .	D11
4.2	Test #2 – realistic cross-section of an electric motor . .	D13

4.3	Uncertainty in the permeability	D13
4.4	Uncertainty in the permittivity	D16
4.5	Uncertainty in the geometry	D17
5	Conclusion	D17
	References	D18
E	Neural Networks for a Stochastic Dielectric	E1
1	Introduction	E3
2	Method	E4
2.1	Field problem	E5
2.2	Collection and organization of data	E6
2.3	Feature extraction	E8
2.4	Regression algorithms	E8
3	Results	E11
3.1	Training, validation and test data set	E11
3.2	Hyperparameter tuning for the neural network	E11
3.3	Performance of the prediction and their comparison	E13
4	Conclusion	E17
	References	E19

Acknowledgments

I would like to thank my supervisors, Thomas, Tomas, and Jan, for their support and advice. In particular, I would like to thank my main supervisor, Thomas, for his endless patience. I realize that your time is extremely valuable, yet you still go above and beyond to support me in my bumpy journey. For this, I am ever grateful.

I want to thank my various collaborators during the years. In particular, I would like to thank Mats for the insightful discussions and advice. I have learned a lot from you.

I also want to thank my friends and family. I want to thank my beautiful wife, Moa, for your helpful advice and for always believing in me. I want to thank my lovely children, Elisabeth and Hilbert, for being bundles of joy and always brightening my day. I want to thank my in-laws, Kerstin and Börje, for always being there for us. Finally, I want to thank my parents, Mia and Bo, for your endless support and love. You are the best parents one could ever wish for. Without all of you, this journey wouldn't have been possible.

Acronyms

AC:	Alternating Current
AttNorm:	Attentive Normalization
BCRLB:	Bayesian Cramér-Rao Lower Bound
Bmse:	Bayesian mean squared error
CEM:	Computational Electromagnetics
CNN:	Convolutional Neural Network
CRLB:	Cramér-Rao Lower Bound
DC:	Direct Current
EMC:	Electromagnetic Compatibility

EMI:	Electromagnetic Interference
FCNN:	Fully-Connected Neural Network
FDTD:	Finite-Difference Time-Domain
FEM:	Finite Element Method
FMM:	Fast Multipole Method
GIANT:	Gradient-Informed Attentive Normalization Training
LVP:	Low-Variance Preserving
MAP:	Maximum A Posteriori
MMSE:	Minimum Mean Square Error
MoM:	Method Of Moments
PDE:	Partial Differential Equation
PDF:	Probability Density Function
PEC:	Perfect Electrical Conductor
PWM:	Pulse-Width Modulation
ReLU:	Rectified Linear Unit
RNN:	Recursive Neural Network
SELU:	Scaled Exponential Linear Unit
SNR:	Signal-To-Noise Ratio
SUT:	Sample Under Test
TL:	Transmission Line

Part I

Overview

CHAPTER 1

Introduction

The advancement of computer hardware makes it possible to solve increasingly complicated tasks in engineering and science, where applications such as design optimization and physically accurate simulations often require significant computational effort. The well-known Moore's law predicts that the transistor density on integrated circuits doubles roughly every two years, yielding exponential growth in computing power. However, it becomes increasingly challenging to continue this growth of computing power as transistors start to reach atomic scales [1], [2], [3]. Simultaneously, new and more sophisticated algorithms are at least as important for handling challenging computational problems. In many scientific domains, the increase in algorithmic performance has significantly outpaced the growth in computing power [4]. In the post-Moore era, these advancements in algorithmic efficiency become more important than ever to enable continued technological progress.

1.1 Computational electromagnetics

Maxwell's equations form the foundation of electromagnetic theory [5], and the ability to solve them efficiently is of great importance for the development

of novel solutions to electromagnetic problems. While analytical solutions of Maxwell's equations often provide valuable insight into the behavior of a problem, they can typically only be derived for special cases. For most practical problems, we have to resort to one of the many numerical methods that form the field of computational electromagnetics (CEM) [6].

The finite element method (FEM) [7], [8], [9] is a popular method for solving Partial Differential Equations (PDEs). Here, we consider the frequency-domain FEM with a single excitation frequency, which is a common formulation for electromagnetic problems. In the FEM, the fields are expanded in basis functions that are defined on finite elements, where curl-conforming basis functions [10] are used to avoid the occurrence of spurious modes. Next, a weighted average of the residual of the PDE is set to zero. This results in a system of linear equations $\mathbf{Ax} = \mathbf{b}$, where the sources of the problem are associated with the right-hand side vector $\mathbf{b}$. Here, the system matrix $\mathbf{A}$ is typically sparse and symmetric, which is beneficial in terms of the memory required to store the matrix. The bandwidth of the matrix is determined by the order of the basis functions that are used for the field expansion, where higher-order basis functions can allow for larger elements and thus fewer unknowns at the cost of a higher matrix bandwidth [11]. For sufficiently small problems, we can factorize the matrix $\mathbf{A}$ and reuse the factorization as a different waveguide port is excited or the angle of incidence changes in a scattering problem. Thus, we can change the sources of the problem at the relatively low computational cost associated with a forward and backward substitution given a new vector $\mathbf{b}$ [12]. However, we need to perform a completely new simulation for any changes to the frequency, materials or geometry of the problem.

A second popular CEM method is the method of moments (MoM) [13], [14]. Based on Maxwell's equations and appropriate information on boundary conditions, a so-called Green's function is used to formulate an integral equation [15], where the sources are expanded in divergence-conforming basis functions [11], [16]. The weighted average of the residual of the integral equation is set to zero, which results in a system of linear equations. The MoM is well suited for situations where the sources are described by relatively few degrees of freedom compared to the degrees of freedom that would be required to describe the corresponding field, such as scattering problems and antenna problems in free space. The MoM is also useful for truncating the computa-

tional domain in hybrid methods such as the finite element - boundary integral method [17], [18], [19]. The frequency-domain MoM shares many properties with the frequency-domain FEM with respect to the dependencies for the system of linear equations $\mathbf{Ax} = \mathbf{b}$. However, the system matrix $\mathbf{A}$ is dense, which is very demanding in terms of memory requirements. Similarly as for the FEM, if we factorize the matrix $\mathbf{A}$, we can reuse the factorization of the matrix for any changes to the vector $\mathbf{b}$, which corresponds to the sources of the problem. However, we need to perform a completely new simulation for any changes to the frequency, materials or geometry of the problem. As an important extension to the MoM, the fast multipole method (FMM) [20], [21] allows grouping of sources in close vicinity to provide a sparse factorization of the matrix $\mathbf{A}$, which can severely decrease memory requirements and the computations needed to solve the linear system of equations. However, the FMM is only appropriate for use with iterative solvers, meaning that we must also perform a completely new simulation for any changes in the vector $\mathbf{b}$.

The finite-difference time-domain (FDTD) [22], [23] scheme is one of the most common methods for microwave problems. In the FDTD scheme, we represent the electric field and magnetic field on uniform grids that are staggered with respect to space. We then use centered finite-differences [12] to approximate the derivatives in Ampère's law and Faraday's law and perform explicit time-stepping by means of the leap-frog time-stepping scheme [9], where the size of the time-step for stable time-stepping is limited by the Courant condition. Common add-ons to the FDTD scheme include the perfectly matched layer [24] for absorbing outgoing waves, near-to-far-field Transformation [22] for computing electromagnetic fields in the far-field region from near-field data and Huygens' surfaces [22] to provide a plane wave excitation. An advantage of the FDTD scheme is that the computer memory can be used almost exclusively for the storage of the electromagnetic field, as no system matrix $\mathbf{A}$ needs to be stored. Another advantage is that we can compute the response for many frequencies in one simulation if we excite the problem with a pulse of large frequency-bandwidth. However, we need to perform a completely new simulation for any change in the materials, geometry, or excitation of the problem. Another major drawback of the FDTD scheme is its difficulty in resolving oblique and curved boundaries due to the discretization of the computational domain by means of brick-shaped cells, which leads to the so-called staircase approximation problem that can severely impact the accuracy

of the results [25]. Approaches to mitigate the staircase approximation issue include conformal FDTD-methods [26] and hybridization with the finite element method [27].

Learning-based surrogate models for Maxwell's equations

In many applications, we wish to know the behavior of the quantities of interest as a function of material and/or geometry parameters for a specific (and limited) region in the parameter space that describes the material and/or geometry of the problem. However, for all the considered CEM methods, it is computationally costly to change the materials or geometry of the problem and we can effectively only sample discrete points in this parameter space. An approach to overcome this limitation is to use a so-called surrogate model [28], a data-driven model that is tuned to fit data that is computed by a conventional CEM method. The surrogate model can then be used to approximate the conventional CEM method for intermediate parameter values. The trained surrogate models can be used by themselves, or as a complement to a detailed model in a space-mapping scheme [29], [30]. Common methods used to construct surrogate models include polynomial response surfaces, radial basis function models, support vector regression, and Kriging [31], [32]. In recent years, deep neural networks have been of particular interest as accurate surrogate models for electromagnetic applications [33] due to their ability to express complicated non-linear relationships between input and output.

Deep neural networks [34] can be structured in a variety of ways depending on the application at hand. Architectures such as convolutional neural networks and auto-encoders [34] display impressive results for tasks such as computer vision and anomaly detection. Transformer-based models [35] achieve state-of-the-art results in tasks such as image classification [36] and machine translation [37], and form the foundation of large language models such as ChatGPT [38] and Claude [39]. Most specialized architectures assume that the data is organized according to a certain structure. For example, convolutional neural networks efficiently treat grid-like data such as images, while recurrent neural networks efficiently treat sequential data such as text or time-domain signals. Data derived from electromagnetic problems often feature more complicated dependencies, which makes more agnostic neural network architectures, such as fully-connected neural networks (FCNNs), attractive.

In between physics-based and learning-based models

The many physics-based methods that comprise the field of CEM allow for the treatment of a wide range of electromagnetic problems. Physics-based models provide an understanding of the underlying problem and allow for the analysis of novel situations for which no observations exist. However, apart from being computationally demanding to evaluate, they typically require simplifications of the problem that can affect the accuracy of the results. Unlike physics-based models that are based on first-order principles, learning-based models allow for learning the behavior of the problem from data. Today's powerful hardware and recent developments in data-driven algorithms open up exciting new possibilities to model problems that are intractable to model using physics-based methods. However, learning-based models typically require large amounts of data to train, which can be challenging to collect and annotate. They also provide a limited understanding of the problem, and a naive implementation may yield non-physical results.

In between purely data-driven and physics-based models, there are hybrid models that include elements of both physics-based and learning-based models. Examples include physics-informed neural networks [40], knowledge-based neural networks [41], [42], and neural networks that predict the poles and residues of pole-residue-based transfer functions [43]. The additional structure granted by the information about the physics can improve the interpolation capabilities of the model and potentially allow extrapolation beyond the range of the training data.

1.2 Aim of thesis

In this thesis, we explore possibilities to complement or replace full-wave solvers with data-driven models that are computationally inexpensive to evaluate. In particular, we exploit neural networks to model the complicated non-linear relationships present in microwave problems. We train the data-driven models to emulate the full-wave solvers in a limited domain of the parameter space, which provides a powerful complement to full-wave solvers. We also explore methods to use physics-based and hybrid models to estimate unknown parameters that caused a set of observations, and use the methods to implement material characterization techniques for inline applications.

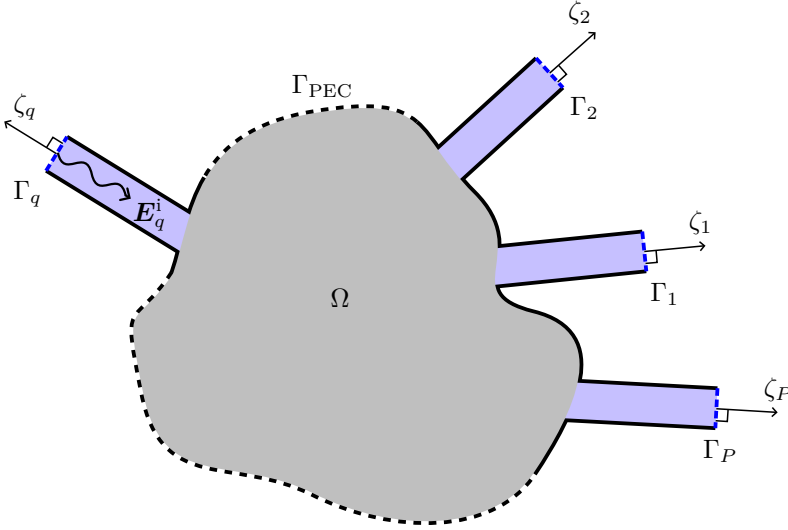


Figure 1.1: The boundary value problem under consideration. The enclosed volume Ω is surrounded by the PEC surface Γ_{PEC} . The volume is connected to P waveguides that are truncated by waveguide ports that coincide with the surfaces $\Gamma_1, \Gamma_2, \dots, \Gamma_P$. Waveguide port q is excited by the incident wave $\mathbf{E}_q^i$. The local longitudinal coordinates $\zeta_1, \zeta_2, \dots, \zeta_P$ are directed normally outwards from $\Gamma_1, \Gamma_2, \dots, \Gamma_P$.

1.3 Canonical electromagnetic problem

We consider the class of problems that can be represented by the general N -port microwave network shown in Fig. 1.1. Here, we have the perfect electric conductor (PEC) surfaces Γ_{PEC} . The ports $\Gamma_1, \Gamma_2, \dots, \Gamma_P$ are connected to waveguides, which have identical, constant cross-sections and are filled with a homogeneous dielectric. We assume that the frequency ω is within the frequency range where only the dominant mode of the waveguides is propagating, and that the length of the waveguide is sufficiently large for all evanescent higher-order modes to have a negligible amplitude at the waveguide port.

