EM-Driven Miniaturization of High-Frequency Structures through Constrained Optimization

Marzieh Mahrokh

Doctor of Philosophy

February 2023

Electrical Engineering

Reykjavík University

PhD Dissertation

EM-Driven Miniaturization of High-Frequency Structures through Constrained Optimization

Dissertation of 30 ECTS credits submitted to the Department of Engineering
at Reykjavík University in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy (PhD) in Electrical Engineering

February 2023

Thesis Supervisors:

Slawomir Koziel, Professor
Reykjavík University, Reykjavík, Iceland

Anna Pietrenko-Dabrowska, Associate Professor
Gdansk University of Technology, Poland

Thesis Committee:

Slawomir Koziel, Professor
Reykjavík University, Reykjavík, Iceland

Anna Pietrenko-Dabrowska, Associate Professor
Gdansk University of Technology, Poland

Ágúst Valfells, Dean of Department of Engineering
Reykjavík University, Reykjavík, Iceland

Ubaid Ullah, Assistant Professor
Al-Ain University, UAE

Thesis Examiner:

Adam Narbudowicz, Senior Research Fellow
Trinity College Dublin, Ireland

EM-Driven Miniaturization of High-Frequency Structures through Constrained Optimization

Short title: Miniaturization of High-Frequency Structures

EM-Driven Miniaturization of High-Frequency Structures through Constrained Optimization

Marzieh Mahrokh

February 2023

Abstract

The trends afoot for miniaturization of high-frequency electronic devices require integration of active and passive high-frequency circuit elements within a single system. This high level of accomplishment not only calls for a cutting-edge integration technology but also necessitates accommodation of the corresponding circuit components within a restricted space in applications such as implantable devices, internet of things (IoT), or 5G communication systems. At the same time, size reduction does not remain the only demand. The performance requirements of the abovementioned systems form a conjugate demand to that of the size reduction, yet with a contrasting nature. A compromise can be achieved through constrained numerical optimization, in which two kinds of constrains may exist: equality and inequality ones. Still, the high cost of electromagnetic-based (EM-based) constraint evaluations remains an obstruction. This issue can be partly mitigated by implicit constraint handling using the penalty function approach. Nevertheless, securing its performance requires expensive guess-work-based identification of the optimum setup of the penalty coefficients. An additional challenge lies in allocating the design within or in the vicinity of a thin feasible region corresponding to equality constraints. Furthermore, multimodal nature of constrained miniaturization problems leads to initial design dependency of the optimization results. Regardless of the constraint type and the corresponding treatment techniques, the computational expenses of the optimization-based size reduction persist as a main challenge. This thesis attempts to address the abovementioned issues specifically pertaining to optimization-driven miniaturization of high frequency structures by developing relevant algorithms in a proper sequence. The first proposed approach with automated adjustment of the penalty functions is based on the concept of sufficient constraint violation improvement, thereby eliminating the costly initial trial-and-error stage for the identification of the optimum setup of the penalty factors. Another introduced approach, i.e., correction-based treatment of the equality constraints alleviates the difficulty of allocating the design within a thin feasible region where designs satisfying the equality constraints reside. The next developed technique allows for global size reduction of high-frequency components. This approach not only eliminates the aforementioned multimodality issues, but also accelerates the overall global optimization process by constructing a dimensionality-reduced surrogate model over a pre-identified feasible region as compared to the complete parameter search space. Further to the latter, an optimization framework employing multi-resolution EM-model management has been proposed to address the high cost issue. The said technique provides nearly 50 percent average acceleration of the optimization-based miniaturization process. The proposed technique pivots upon a newly-defined concept of model-fidelity control based on a combination of algorithmic metrics, namely convergence status and constraint violation level. Numerical validation of the abovementioned algorithms has also been provided using an extensive set of high-frequency benchmark structures. To the best of the author´s knowledge, the presented study is the first investigation of this kind in the literature and can be considered a contribution to the state of the art of automated high-frequency design and miniaturization.

The undersigned hereby certify that they recommend to the School of Science and Engineering at Reykjavík University for acceptance this Dissertation entitled **EM-Driven Miniaturization of High-Frequency Structures through Constrained Optimization** submitted by **Marzieh Mahrokh** in partial fulfilment of the requirements for the degree of **Doctor of Philosophy (PhD) in Electrical Engineering**

...................▬▬▬▬▬▬▬▬▬▬▬▬
Date: 30-02-2023

...............▬▬▬▬▬▬▬▬▬▬▬▬▬▬...
Slawomir Koziel, Professor, Supervisor
Reykjavík University, Iceland

...............▬▬▬▬▬▬▬▬▬▬▬▬▬...
Anna Pietrenko-Dabrowska, Associate Professor, Supervisor
Gdansk University of Technology, Poland

...............▬▬▬▬▬▬▬▬▬▬▬▬▬...
Ágúst Valfells, Department of Engineering
Reykjavík University, Iceland

...............▬▬▬▬▬▬▬▬▬▬▬▬▬...
Ubaid Ullah, Assistant Professor
Al-Ain University, UAE

...............▬▬▬▬▬▬▬▬▬▬▬▬...
Adam Narbudowicz, Senior Research Fellow
Trinity College Dublin, Ireland

The undersigned hereby grants permission to the Reykjavík University Library to reproduce single copies of this Dissertation entitled **EM-Driven Miniaturization of High-Frequency Structures through Constrained Optimization** and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the Dissertation, and except as herein-before provided, neither the Dissertation nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Date: 30-02-2023

...................━━━━━━━━━━━━━━━━━━━━━...

Marzieh Mahrokh
Doctor of Philosophy

To my beloved parents, siblings, and friends

For their endless love, support and encouragement

Acknowledgments

The accomplishment of this undertaking would not have been possible without the assistance and participation of many people; the names of only a few of them are given a particular mention here.

First and foremost, I would like to acknowledge the constant support and guidance offered by Prof. Sławomir Kozieł, the supervisor of this doctoral thesis. His regular, invaluable discussions and iterative feedbacks on my research progress did play the leading part in the in time accomplishment of this work.

I would also like to express my gratitude to Prof. Anna Pietrenko-Dabrowska, the co-supervisor of this work, for her helpful suggestions and feedbacks throughout my PhD studies.

I would like to thank Dassault Systemes, France, for making CST Microwave Studio available.

Finally, it is time to express my wholehearted gratitude to my family and beloved friends, Virginia, Breno, Qazal, Delaram, Iban, Sibba, and Bruno, for their unconditional love and support throughout my vulnerable times.

Preface

This dissertation is the original work of the author, Marzieh Mahrokh. A portion of the thesis (Chapter 4 through Chapter 8) is based upon the journal papers, published during the Ph.D. study. Note that these papers appear as per their online version in the chapters. The overall list of journal papers throughout this study is listed below.

Journal papers:

1. M. Mahrokh and S. Koziel, "Optimization-based antenna miniaturization using adaptively-adjusted penalty factors," *Electronics*, vol. 10, no. 15, paper no. 1751, 2021.

2. M. Mahrokh and S. Koziel, "Explicit size-reduction of circularly polarized antennas through constrained optimization with penalty factor adaptation," *IEEE Access*, vol. 9, pp. 132390-132396, 2021.

3. M. Mahrokh and S. Koziel, "Improved-efficacy EM-based antenna miniaturization by multi-fidelity simulations and objective function adaptation," *Energies*, vol. 15, no. 2, paper no. 403, 2021.

4. S. Koziel, A. Pietrenko-Dabrowska, and M. Mahrokh, „On decision-making strategies for improved-reliability size reduction of microwave passives: intermittent correction of equality constraints and adaptive handling of inequality constraints," *Knowledge-Based Syst.*, vol. 255, paper no. 109745, 2022.

5. A. Pietrenko-Dabrowska, S. Koziel, and M. Mahrokh, „Optimization-based high-frequency circuit miniaturization through implicit and explicit constraint handling: recent advances," *Energies*, vol. 15, no. 19, paper no. 6955, 2022.

6. S. Koziel, A. Pietrenko-Dabrowska, and M. Mahrokh, „Globalized simulation-driven miniaturization of microwave circuits by means of dimensionality-reduced constrained surrogates," *Scientific Reports*, vol. 12, paper no. 16418, 2022.