Given this setting, we consider the boundary value problem

$$\nabla \times \left(\frac{1}{\mu} \nabla \times \mathbf{E} \right) - (\omega^2 \epsilon - j\omega\sigma) \mathbf{E} = \mathbf{0} \quad \text{in } \Omega, \quad (1.1)$$

$$\hat{\mathbf{n}} \times \mathbf{E} = \mathbf{0} \quad \text{on } \Gamma_{\text{PEC}}, \quad (1.2)$$

$$\hat{\mathbf{n}} \times \nabla \times \mathbf{E} + \gamma \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathbf{E} = 2\gamma \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathbf{E}_q^i \quad \text{on } \Gamma_q, \quad (1.3)$$

$$\hat{\mathbf{n}} \times \nabla \times \mathbf{E} + \gamma \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathbf{E} = \mathbf{0} \quad \text{on } \Gamma_p \quad \forall p \neq q. \quad (1.4)$$

Here, the electric field $\mathbf{E}$ is governed by the three-dimensional vector Helmholtz' equation (1.1). The materials in the computational volume Ω are modeled by the permeability μ , permittivity ϵ and conductivity σ , all of which are allowed to be inhomogeneous and frequency dependent. The unit normal $\hat{\mathbf{n}}$ is directed outwards from the volume Ω . We use a Dirichlet boundary condition (1.2) on the PEC surfaces Γ_{PEC} . On the surfaces $\Gamma_1, \Gamma_2, \dots, \Gamma_P$ that coincide with the waveguide ports, we use Robin boundary conditions (1.3) and (1.4). Here, γ is the complex propagation constant of the dominant waveguide mode $\mathbf{m}$ [44], which is used to excite port q by the incident electric field

$$\mathbf{E}_q^i = E_{0,q}^i e^{+\gamma \zeta_q} \mathbf{m}. \quad (1.5)$$

The coordinate ζ_q is the local longitudinal coordinate associated with port q , as shown in Figure 1.1. This general formulation encapsulates a wide range of wave-propagation problems within an enclosed computational domain.

Scattering parameters

We define the *scattering matrix* as the matrix $\mathbf{S}$ that contains the *scattering parameters*

$$S_{pq} = \frac{E_{0,p}^r}{E_{0,q}^i} \quad (1.6)$$

where $\mathbf{E}_p^r = E_{0,p}^r e^{-\gamma \zeta_p} \mathbf{m}$ is the reflected electric field at port S_p . To compute all scattering parameters, we need to excite each port once, meaning that we need to solve P boundary value problems.

Solution by finite element method

Here, we use the FEM to solve the boundary value problem (1.1)-(1.4). The weak formulation is stated as: find the electric field $\mathbf{E} \in H_0(\text{curl}; \Omega)$ such that

$$a(\mathbf{v}, \mathbf{E}) = b(\mathbf{v}), \quad \forall \mathbf{v} \in H_0(\text{curl}; \Omega) \quad (1.7)$$

where

$$\mathbf{H}(\text{curl}, \Omega) = \left\{ \mathbf{E} : \left(\int_{\Omega} |\mathbf{E}|^2 d\Omega < \infty \right) \text{ and } \left(\int_{\Omega} |\nabla \times \mathbf{E}|^2 d\Omega < \infty \right) \right\} \quad (1.8)$$

$$\mathbf{H}_0(\text{curl}, \Omega) = \{ \mathbf{E} \in \mathbf{H}(\text{curl}, \Omega) : \hat{\mathbf{n}} \times \mathbf{E} = \mathbf{0} \text{ on } \Gamma_{\text{PEC}} \} \quad (1.9)$$

For the boundary value problem (1.1)-(1.4), we have the bilinear form a and linear form b as given by

$$\begin{aligned} a(\mathbf{v}, \mathbf{E}) = & \int_{\Omega} \left[\frac{1}{\mu} (\nabla \times \mathbf{v}) \cdot (\nabla \times \mathbf{E}) - (\omega^2 \epsilon - j\omega\sigma) \mathbf{v} \cdot \mathbf{E} \right] d\Omega \\ & + \gamma_m \sum_{p=1}^P \int_{\Gamma_p} (\hat{\mathbf{n}} \times \mathbf{v}) \cdot (\hat{\mathbf{n}} \times \mathbf{E}) d\Gamma \end{aligned} \quad (1.10)$$

$$b(\mathbf{v}) = 2\gamma_m \int_{\Gamma_q} (\hat{\mathbf{n}} \times \mathbf{v}) \cdot (\hat{\mathbf{n}} \times \mathbf{E}_q^i) d\Gamma. \quad (1.11)$$

Next, we expand $\mathbf{E}$ in the curl-conforming basis functions $\mathbf{N}_j \in \mathbf{H}_0(\text{curl}, \Omega)$ [7] according to

$$\mathbf{E}(\mathbf{r}) = \sum_j E_j \mathbf{N}_j(\mathbf{r}). \quad (1.12)$$

We use Galerkin's method and choose the weighting functions from the set of basis functions $\mathbf{N}_j$. Finally, Eq. (1.7) yields a system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ formulated in terms of the unknown vector $\mathbf{x}$ with the sought coefficients E_j .

Sensitivity analysis for the scattering parameters

Given a parameter vector $\boldsymbol{\xi}$ that contains parameters describing the materials and/or geometry of the BVP (1.1) - (1.4), we can expand the scattering

parameter (1.6) around a linearization point ξ_{lin} as

$$S_{pq}(\xi_{\text{lin}} + \delta\xi) = S_{pq}(\xi_{\text{lin}}) + \delta S_{pq}(\xi_{\text{lin}}, \delta\xi) + \dots \quad (1.13)$$

Here, we have the *sensitivity* $\delta S_{pq}(\xi_{\text{lin}}, \delta\xi)$ of the scattering parameter S_{pq} with respect to the linearization point ξ_{lin} and perturbation $\delta\xi$, which is an arbitrary infinitesimal perturbation of the parameter vector ξ . If we neglect the higher order terms, we obtain the linear model

$$S_{pq}(\xi_{\text{lin}} + \delta\xi) \approx S_{pq}(\xi_{\text{lin}}) + \delta S_{pq}(\xi_{\text{lin}}, \delta\xi). \quad (1.14)$$

If we have access to the sensitivity $\delta S_{pq}(\xi_{\text{lin}}, \delta\xi)$, we can use (1.14) as a locally applicable model of S_{pq} .

We can apply continuum sensitivity analysis to compute the sensitivity $\delta S_{pq}(\xi_{\text{lin}}, \delta\xi)$ from the field solutions of the *original* problem and of an *adjoint* problem. To exemplify, we consider the situation where the waveguides of the BVP (1.1) - (1.4) consist of coaxial cables, which is an example typical of situations where scattering parameters are measured using a vector network analyzer. Then, the sensitivity with respect to a perturbation $\delta\xi$ is given by [45]

$$\delta S_{pq}(\xi_{\text{lin}}, \delta\xi) = -\frac{1}{\gamma_m \kappa E_{0,p}^{\text{adj}} E_{0,q}^{\text{ori}}} \delta a(\mathbf{E}^{\text{adj}}, \mathbf{E}^{\text{ori}}; \delta\xi) \quad (1.15)$$

where κ is a constant that is computed from the geometry of the waveguide port, and δa refers to the perturbation of the bilinear form (1.10) that is associated with the field solutions $\mathbf{E}^{\text{adj}} = \mathbf{E}^{\text{adj}}(\xi_{\text{lin}})$ and $\mathbf{E}^{\text{ori}} = \mathbf{E}^{\text{ori}}(\xi_{\text{lin}})$ of the adjoint and of the original problem, as well as the perturbation $\delta\xi$. For example, the perturbation δa that is associated with a perturbation $\delta\epsilon = \delta\epsilon(\delta\xi)$ of the permittivity ϵ is given by

$$\delta a(\mathbf{E}^{\text{adj}}, \mathbf{E}^{\text{ori}}; \delta\xi) = -\omega^2 \int_{\Omega} \delta\epsilon \mathbf{E}^{\text{adj}} \cdot \mathbf{E}^{\text{ori}} d\Omega. \quad (1.16)$$

For the reciprocal microwave problems that we consider, the field solutions to all adjoint problems are available once the full scattering matrix is computed. Then, we can compute all the corresponding sensitivities without the need to solve any additional field problems. In comparison to the computational cost associated with the solution of the field problem, the additional

computational cost associated with the evaluation of the sensitivity is negligible since it only involves a weighted inner product between two (already available) field solutions.

1.4 Abstract formulation of the canonical electromagnetic problem

We introduce the input vector

$$\mathbf{x} = [\omega, \boldsymbol{\xi}] \quad (1.17)$$

of length Q where we have the parameter vector $\boldsymbol{\xi}$, that contains parameters describing the materials and/or geometry of the BVP (1.1) - (1.4), and the excitation frequency ω . We also introduce an output vector $\mathbf{y}$ of length P that contains quantities of interest for the specific problem, such as the scattering parameters (1.6).

The input vector $\mathbf{x}$ and output vector $\mathbf{y}$ are related through a model $\mathbf{f}$ as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{P}. \quad (1.18)$$

Here, the model $\mathbf{f}$ consists of the CEM method that is chosen to solve the boundary value problem (1.1) - (1.4), as well as any additional post-processing steps that are needed to compute the quantities of interest. We refer to the set $\mathcal{P}$ as the *domain* of the model, and the set of attainable $\mathbf{y}$ for $\mathbf{x} \in \mathcal{P}$ as the *range* of the model. While this thesis only considers electromagnetic problems that can be modelled by the boundary value problem (1.1) - (1.4), this formulation encapsulates any type of frequency-domain problem where one or more quantities $\mathbf{y}$ are deterministically dependent on the excitation frequency ω and the parameters $\boldsymbol{\xi}$ that define the problem.

CHAPTER 2

Learning-Based Models for Electromagnetic Problems

Most conventional CEM methods are computationally costly, especially in three space-dimensions. This can be a severely limiting factor in applications where the model $\mathbf{f}$ must be evaluated a large number of times. For computationally demanding applications, an attractive alternative is to complement or replace the CEM method with a data-driven model that is trained to emulate the CEM method, a so-called *surrogate model* [28]. We define the surrogate model $\hat{\mathbf{f}}$ as

$$\hat{\mathbf{y}} = \hat{\mathbf{f}}(\mathbf{x}; \boldsymbol{\theta}) \quad (2.1)$$

where $\hat{\mathbf{y}}$ is an estimate of $\mathbf{y}$ and $\boldsymbol{\theta}$ is a set of tunable parameters. When we train the model, we tune the parameters $\boldsymbol{\theta}$ of the model such that

$$\hat{\mathbf{f}}(\mathbf{x}; \boldsymbol{\theta}) \approx \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \hat{\mathcal{P}} \subseteq \mathcal{P} \quad (2.2)$$

where $\hat{\mathcal{P}}$ is a subset of the full parameter domain $\mathcal{P}$. If we choose a surrogate model $\hat{\mathbf{f}}$ that is inexpensive to evaluate, it can then be a very powerful and useful complement to the conventional CEM method.

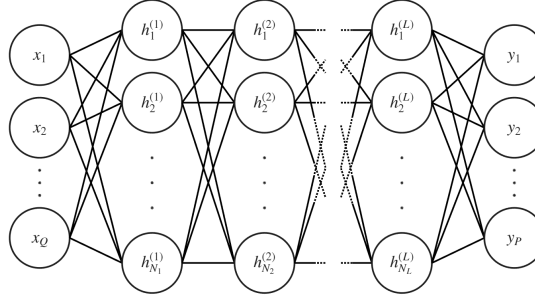


Figure 2.1: Schematic overview of a fully-connected neural network. The input $\mathbf{x}$ has Q elements and is followed by L fully-connected layers, where layer (l) has the width N_l . The output $\mathbf{y}$ has P elements and is given by the final hidden layer with index (L) .

2.1 Supervised learning workflow

In this work, we use supervised learning [46] to construct surrogate models. In a typical workflow to construct a surrogate model using supervised learning, we use a conventional CEM method to create a dataset that consists of input-output pairs, or *samples*. To reduce numerical issues in the training process, we preprocess the data to *normalize* the input and output features such that each feature has zero mean and unit variance [12]. The dataset is split into a *training set*, a *validation set*, and a *test set*. We then train a data-driven model on the training set to make the response of the model match the provided data. We use the validation set to evaluate the interpolation performance of the trained model and tune the hyper-parameters of the model until satisfactory performance is achieved. To obtain an unbiased estimate of the performance on unseen data, we finally evaluate the trained model on the unseen data in the test set.

2.2 Foundations of fully-connected neural networks

A neural network is a numerical model that consists of *neurons*, where each neuron is a computational unit that produces a numerical output from one or more numerical inputs. As illustrated in Fig. 2.1, in a fully-connected neural network (FCNN) the neurons are organized into *layers* and the output of each

neuron is connected to the input of every neuron in the subsequent layer. The number of layers L is referred to as the *depth* of the neural network, whereas the number of neurons in a layer is referred to as the *width* of the layer. Given a sufficient number of neurons, FCNNs can approximate any Borel measurable function from one finite dimensional space to another to an arbitrary degree of accuracy [47]. Contrary to more specialized neural network architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [34], the FCNN does not require the data to be organized according to any particular structure. This property makes FCNNs well suited for general regression tasks.

We can describe an L -layer FCNN as shown in Fig. 2.1 as the function composition

$$\hat{\mathbf{f}}(\mathbf{x}) = \left(\mathbf{h}^{(L)} \circ \mathbf{h}^{(L-1)} \circ \dots \circ \mathbf{h}^{(1)} \right) (\mathbf{x}) \quad (2.3)$$

where $\mathbf{h}^{(l)}$ denotes the fully connected layer l . Here and in the following, we use the super-index (l) in parentheses to indicate that a quantity is associated with fully connected layer l . The operation $\mathbf{f} = \mathbf{g} \circ \mathbf{h}$ is the function composition operation that takes two functions $\mathbf{g}$ and $\mathbf{h}$ and returns a function $\mathbf{f}$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{g}(\mathbf{h}(\mathbf{x}))$. The fully connected layer l is given by

$$\mathbf{h}^{(l)}(\mathbf{z}) = g^{(l)} \left(\mathbf{W}^{(l)} \mathbf{z} + \mathbf{b}^{(l)} \right) \quad (2.4)$$

where $g^{(l)}$ is an element-wise non-linear *activation function*, $\mathbf{W}^{(l)}$ a *weight matrix* and $\mathbf{b}^{(l)}$ a *bias vector*. When we train an FCNN, we tune the elements in the weight matrices and bias vectors to minimize a *loss function* and make the function (2.3) described by the FCNN match the input-output pairs in the training set. Today, the availability of powerful GPUs and other massively parallel computing architectures makes it possible to train very deep FCNNs with the capability to express complicated non-linear relationships between input and output.

The choice of activation function $g^{(l)}$ has important implications for the training and performance of the FCNN. The Rectified Linear Unit (ReLU) [34] is a common choice. However, the derivative of the ReLU is discontinuous and piecewise constant, which is prohibitive in applications where the second-order derivative of the activation function must be evaluated. Other common choices include the sigmoid and the hyperbolic tangent functions [34],

which are smooth but generally offers worse training performance than the ReLU [48]. The Swish activation function [49] was found by means of reinforcement learning and offers improved performance when compared to the ReLU on a number of tests. Klambauer et al. [50] proposed the Scaled Exponential Linear Unit (SELU) activation function, which has self-normalizing properties. In the context of surrogate models for electromagnetic problems, Jin et al. [51] proposed a smooth ReLU activation function which they exploit to successfully train fully-connected neural networks that are up to 14 layers deep.

Increasing the depth of the neural network increases its expressivity much faster than increasing its width [52], which is why much research effort has concerned the training of increasingly deep neural networks. However, there are several complications associated with training very deep FCNNs. For sufficiently deep neural networks, we face the problem of vanishing or exploding gradients [53], in particular when we use activation functions g for which the derivative g' can become much larger or smaller than unity [48]. An important consideration in the training of neural networks is the choice of weight optimizer, where optimizers that feature adaptive learning rates such as Adam [54] can help mitigate the effects of vanishing or exploding gradients [55]. Another technique that can mitigate the issue of vanishing or exploding gradients is to introduce normalization between the hidden layers of the neural network, where common examples of such techniques include BatchNorm [56] and LayerNorm [57]. However, it has been noted that FCNNs in particular are difficult to train using these techniques, which Klambauer et al. attribute to perturbations of the training process that are associated with updates of the normalization parameters [50]. For deep FCNNs with very many degrees of freedom, we also risk *overfitting* [46] the neural network to the training data, which means that the neural network memorizes the samples of the training set and does not generalize to unseen data. A clear indicator of overfitting is when the error on the training set is much smaller than the error on the validation set. To some extent, *regularization* can decrease the problem of overfitting. Commonly used regularization techniques include explicit techniques such as penalizing large weight matrices $\mathbf{W}^{(l)}$ in the loss function [34] and dropout [58], but also implicit techniques such as early stopping [59]. However, the typical solution to the problem of overfitting is to add more samples to the training set. These samples can be computationally expen-

sive to obtain, which means that the number of samples is often the primary limiting factor for how expressive FCNNs we can train.

2.3 The GIANT neural network training framework

In Paper A, we present *Gradient Informed Attentive Normalization Training (GIANT)*, a neural network training framework that enables efficient training of very deep FCNNs. GIANT features (i) a weight-bias reparametrization procedure that is referred to as Attentive Normalization (AttNorm) and is designed to reduce numerical issues that are associated with the training of very deep neural networks, and (ii) a loss function that incorporates the derivatives of the output $\mathbf{y}$ with respect to the input $\mathbf{x}$ in the training of the neural network to significantly decrease the number of samples necessary for training. The two methods together makes it possible to train very deep neural networks with comparatively few samples. We refer to an FCNN that was trained using the GIANT framework as a *GIANT network*.

Attentive Normalization

In Paper A, we present a novel normalization procedure that we refer to as *Attentive Normalization (AttNorm)*. As illustrated in Fig. 2.2, AttNorm consists of normalization layers that are inserted between the hidden layers of the FCNN and transform each element of the layer output as

$$\tilde{h}_j^{(l)} = s_j^{(l)} \left(h_j^{(l)} - \mu_j^{(l)} \right). \quad (2.5)$$

Here, the centering term $\mu_j^{(l)}$ and the scale factor $s_j^{(l)}$ are normalization parameters and $\tilde{h}_j^{(l)}$ is element j of the normalized output of layer l . The statistics of $h_j^{(l)}$ shift during training of the neural network which means that $\mu_j^{(l)}$ and $s_j^{(l)}$ need to be repeatedly updated during training for the normalization to be effective.

To choose the scale factors, we introduce a novel low-variance preserving (LVP) scaling function

$$s(\sigma) = \frac{\sigma + \epsilon}{\sigma^2 + \epsilon}. \quad (2.6)$$

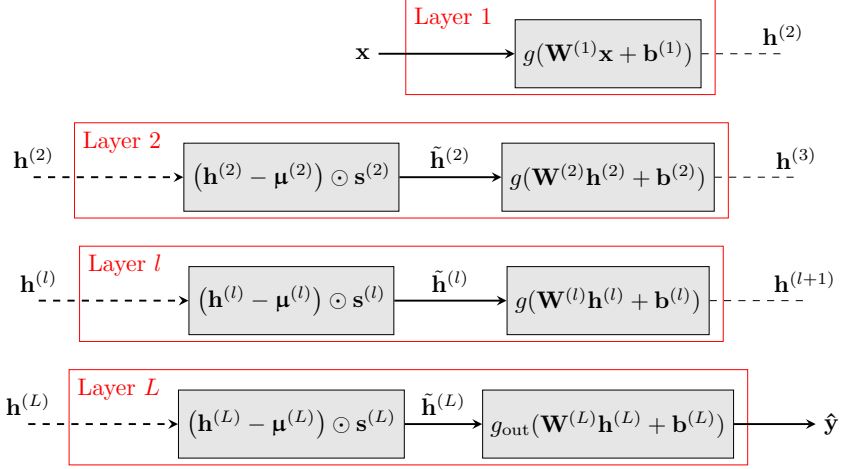


Figure 2.2: Architecture of an AttNorm-equipped neural network. We have the normalization layers which normalize the input to each layer, where the normalization layer l is characterized by the sample mean values $\boldsymbol{\mu}^{(l)}$ and the scale factors $\mathbf{s}^{(l)}$. The operator " $\odot$ " denotes an element-wise product. The remaining layers are the fully-connected layers. Note that the output fully-connected layer $l = L$ uses the activation function g_{out} which is chosen for the application at hand.

to choose the scale factors $s_j^{(l)}$ based on the sample standard deviations $\sigma_j^{(l)}$. The hyper-parameter ϵ , which controls the scaling for small standard deviations σ , needs to be tuned for the application at hand. The scaling function has its maximum at $\sigma = \sqrt{\epsilon(1 + \epsilon)} - \epsilon$ and reduces to unit scaling $s(\sigma) \simeq 1$ for $\sigma \ll \epsilon$. Fig. 2.3 shows a comparison between the LVP scaling function and a conventional scaling function $s_{\text{conv}}(\sigma) = 1/\sqrt{\sigma^2 + \epsilon}$. The LVP scaling function is designed to preserve the variance of nearly constant signals with a very small standard deviation and in Paper A we demonstrate that it improves the stability of the neural network training process compared to a conventional scaling function.

After updating the normalization parameters, the neural network as shown in Fig. 2.2 will represent a different function $\hat{\mathbf{f}}$. To lessen the perturbation of the training process that is associated with the update of the normalization parameters, we also update the weight matrices and bias vectors of the neural network such that the function $\hat{\mathbf{f}}$ remains unchanged by the update. AttNorm

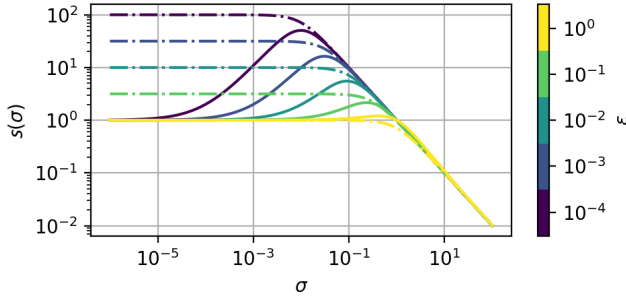


Figure 2.3: Low-variance preserving scaling function compared to a conventional scaling function for different values of ϵ . Solid lines: low-variance preserving scaling function; and dash-dotted lines: conventional scaling function.

is designed to be used with the Adam optimizer [54], and as part of the procedure, we also update the internal variables of the optimizer to remain consistent with the reparameterized weight-bias space. The procedure allows us to efficiently train very deep FCNNs, and in Paper A we demonstrate successful training of FCNNs that are 30 layers deep.

Gradient-informed training of neural networks

In the standard approach, the neural network is trained only to match the function value $\mathbf{y}$ for a given input $\mathbf{x}$. Here, we extend the output data to also contain the derivatives of the output with respect to the input, which is a special case of so-called Sobolev training [60]. In this approach, we include the Jacobian matrix $\mathbf{J}$ with the elements

$$J_{pq} = \frac{\partial y_p}{\partial x_q} \quad (2.7)$$

for the samples in our data sets. Then, we include the misfit with respect to $\mathbf{J}$ in the loss function that we use when we train the neural network. This can massively reduce the number of samples needed to successfully train the network. In situations where the derivatives are computationally inexpensive to obtain, this can allow training much more expressive neural networks for a

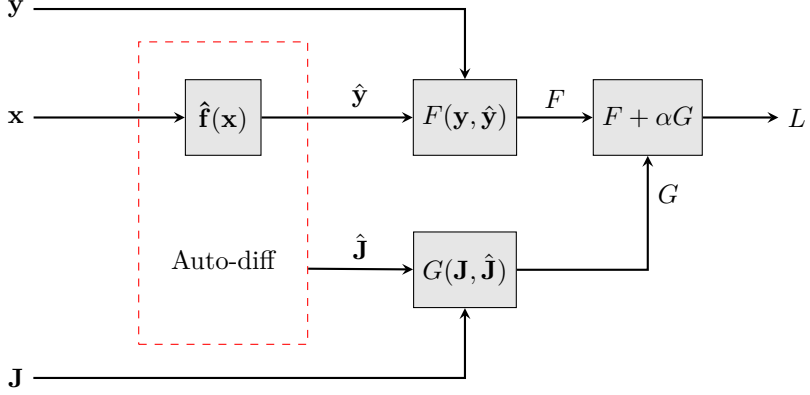


Figure 2.4: Flowchart illustrating how the loss L is computed from the input $\mathbf{x}$, the output $\mathbf{y}$ and the Jacobian matrix $\mathbf{J}$. Here, the neural network is denoted by $\hat{\mathbf{f}}(\mathbf{x})$. The Jacobian $\hat{\mathbf{y}}$ is computed using automatic differentiation.

given computational budget.

In Paper A, we exploit automatic differentiation [61] to differentiate the FCNN (2.3) with respect to the input $\mathbf{x}$ as shown in Fig. 2.4. As we evaluate the neural network, we then obtain both the function values $\hat{y}_p$ and the Jacobian matrix

$$\hat{J}_{pq} = \frac{\partial \hat{y}_p}{\partial x_q} \quad (2.8)$$

simultaneously. To include the derivatives in the training of the neural network, we formulate the total loss function

$$L = F(\mathbf{y}, \hat{\mathbf{y}}) + \alpha G(\mathbf{J}, \hat{\mathbf{J}}) \quad (2.9)$$

which is a linear combination of two loss terms: $F(\mathbf{y}, \hat{\mathbf{y}})$ - the misfit of the function values; and $G(\mathbf{J}, \hat{\mathbf{J}})$ - the misfit of their derivatives. The weight α is considered a hyper-parameter that needs to be tuned, where $\alpha = 0$ yields the conventional loss function based only on the misfit of the function values. Here, we note that differentiating the loss function 2.9 with respect to the weights or bias terms of the neural network includes computing the second

derivatives of the activation functions $g^{(l)}$. Thus, the activation functions that are used for gradient-informed neural networks must be twice differentiable, which excludes piece-wise linear activation functions such as the ReLU. For the test-problems that we consider, the number of samples that are required to train the neural network is significantly reduced when we include the derivatives in the loss function. As discussed in Section 1.3, continuum sensitivity analysis can be used to compute these derivatives at a very low computational cost for the reciprocal electromagnetic problems that we consider.

2.4 Surrogate models based on Padé approximants

Another important class of surrogate models for scattering parameters is the *Padé approximant*. Given the function $f(x)$, a Padé approximant $R_{M/N}(x)$ of order $[M/N]$ is a rational function that approximates $f(x)$ as

$$f(x) \approx R_{M/N}(x) = \frac{\sum_{m=0}^M b_m x^m}{1 + \sum_{n=1}^N a_n x^n}, \quad (2.10)$$

where $M \geq 0$ and $N \geq 1$. The tunable parameters θ consist of the coefficients b_m and a_n , which are learned from data. Padé approximants are useful in the study of linear time-invariant systems, as their transfer functions can be represented in the form of rational functions [62]. Here, the denominator represents the differential operator while the excitation is contained in the numerator. As opposed to Taylor expansions, Padé approximants have the ability to accurately model poles, which makes them more suitable for global approximation. Due to their strong non-linearity and ability to represent poles, Padé approximants are also useful as learnable activation functions for neural networks [63], [64].

In applications where we wish to interpolate scattering parameters with respect to frequency, it is natural to set $m = n - 1$ [65]. In this case, we can write the Padé approximant of a scattering parameter S_{pq} as the pole-residue expansion

$$S_{pq}(\omega) \approx \sum_{k=1}^K \frac{r_{pq,k}}{\omega - \omega_k}. \quad (2.11)$$

where the scattering parameter $S_{pq}(\omega)$ is characterized by the K poles ω_k and the residues $r_{pq,k}$. We assume that all scattering parameters share a

common set of poles [66] and thus omit the indices pq in the denominator of Eq. (2.11). With the help of algorithms such as FSID [67], vector fitting [68] or Padé-via-Lanczos [65], we can identify the residues and poles from the samples in the training set. The number of samples that are needed for identification increases with model order K . However, it is often sufficient to identify the dominant pole-residue pairs in the frequency range of interest, which presents the opportunity for substantial reductions in model order K . For noisy data, we can increase the number of samples that we use for identification to achieve a more robust approximation.

Multi-variate Padé approximants can also be used for interpolation with respect to parameters that describe the geometry and materials of a problem [69]. In Paper C, we use single-pole Padé approximants to locally interpolate scattering parameters with respect to the permittivity of a homogeneous dielectric within a metal pipe. The coefficients of the local Padé approximants are learned from samples that are computed for permittivity values that are uniformly sampled on a grid. Given a realization of an inhomogeneous dielectric in the pipe, we then exploit the local Padé approximants in a non-linear optimization algorithm to estimate the average permittivity of the inhomogeneous dielectric.