Contents

List of Figures

List of Acronyms

ANN   Artificial Neural Network

AR   Axial Ratio

ATL   Artificial Transmission Line

AVA   Antipodal Vivaldi Antenna

BFGS   Broyden-Fletcher-Goldfarb-Shanno

BW   Bandwidth

CAD   Computer Aided Design

CP   Circularly Polarized

CST   Computer Simulation Technology

EA   Evolutionary Algorithms

EM   Electromagnetic

ES   Evolution Strategy

GA   Genetic Algorithm

GP   Gauss Process

HMSIW Half-Mode Surface Integrated Waveguide

IC   Integrated Circuit

IoT   Internet of Things

KKT   Karush-Kuhn-Tucker

NN   Neural Net

PSO   Particle Swarm Optimization

RBF   Radial Basis Function

RF   Radio Frequency

RFID   Radio Frequency Identification

RSA   Response Surface Approximation

SBO   Surrogate-Based Optimization

SISS   Step Impedance Shunt Stub

SIW   Surface Integrated Waveguide

TR   Trust Region

List of Symbols

$A(\boldsymbol{x})$	Circuit footprint area
AR	Axial ratio
BW	Bandwidth
$c_{eq}(\boldsymbol{x})$	Equality constraint penalty function
$c_{ineq}(\boldsymbol{x})$	Inequality constraint penalty function
Cov	Covariance matrix
dB	Decibels
f	Frequency
f_c	Centre frequency
f_l	Lower frequency
f_u	Upper frequency
$f(\boldsymbol{x})$	Objective function
GHz	Gigahertz, 10^9 Hertz
$\boldsymbol{H}(\boldsymbol{x})$	Hessian matrix
$i^2(\boldsymbol{x})$	Variance
$I(\boldsymbol{x})$	Localized variations (Kriging)
J	Jacobian matrix
l	Response level
$LB(\boldsymbol{x})$	Statistical lower bound
n	Number of design variables
n_c	Total number of constraints
P	Population (in metaheuristic algorithms)
r	Gain ratio for trust-region framework
$R(\boldsymbol{x})$	Response vector of an EM-simulation model
$\boldsymbol{R}$	Correlation matrix
$\boldsymbol{R}_c(\boldsymbol{x})$	Low-fidelity EM-simulated response
$\boldsymbol{R}_f(\boldsymbol{x})$	High-fidelity EM-simulated response
$\boldsymbol{R}_s(\boldsymbol{x})$	Surrogate model response
$\hat{s}(\boldsymbol{x})$	Surrogate prediction of the objective function
$S_{eq}(\boldsymbol{x})$	Equality constraint
$S_{ineq}(\boldsymbol{x})$	Inequality constraint
S_{11}	Reflection coefficient
$S_{pq}(\boldsymbol{x})$	Scattering parameters
$U_p(\boldsymbol{x})$	Objective function
$v(\boldsymbol{x})$	Basis function (polynomial regression)
$\boldsymbol{w}$	Basis function centre (radial basis function)
$\boldsymbol{x}$	Vector of design variables
$\boldsymbol{x}^*$	Optimum design
X	Design space
$\boldsymbol{X}^+$	Pseudoinverse of $\boldsymbol{X}$
$\boldsymbol{\alpha}$	Vector of model parameters (polynomial regression)
β_{eq}	Penalty coefficient (equality constraint)
β_{ineq}	Penalty coefficient (inequality constraint)
ε_r	Relative permittivity
ζ	Lagrange multiplier
η	Lagrange multiplier

θ Model parameter (radial basis function)

λ Vector of unknown model parameters (Kriging)

σ Scaling coefficient (Gaussian function)

φ Radially symmetric function

Chapter 1

1 Introduction

1.1 Background

The continuous tendency of shrinking the size of the high-frequency electronic devices, has led to emerging of high-level system integration technologies aimed at solving size restrictive accommodation challenges. Active and passive high-frequency circuit components are employed in applications such as implantable medical devices [1]- [2], autonomous vehicles [3], RF ICs, or wireless communication systems [4]. The examples of the relevant space-occupying passive components include phase shifters [5], beamforming antenna arrays [6], artificial transmission lines [7], and couplers/dividers [8]. Innovative circuit-level modifications of the conventional design procedures have been successful in obtaining compact footprints through the incorporation of electromagnetic wave theory along with additional geometrical variations. Among popular circuit-level miniaturization techniques the incorporation of meandered traces [9], the use of slotted substrate integrated waveguide (SIW) [10], the introduction of half-mode SIW (HMSIW) evanescent mode [11], or the employment of shared partial patches [12] may be listed.

The aforementioned techniques provide reliable means of rendering compact structures based on solid grounds of fundamental electromagnetic wave theory. Nevertheless, their implementation may require a unique intermingling of design concepts for each case, with considerable geometrical alterations. Design complexity and manipulation of additional design parameters, being a result of topology modifications, remains a challenge. At the same time, a further challenge lies in the contrasting nature of the demands for system performance requirements with those for compact size. Considering the trade-off between the latter and the former, optimum arrangement of the numerous geometry parameters can be identified through constrained multi-objective local [13]- [14], or global search optimization methods [15], [16], [17].

Fulfilling the system performance requirements (electrical and field performance figures) embedded in the form of design constraints, results in algorithmic complexity. The penalty function approach [18], [19], [20], [21], [22] offers a convenient workaround by controlling the design constraints in an implicit manner. In [19], the penalty function approach was accompanied by an adaptive reflection coefficient acceptance threshold. As a result, the performance of the algorithm was improved both in terms of controlling the reflection coefficient constraint and the achievable size reduction rates as compared to the fixed-

threshold penalty function setup. Yet another algorithmic performance challenge lies in the nonlinearity of the penalty function due to the parameter sensitivity inconsistencies between the structure size and performance-related constraints. The issue has been addressed in [20] by introducing an objective relaxation scheme between size and reflection coefficient constraints. Other aspects concerning the time and computational expenses are improved by utilization of coarse-discretization EM models in [21], or the employment of variable-fidelity EM model in [22].

Each of the aforementioned variations of the penalty function approach provides specific performance enhancements as compared to the conventional version. At the same time, a common shortcoming resides in the fact that the performance of the optimization process is highly dependent on the setup of the penalty terms. Typically, the optimum arrangement of the penalty factors is conventionally identified through time and computationally expensive trial-and-error attempts. An additional challenge appears when applying the penalty function approach to specific high-frequency structures with either equality, or a combination of equality and inequality constraints. The corresponding individual feasible regions intersect over a very thin set of the parameter space resulting in exploration difficulties. The issue of appropriate handling of equality constraints is of great significance, as, in high-frequency design, this type of constraints is frequently met, and the examples of equality constraints include power split ratio or phase shift in power dividers/combiners, couplers, or phase shifters.

Many local optimization-based miniaturization techniques [23], [24], [25], [26], [27] are prone to multimodality and thus initial design dependency. Whereas their counterpart global methods [28]- [29] exhibit robustness to these issues. Conventional global optimization procedures including simulated annealing [30], or population-based metaheuristics (evolutionary strategies [31], and particle swarm optimization [32]) are likely to encounter a prohibitive computational time and burden of the evaluation of the expensive objective functions. An alternative approach is surrogate-based optimization (SBO) [13], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], where the burden is reduced by switching to a computationally cheaper surrogate model. In [33], the computational efficiency of expensive multi-objective PSO problems was significantly reduced by approximation of the Pareto front using a Gaussian process (GP) surrogate model. An acceleration of multi-objective microwave device design was proposed in [34] through the employment of nested kriging methodology for the identification of the Pareto front. A novel technique developed in [13] attempted to address the curse of dimensionality by the construction of a two-level kriging surrogate model, resulting in modelling efficiency within wide ranges of both the parameters, and the performance figures space. Nevertheless, techniques exploiting surrogate models face the problems on their own, such as the aforementioned curse of dimensionality due to which models' usefulness is restricted to components featuring few geometry parameters number of relatively narrow ranges. Considerable nonlinearity of responses of high-frequency devices poses yet another challenge. Moreover, the additional computational expenses associated with the increased number of geometry parameters in unconventional structures, remain a practical obstacle.

Regardless of either the specific optimization technique, or constraint type and their corresponding treatment, size reduction remains a time and computationally expensive task. A number of algorithmic acceleration techniques has been proposed to improve the computational efficiency of the optimization process. These include the incorporation of

adjoint sensitivities [46], [47], [48], [49], [50], [51], utilization of sparse sensitivity updates [52] and restricted sensitivity updates [53], or the employment of Broyden updating formulas [54]. Significant acceleration rates of up to forty [55] and sixty percent [56] have been achieved using the abovementioned techniques. An alternative method is the incorporation of variable-fidelity EM simulation models [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67]. Over forty percent CPU time savings were obtained using a set of discrete EM-model fidelities [61]. However, the quality of the final optimization results both in terms of constraint satisfaction, and achievable size reduction rates is highly dependent on a meticulous selection of the EM-model fidelities [65]- [66].

In light of the abovementioned performance aspects of the optimization-based miniaturization—specifically concerning fixed-geometry structures where no topology optimization is required—this thesis attempts to address the corresponding challenges of simulation-driven size reduction of high-frequency structures in the context of rigorous numerical methods. This includes the development of more efficient algorithms which are intended to advance the state of the art of high-frequency CAD.

The consecutive algorithms developed within this thesis form a proper sequence. The first technique, i.e., automated adjustment of the penalty functions, based on the concept of sufficient constraint violation improvement, eliminates the costly initial trial-and-error stage for the identification of the optimum setup of the penalty factors. Further, the correction-based treatment of the equality constraints alleviates the difficulty of allocating the design within a thin feasible region. Another developed technique for global size reduction of high-frequency components not only eliminates the multimodality issue, but also accelerates the overall global optimization process by constructing a dimensionality-reduced surrogate model over a pre-identified feasible region as compared to the complete parameter search space. Further to the latter, a multi-resolution EM-model management provides an average 43% acceleration of the optimization-based miniaturization process. The proposed technique pivots upon a newly-defined concept of model-fidelity control based on a combination of algorithmic metrics, namely convergence status, and constraint violation level. Numerical validation of the abovementioned algorithms has also been provided through an analysis of the results obtained for a set of high-frequency benchmark structures.

1.2   Thesis Contribution

The contributions of this thesis have been described in five journal publications that constitute Chapters 4 through 8. These serve as the overall record of the candidate, prepared during her Ph.D. study. Their common theme is the development of more efficient algorithms with the major focus on constrained optimization-based size reduction of high-frequency structures.

- **Paper #1:**   The contrasting natures of the demands for compact size and those for the electrical and field performance of the high-frequency structures can be mitigated through constrained numerical optimization. Although the implicit constraint handling provided by the penalty function approach reduces the high computational costs of EM-based constraint evaluations, the constraint control and size reduction performance aspects of the said approach are strongly dependent on the setup of the penalty factors, conventionally identified through costly trial-and-error efforts. This paper proposed a novel procedure featuring an automated penalty

factor adjustment throughout the optimization run. The adjustment process is based on a notion of sufficient constraint violation improvement. The numerical validations demonstrated that the proposed algorithm outperforms the fixed penalty function approach in terms of constraint satisfaction and achievable size reduction rates.

- o **Bibliographic Note:**
  M. Mahrokh and S. Koziel, "Optimization-based antenna miniaturization using adaptively-adjusted penalty factors," *Electronics*, vol. 10, no. 15, paper no. 1751, 2021.

- ▪ **Paper #2:** This study was focused on explicit size reduction of high-frequency structures with multiple design constraints. The appealing feature of the proposed procedure lies in the identification of the most appropriate penalty coefficient levels based on the monitored feasibility status and constraint violation levels of the consecutive designs in the course of the optimization process. The methodology was validated using two circularly polarized (CP) antenna structures. The dynamically adjusted penalty coefficients exercised a precise control over multiple design constraints while outperforming the manual penalty function setup in terms of the final miniaturization rates.

  - o **Bibliographic Note:**

    M. Mahrokh and S. Koziel, "Explicit size-reduction of circularly polarized antennas through constrained optimization with penalty factor adaptation," *IEEE Access*, vol. 9, pp. 132390-132396, 2021.

- ▪ **Paper #3:** This study was concerned with an additional challenge of applying the penalty function approach to specific high-frequency structures involving equality constraints. The corresponding feasible region being a very thin subset of the parameter space, results in impediments in the exploration of the feasible region boundary. The proposed optimization scheme alleviated this difficulty by a correction-based treatment of the equality constraints in conjunction with implicit handling of the inequality constraints using the adaptive penalty function approach. Numerical verification of the procedure was performed using four microstrip couplers involving power split ratio as the equality constraint. The obtained results demonstrated superior miniaturization rates along with a precise control of the design constraints as compared to the benchmark methods.

  - o **Bibliographic Note:**

    S. Koziel, A. Pietrenko-Dabrowska, and M. Mahrokh, „On decision-making strategies for improved-reliability size reduction of microwave passives: intermittent correction of equality constraints and adaptive handling of inequality constraints," *Knowledge-Based Syst.*, vol. 255, paper no. 109745, 2022.

- ▪ **Paper #4:** The multimodality issues inherent in constrained optimization-based miniaturization result in initial design dependency of the optimization results. At the

same time, conducting a global search for the identification of the global optima is both computationally time and cost prohibitive. The proposed global size reduction technique worked around the abovementioned issues by: (i) eliminating the multimodality issue by means of a global search, (ii) expediting the overall optimization process by adopting a dimensionality-reduced surrogate model over a pre-identified feasible region as compared to the complete parameter space search. The framework was benchmarked against gradient-based miniaturization, as well as nature-inspired optimization. The obtained results indicated superior size reduction rates along with reduced time and computational cost.

- o **Bibliographic Note:**

  S. Koziel, A. Pietrenko-Dabrowska, and M. Mahrokh, „Globalized simulation-driven miniaturization of microwave circuits by means of dimensionality-reduced constrained surrogates," *Scientific Reports*, vol. 12, paper no. 16418, 2022.

- **Paper #5:** The focus of this study was an acceleration of optimization-based miniaturization of high-frequency structures. To this end, a variable-fidelity EM-model fidelity management scheme was proposed. Therein, the model fidelity was continuously adjusted based upon a properly defined unified metric for the convergence-feasibility status of the design throughout the optimization process. The immediate benefit lied in the efficiency of the proposed acceleration procedure in terms of the achievable miniaturization rates, as well the quality of the constraint control. The proposed procedure was validated using five microstrip antennas including broadband and CP structures. The obtained results demonstrated an average 43% improvement in the CPU time saving as compared to the fixed-fidelity adaptive penalty function approach.

  - o **Bibliographic Note:**

    M. Mahrokh and S. Koziel, "Improved-efficacy EM-based antenna miniaturization by multi-fidelity simulations and objective function adaptation," *Energies*, vol. 15, no. 2, paper no. 403, 2021.

1.3   Thesis Outline

The remainder of the thesis is organized into eight chapters. Chapter 2 delivers a brief background and motivation for high-frequency circuit miniaturization, discusses circuit-level miniaturization techniques followed by optimization-based miniaturization, describes how circuit characteristics are included as constraints in the optimization problem, and finally provides details of the problem formulation for the optimization-based size reduction using the penalty function approach. Chapter 3 highlights the algorithmic tools employed in solving EM-driven miniaturization problems including local and global optimization methods, surrogate-based optimization and modelling techniques. In Chapter 4, the automated penalty function approach is described, and the employed algorithmic tools are discussed. A detailed discussion of the sufficient constraint violation improvement is also provided. Moreover, a statistical

analysis of the results is conducted to verify the efficiency of the proposed methodology both in terms of the achievable miniaturization rates, and satisfactory constraint control. Chapter 5 extends the previous design procedure, and introduces an automated penalty factor adjustment scheme for the explicit size reduction of high frequency structures with multiple constraints. The adjustment procedure involves monitoring of the constraint violation levels and the feasibility status of the consecutive designs in the course of the optimization process. The practical utility of the proposed procedure is illustrated through miniaturization of two CP antenna structures with constraints defined as maximum in-band axial ratio (AR), and reflection coefficient. Chapter 6 introduces a systematic and efficient approach to miniaturization of specific high-frequency structures involving equality constraints. The included algorithmic details provide a description of how the proposed approach overcomes the impediments in the exploration of the thin feasible region where the optimal solutions reside. Towards the end, a discussion about the application of the new procedure, as well as a comparison of its performance against the standard benchmark methods is provided. Chapter 7 addresses the multimodality issue inherent to constrained optimization-based miniaturization problems by presenting a surrogate-based global optimization procedure. Further to the later, an alleviation of the time and computational burden of the said global procedure is suggested through the employment of a dimensionality-reduced surrogate model, as well as proper incorporation of multi-fidelity models. In the latter part, a verification of the reliability and the efficacy of the proposed methodology is provided by proper benchmarking against gradient-based size reduction, as well as nature-inspired optimization frameworks. Finally, Chapter 8 focuses on the acceleration of optimization-based miniaturization problems. Therein, a description of a variable-fidelity EM-model management is provided. The underlying acceleration mechanism is described as a continuous adjustment of the model fidelity based on a proper unification of the defined metrics for the convergence status, and the feasibility status of the current design in the course of the optimization process. The significance of the proposed acceleration procedure is demonstrated through benchmarking using a number of single and multiple-constraint antenna structures.