2.5 Towards global optimization

Efficient, accurate and robust global optimization is very desirable in many engineering design challenges and inverse problems. Here, we discuss how computationally efficient and sufficiently accurate surrogate models can be used to achieve this objective. We formulate an *optimization problem* [70] as

$$\begin{aligned} \mathbf{x}_{\text{opt}} &= \arg \min_{\mathbf{x} \in \mathcal{P}} G(\mathbf{f}(\mathbf{x})) \\ \text{s.t. } &\begin{cases} g_1(\mathbf{x}) \leq 0, g_2(\mathbf{x}) \leq 0, \dots, g_M(\mathbf{x}) \leq 0 \\ h_1(\mathbf{x}) = 0, h_2(\mathbf{x}) = 0, \dots, h_N(\mathbf{x}) = 0 \end{cases} \end{aligned} \quad (2.12)$$

where we seek to find the *global minimum* $\mathbf{x}_{\text{opt}}$ that minimizes the objective function G under a set of inequality constraints g_m for $m = 1, 2, \dots, M$ and equality constraints h_n for $n = 1, 2, \dots, N$. An optimization problem is called *feasible* if there is a vector $\mathbf{x}$ such that all constraints are fulfilled simultaneously, otherwise the optimization problem is called *infeasible*. For feasible problems, the set of parameter vectors $\mathbf{x}$ where all constraints are fulfilled is then called the *feasible region*. Furthermore, we distinguish between *linear* and *non-linear* optimization problems, as they typically require different solution methods. An optimization problem is only linear if the loss function and all the constraints are linear functions of $\mathbf{x}$. Many methods for non-linear optimization are local optimization methods, meaning that they locate a minimum in the vicinity of an initial guess [71]. For *convex* optimization problems, the global minimum $\mathbf{x}_{\text{opt}}$ is the only minimum, but in general, there may be *local minima* that does not correspond to the global minimum $\mathbf{x}_{\text{opt}}$. To avoid poor local minima, it is therefore necessary to find a sufficiently good initial guess for the optimization method, or to apply the method starting from sufficiently many initial guesses.

For an optimization method supported by a surrogate model, the surrogate model must be trained on the feasible region. For satisfactory interpolation performance, we should also ensure that the parameter domain $\hat{\mathcal{P}}$ is densely populated in the vicinity of the minimum that we seek. For problems with small parameter domains, we can simply densely populate the entirety of the parameter domain. However, for problems with high-dimensional parameter domains, the number of samples needed to densely populate the entire parameter domain may be prohibitively large. In some situations, we can use

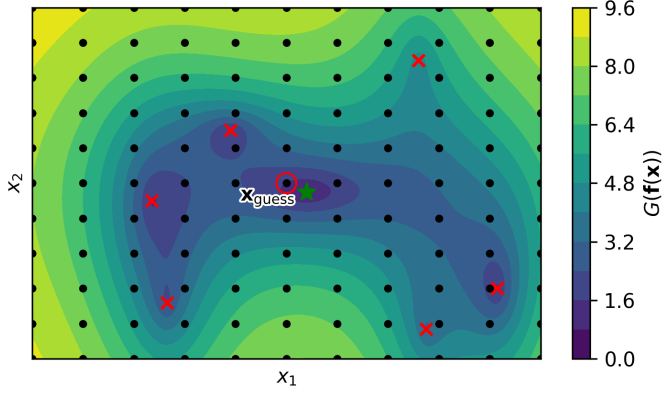


Figure 2.5: Example of a grid-search method given a two-dimensional parameter vector $\mathbf{x}$. The local minima of the objective function is shown with red crosses, while the global optimum is shown with a green star. The black dots indicate the set of candidates $\mathbf{x}^{(i)}$ for $i = 1, 2, \dots, N$ which are spaced on an equi-distant grid. The red circle shows the found initial guess $\mathbf{x}_{\text{guess}}$, which is the candidate $\mathbf{x}^{(i)}$ that gives the lowest value of $G(\mathbf{f}(\mathbf{x}^{(i)}))$.

conventional methods to obtain an initial guess for the minimum and populate the data set densely in the vicinity of this guess. Another powerful alternative is to employ iterative resampling schemes where the data set is iteratively populated in the vicinity of minima found by an optimization algorithm [72], [73].

In the following, we discuss how a computationally efficient surrogate model can be used to massively sample the parameter domain to find a satisfactory initial guess for local optimization methods. We also discuss how the Jacobian matrix (2.8) and higher-order derivatives, which can be readily computed given a differentiable surrogate model, can be used to formulate iterative methods for local optimization.

Sampling strategies and zeroth-order methods

When we solve a non-linear optimization problem, we often have little to no prior knowledge of the behavior of the loss function that we wish to minimize.

When this is the case, we must sufficiently explore the parameter domain to find a good initial guess before we apply any local optimization methods. A simple and robust approach is to find an initial guess $\mathbf{x}_{\text{guess}}$ for the global minimum as

$$\mathbf{x}_{\text{guess}} = \arg \min_{\mathbf{x}^{\{i\}}} G(\mathbf{f}(\mathbf{x}^{\{i\}})), \quad i \in \{1 \dots N\} \quad (2.13)$$

for some suitable set of candidates $\mathbf{x}^{\{i\}}$ for $i = 1, 2, \dots, N$. Fig. 2.5 shows an example of this, where the candidates $\mathbf{x}^{\{i\}}$ for a two-input model are chosen on a Cartesian grid. With a sufficiently small grid-spacing, grid-search methods allow thorough exploration of the parameter domain. However, the number of function evaluations grows exponentially with the number of parameters. There are a number of alternative strategies to select the set of candidates $\mathbf{x}^{\{i\}}$ that aim to mitigate the curse of dimensionality. Monte Carlo methods [74] employ random sampling of the parameter domain, whereas quasi-Monte Carlo methods base the sampling on pseudo-random sequences such as Sobol sequences [75] to avoid clustering of the samples. Latin hypercube sampling [76] divides the range of each parameter into equiprobable intervals and selects one sample in each interval for each parameter. Adaptive sampling strategies [77] dynamically assess already observed samples to make new samples target high-uncertainty regions of the parameter space.

The grid-search strategies discussed above are examples of *zeroth-order* optimization methods [78]. Zeroth-order optimization methods are based only on the value of the objective function G and do not require any explicit derivative information. They are particularly useful when derivatives are computationally expensive or unavailable, e.g. for discrete parameters. Global zeroth-order optimization methods aim to find an approximation of the global minimum of G , which can then be refined by means of local optimization methods. Common examples include genetic algorithms [79], particle swarm optimization [80] and simulated annealing [81]. A disadvantage of zeroth-order optimization methods is that they typically require a very large number of evaluations of the objective function G . To employ such methods, it is therefore very beneficial to have access to a computationally efficient surrogate model.

Iterative methods for local optimization

Iterative optimization methods based on a search direction are an important kind of local optimization methods that exploit the updating scheme

$$\mathbf{x}^{[n+1]} = \mathbf{x}^{[n]} - \gamma^{[n]} \mathbf{d}^{[n]} \quad (2.14)$$

where we, for each iteration n , take a step of length $\gamma^{[n]}$ in direction $\mathbf{d}^{[n]}$ [82]. The choice of optimization method dictates how $\gamma^{[n]}$ and $\mathbf{d}^{[n]}$ are chosen. The methods are initialized with an initial guess $\mathbf{x}^{[0]}$ and attempt to locate a minimum in the vicinity of this guess. Here, it is advantageous to use a global search strategy as discussed in Section 2.5 to obtain a good initial guess for the global minimum.

First-order methods

In *first-order* optimization methods [83], the choice of direction $\mathbf{d}^{[n]}$ is based on the gradient of the objective function $\nabla_{\mathbf{x}}G$. Examples include line-search methods and trust-region-based methods such as the Levenberg-Marquardt algorithm and Powell’s dogleg method. In the Gauss-Newton method, which is used to solve non-linear least squares problems, the gradient $\nabla_{\mathbf{x}}G$ is used to compute an approximation of the Hessian matrix for the goal function G . We note that

$$\nabla_{\mathbf{x}}G = \mathbf{J}^T \nabla_{\mathbf{y}}G \quad (2.15)$$

and thus, as long as we can differentiate G with respect to the elements of $\mathbf{y}$ to compute $\nabla_{\mathbf{y}}G$, we can use the Jacobian matrix $\mathbf{J}$ to compute $\nabla_{\mathbf{x}}G$.

For conventional CEM methods, the derivatives in $\mathbf{J}$ can be computationally costly to obtain. Some derivatives can be obtained by means of continuum sensitivity analysis as discussed in Section 1.3, but this is not always possible. Another alternative is to estimate $\mathbf{J}$ through finite-difference approximations of the derivatives, but this approach is computationally costly and prone to numerical errors. Here, it is advantageous to use surrogate models for which $\mathbf{J}$ can be evaluated cheaply, for example by means of automatic differentiation [61] of a neural network.

Second-order and higher-order methods

Second-order optimization methods include second-order derivatives in their update schemes to improve the rate of convergence compared to lower-order methods. It is possible to explicitly formulate such methods for Maxwell's equations [84], but it involves significant effort. Here, it is attractive to have access to an easily differentiable surrogate model. A common example of a second-order method is the Newton-Raphson method [83], which uses the inverse of the Hessian matrix to scale the gradient step. Should even higher-order derivatives be available, *higher-order* methods may further improve the rate of convergence [85], [86].

The Hessian matrix is also useful to characterize stationary points located by a first-order optimization method. By studying the eigenvalues of the Hessian matrix, we can determine if the stationary point is a minimum, maximum or saddle point. Should the Hessian feature zero (or small) eigenvalues, the test based on only second-order derivatives is inconclusive and higher-order derivatives should be evaluated and assessed. Again, automatic differentiation applied to a neural network may be an attractive alternative for such a task.

CHAPTER 3

Methods for Inline Material Characterization

Most conventional techniques for electromagnetic characterization of materials are offline methods, where a sample under test (SUT) is placed in a custom measurement device that is designed for the characterization task at hand. The measurement device is typically designed such that the quantity of interest is clearly observable and simple to calculate from the observed quantities. For the characterization of solid dielectrics, common methods include cavity perturbation techniques [87] and waveguide measurements [88], while liquids are typically characterized using coaxial probes [89]. For the characterization of magnetic materials, common measurement setups include the Epstein frame, single sheet testers, and ring sample methods [90]. However, offline characterization methods are inappropriate for many applications. In the manufacturing industry, discrepancies due to manufacturing tolerances mean that the electric or magnetic properties of every individual product may vary significantly from their desired values. Moreover, in process industries where liquid or gaseous materials are continuously processed, it is desirable to continuously monitor the properties of the materials to quickly identify any potential issues.

Unlike offline characterization methods, inline characterization methods are

integrated directly into the manufacturing or processing pipeline. This typically means that we have less control over the design of the measurement setup and that potential issues with observability must be addressed. The methods may also include the need to model complicated electromagnetic systems, and significant calculations may be required to compute the parameters of interest from the observed data. In this chapter, we describe how statistical signal processing can be used to formulate material characterization methods that exploit prior knowledge about the sought parameters.

3.1 Formulating the material characterization problem

We formulate the material characterization problem as a *parameter estimation* problem [91], where we wish to determine the underlying parameters $\boldsymbol{\xi}$ that caused a set of observations of $\mathbf{y}$. We complete the model problem (1.18) to form the signal model for the frequency-wise observation

$$\tilde{\mathbf{y}}_n = \mathbf{f}(\omega_n, \boldsymbol{\xi}) + \mathbf{w}_n, \quad n = 1, 2, \dots, N \quad (3.1)$$

where we may have observations for many different frequencies ω_n for the same geometry and material description $\boldsymbol{\xi}$. Here, we have the stochastic noise vector $\mathbf{w}_n$ that contains the discrepancy between the output of the model and the observed data, typically due to measurement noise. We combine the frequency-wise observations to form the full observation

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \tilde{\mathbf{y}}_2 \\ \vdots \\ \tilde{\mathbf{y}}_N \end{bmatrix} \quad (3.2)$$

which is characterized by the probability density function (PDF) $p(\tilde{\mathbf{Y}}; \boldsymbol{\xi})$ that is parameterized by the parameter vector $\boldsymbol{\xi}$. Depending on which parameters are contained in $\boldsymbol{\xi}$, this general formulation also includes other applications, such as direction of arrival estimation and calibration of measurement systems.

For the following, we assume a simple noise model where we have complex Gaussian noise $\mathbf{w}_n \sim \mathcal{CN}(\mathbf{0}, \sigma_y^2 \mathbf{I}_P)$ where we use $\mathbf{I}_n$ to denote the identity

matrix of size $n \times n$. Thus, for a fixed parameter vector $\boldsymbol{\xi}$, we have

$$\tilde{\mathbf{Y}} \sim \mathcal{CN}(\boldsymbol{\mu}, \sigma_y^2 \mathbf{I}_{NP}) \quad (3.3)$$

where the mean is determined by

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{f}(\omega_1, \boldsymbol{\xi}) \\ \mathbf{f}(\omega_2, \boldsymbol{\xi}) \\ \vdots \\ \mathbf{f}(\omega_N, \boldsymbol{\xi}) \end{bmatrix}. \quad (3.4)$$

Estimators

An *estimator* of $\boldsymbol{\xi}$ is a mapping that assigns an *estimate* $\hat{\boldsymbol{\xi}}$ of the parameter vector $\boldsymbol{\xi}$ to a realization of the observation $\tilde{\mathbf{Y}}$. As the observation $\tilde{\mathbf{Y}}$ is a stochastic variable, the estimate $\hat{\boldsymbol{\xi}}$ will also be stochastic. An estimator is *unbiased* if

$$\mathbb{E}_{\mathbf{w}}\{\hat{\boldsymbol{\xi}}\} = \boldsymbol{\xi} \quad (3.5)$$

where $\mathbb{E}_{\mathbf{w}}\{\hat{\boldsymbol{\xi}}\}$ denotes the expected value of the estimate $\hat{\boldsymbol{\xi}}$ with respect to the random noise $\mathbf{w}$. When we attempt to find an estimator, there are potential complications that we must consider. All parameters may not be *identifiable*, which means that there are multiple parameter vectors $\boldsymbol{\xi}$ that produce the same probability distribution $p(\tilde{\mathbf{Y}}; \boldsymbol{\xi})$. To obtain useful estimates for such problems, we can apply a Bayesian framework to exploit prior knowledge about the sought parameters as discussed in Section 3.2.

Cramér-Rao lower bounds

It is often of interest to know the best possible performance of an estimator. For unbiased estimators, the *Cramér-Rao lower bound* (CRLB) provide lower bounds on the variances of the estimates $\hat{\xi}_q$ as

$$\mathbb{E}_{\mathbf{w}}\{(\hat{\xi}_q - \xi_q)^2\} \geq \{\mathbf{F}^{-1}(\xi_q)\}_{qq} \triangleq \text{CRLB}_q, \quad \forall q \quad (3.6)$$

where $\mathbf{F}$ is known as the *Fisher information matrix* [91]. For the simple noise model in (3.3), the frequency-wise observations are independent, which makes

the Fisher information matrices additive for each frequency point yielding

$$\mathbf{F} = \sum_{n=1}^N \mathbf{F}_n. \quad (3.7)$$

Here, we have the frequency-wise Fisher information matrix $\mathbf{F}_n$ for the frequency point ω_n . Further, the elements in $\mathbf{F}_n$ are given by

$$F_{n,ij} = \frac{2}{\sigma_y^2} \Re \left\{ \frac{\partial \mathbf{f}^H}{\partial \theta_i} \frac{\partial \mathbf{f}}{\partial \theta_j} \right\} \bigg|_{\omega=\omega_n}. \quad (3.8)$$

The CRLB can be used to derive theoretical bounds for the accuracy of experimental setups [92], which is useful for experiment design [93], [94]. They also allow us to make informed choices about which frequency points ω_n to use in the parameter estimation problem. In particular, this is valuable for situations where the model $\mathbf{f}$ is computationally expensive to evaluate. Here, a computational budget may limit the number of times we can evaluate $\mathbf{f}$ and, in turn, how many frequency points N we are able to use for the parameter estimation problem. Then, we can use the CRLB to choose the N frequency points that provide the most information about the parameters of interest.