Chapter 9 concludes the thesis by stating the main findings of the conducted research over the course of this Ph.D. work, as well as discussing the corresponding challenges for potential future research directions.

Chapter 2

2 Motivation and Literature Review

This chapter discusses the motivation for the thesis, including a practical necessity for and applications of high-frequency circuit miniaturization. A brief summary of circuit-level miniaturization techniques including the incorporation of geometrical modifications, and the employment of electromagnetic wave theory constitute a part of the discussion. Subsequently, the transition phase from circuit-level to optimization-based miniaturization and the details of corresponding problem formulation are described.

2.1 High-Frequency Circuit Miniaturization

Proliferation of the electronic systems has led to accommodation, limited power source, and maintenance issues within the emerging application areas including wireless communication networks, medical and healthcare systems, manufacturing industry, construction management, and home automation. The accommodation issue directly imposes size restrictions on the individual modules of the electronic devices, while the power loss can be reduced through elimination or shortening of the corresponding interconnection wirings or transmission lines. Integration of miniaturized system modules within a unified system not only helps further reduce the overall size, but also contributes to power loss reduction. An additional benefit lies in facilitation of the system maintenance and troubleshooting resulting from a reduced number of externally-connected individual modules. Figure .1 illustrates forty percent size reduction of multi-protocol integrated wireless module provided by Innophase Incorporation [68]. INP1013 is a smaller version of its counterpart INP1010, including a microcontroller, clocks, antenna and passive components; all integrated within a single module.
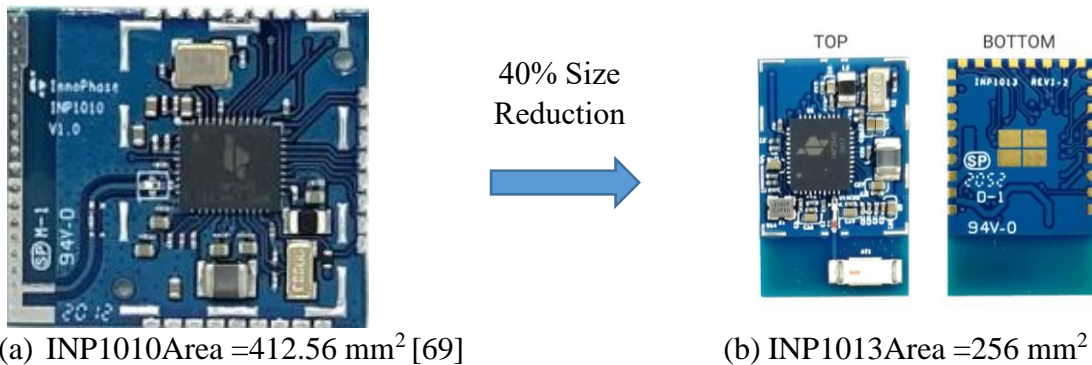

(a) INP1010Area $=412.56$ mm^2 [69] (b) INP1013Area $=256$ mm^2 [68]

Figure 2.1: 40 % size reduction of multi-protocol wireless integrated module; (a) INP1010: area $= 412.56$ mm^2, (b) INP1013: area $= 256$ mm^2 [68].

2.1.1 Miniaturization: A Practical Necessity

The emerging trends in size reduction of high-frequency electronic devices give rise to higher-level integration requirements of the system components. The resulting size restrictions demand not only compact size, but also a juxtaposition of active and passive high-frequency circuit elements within a single integrated system. This undertaking has led to the development of the state-of-the-art integration technologies in applications such as picosatellites [70], implantable medical devices [71], internet of things (IoT) [72], radio frequency identification (RFID) [73], or 5G communication systems [74]. Some of the corresponding voluminous passive components include power combiners [75], couplers/dividers [76], antenna arrays and their feeding networks [77], or filters and phase shifters [78]. While system performance requirements persist as a contrasting demand to that of compact size, size reduction becomes a practical necessity.

2.1.2 Circuit-Level Miniaturization Techniques

Realization of high-frequency circuit miniaturization may be carried out by resorting to alterations of the conventional designs. This can be achieved through the employment of electromagnetic wave theory along with additional degrees of freedom provided by geometrical modifications. Some of the relevant size reduction methods (referred to as circuit-level miniaturization techniques in this context) include the incorporation of rectangular corrugations in the flares of antipodal Vivaldi antenna (AVA) arrays [79], the insertion of parasitic patches in close proximity of the driven patch in broadband endfire antennas [80], the utilization of multimode SIW beamforming networks in millimeter-wave (MM-wave) endfire antenna arrays [81], the introduction of filtering structures with vertical slots in 5G base-station antennas [82], the incorporation of slotted SIW in wideband phase shifters [10], the utilization of half-mode SIW (HMSIW) in broadband phase shifters [11], the simultaneous employment of quarter-wavelength coupled lines and quarter-wavelength transmission lines in phase shifter structures [78], the utilization of dual-mode folder circular SIW cavities in filter structures [83], the employment of artificial transmission lines (ATL) in MM-wave quadrature hybrid coupler structures [84], or the introduction of non-periodic step impedance shunt stub (SISS) loaded lines in branch-line coupler structures [85].

The abovementioned miniaturization techniques have been proven to be efficient for the development of the initial circuit topology based on well-established grounds of electromagnetic wave propagation and radiation theory. Nevertheless, they lead to design complexities both in terms of a need for innovative design procedures specific to the structure type, and a following manipulation of additional geometry parameters for the optimum system performance. If only a few (say, one or two) variables are of importance, sub-optimal designs can be identified even after fabrication, by having implemented adjustable mechanical parts based on engineering insight to control the reflection coefficient or other antenna characteristics. However, as the structures evolve into more complex geometries described by numerous parameters, the tuning process becomes considerably more intricate. The unintuitive interconnection between the structure dimensions and the electrical and field performance figures, creates critical obstructions in identification of an optimum design using conventional tuning techniques. Therefore, the only way to achieve the optimum design is through optimization-based size reduction of the structure with all the geometrical parameters being

optimized simultaneously. This issue will be addressed in more detail in the remaining sections of this chapter.

2.2 Optimization-Based Miniaturization

Topological complexity of miniaturized components leads to new design challenges. These are mainly associated with the necessity of appropriate and simultaneous adjustment of multiple geometry parameters. Traditional trial-and-error strategy falls short of unraveling the unintuitive interconnection between the structure dimensions and system performance figures. Identification of the optimum design can only be accomplished through rigorous numerical optimization of all the structure dimensions in a simultaneous manner. This section provides an overview of high-frequency circuit characteristics referred to as performance figures, recalls the formulation of constrained optimization-based high-frequency circuit miniaturization, and subsequently presents the penalty function approach, as well as the corresponding performance challenges.

2.2.1 High-Frequency Circuit Characteristics

The primary purpose of this work is to reduce the physical size of the system while maintaining satisfactory levels of the performance figures. The performance figures are actually defined as circuit properties relying on circuit responses. This section discusses typical types of high-frequency circuit responses within the scope of the thesis.

Considering an m-port device, the corresponding scattering parameters can be denoted as

$$|S_{pq}|, \, p = 1, \dots, m, \; q = 1, \dots, m. \tag{2.1}$$

Some of the popular circuit figures of interest within the scope of this thesis include reflection coefficient, power split ratio, axial ratio (AR), bandwidth (BW), and center frequency f_c. They are either explicitly, or implicitly related to the scattering parameters given by (2.1).

- Reflection coefficient: The ratio of the power reflected back from the circuit element to the source

$$|S_{pp}| = -20 \, log \, |\varGamma| \tag{2.2}$$

where $\varGamma$ is a measure of the impedance mismatch between the source output impedance Z_C, and the corresponding circuit input impedance Z_A, defined as

$$\varGamma = (Z_A - Z_C) / (Z_A + Z_C) \tag{2.3}$$

- Power split ratio: The ratio of the output to input power defined as $|S_{31} - S_{21}|$ (applicable to a coupling structure)
- Axial ratio: The ratio of the orthogonal components of the electric field denoted as AR (applicable to CP antennas)
- Bandwidth: The range of frequencies delimited by a lower f_l, and upper frequency f_u, within which the circuit characteristics needs to meet the desired specifications; given by $BW = f_u - f_l$.
- Center frequency: The arithmetic mean of f_u and f_l, formulated as $f_c = (f_u + f_l)/2$

2.2.2 Problem Formulation

One of the main challenges of high-frequency circuit miniaturization stems from of a close interconnection between the optimal system performance and a specific arrangement of the geometry parameters. A large number of geometry parameters resulting from the circuit-level miniaturization techniques makes manual or guess-work identification of the optimal arrangement of design variables grossly ineffective. Evaluation of the performance figures pertinent to the electrical and field characteristics of the high-frequency circuits—later on formulated as design constraints— is usually carried out using EM simulations of the scaled structure described by the n-element vector of the geometry parameters $x = [x_1 \dots x_n]^T$. The optimization-based miniaturization problem can be formulated as

$$x^* = \arg \min_{x \in X} A(x)$$

$$(2.4)$$

where $A(x)$ is the footprint area of the circuit of interest (in a more generic form it can be defined as the corresponding volume $V(x)$), X is the n-dimensional parameter space, and x^* is the optimum set of geometry parameters to be found. The latter corresponds to the final structure dimensions ensuring a satisfaction of all the design constraints which may be of equality or inequality type. The equality constraints are formulated as

$$S_{ineq,k}(x) \le s_k, \ k = 1, \dots, n_{ineq} \qquad (2.5)$$

where s_k are the acceptance thresholds for the corresponding constraints.

Whereas the inequality ones are given as

$$S_{eq,k}(x) = 0, \ k = 1, \dots, n_{eq} \qquad (2.6)$$

where n_{ineq} and n_{eq} count for the number of inequality and equality constraints, respectively. A summation of the former and the latter is considered as the total number of constraints, $n_c = n_{ineq} + n_{eq}$.

For illustration purposes, the examples of the abovementioned constraint formulations applied to some of the high-frequency circuit characteristics previously mentioned in Section 2.2.1 may be described as:

- Reflection coefficient not exceeding –10 dB defined as an inequality constraint over the operational BW, i.e., $|S_{pp}(x,f)| \le -10$ dB for $f \in [f_l \ f_u]$
- Axial ratio not exceeding 3 dB defined as an inequality constraint over the operational BW, i.e., $|AR(x,f)| \le 3$ dB for $f \in [f_l \ f_u]$
- Power split ratio equal to zero defined as an equality constraint over the operational BW, i.e., $|S_{31}(x,f_0)| - |S_{21}(x,f_0)| = 0$ dB for $f \in [f_l \ f_u]$

2.2.3 Constraint Handling

A prerequisite for handling of the design constraints of any type lies in their proper evaluations. In the context of this thesis, the task is normally carried out using the EM-simulated response of the circuit structure of interest, denoted as $R(x)$. Accordingly, a violation of the inequality constraints is measured as

$$\gamma_{ineq,k} = S_{ineq,k}(x) - s_k, \ k = 1, \dots, n_{ineq} \qquad (2.7)$$

while a violation of the equality constraints is measured as

$$\gamma_{eq,k} = S_{eq,k}(\boldsymbol{x}), k = 1, \ldots, n_{eq} \tag{2.8}$$

A need for satisfaction of these constraints within an optimization process not only imposes algorithmic complexity, but also requires numerous system evaluations which usually turn out to be numerically cost-prohibitive. Constraint handling methods include explicit, and implicit treatment. Among popular explicit techniques the quadratic penalty function methods [86]-[87] , augmented Lagrangian methods [87]-[88], and sequential quadratic programming may be mentioned. The quadratic penalty function methods formulate the constraints as penalty terms consisting of a multiple of the square of the corresponding violations. Besides the attraction due to simplicity and intuition, the quadratic penalty functions suffer from ill conditioning. This disadvantage is overcome by the succeeding method, namely, augmented Lagrangian method. Sequential quadratic programming [87] is one of the most efficient explicit constraint handling methods especially when the problem inherits significant nonlinearities.

In this thesis, we specifically adopt an implicit approach offered by the penalty function approach [18], [19], [20], [21], [22], where the design constraints are handled in an implicit manner not requiring costly system response evaluations. A detailed formulation of the method will be provided in Section 2.2.4.

2.2.4 Penalty Function Approach

The penalty function approach reformulates the constrained optimization-based miniaturization problem (1) into an unconstrained one as

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x} \in X} U_p(\boldsymbol{x}) \tag{2.9}$$

where the objective function takes the form of

$$U_P(\boldsymbol{x}) = A(\boldsymbol{x}) + \sum_{k=1}^{n_{ineq}} \beta_{ineq,k} c_{ineq,k}(\boldsymbol{x})^2 + \sum_{k=1}^{n_{eq}} \beta_{eq,k} c_{eq,k}(\boldsymbol{x})^2 \tag{2.10}$$

The immediate benefit of (2.10) is the smart distinguishability between the cheap geometric-based constraint $A(\boldsymbol{x})$, and the expensive EM-based electrical-and-field-related ones. Therein, an explicit treatment of the former as the main objective is accompanied by an implicit treatment of the latter using penalty functions $c_k(\boldsymbol{x})$, $k = 1, \ldots, n_c$. These are typically defined as quantifications of the relative constraint violations for the inequality constraints as

$$c_{ineq,k}(\boldsymbol{x}) = \max\{0, \gamma_{ineq,k}/s_k\}, k = 1, \ldots, n_{ineq} \tag{2.11}$$

and the same way for the equality constraints

$$c_{eq,k}(\boldsymbol{x}) = \max\{0, \gamma_{eq,k}\}, k = 1, \ldots, n_{eq} \tag{2.12}$$

The proper contributing proportion of $c_{ineq,k}(\boldsymbol{x})$ and $c_{eq,k}(\boldsymbol{x})$ to the main objective being the circuit size is determined by the penalty coefficients $\beta_{ineq,k}$ and $\beta_{eq,k}$, respectively.

2.2.5 Challenges

Simulation-driven miniaturization of high-frequency structures is a computationally challenging task. Evaluation of the system responses requires CPU-intensive full-wave EM analysis. The corresponding time and computational cost multiplies very fast with the increased number of the geometry parameters in more complex structures. The expensive electrical and field related constraints inherent to size reduction task, requires an intricate treatment. This complexity is further intensified for the design cases involving equality constraints. Furthermore, the multimodal nature of constrained miniaturization problems results in initial design dependency of the optimization results. Thus, conventional optimization-based miniaturization methods are either inefficient or unreliable.

The penalty function approach discussed in Section 2.2.4 has made an attempt to simplify constraint handling procedures. The said approach not only ensures the efficacy of the miniaturization problem by an explicit treatment of the size objective, but also facilitates handling of the costly EM-based electrical-and-field-related constraints through their implicit treatment. Numerous variations of the penalty function approach have been proposed targeting specific performance improvements. Implementation of an adaptive reflection coefficient acceptance threshold in [19] not only allowed for a precise control of the reflection coefficient constraint, but also resulted in superior miniaturization rates as compared to the fixed-threshold penalty function approach. Another performance shortcoming of the penalty function approach lies in its nonlinear behavior due to the inconsistency of the parameter sensitivities concerning the size and reflection coefficient constraints. The issue has been addressed in [20] by incorporation of an objective relaxation scheme between the two types of constraints. Further efforts include the improvement of the time and computational expenses of the penalty function approach through the employment of coarse-discretization EM models [21], or the incorporation of variable-fidelity EM model in [22]. Nevertheless, algorithmic performance reliability cannot be ensured without a costly trial-and-error-based identification procedure for the optimum arrangement of the penalty factors. Typical performance indicators include quality of the final design in terms of the obtainable miniaturization rates, and constraint satisfaction. This is because using too high values of the penalty coefficients may result in poor size reduction rates, while too low values may lead to poor constraint control.

Another performance challenge of the aforesaid approach comes into light when targeting miniaturization of specific high-frequency structures involving equality constraints. The corresponding feasible region that encompasses the optimal solutions is a very thin set of the parameter space, which results in impediments in the exploration of the feasible region boundary. Examples of the equality constraints include power split ratio or phase shift in power dividers/combiners, couplers, or phase shifters.