3.2 Bayesian estimation

In classical estimation, the parameters in $\boldsymbol{\xi}$ are viewed as fixed and deterministic. However, in the context of inline material characterization methods, we typically wish to estimate $\boldsymbol{\xi}$ repeatedly for many different samples. In *Bayesian estimation*, we view the parameters in $\boldsymbol{\xi}$ as stochastic variables, which provides a way to make use of prior knowledge about the parameters in the form of a prior PDF $p(\boldsymbol{\xi})$.

In Paper B, we present a Bayesian estimation framework for electric machines. Here, we discuss how we can use this Bayesian estimation framework to exploit prior knowledge about the sought parameters and approach otherwise intractable estimation problems.

Choice of prior distribution

To make use of the prior knowledge, we must choose a prior PDF $p(\boldsymbol{\xi})$. Typically, the prior knowledge is obtained by means of physical arguments or empirical observations. In this work, we consider the parameters in $\boldsymbol{\xi}$ to be independently normally distributed according to

$$p(\boldsymbol{\xi}) \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\xi}}, \mathbf{C}_{\boldsymbol{\xi}}) \quad (3.9)$$

where

$$\mathbf{C}_{\boldsymbol{\xi}} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_Q^2 \end{bmatrix}, \quad (3.10)$$

but other choices are also possible. Here, we can choose a very large variance σ_q^2 for parameters ξ_q for which we have little or no prior knowledge.

Bayesian Cramér-Rao lower bounds

For a stochastic $\boldsymbol{\xi}$, we can define a Bayesian Fisher information matrix [95]

$$\mathbf{F}_B = \mathbf{F} + \mathbf{F}_P. \quad (3.11)$$

Here, the Fisher information matrix (3.7) represents the information provided by the observed data, while $\mathbf{F}_P$ represents the information provided by the prior knowledge. Given the Bayesian Fisher information matrix $\mathbf{F}_B$, we can compute a Bayesian version of the CRLB for the Bayesian mean squared error (Bmse) of the estimate $\hat{\xi}_q$ as

$$\text{Bmse} \left\{ \hat{\xi}_q \right\} = \mathbb{E}_{\mathbf{w}, \boldsymbol{\xi}} \left\{ (\xi_q - \hat{\xi}_q)^2 \right\} \geq \left\{ \mathbf{F}_B^{-1}(\boldsymbol{\mu}_{\boldsymbol{\xi}}, \mathbf{C}_{\boldsymbol{\xi}}) \right\}_{qq} \triangleq \text{BCRLB}_q, \quad \forall q \quad (3.12)$$

where the joint expectation $\mathbb{E}_{\mathbf{w}, \boldsymbol{\xi}}$ is taken both with respect to the noise vector $\mathbf{w}$ and the parameter vector $\boldsymbol{\xi}$. For the Gaussian prior (3.9), the prior information matrix $\mathbf{F}_P$ is given by $\mathbf{F}_P = \mathbf{C}_{\boldsymbol{\xi}}^{-1}$.

In Paper B, we compute the BCRLB for parameters that describe the materials of an electric motor. First, we use the BCRLB to assess the variance

of the estimates for different numbers of frequency points N to make an informed choice about how many frequency points to use in the estimation procedure. We also compare the performance of an estimator to the BCRLB for the sought parameters, which allows us to demonstrate that the estimator performs as expected.

Maximum a posteriori estimator

Given a prior PDF $p(\boldsymbol{\xi})$, we apply Bayes' theorem to find the *posterior PDF*

$$p(\boldsymbol{\xi}|\tilde{\mathbf{Y}}) = \frac{p(\tilde{\mathbf{Y}}|\boldsymbol{\xi})p(\boldsymbol{\xi})}{p(\tilde{\mathbf{Y}})}. \quad (3.13)$$

The mean of the posterior PDF gives the minimum mean square error (MMSE) estimator $\hat{\boldsymbol{\xi}}_{\text{MMSE}} = \mathbb{E}_{\boldsymbol{\xi}}\{p(\boldsymbol{\xi}|\tilde{\mathbf{Y}})\}$, which minimizes the Bmse $\mathbb{E}_{\mathbf{w}, \boldsymbol{\xi}}\{\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\|_{\text{F}}^2\}$. However, computing the mean of the posterior PDF typically involves integration, which can be very computationally demanding to implement.

In this work, we instead formulate the maximum a posteriori (MAP) estimator, which is a natural Bayesian generalization of maximum likelihood estimation (MLE). Here, we maximize the logarithm of the posterior PDF (3.13) to obtain the MAP estimator

$$\hat{\boldsymbol{\xi}}_{\text{MAP}} = \arg \max_{\boldsymbol{\xi}} \{\log p(\tilde{\mathbf{Y}}|\boldsymbol{\xi}) + \log p(\boldsymbol{\xi})\}. \quad (3.14)$$

The MAP estimator only includes a maximization, which is attractive for numerical implementation. Should the posterior PDF be symmetric and unimodal, the MMSE estimator and the MAP estimator will be identical [91].

For the simple noise model (3.3) and the Gaussian prior, the MAP estimator (3.14) can be written as

$$\hat{\boldsymbol{\xi}}_{\text{MAP}} = \arg \min_{\boldsymbol{\xi}} \{G(\boldsymbol{\xi}; \tilde{\mathbf{Y}})\} \quad (3.15)$$

where

$$G(\boldsymbol{\xi}; \tilde{\mathbf{Y}}) = \sum_{n=1}^N \|\tilde{\mathbf{y}}_n - \mathbf{f}(\omega_n, \boldsymbol{\xi})\|_{\text{F}}^2 + \frac{\sigma_y^2}{2} \sum_{q=1}^Q \frac{1}{\sigma_q^2} (\xi_q - \mu_q)^2. \quad (3.16)$$

We can interpret this as the minimization of a cost function that is the weighted sum of two terms: (i) the mismatch between the model and the observation; and (ii) a regularization term based on the deviation between the parameters ξ_q and their prior mean values μ_q . The weight between the two terms is determined by the power of the measurement noise σ_y^2 , while the weight of term q in the regularization term is determined by the variance σ_q^2 .

It is often difficult or impossible to derive closed-form expressions for estimators such as the MAP estimator (3.14). In these situations, we have to resort to numerical methods such as the Expectation-Maximization algorithm or Fisher's scoring algorithm [91] to compute the estimate. In Paper B, we instead formulate the MAP estimator in terms of an optimization problem (2.12). Here, we use the constraints g_i and h_j to limit the search space of the optimization method to avoid non-physical parameter values.

3.3 Choice of model parameterization

An important consideration for the parameter estimation problem is how we *parameterize* our model. Often, we need to estimate other unknown parameters in addition to those in which we are interested. We refer to parameters that affect the output of the model but are not the objective of the estimation as *nuisance parameters*. In Bayesian estimation, one possibility is to marginalize the nuisance parameters out of the posterior PDF to obtain a marginal PDF for only the sought parameters [91]. For estimation problems with many parameters, it is often the case that only a few linear combinations of parameters are identifiable [96]. In these situations, reducing the dimensionality of the parameter space or including prior information as described in Section 3.2 can help make the estimation problem tractable [97].

In Paper B, we wish to estimate the effective permeability of the laminates of an electric motor. Here, we must also estimate the unknown magnetic couplings of the parts of the motor windings that protrude outside of the motor, which we consider to be nuisance parameters. Based on geometrical arguments, we identify couplings that are of the same magnitude to significantly reduce the number of magnetic couplings that we must estimate. The model in Paper B is an example of a *gray-box model*, which is a physics-based model with learning-based elements for parts of the model that are difficult to represent accurately using conventional physics-based methods.

CHAPTER 4

Applications and Results

Here, we present several applications for the proposed methods in material characterization, computationally efficient modeling and design optimization; along with results that validate the effectiveness of the methods.

4.1 Monitoring of inhomogeneous media in a metal pipe

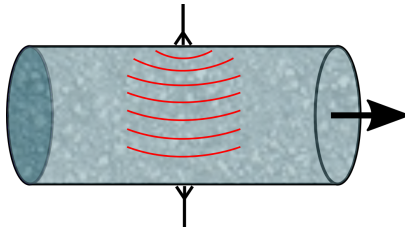


Figure 4.1: Inhomogeneous dielectric media that is pumped through a metal pipe equipped with two microwave antennas.

Pumped dielectric media are ubiquitous in process industry. Microwave sensors provide a cheap and non-destructive method for monitoring material properties of dielectric media, with important applications in quality control and process monitoring [98].

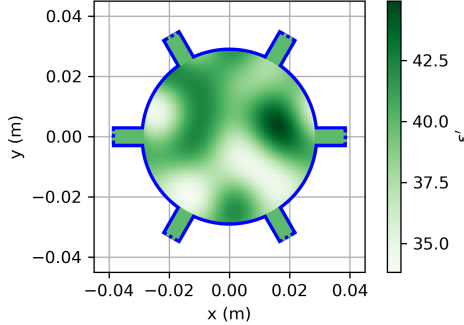


Figure 4.2: Geometry for a model of a six-port measurement device connected to a circular cavity with a realization of a random dielectric medium.

In Papers A, C, and E, we consider an inhomogeneous dielectric medium that flows through a metal pipe, as illustrated in Fig. 4.1. Fig. 4.2 shows a cross-section of the pipe, where we model the antennas as parallel-plate waveguides connected to the pipe. The flowing dielectric medium is modeled by a random permittivity, where each realization of the random permittivity corresponds to a new snapshot of the medium that is currently inside the pipe.

Estimation of statistical properties of dielectric medium

In Paper E, we train a fully-connected neural network to estimate the point-wise mean and standard deviation of the permittivity of the dielectric medium given observations of the scattering parameters at the waveguide ports. As seen in Fig. 4.3, the neural network significantly outperforms a linear model in estimating the standard deviation of the permittivity.

High-dimensional histograms for detection problems

In Paper A, we first use a FEM solver to create a data set that contains the scattering parameters at the waveguide ports for many random realizations of

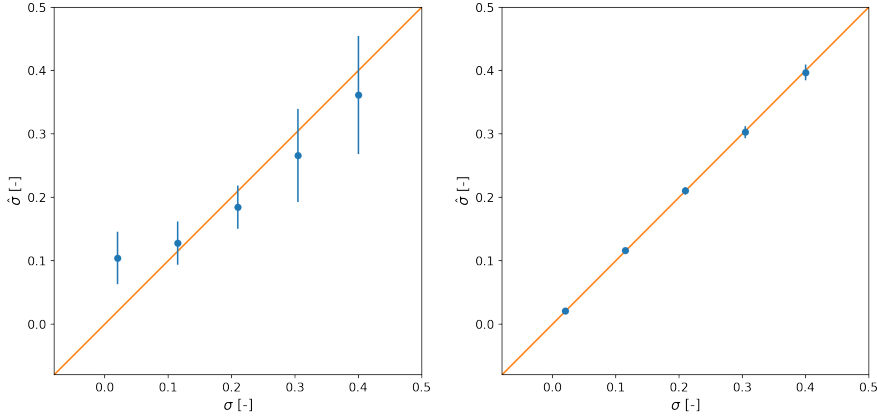


Figure 4.3: Estimated point-wise standard deviation of permittivity of dielectric medium - left: linear model; and right: neural network. The x-axis corresponds to the true standard deviation σ , while the y-axis corresponds to the estimated standard deviation $\hat{\sigma}$. Discs - mean value of the estimates; and vertical bars - standard deviation of the estimates.

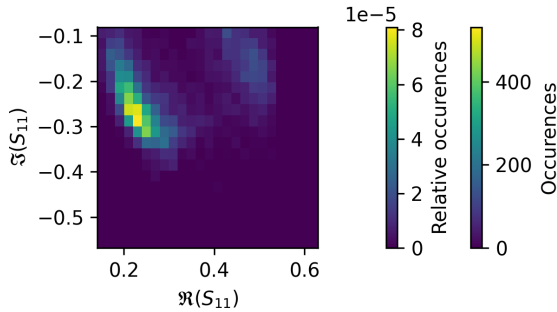


Figure 4.4: Slice of a high-dimensional histogram.

the random permittivity. We use continuum sensitivity analysis to compute gradients of the scattering parameters with respect to the parameters that describe the random permittivity, which we also include in the data set. As described in Section 2.3, we then use this data set to train a GIANT networks that are up to 30 layers deep to act as accurate and computationally efficient

surrogate models for the FEM solver. The gradient-informed training allows us to efficiently train GIANT networks that serve as very accurate surrogate models with only a small fraction of the number of samples that are needed to train a corresponding conventional neural network.

We use the GIANT networks to estimate high-dimensional histograms, as seen in Fig. 4.4. High-dimensional histograms are useful in detection applications [99]; however, they require large amounts of data to construct. With the FEM solver, it would take more than 6 years to compute all the samples that are needed to construct a histogram of acceptable resolution. For a neural network trained by the GIANT framework, it takes less than 6 hours for the full process of creating training and test sets, training the neural network, and computing the samples that are needed to construct the histogram.

Inline calibration and permittivity estimation

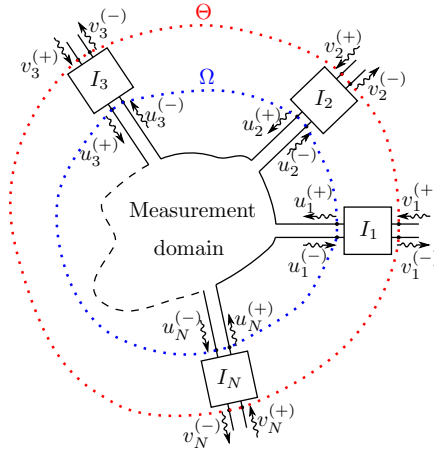


Figure 4.5: Near-field microwave measurement system with N ports, where each port p is equipped with a two-port microwave instrument I_p .