A common challenge among the local search routines [23], [24], [25], [26], [27], is multimodality or initial design dependency. The penalty function approach as a local optimization-based size reduction technique is prone to the same issue. A workaround can be resorting to the global methods [28]- [29]. Relevant conventional optimization routines include population-based metaheuristics (evolutionary strategies [31], particle swarm optimization [32]), or simulated annealing [30]. Nevertheless, the associated time and computational expenses persist as a main obstacle. The computational burden arising mainly from numerous expensive EM-based evaluations of the objective functions, can be alleviated through the incorporation of SBO routines [13], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43],

[44], [45]. Here, the principal idea lies in reducing the computational burden by switching to a cheaper surrogate model. However, surrogate modeling techniques are not a common choice for solving miniaturization tasks.

In [38], the computational efficiency of yield estimation for antenna structures was significantly improved by the incorporation of polynomial chaos-Kriging (PC-Kriging) surrogate modeling technique. In [34], an expedited procedure for the multi-objective microwave device design was proposed, where the Pareto front was approximated using nested kriging surrogate methodology. A novel algorithm presented in [33] attempted to reduce the computational burden of multi-objective PSO problems through the identification of the Pareto front using a Gaussian process (GP) surrogate model. The curse of dimensionality was addressed in [13] by the employment of a two-level kriging surrogate model, where modeling efficacy was ensured within wide ranges of the parameters, as well as the performance figures space. Nevertheless, SBO encounters additional computational challenges due to the increased number of geometry parameters associated with topologically modified structures. Moreover, nonlinear characteristics inherent to high-frequency devices tend to impose restrictions in practical utility of the surrogate models.

Restrictions of time and computational resources associated with high-frequency size reduction, tend to be generalized to any optimization technique (either local or global search). Various algorithmic techniques have been developed to lower the computational burden and expedite the optimization process. Among popular acceleration techniques the incorporation of sparse sensitivity updates [50]-[54], or the employment of adjoint sensitivities [46], [47], [48], [49], [50], [51], may be mentioned. These techniques have successfully reduced the CPU time up to forty [55] and sixty percent [56]. Another acceleration technique is the employment of variable-fidelity EM simulation models [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67]. Computational time savings of over forty percent have been obtained using a set of discrete EM-model fidelities [61]. Nevertheless, availability of only a few preselected model fidelities can lead to shortcomings in constraint satisfaction, and poor miniaturization rates [65]- [66].

In view of the abovementioned performance challenges of the optimization-based size reduction concerning CPU time, computational cost, constraint handling intricacy, and initial design dependency, this thesis attempts to address the corresponding issues by the development of novel, reliable, and cost-effective algorithms, which are intended to advance the state of the art of automated design and miniaturization of high-frequency structures.

Chapter 3

3 Algorithmic Tools Employed in Solving EM-Driven Size Reduction Problems

The problem considered in this work is an automated design optimization of high-frequency circuits with emphasis on size reduction. In practice, the design task may be handled through the employment of local and global optimization techniques, the computational efficiency of either can be significantly improved through utilization of a substitute surrogate model and a subsequent surrogate-based optimization (SBO). Depending on the structure type and performance characteristics, the existent constraints are embedded in the optimization problem as either equality or inequality types. Accordingly, the objective function is formulated in a problem-specific manner considering the former and the latter constraint types. Below, a brief description of the mathematical and algorithmic tools employed in the development of new miniaturization algorithms that exhibit improved efficacy and enable automation of the process is provided. The details of the relevant contributions can be in the following chapters of this thesis.

3.1 Local Optimization Methods

Local optimization refers to identifying local optima of an objective function within the feasible region of the search space. Local search routines principally operate on a single initial solution with an iterative refinement towards the optimum solution. They are usually efficient in finding reasonable solutions within reasonably defined ranges of the parameter space. The examples of popular local optimization techniques include descent methods [89], trust-region gradient-based algorithm [87], or Newton and Quasi-Newton methods [88]. The remaining part of this section provides a generic formulation of the abovementioned techniques applied to the unconstrained optimization problems. Subsequently, the inclusion of the design constraints within the corresponding formulations that enables an expansion of the applicability of these methods to many practical constrained optimization problems is provided.

3.1.1 Descent Methods

We will consider iterative methods producing a sequence of vectors $\boldsymbol{x}^{(i)}$, $i = 0, 1, 2, \ldots$ [89]. They are supposed to converge to a local minimizer $\boldsymbol{x}^*$

$$\boldsymbol{x}^* = \arg \min_{x \in X} f(\boldsymbol{x}) \tag{3.1}$$

where $f(\boldsymbol{x})$ is the given objective function. The descending property is enforced with respect to the objective function value, i.e., $f(\boldsymbol{x}^{(i+1)}) < f(\boldsymbol{x}^{(i)})$. The search for $\boldsymbol{x}^*$ starts from the initial solution $\boldsymbol{x}^{(0)}$. Subsequent vectors $\boldsymbol{x}^{(i+1)}$ are found along a descent direction such that there is at least a local decrease in the value of f.

A critical performance metric of an iterative method is the convergence rate. It provides a description of how quickly the minimizer $\boldsymbol{x}^*$ is approached. The convergence rate measures include:

- Linear convergence defined by

$$\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^*\| \leq c_1 \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^*\| \tag{3.2}$$

where $0 < c_1 < 1$

- Quadratic convergence defined by

$$\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^*\| \leq c_2 \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^*\|^2 \tag{3.3}$$

where $0 < c_2 < 1$

- Superlinear convergence defined by

$$\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^*\| / \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^*\| \to 0 \text{ for } i \to \infty \tag{3.4}$$

The superlinear convergence is faster than the linear one. However, it is not as good as quadratic convergence.

Given $\boldsymbol{x}^{(0)}$ as an initial solution, a generic descent procedure operates as follows [84]:

1. Set $i = 0$;
2. Find a search direction $\boldsymbol{h}_d$;
3. Find a step length α;
4. Set $\boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i)} + \alpha \boldsymbol{h}_d$;
5. Set $i = i + 1$;
6. If the termination condition is not satisfied, go to 2;
7. END

Among the abovementioned steps of the iteration, Steps 2 and 3 are considered as the most critical stages of the procedure. These include finding a descent direction $\boldsymbol{h}_d$, and the following appropriate step length $\alpha \boldsymbol{h}_d$, to be taken along. Therein, $\boldsymbol{h}_d$ represents a descent direction vector for which $\nabla f(\boldsymbol{x})^T \cdot \boldsymbol{h}_d < 0$. The value of $\alpha \boldsymbol{h}_d$ needs to be determined such that the descending property of the objective function value is ensured, i.e. $f(\boldsymbol{x}^{(i+1)}) < f(\boldsymbol{x}^{(i)})$. The algorithm can be terminated based on various convergence criteria including:

- Convergence in argument defined as

$$\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}\| \leq \varepsilon_1, \tag{3.5}$$

- Vanishing of the gradient defined as

$$\|\nabla f(\boldsymbol{x}^{(i)})\| \leq \varepsilon_2, \tag{3.6}$$

- Convergence in the function value defined as

$$f(\boldsymbol{x}^{(i)}) - f(\boldsymbol{x}^{(i+1)}) \leq \varepsilon_3, \tag{3.7}$$

- Combination of the above.

3.1.2 Trust-Region Gradient-Based Optimization

An alternative iterative approach to solving the local optimization problem (3.1) is offered by trust-region gradient-based framework [87]. Therein, a series of candidate solution vectors

$x^{(i+1)}$ is identified as

$$x^{(i+1)} = \arg \min_{x;\, \|x-x^{(i)}\| \le \delta} L^{(i)}(x), i = 0,1,\dots \tag{3.8}$$

where δ is the trust-region search radius, and $L^{(i)}(x)$ is a first-order Taylor approximation of $f(x)$ at $x^{(i)}$ formulated as

$$L^{(i)}(x) = f(x^{(i)}) + \nabla f(x^{(i)})(x - x^{(i)})^T \tag{3.9}$$

The search radius δ is adjusted based on a gain ratio [87] calculated as

$$r = \frac{f(x^{(i+1)}) - f(x^{(i)})}{L(x^{(i+1)}) - L(x^{(i)})}$$

(3.10)

In the case of design improvement, i.e., if $r > 0$, the candidate design $x^{(i+1)}$ is accepted. The standard rules pertaining to the adjustment scheme of the trust region size throughout the optimization run are delineated in [87], a pseudocode of which may be described as follows:

if $r > 0.75$

$\quad \delta \leftarrow 2\delta;$

else if $r < 0.25$

$\quad \delta \leftarrow \delta/3;$

 else if $0.25 < r < 0.75$

$\quad \delta \leftarrow \delta;$

end

3.1.3 Newton and Quasi-Newton Methods

The exploitation of higher-order derivatives can lead to the improvement of the efficacy of the trust-region gradient-based search routine. Specifically, the incorporation of the Hessian of the objective function, instead of the first-order Taylor approximation as in (3.9), has led to the development of a family of Newton and quasi-Newton methods which are described in the following.