There are always errors associated with measurements. The errors can be divided into three main types [100]: (i) random errors; (ii) systematic errors; and (iii) drift errors. Random errors are typically caused by electronic noise. Systematic errors come from imperfections in the measurement setup, for ex-

ample, from cables and connectors. Drift errors are systematic errors that vary over time, often caused by temperature fluctuations or other environmental changes. We can compensate for systematic errors and drift errors through *calibration* of the measurement system. To account for drift errors, the calibration must be repeated in regular intervals. This can be problematic in settings with limited access to the measurement system, such as production settings.

In Paper C, we present an auto-calibration method for a near-field microwave measurement system, as illustrated in Fig. 4.5, that eliminates the need for a separate calibration phase. Here, we use observations of scattering parameters that are corrupted by systematic/drift errors, which we model using unknown calibration parameters. Given the corrupted scattering parameter observations, we wish to determine the average permittivity of an unknown medium under test. We assume that we have access to a series of a-priori characterization observations of media with known permittivities, where we use Padé approximants as described in Section 2.4 for interpolation in the permittivity domain between the characterization observations. Once we collect observations on an unknown medium under test, we estimate the calibration parameters and the average permittivity of the medium under test simultaneously. As shown in Fig. 4.6, the method works for a wide range of Signal-to-Noise Ratios (SNR) and standard deviations σ of the inhomogeneous permittivity. Since we gather the characterization observations in advance, the method is well suited for inline applications with limited access to the measurement system.

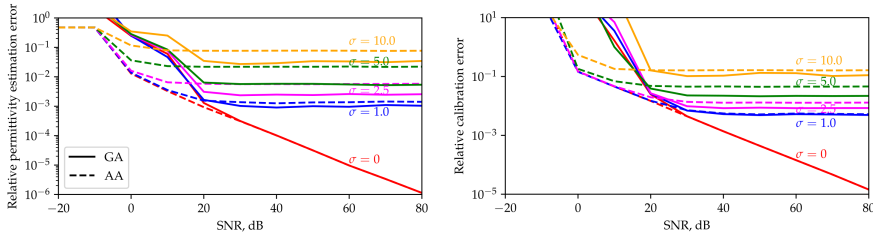


Figure 4.6: Relative permittivity-estimation and calibration errors as functions of SNR for true permittivity value for $\bar{\epsilon} = 60.4 - j13.2$ and a series of values of standard deviation of permittivity σ' .

4.2 Robust geometry optimization for microwave waveguide filter

Waveguide filters are important components of communication systems that can be realized by modifying the geometry of a waveguide. However, the characteristics of the filters can be very sensitive to perturbations of the waveguide geometry within typical manufacturing tolerances.

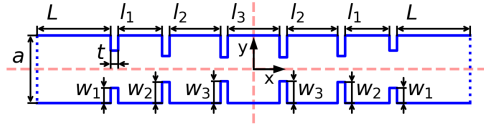


Figure 4.7: Geometry of waveguide filter.

In Paper A, we study a pass-band H-plane waveguide filter as shown in Fig. 4.7. First, we use a FEM solver to create a data set that contains the reflection coefficient S_{11} of the waveguide filter for many random realizations of the parameters that describe its geometry. We use continuum sensitivity analysis to compute gradients of the scattering parameters with respect to the geometry parameters, which we also include in the data set. As described in Section 2.3, we then use this data set to train a GIANT network as a surrogate model for the FEM solver. Finally, we use the GIANT network to find filter geometries that match given filter specifications and are robust with respect to perturbations of the geometry. We iteratively augment the data set with geometries in the vicinity of the optimized geometries to ensure good performance of the GIANT network.

Fig. 4.8 shows the magnitude of the reflection coefficient $|S_{11}|$ for an optimized filter geometry. The red shaded areas indicate the filter specification, where the reflection coefficient should be above the shaded area in the *stop-bands* and below the shaded area in the *pass-band* of the filter. The blue solid curve shows $|S_{11}|$ as estimated by the neural network. The blue dashed curve shows the corresponding $|S_{11}|$ as computed by the FEM solver, which is considered the ground truth. The red dash-dotted curve indicates the magnitude of the error between the neural network and the FEM solver. Additionally, the green dotted curve shows the $|S_{11}|$ that best matches the filter specification

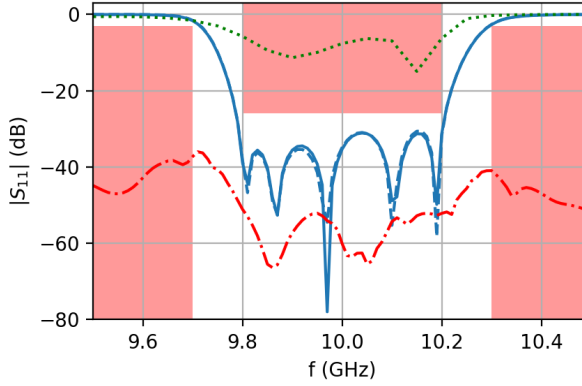


Figure 4.8: Magnitude of reflection coefficient for a filter that is robust to perturbations of the filter geometry where the filter specification is indicated by the red shaded areas. Solid curve – output of neural network; dashed curve – output of FEM solver; red dash-dotted curve – magnitude of error between output of neural network and FEM solver; and green dotted curve – best match of training set.

out of all the filters in the training set. We see that even the best filter of the training set shows a very poor match to the filter specification, which indicates that the neural network is capable of interpolating between the geometries of the training set to find novel geometries of interest. To ensure that the filters are robust to perturbations of the geometry, we also evaluate perturbed filter geometries within typical manufacturing tolerances of a waveguide filter. To rapidly evaluate the many perturbed filter geometries, it is of great benefit to have access to a computationally efficient surrogate model.

4.3 Electromagnetic compatibility for electric drivelines

Electromagnetic Compatibility (EMC) is the ability of electric systems to coexist without causing or suffering from Electromagnetic Interference (EMI). It is desirable to detect and treat EMC-related issues in the early stages of product development, where costly redesigns can be avoided. Here, it is of

great benefit to have access to efficient and accurate models to easily evaluate component designs.

Due to the ongoing shift towards electric drivelines, EMC considerations are of increasing concern for the automotive industry [101]. If not handled properly, EMI associated with the large currents and high voltages present in electric drivelines may cause malfunctions in the many sensitive electrical components that are part of modern control and entertainment systems. In a typical electric driveline, a power inverter uses pulse width modulation (PWM) to convert the direct current (DC) power of a battery to alternating current (AC), which feeds the electric motor. The battery, along with the power inverter, constitutes a source of EMI due to the rapid switching of the power inverter. Common-mode (CM) currents are particularly problematic with regard to radiated emissions [102]. The electric motor constitutes a coupling path through which the CM currents can propagate to the vehicle chassis and cause issues with EMI, overvoltages and bearing discharging currents. Here, the CM impedance Z_{CM} of the motor is an important quantity for predicting and understanding potential issues with CM currents.

Perturbation analysis for model of hairpin electric motor

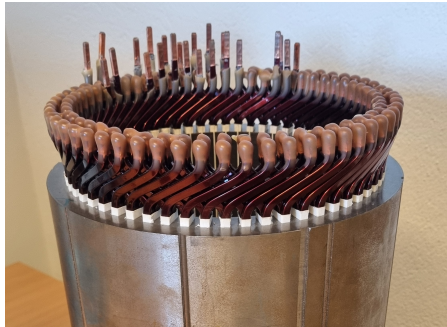


Figure 4.9: Welded end windings of an electric motor with hairpin windings. The hairpin windings consist of so-called hairpins, rectangular copper wires with a hairpin shape, that are welded together to form the windings.

In Paper D, we model an electric motor that is equipped with hairpin windings as shown in Fig. 4.9. Hairpin windings have rectangular cross-sections

and offer improved performance and efficiency when compared to conventional windings [103]. The objective of the model is to compute the common-mode impedance Z_{CM} of the motor for frequencies up to 100 MHz.

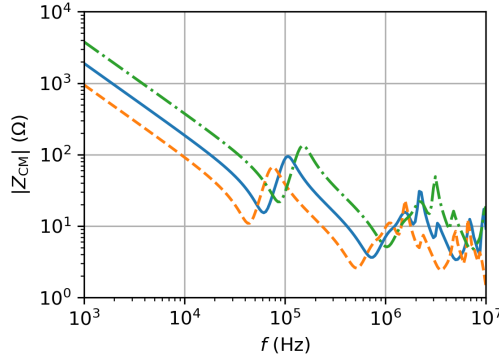


Figure 4.10: Common-mode impedance Z_{CM} for uncertainty in the permittivity of the insulation material of an electric motor: solid blue curve – $\epsilon_r = 4.5$; dashed orange curve – $\epsilon_r = 9.0$; and dashed-dotted green curve – $\epsilon_r = 2.3$.

A full 3D model is costly, both in terms of labor to construct and computational resources to evaluate. Here, we instead exploit the cylindrical symmetry of the motor to decompose the model into a 2D model of the cross-section together with a 1D model along the motor axis, where the details of the model are discussed below in relation to Paper B. We use the model to determine the sensitivity of the CM impedance Z_{CM} to perturbations in the parameters that describe the materials and geometry of the motor. We conclude that the CM impedance is sensitive to perturbations in the permeability of the motor laminations and in the permittivity of the insulation materials, as shown in Fig. 4.10. However, the CM impedance is robust with respect to perturbations of the positions of the hairpin conductors within typical manufacturing tolerances, which we determine to be an advantage of hairpin electric motors.

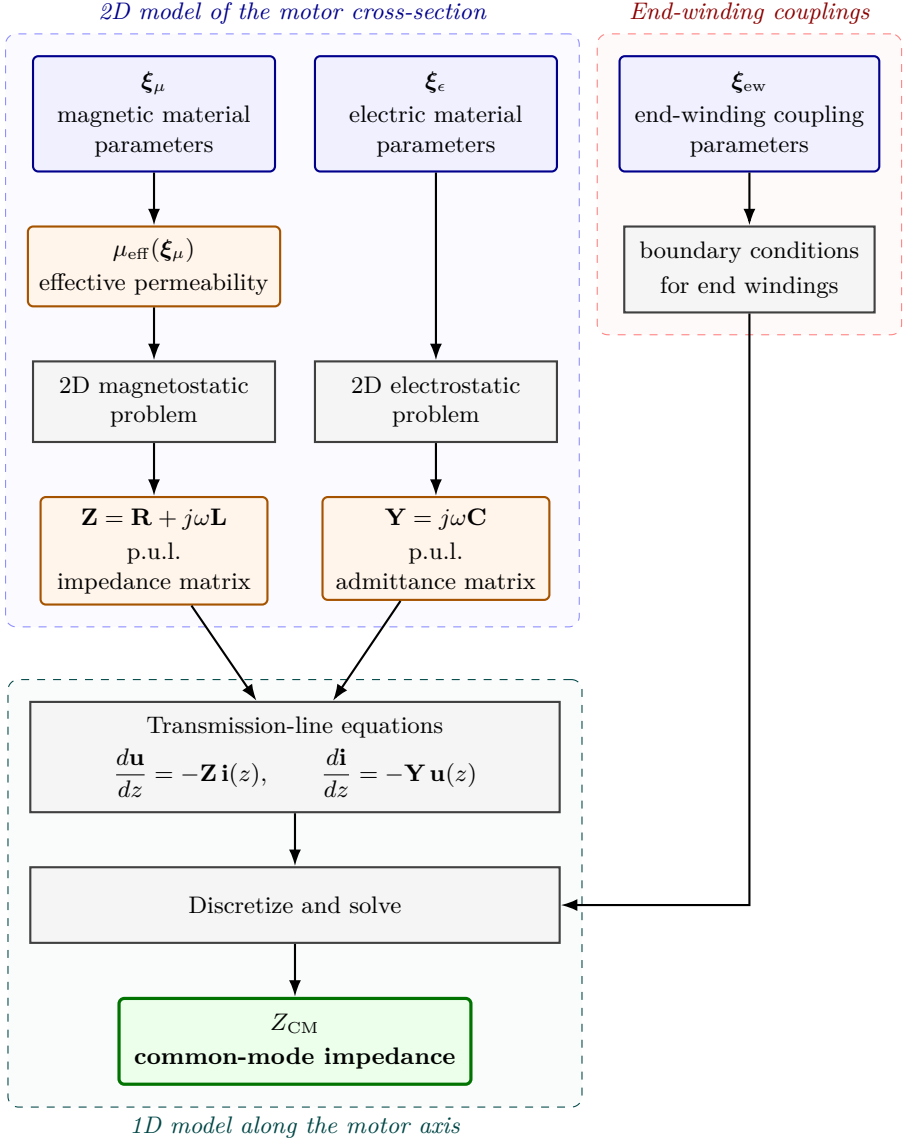


Figure 4.11: Flowchart of the electric motor model.

Parameter estimation for model of hairpin electric motor

In Paper B, we complete the model of Paper D with a description of the couplings between the parts of the windings that are external to the motor. Fig. 4.11 shows a flowchart that describes the complete model of the electric motor. The model is parametrized by parameters (i) ξ_μ that describe the magnetic properties of the materials and yields an effective permeability $\mu_{\text{eff}}(\xi_\mu)$ of the laminations of the motor; (ii) ξ_ϵ that describe the electric properties of the materials; and (iii) ξ_{ew} that describe the magnetic couplings between the end windings of the motor. The per-unit length impedance matrix $\mathbf{Z}$ and admittance matrix $\mathbf{Y}$ are computed from two-dimensional magnetostatic and electrostatic problems in the cross-section of the motor. The per-unit length matrices $\mathbf{Z}$ and $\mathbf{Y}$ are then used in the transmission line equations

$$\frac{d\mathbf{u}}{dz} = -(\mathbf{R} + j\omega\mathbf{L})\mathbf{i}(z) = -\mathbf{Z}\mathbf{i}(z) \quad (4.1)$$

$$\frac{d\mathbf{i}}{dz} = -(\mathbf{G} + j\omega\mathbf{C})\mathbf{u}(z) = -\mathbf{Y}\mathbf{u}(z) \quad (4.2)$$

which describe the currents $\mathbf{i}$ and the voltages $\mathbf{u}$ along the windings of the motor. We complete the transmission line equations with boundary conditions that are derived from the parameters ξ_{ew} that describe the couplings of the end windings. Finally, we discretize and solve the transmission line equations, which allows us to compute the common-mode impedance Z_{CM} of the motor.

We implement an MAP estimator as described in Section 3.2 to estimate the parameters in ξ_μ , ξ_ϵ and ξ_{ew} given measurements of scattering parameters between the phases of an electric hairpin motor. Here, we use separate low-frequency measurements and simple analytical models to obtain prior values for the parameters that are to be estimated. As shown in Fig. 4.12, the CM impedance Z_{CM} computed by the model displays good agreement with the measurements for the whole frequency range under consideration.

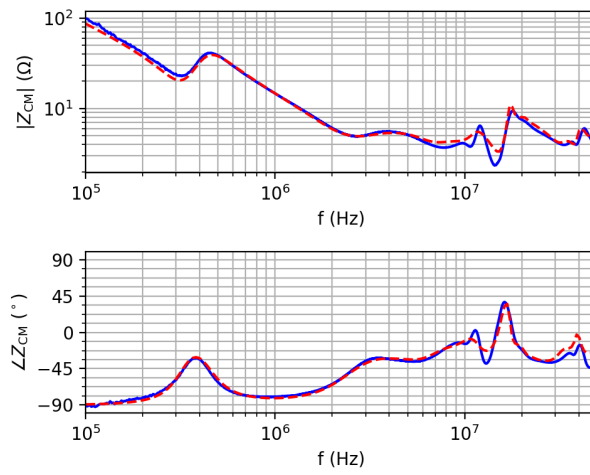


Figure 4.12: Common-mode impedance estimated by the motor model compared to measurements

CHAPTER 5

Author Contributions to Included Papers

This chapter provides a summary of the author contributions to the included papers.

5.1 Paper A

Simon Stenmark, Thomas Rylander, Tomas McKelvey, Andrei Ludvig-Osipov

GIANT Networks: Very Deep Fully-Connected Neural Networks Applied to Microwave Problems

Published in IET Microwaves, Antennas & Propagation,
vol. 20, e70077, 2026.

DOI: 10.1049/mia2.70077 .

Simon Stenmark invented the low-variance preserving scaling function and derived the update of the Adam optimizer parameters. **Simon Stenmark** and Tomas McKelvey jointly developed the update scheme for the normalization parameters. **Simon Stenmark**, Andrei Ludvig-Osipov and Thomas Rylander jointly developed the Sobolev-training technique based on

continuum sensitivity analysis. **Simon Stenmark** performed all implementation, developed all test cases, tuned the hyper-parameters, trained the neural networks and generated all results of the paper, which involved analyzing the results in collaboration with the other authors. **Simon Stenmark** and Thomas Rylander co-wrote the paper with input from the other authors and revised the paper based on feedback from reviewers.

5.2 Paper B

Simon Stenmark, Thomas Rylander, Jan Carlsson, Mats Viberg,
Matthys M. Botha
Parameter Estimation for Electric Machines with Hairpin Conductors
Utilizing Hybrid Modeling
Submitted for publication in IEEE Transactions on Energy Conversion.

Simon Stenmark, Thomas Rylander, Jan Carlsson and Matthys M. Botha jointly developed the mathematical description of the model for the electric machine. **Simon Stenmark** conceived the idea to apply the Bayesian estimation framework and the mathematical description of the MAP estimator was developed by **Simon Stenmark** and Mats Viberg. **Simon Stenmark** implemented the model of the electric machine and the MAP estimator, created the synthetic data sets, developed the test cases and generated all the results of the paper. The novel measurement setup was conceptualized **Simon Stenmark**, Thomas Rylander and Jan Carlsson. All the measurements were performed by **Simon Stenmark** and Jan Carlsson. **Simon Stenmark** and Thomas Rylander co-wrote the paper with input from the other authors.

5.3 Paper C

Andrei Ludvig-Osipov, **Simon Stenmark**, Thomas Rylander,
Tomas McKelvey
Auto-calibration of near-field microwave measurements for complex permittivity estimation
Published in Progress In Electromagnetics Research M,
vol. 136, pp. 46-56, 2025.
DOI: 10.2528/PIERM25090302 .

Simon Stenmark selected and computed all data sets used in the paper based on relevant applications in process industry. This included the implementation of parallelized code for the parameter sweeps and performing the computations on the Vera compute cluster at Chalmers. **Simon Stenmark** and Thomas Rylander developed the model for the measurement system. All co-authors of the paper jointly developed the estimation theory with the generalized average and Möbius transforms. Andrei Ludvig-Osipov implemented the methods and generated all the estimation results of the paper. Andrei Ludvig-Osipov wrote the original version of the paper with input from the other authors. **Simon Stenmark** and Thomas Rylander revised the paper based on feedback from the reviewers.

5.4 Paper D

Simon Stenmark, Thomas Rylander, Matthys M. Botha, Jan Carlsson
Common-Mode Impedance of an Electric Motor and the Impact of Material and Geometry Uncertainties
2024 International Symposium on Electromagnetic Compatibility – EMC Europe,
Brugge, Belgium, 2024, pp. 895–900.
DOI: 10.1109/EMCEurope59828.2024.10722688 .

All co-authors jointly developed the mathematical description of the model for the electric machine. **Simon Stenmark** implemented and validated the model of the electric machine, developed the test cases and generated all the results, which involved analyzing the results in collaboration with the other authors. Thomas Rylander wrote the paper with input from the other authors.

5.5 Paper E

Simon Stenmark, Thomas Rylander, Tomas McKelvey
Neural Networks for the Estimation of Low-Order Statistical Moments of a Stochastic Dielectric
2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC),

Glasgow, United Kingdom, pp. 1–6.

©2021 IEEE DOI: 10.1109/I2MTC50364.2021.9459996 .

Simon Stenmark created the data sets, implemented the methods, tuned the hyper-parameters, trained the neural networks and generated all results in the paper, which involved analyzing the results in collaboration with the other authors. All co-authors jointly developed the feature extraction scheme. **Simon Stenmark** and Thomas Rylander co-wrote the paper with input from Tomas McKelvey.

Summary and Concluding Remarks

Machine learning has revolutionized many technological fields, where today's powerful hardware and recent developments in data-driven algorithms open up exciting new possibilities to approach problems that are difficult to handle with conventional physics-based methods. However, learning-based methods typically require large amounts of data to train. As a powerful alternative, hybrid models that combine learning-based methods with conventional physics-based methods benefit from the strengths of both types of methods. In this thesis, we explore possibilities to combine and complement the methods of Computational Electromagnetics (CEM) with learning-based methods, with applications in computationally efficient surrogate modeling and methods for inline material characterization.

The wide array of CEM methods can treat many types of electromagnetic problems. However, a downside of most CEM methods is that they are computationally demanding, especially for three-dimensional problems. In the first part of the thesis, we present Gradient-Informed Attentive Normalization Training (GIANT). GIANT is a novel neural network training framework that allows us to efficiently train very deep fully-connected neural networks (FCNNs) that we use as accurate and computationally efficient surrogate models

for CEM methods. The first component of GIANT is Attentive Normalization (AttNorm). AttNorm consists of normalization layers that are updated by means of a novel update procedure, where the normalization parameters are updated intermittently throughout the training by means of a novel low-variance preserving scaling function. After the update, the weight-bias parameters of the FCNN are updated such that the output for any given input remains unchanged. AttNorm decreases problems that are associated with the training of deep FCNNs and allows for efficient training of very deep FCNNs. The second component of GIANT is gradient-informed training, where we use a loss function that is a linear combination of the misfit in (i) the output of the FCNN and (ii) the derivatives of the outputs of the FCNN with respect to its inputs. By including the misfit of the derivatives in the training of the neural networks, the number of samples required for training is substantially reduced. For the reciprocal microwave problems that we consider, we use continuum sensitivity analysis to compute these derivatives at a very low computational cost.

In the second part of the thesis, we develop and evaluate inline material characterization methods. Most conventional methods for the electromagnetic characterization of materials are offline methods, in which a sample of the material under test is removed and placed in a specialized measurement device. In applications where continuous monitoring of material properties is of interest, offline material characterization methods are of limited use. In particular, we develop and evaluate a method for the characterization of the effective permeability of the laminations of an electric machine. The method is based on a computationally efficient hybrid model of the machine, which consists of three main parts: (i) a one-dimensional transmission-line (TL) model along the machine axis; (ii) two-dimensional models of the cross-section of the machine, which are based on the finite element method (FEM); and (iii) circuit models that are used to represent external circuits and the magnetic couplings of the end windings of the machine. Then, we apply a Bayesian estimation framework to the model and obtain a maximum a posteriori (MAP) estimator for the sought parameters. The MAP estimator allows us to include prior knowledge about the electric and magnetic properties of the machine in the form of a prior distribution. We also develop an auto-calibration method that simultaneously determines (i) a set of unknown calibration parameters of a measurement system and (ii) the mean permittivity of an unknown medium

under test. We assume that we have access to a set of a-priori characterization observations to which we compare uncalibrated scattering parameter observations for the medium under test. As all characterization observations are gathered in advance, the method is well-suited for in-line measurement applications with limited access to the measurement system.

We demonstrate the methods on several applications. The first two applications consider a type of microwave measurement device intended for an inhomogeneous dielectric medium transported through a metal pipe [99], [104]. The test-problem consists of a circular cavity connected to a number of parallel-plate waveguides, where the circular cavity contains an inhomogeneous dielectric. As the first application for this test-problem, we use a GIANT network to model the device and estimate high-dimensional histograms, which are useful in detection problems [99]. We demonstrate that using the CEM method to train a GIANT network provides significant computational savings compared to using the CEM method to directly populate the histograms. As the second application for this test-problem, we apply the auto-calibration method to determine the mean permittivity of the inhomogeneous dielectric, given observations of the scattering parameters that are corrupted by unknown amplification factors. We demonstrate the robustness of the method for several levels of measurement noise and stochastic perturbations of the permittivity. For the third application, we train a GIANT network to model the frequency response of an H-plane waveguide filter. We use the GIANT network to optimize the geometry of the filter to achieve a band-pass filter that meets the given filter specifications and is robust to perturbations of the geometry within typical manufacturing tolerances. Here, we benefit from the ability to massively parallelize the evaluation of the GIANT network, which allows us to thoroughly explore the design space to find a good starting guess for the optimization algorithm. Fourth, we apply the Bayesian estimation procedure to the stator of an electric motor. In addition to the effective permeability of the laminations, we also estimate properties related to the dielectric materials and magnetic couplings of the end windings of the stator. We determine the parameters of the prior distribution using a combination of low-frequency measurements and simple analytical models. The method is evaluated on both synthetic and experimental data and is found to yield models that are accurate for frequencies up to 50 MHz, which makes it suitable for applications in Electromagnetic Compatibility (EMC).

In conclusion, the methods in this thesis provide tools for computationally efficient surrogate modeling and pave the way for inline methods for quality control in an industrial context. While at first glance, the many methods developed throughout this thesis may, to some extent, appear disjointed; they all include tuning the parameters of a model so that it matches observed data. For learning-based surrogate models, the parameters correspond to abstract weights of a generic function; while for the material characterization methods, the parameters correspond to physically meaningful material parameters. The many similarities between the methods provide the opportunity for further hybridization and novel combinations of methods for the research community to explore. For example, parameter estimation methods that exploit computationally efficient surrogate models may present opportunities for significant computational savings. By continuing to explore the domain between learning-based and physics-based methods, we can benefit from the rapid developments in machine learning, while still being able to exploit the vast amount of experience and knowledge in the research community about the physics of electromagnetic waves.

References

- [1] M. S. Lundstrom and M. A. Alam, “Moore’s law: The journey ahead,” *Science*, vol. 378, no. 6621, pp. 722–723, Nov. 2022.
- [2] K. S. Kim et al., “The future of two-dimensional semiconductors beyond Moore’s law,” *Nat. Nanotechnol.*, vol. 19, no. 7, pp. 895–906, Jul. 2024.
- [3] M. Morsy, F. Znid, and A. Farraj, “A critical review on improving and moving beyond the 2 nm horizon: Future directions and impacts in next-generation integrated circuit technologies,” *Mater. Sci. Semicond. Process.*, vol. 190, May 2025, Art. no. 109376.
- [4] Y. Sherry and N. C. Thompson, “How fast do algorithms improve? [Point of View],” *Proc. IEEE*, vol. 109, no. 11, pp. 1768–1777, Nov. 2021.
- [5] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. New York, NY: Wiley, 1999.
- [6] J.-M. Jin, *Theory and Computation of Electromagnetic Fields*. Hoboken, NJ: John Wiley & Sons, 2015.
- [7] J.-M. Jin, *The Finite Element Method in Electromagnetics*, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2014.
- [8] P. Monk, *Finite Element Methods for Maxwell’s Equations*. Oxford, England: Oxford University Press, 2003.
- [9] T. Rylander, A. Bondeson, and P. Ingelström, *Computational Electromagnetics*, 2nd ed. New York, NY: Springer, 2013.

- [10] J.-C. Nédélec, “Mixed finite elements in $\mathbb{R}^3$,” *Numer. Math.*, vol. 35, no. 3, pp. 315–341, 1980.
- [11] R. D. Graglia and A. F. Peterson, *Higher-Order Techniques in Computational Electromagnetics*. Raleigh, NC: SciTech Publishing, Aug. 2015.
- [12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, England: Cambridge University Press, 2007.
- [13] W. C. Chew, J.-M. Jin, E. Michielssen, and J. M. Song, *Fast and Efficient Algorithms in Computational Electromagnetics*. Boston, MA: Artech House, 2001.
- [14] W. C. Gibson, *The Method of Moments in Electromagnetics*, 3rd ed. Boca Raton, FL: CRC Press, 2022.
- [15] G. Kristensson, *Scattering of Electromagnetic Waves by Obstacles*. Raleigh, NC: SciTech Publishing, 2016.
- [16] S. Rao, D. Wilton, and A. Glisson, “Electromagnetic scattering by surfaces of arbitrary shape,” *IEEE Trans. Antennas Propag.*, vol. 30, no. 3, pp. 409–418, 1982.
- [17] P. Silvester and M.-S. Hsieh, “Finite-element solution of 2-dimensional exterior-field problems,” *Proc. Inst. Electr. Eng.*, vol. 118, no. 12, pp. 1743–1747, 1971.
- [18] B. H. McDonald and A. Wexler, “Finite-element solution of unbounded field problems,” *IEEE Trans. Microw. Theory Tech.*, vol. 20, no. 12, pp. 841–847, 1972.
- [19] M. M. Botha and J.-M. Jin, “On the variational formulation of hybrid finite element-boundary integral techniques for electromagnetic analysis,” *IEEE Trans. Antennas Propag.*, vol. 52, no. 11, pp. 3037–3047, 2004.
- [20] N. Engheta, W. Murphy, V. Rokhlin, and M. Vassiliou, “The fast multipole method (FMM) for electromagnetic scattering problems,” *IEEE Trans. Antennas Propag.*, vol. 40, no. 6, pp. 634–641, Jun. 1992.
- [21] R. Coifman, V. Rokhlin, and S. Wandzura, “The fast multipole method for the wave equation: A pedestrian prescription,” *IEEE Antennas Propag. Mag.*, vol. 35, no. 3, pp. 7–12, Jun. 1993.