Let $f(x)$ be at least twice continuously differentiable objective function. A local representation of $f(x)$ using its second-order Taylor approximation is as follows

$$f(x + h) \approx q(h) \tag{3.11}$$

$$q(h) = f(x) + \nabla f(x) \cdot h^T + \frac{1}{2} h^T H(x) h \tag{3.12}$$

If the Hessian of f at x meets the conditions for positive-definiteness property, the model $q(h)$ leads to a unique minimizer for h described as

$$\nabla q(h) = 0, \tag{3.13}$$

in other words

$$\nabla f(\boldsymbol{x}) + \boldsymbol{H}(\boldsymbol{x})\boldsymbol{h} = 0 \tag{3.14}$$

Equation (3.14) provides the basis of the Newton's method [89]. Therein, the next solution is obtained by solving the abovementioned equation as:

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} - [\boldsymbol{H}(\boldsymbol{x}^{(i)})]^{-1}\nabla f(\boldsymbol{x}^{(i)}) \tag{3.15}$$

The procedure (3.15) is well defined if the Hessian $\boldsymbol{H}(\boldsymbol{x})$ is non-singular. If positive definiteness of the Hessian is ensured for all iterations, and the initial solution is sufficiently close to the minimizer, the method usually has a very quick quadratic convergence rate. At the same time, the basic Newton algorithm suffers from some practical shortcomings including lack of global convergence for many problems, possibility of convergence to a saddle point or a maximum. Furthermore, besides requiring analytical second-order derivatives, the procedure (3.15) may be ill-conditioned.

The bottleneck in solving (3.15) appears when the Hessian $\boldsymbol{H}(\boldsymbol{x})$ shows a lack of positive definiteness. The damped Newton method works around this problem by a reformulation of (3.15) as

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} - [\boldsymbol{H}(\boldsymbol{x}^{(i)}) + \mu\boldsymbol{I}]^{-1}\nabla f(\boldsymbol{x}^{(i)})$$
$$\tag{3.16}$$

where $\boldsymbol{I}$ represents the identity matrix. An adequately large $\mu > 0$ can make the matrix $\boldsymbol{H}(\boldsymbol{x}) + \mu\boldsymbol{I}$ positive definite. The solution of the problem

$$[\boldsymbol{H}(\boldsymbol{x}) + \mu\boldsymbol{I}]\boldsymbol{h}_\mu = -\nabla f(\boldsymbol{x}) \tag{3.17}$$

denoted as $\boldsymbol{h}_\mu$, serves as a minimizer of the model

$$q_\mu(\boldsymbol{h}) = q(\boldsymbol{h}) + \frac{1}{2}\mu\boldsymbol{h}^T\boldsymbol{h} = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})\cdot\boldsymbol{h}^T + \frac{1}{2}\boldsymbol{h}^T[\boldsymbol{H}(\boldsymbol{x}) + \mu\boldsymbol{I}]\boldsymbol{h} \tag{3.18}$$

The damped Newton's method utilizes large values of μ in the early stages of the optimization process and gradually decreases them as it proceeds towards convergence. Sufficiently small value of μ occurs when the algorithm is close to the minimizer, i.e. defaulting to the original Newton's method to identify the minimizer of *f*. Levenberg-Marquardt procedure can be referred to as one of the most popular algorithms implementing the aforementioned concept.

While damped Newton methods work around all difficulties of the basic Newton's method, the high evaluation cost concerning the objective function's second-order derivatives remains a critical obstacle. Quasi-Newton methods attemp to overcome this shortcoming by an approximation of the exact Hessian using proper updating formulas. Broyden-Fletcher-Goldfarb-Shanno (BFGS) has been considered as the most efficient updating formula particularly for this purpose. Therein, the properties of positive definiteness and symmetricity of the approximated Hessian is preserved. The BFGS approximation of the Hessian denoted by $\boldsymbol{B}$ is formulated as follows:

$$B_{new} = B + \frac{1}{h^T y} yy^T - \frac{1}{h^T u} uu^T, \tag{3.19}$$

where

$$h = x_{new} - x, \ y = \nabla f(x_{new}) - \nabla f(x), \tag{3.20}$$

and $u = Bh$.

Whereas the inverse of the Hessian is formulated as:

$$D_{new} = D + \kappa_1 hh^T - \kappa_2 (hv^T + vh^T),$$

$$\tag{3.21}$$

where

$$h = x_{new} - x, \quad y = \nabla f(x_{new}) - \nabla f(x), \quad v = By$$

$$\kappa_2 = \frac{1}{h^T y}, \quad \kappa_1 = \kappa_2 (1 + \kappa_2 (y^T v))$$

$$\tag{3.22}$$

3.1.4 Constrained Optimization

The previously described algorithms permit solving unconstrained optimization tasks. At the same time, most practical problems are constrained ones with inequality/equality constraints, or lower and upper bounds on the design variables. This section attempts to provide some insight into the constrained optimization techniques. For a more detailed discussion, the reader can be referred to the numerical optimization textbooks [88] - [90].

The optimization problem can be formulated similarly to the problem (3.1), however with explicitly given constraints as

$$x^* = \arg\min f(x) \tag{3.23}$$

which is subject to inequality constraints

$$c_{ineq,k}(x) \le 0, \ k = 1, \ \ldots, \ N \tag{3.24}$$

and equality constraints

$$c_{eq,k}(x) = 0, \ k = 1, \ \ldots, \ M \tag{3.25}$$

Typical constraint types include box constraints, $l_b \le x \le u_b$, i.e. lower and upper bounds on the design variables, geometry constraints that permit avoiding geometrical inconsistency of the design along with controlling the circuit footprint area. Moreover, electrical and field performance figures can also be controlled through equality or inequality constraints depending on their corresponding nature. The examples of the relevant performance figures include power split ratio in coupler/divider structures, axial ratio AR in CP antenna structures, or reflection coefficient in coupler/divider and antenna structures.

The set of vectors satisfying the constraints (3.24) and (3.25) is defined as a feasible region where constrained optimization is supposed to identify the minimizer of the function $f(\boldsymbol{x})$.

The optimality of the solution of a constrained problem (3.23)-(3.25) is ensured through the first derivative tests known as Karush-Kuhn-Tucker (KKT) conditions [91]. Therein, the local minimizer $\boldsymbol{x}^*$ of $f(\boldsymbol{x})$ is conditioned to the existence of the constants $\zeta_1, \ldots, \zeta_N$, and $\eta_1, \eta_2, \ldots, \eta_M$, such that

$$\nabla f(\boldsymbol{x}^*) + \sum_{k=1}^{N} \zeta_k \nabla c_{ineq,k}(\boldsymbol{x}^*) + \sum_{k=1}^{M} \eta_k \nabla c_{eq,k}(\boldsymbol{x}^*) = 0 \tag{3.26}$$

$$c_{ineq,k}(\boldsymbol{x}^*) \le 0, \qquad \zeta_k c_{ineq,k}(\boldsymbol{x}^*) = 0, \qquad k = 1, \ldots, N$$
$$\text{(3.27)}$$

where

$$\zeta_k \ge 0, \qquad k = 1, \ldots, N \tag{3.28}$$

provided continuous differentiability of all the objective and constraint related functions.

The function at the left-hand side of (3.28) is called the Lagrangian function, whereas the coefficients ζ and η are referred to as Lagrange multipliers.