-
- [22] A. Taflové and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed. Boston, MA: Artech House, 2005.
 - [23] K. Yee, “Numerical solution of initial boundary value problems involving Maxwell’s equations in isotropic media,” *IEEE Trans. Antennas Propag.*, vol. 14, no. 3, pp. 302–307, 1966.
 - [24] J.-P. Bérenger, *Perfectly Matched Layer (PML) for Computational Electromagnetics*. Cham, Switzerland: Springer, 2025.
 - [25] A. C. Cangellaris and D. B. Wright, “Analysis of the numerical error caused by the stair-stepped approximation of a conducting boundary in FDTD simulations of electromagnetic phenomena,” *IEEE Trans. Antennas Propag.*, vol. 39, no. 10, pp. 1518–1525, 1991.
 - [26] I. Zagorodnov, R. Schuhmann, and T. Weiland, “A uniformly stable conformal FDTD-method in cartesian grids,” *Int. J. Numer. Model.*, vol. 16, no. 2, pp. 127–141, Mar. 2003.
 - [27] J. Wings and T. Rylander, “Higher-order brick-tetrahedron hybrid method for maxwell’s equations in time domain,” *J. Comput. Phys.*, vol. 321, pp. 698–707, 2016.
 - [28] A. Forrester, A. Sobester, and A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*. Chichester, England: John Wiley & Sons, 2008.
 - [29] J. Bandler, R. Biernacki, S. H. Chen, P. Grobelny, and R. Hemmers, “Space mapping technique for electromagnetic optimization,” *IEEE Trans. Microw. Theory Tech.*, vol. 42, no. 12, pp. 2536–2544, Dec. 1994.
 - [30] J. Bandler et al., “Space mapping: The state of the art,” *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 1, pp. 337–361, Jan. 2004.
 - [31] P. Jiang, Q. Zhou, and X. Shao, *Surrogate Model-Based Engineering Design and Optimization*. New York, NY: Springer, 2020.
 - [32] A. I. J. Forrester and A. J. Keane, “Recent advances in surrogate-based optimization,” *Prog. Aerosp. Sci.*, vol. 45, no. 1, pp. 50–79, Jan. 2009.
 - [33] C. Roy and K. Wu, “A review of electromagnetics-based microwave circuit design optimization,” *IEEE Microw. Mag.*, vol. 25, no. 7, pp. 16–40, Jul. 2024.

- [34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.
- [35] A. Vaswani et al., “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [36] C.-F. R. Chen, Q. Fan, and R. Panda, “CrossViT: Cross-attention multi-scale vision transformer for image classification,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 357–366.
- [37] Q. Wang et al., “Learning deep transformer models for machine translation,” in *Proc. 57th Annu. Meet. Assoc. Comput. Linguistics (ACL)*, Florence, Italy, Jul. 2019, pp. 1810–1822.
- [38] OpenAI, *ChatGPT*, <https://chatgpt.com/>, Accessed: 2026-09-09.
- [39] Anthropic, *Claude*, <https://claude.ai/>, Accessed: 2026-09-09.
- [40] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.
- [41] Y. Chen, Y.-B. Tian, Z. Qiang, and L. Xu, “Optimisation of reflection coefficient of microstrip antennas based on KBNN exploiting GPR model,” *IET Microw. Antennas Propag.*, vol. 12, no. 4, pp. 602–606, Mar. 2018.
- [42] Z. Ye, W. Shao, X. Ding, B.-Z. Wang, and S. Sun, “Knowledge-based neural network for multiphysical field modeling,” *IEEE Trans. Microw. Theory Tech.*, vol. 71, no. 5, pp. 1967–1976, May 2023.
- [43] F. Feng, C. Zhang, J. Ma, and Q.-J. Zhang, “Parametric modeling of EM behavior of microwave components using combined neural networks and pole-residue-based transfer functions,” *IEEE Trans. Microw. Theory Tech.*, vol. 64, no. 1, pp. 60–77, Jan. 2016.
- [44] D. M. Pozar, *Microwave Engineering*, 4th ed. Hoboken, NJ: John Wiley & Sons, 2012.
- [45] T. Rylander, P. Hashemzadeh, and M. Viberg, “Reconstruction of metal protrusion on flat ground plane,” *IET Microw. Antennas Propag.*, vol. 4, pp. 1746–1755, 2010.
- [46] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006, vol. 4.

-
- [47] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
 - [48] A. D. Rasamoelina, F. Adjailia, and P. Sinčák, “A review of activation function for artificial neural network,” in *Proc. IEEE 18th World Symp. Appl. Mach. Intell. Inform. (SAMI)*, Jan. 2020, pp. 281–286.
 - [49] P. Ramachandran, B. Zoph, and Q. V. Le, *Searching for activation functions*, arXiv:1710.05941 [cs], Oct. 2017.
 - [50] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, *Self-normalizing neural networks*, arXiv:1706.02515 [cs], Sep. 2017.
 - [51] J. Jin, C. Zhang, F. Feng, W. Na, J. Ma, and Q.-J. Zhang, “Deep neural network technique for high-dimensional microwave modeling and applications to parameter extraction of microwave filters,” *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 10, pp. 4140–4155, Oct. 2019.
 - [52] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein, “On the expressive power of deep neural networks,” in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, Jul. 2017, pp. 2847–2854.
 - [53] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proc. 13th Int. Conf. Artif. Intell. Stat. (AISTATS)*, May 2010, pp. 249–256.
 - [54] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv:1412.6980 [cs], Jan. 2017.
 - [55] H. H. Tan and K. H. Lim, “Vanishing gradient mitigation with deep learning neural network optimization,” in *Proc. 7th Int. Conf. Smart Comput. Commun. (ICSCC)*, 2019, pp. 1–4.
 - [56] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, arXiv:1502.03167 [cs], Mar. 2015.
 - [57] J. L. Ba, J. R. Kiros, and G. E. Hinton, *Layer normalization*, arXiv:1607.06450 [stat.ML], Jul. 2016.
 - [58] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

- [59] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, *Understanding deep learning requires rethinking generalization*, arXiv:1611.03530 [cs], Feb. 2017.
- [60] W. M. Czarnecki, S. Osindero, M. Jaderberg, G. Świrszcz, and R. Pascanu, *Sobolev training for neural networks*, arXiv:1706.04859 [cs], Jul. 2017.
- [61] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, “Automatic differentiation in machine learning: A survey,” *J. Mach. Learn. Res.*, vol. 18, no. 153, pp. 1–43, 2018.
- [62] C.-T. Chen, *Linear System Theory and Design*. New York, NY: Holt, Rinehart and Winston, 1984, vol. 301.
- [63] A. Molina, P. Schramowski, and K. Kersting, *Padé activation units: End-to-end learning of flexible activation functions in deep networks*, arXiv:1907.06732 [cs], Feb. 2020.
- [64] O. Keleş and A. M. Tekalp, “Padé neurons for efficient neural models,” *IEEE Trans. Image Process.*, vol. 35, pp. 1508–1520, 2026.
- [65] P. Feldmann and R. W. Freund, “Efficient linear circuit analysis by Padé approximation via the Lanczos process,” *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 14, no. 5, pp. 639–649, 1995.
- [66] C. C. Chou and J. E. Schutt-Aine, “Equivalent circuit synthesis of multiport S parameters in pole-residue form,” *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 11, no. 11, pp. 1971–1979, Nov. 2021.
- [67] T. McKelvey and M. Gibanica, “FSID - a frequency weighted MIMO frequency domain identification and rational matrix approximation method for Python, Julia and Matlab,” *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 403–408, Jan. 2021.
- [68] B. Gustavsen and A. Semlyen, “Rational approximation of frequency domain responses by vector fitting,” *IEEE Trans. Power Deliv.*, vol. 14, no. 3, pp. 1052–1061, Jul. 1999.
- [69] O. Farle, V. Hill, P. Ingelström, and R. Dyczij-Edlinger, “Multi-parameter polynomial order reduction of linear finite element models,” *Math. Comput. Model. Dyn. Syst.*, vol. 14, no. 5, pp. 421–434, Oct. 2008.

-
- [70] I. Griva, S. G. Nash, and A. Sofer, *Linear and Nonlinear Optimization*. Philadelphia, PA: SIAM, Jan. 2009.
 - [71] L. E. Scales, *Introduction to Non-Linear Optimization*. Heidelberg, Germany: Springer-Verlag, 1985.
 - [72] K. O. Lye, S. Mishra, D. Ray, and P. Chandrasekhar, “Iterative surrogate model optimization (ISMO): An active learning algorithm for PDE constrained optimization with deep neural networks,” *Comput. Methods Appl. Mech. Eng.*, vol. 374, Feb. 2021, Art. no. 113575.
 - [73] J. Zhao, Y. Yang, X. Chen, and F. Gao, “An iterative modeling and trust-region optimization method for batch processes,” *Ind. Eng. Chem. Res.*, vol. 54, no. 12, pp. 3186–3199, Apr. 2015.
 - [74] N. Metropolis and S. Ulam, “The Monte Carlo method,” *J. Am. Stat. Assoc.*, vol. 44, no. 247, pp. 335–341, Sep. 1949.
 - [75] S. Burhenne, D. Jacob, and G. P. Henze, “Sampling based on Sobol’ sequences for Monte Carlo techniques applied to building simulations,” in *Proc. Building Simulation 2011*, vol. 12, 2011, pp. 1816–1823.
 - [76] M. D. McKay, R. J. Beckman, and W. J. Conover, “A comparison of three methods for selecting values of input variables in the analysis of output from a computer code,” *Technometrics*, vol. 42, no. 1, pp. 55–61, Feb. 2000.
 - [77] Y. Zhang, N. H. Kim, and R. T. Haftka, “General-surrogate adaptive sampling using interquartile range for design space exploration,” *J. Mech. Des.*, vol. 142, Nov. 2019, Art. no. 051402.
 - [78] J. Larson, M. Menickelly, and S. M. Wild, “Derivative-free optimization methods,” *Acta Numer.*, vol. 28, pp. 287–404, 2019.
 - [79] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA: MIT Press, 1992.
 - [80] R. Eberhart, J. Kennedy, et al., “Particle swarm optimization,” in *Proc. IEEE Int. Conf. Neural Netw. (ICNN)*, Perth, Australia, vol. 4, 1995, pp. 1942–1948.
 - [81] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.

- [82] A. Ruszczyński, *Nonlinear Optimization*. Princeton, NJ: Princeton University Press, 2011.
- [83] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY: Springer, 2006.
- [84] M. Norgren and T. Takenaka, “Full Newton method for inverse transmission line problems, utilising explicit second order derivatives,” *Inverse Probl. Sci. Eng.*, vol. 15, no. 8, pp. 827–853, Dec. 2007.
- [85] B. Jiang, H. Wang, and S. Zhang, *An optimal high-order tensor method for convex optimization*, arXiv:1812.06557 [math.OC], Apr. 2020.
- [86] G. Gundersen and T. Steihaug, “On large-scale unconstrained optimization problems and higher order methods,” *Optim. Methods Softw.*, vol. 25, no. 3, pp. 337–358, Jun. 2010.
- [87] B. Meng, J. Booske, and R. Cooper, “Extended cavity perturbation technique to determine the complex permittivity of dielectric materials,” *IEEE Trans. Microw. Theory Tech.*, vol. 43, no. 11, pp. 2633–2636, 1995.
- [88] M. D. Deshpande, C. J. Reddy, P. I. Tiemsin, and R. Cravey, “A new approach to estimate complex permittivity of dielectric materials at microwave frequencies using waveguide measurements,” *IEEE Trans. Microw. Theory Tech.*, vol. 45, no. 3, pp. 359–366, 1997.
- [89] D. V. Blackham and R. D. Pollard, “An improved technique for permittivity measurements using a coaxial probe,” *IEEE Trans. Instrum. Meas.*, vol. 46, no. 5, pp. 1093–1099, 1997.
- [90] V. Franco and B. Dodrill, Eds., *Magnetic Measurement Techniques for Materials Characterization*. Cham, Switzerland: Springer, 2021.
- [91] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1993.
- [92] D. Sjöberg and C. Larsson, “Cramér–Rao bounds for determination of permittivity and permeability in slabs,” *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 11, pp. 2970–2977, 2011.
- [93] A. Abou-Elyazied Abdallah, G. Crevecœur, and L. Dupré, “A priori experimental design for inverse identification of magnetic material properties of an electromagnetic device using uncertainty analysis,” *COMPEL*, vol. 31, no. 3, pp. 972–984, 2012.

-
- [94] X. Liao and L. Carin, “Application of the theory of optimal experiments to adaptive electromagnetic-induction sensing of buried targets,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 961–972, 2004.
 - [95] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. New York, NY: John Wiley & Sons, 2001.
 - [96] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, “Parameter space compression underlies emergent theories and predictive models,” *Science*, vol. 342, no. 6158, pp. 604–607, Nov. 2013.
 - [97] M. K. Transtrum, B. B. Machta, and J. P. Sethna, “Geometry of nonlinear least squares with applications to sloppy models and optimization,” *Phys. Rev. E*, vol. 83, no. 3 Pt 2, Mar. 2011, Art. no. 036701.
 - [98] E. Nyfors and P. Vainikainen, *Industrial Microwave Sensors*. London, England: Artech House, 1989.
 - [99] J. Nohlert, T. Rylander, and T. McKelvey, “Microwave measurement system for detection of dielectric objects in powders,” *IEEE Trans. Microw. Theory Tech.*, vol. 64, no. 11, pp. 3851–3863, 2016.
 - [100] Agilent Technologies, “AN 1287-3 applying error correction to network analyzer measurements,” Agilent Technologies, Tech. Rep., 2002.
 - [101] K. Ramya, J. Gopalakrishnan, B. Chokkalingam, R. Verma, and L. Mihet-Popa, “A comprehensive evaluation and assessment practices of electromagnetic interferences in electric vehicle,” *IEEE Access*, vol. 13, pp. 40 520–40 560, 2025.
 - [102] C. R. Paul, R. C. Scully, and M. A. Steffka, *Introduction to Electromagnetic Compatibility*, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2022.
 - [103] A. Kampker, H. H. Heimes, B. Dorn, F. Brans, and C. Stäck, “Challenges of the continuous hairpin technology for production techniques,” *Energy Rep.*, vol. 9, pp. 107–114, Mar. 2023.
 - [104] J. Wings, L. Cerullo, T. Rylander, T. McKelvey, and M. Viberg, “Compressed sensing for the detection and positioning of dielectric objects inside metal enclosures by means of microwave measurements,” *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 462–476, 2017.