A simple method to handle constrained optimization is the penalty function approach discussed in Section 2.2.4. Therein, the original problem (3.23)-(3.25) is reformulated as

$$\arg \min_{\boldsymbol{x}} U(\boldsymbol{x}, \beta_{ineq}, \beta_{eq}) = \arg \min_{\boldsymbol{x}} \left\{ f(\boldsymbol{x}) + \sum_{k=1}^{n_{ineq}} \beta_{ineq,k} c_{ineq,k}^2(\boldsymbol{x}) + \sum_{k=1}^{n_{eq}} \beta_{eq,k} c_{eq,k}^2(\boldsymbol{x}) \right\} \tag{3.29}$$

As previously mentioned in Section 2.2.4, $\beta_{ineq,k}$ and $\beta_{eq,k}$ are the penalty coefficients determining the proper proportion of the corresponding penalty functions $c_{ineq,k}(\boldsymbol{x})$ and $c_{eq,k}(\boldsymbol{x})$, respectively.

3.2 Global Optimization Methods

As opposed to local optimization, global optimization algorithms seek to locate the global optimum of the objective function within the entire search space. Depending on the particular methods, their principal operation can be either based on a single initial solution as in simulated annealing [92], a cluster of initial solutions as in random search [93], or may not require the initial solution whatsoever as in population-based metaheuristics [96]. In global optimization algorithms, new candidate solutions are generated, evaluated, and updated in an iterative manner until the convergence criterion is met. These methods exhibit no restrictions regarding properties of the objective functions such as continuity or differentiability, in contrast to the counterpart local optimization methods which require continuously differentiable objective functions. This section provides a description of the abovementioned global optimization techniques along with their detailed operational flow.

3.2.1 Random Search

Many local optimization techniques are prone to multimodality due to the presence of the objective functions with multiple optima. In other words, the final optimization results are dependent on the initial design vector. A global optimization method known as random search, avoids being stuck in a local minimum by sampling the parameter space according to a uniform distribution. A pseudocode of the method is provided in the Fig. 3.1.

```
Random search algorithm:
f_best = Inf;

j = 0;

while j < j_max

    x_new = generate_random_point();

    f_new = f(x_new);

    if f_new < f_best

        x_best = x_new;

        f_best = f_new;

    end

    j = j + 1;

end

where:

 j – iteration number

 j_max – maximum number of iterations

 f_best – current value of the objective function

 f_new – new value of the objective function
```

Figure 3.1: Operational flow of the random search algorithm

A modified version of the random search method attempts to enhance its exploitation capabilities through biasing of the random samples by the best solution identified so far. Therein, the new sample is formulated as

$$x_{new} = \alpha x + (1-\alpha)x_{best} \tag{3.30}$$

with $\alpha \to 0$ for $j \to j_{max}$ (maximum number of function evaluations).

Therefore, the search tends to become progressively more local in the vicinity of x_{best} as the algorithm moves towards convergence. A pseudocode of the procedure is described in Fig. 3.2.

3.2.2 Simulated Annealing

Simulated annealing is a probabilistic global optimization technique often applicable to problems with a discrete parameter space. The principal operation of this algorithm is based on replacing the current solution by randomly selected nearby solution. The selection criterion is a transition probability $P(f, f_{new}, T)$, that depends on two factors: (i) the difference between

the corresponding objective function values i.e. $f - f_{new}$, (ii) a global variable referred to as temperature T with gradually decreasing trend throughout the optimization process. The corresponding formulation is as follows

```
Modified random search algorithm:
x_best = generate_random_point();

f_best = f(x_best);

j = 0;

while j < j_max

    x_new = generate_random_point();

    x_new = αx_new + (1- α)x_best;

    f_new = f(x_new);

    if f_new < f_best

        x_best = x_new;

        f_best = f_new;

    end

    α = update_alpha(j,j_max);

    j = j + 1;

end

where:

 j – iteration number

 j_max – maximum number of iterations

 f_best – current value of the objective function

 f_new – new value of the objective function

 α – bias coefficient
```

Figure 3.2: Operational flow of the modified random search algorithm

$$P(f, f_{new}, T) = \begin{cases} 1 & if \ f < f_{new} \\ \exp((f - f_{new})/T) & otherwise \end{cases} \tag{3.31}$$

Random changes are observed in the current solution at the initial stages of the optimization process where T is large, while there is an increasingly downhill behavior as T approaches zero. The immediate benefit of the initial random behavior lies in a possibility of moving uphill, therefore, reducing the chances of being trapped in a local optimum. Figure 3.3 provides the standard procedure for the implementation of the simulated annealing algorithm.

```
Simulated annealing algorithm:
x_best = x = x^(0); f_best = f = f(x);

j = 0;

while j < j_max

    x_new = neighbor (x);

    f_new = f(x_new);

    if f_new < f_best

        x_best = x_new; f_best = f_new;

    end

    if P(f,f_new,temp(j/j_max)) > rand()

        x = x_new; f = f_new;

    end

end

where:

 j – iteration number

 j_max – maximum number of iterations

 f – current value of the objective function

 f_new – new value of the objective function
```

Figure 3.3: Standard operating procedure of the simulated annealing algorithm

3.2.3 Population-Based Metaheuristics

Population-based metaheuristics are derivative-free global optimization methods inspired by natural phenomena such as natural or biological systems involving a cluster of interacting individuals. Immediate benefits of this inspiration include the capability of handling discontinuous or non-differentiable objective functions, preventing multimodality issues due to multiple local optima, and eventually convergence to the global optimum of the system at hand. Among the most popular population-based techniques, evolution strategies (ES) [97], genetic algorithms (GAs) [98], particle swarm optimization (PSO) [99], evolutionary algorithms (EAs) [97], or ant systems (AS) [98] may be mentioned.

A general flow of a population-based search routine, specifically concerning algorithms such as EAs or GAs, is provided in Fig. 3.4 [99].

The terminology utilized in the above algorithm description is listed in the following.

- Evaluation: The process of evaluating the objective function values for the individuals,
- Parents: Individuals creating a new individual,
- Recombination: The process in which new individuals are created,
- Mutation: The process in which individuals undergo random modification through local perturbations.

```
Population-based search algorithm:
initialize population P_μ = {x^(1), …, x^(μ) }

while ~termination_condition

    evaluate individuals in P_μ;

    choose parent individuals P_λ from P_μ;

    P_μ.new = recombination(P_λ);

    P_μ = mutation(P_μ.new)

end

where:

  recombination() – recombination operator

  mutation()  – mutation operator
```

Figure 3.4: Standard operating procedure of the population-based search algorithm

Population-based metaheuristics are considered as the preferred techniques for handling multimodal design problems with either limited computational resources, or low evaluation time of the objective function. A typical example may be, e.g., the automated design and optimization of antenna arrays where the objective function is evaluated using analytical models for array factor analysis [102], [103]. However, for the design problems where the evaluation of the objective function requires performing costly EM simulations, population-based metaheuristics encounter prohibitive computational burdens.

3.3   Surrogate Modeling

3.3.1 Introduction

The reliability as well as the efficacy of the optimization algorithms discussed so far has a strong dependency on the accuracy of the computational model evaluating the objective function values. In the context of this thesis, i.e., miniaturization of high-frequency structures, the computational model is most often based on full-wave EM analysis. The costly nature of EM simulations, intensified by numerous evaluations inherent to conventional optimization algorithms, tends to create a significant computational burden. A workaround is the employment of fast replacement models a.k.a. surrogate models [23], [104]. Nevertheless, in order to ensure the reliability of the optimization process with the original computational model being replaced by a cheaper one, the underlying surrogate needs to exhibit specific features including fast computational capability, sufficient accuracy, and analytical tractability. While there is a vast variety of surrogate modeling techniques in the literature, they can be categorized into two major types including approximation (or data-driven) models, and physics-based models, a detailed description of each type is provided in this section.

3.3.2 Data-Driven Surrogate Modeling

Data-driven models do not require any physical insight into the system of interest. They are mainly constructed using sampled high-fidelity system evaluations. Therefore, they are

considered as generic surrogate models with the potential to be applied to any type of optimization problems. Besides versatility, the approximation models are cheap to evaluate, as they are basically constructed in the form of analytical formulas. Nevertheless, their predictive power is closely related to the number of high-fidelity training data samples. This section provides a detailed mathematical formulation of a number of the selected approximation surrogates including polynomial regression [37], radial basis functions [105], kriging [104], and artificial neural networks [105].

3.3.2.1 Polynomial Regression

A simple, yet popular, data-driven surrogate modeling technique is polynomial regression. Therein, low-order polynomials are utilized for either local surrogate modeling, or a construction of the initial surrogate models with low accuracy to indicate the trends of the objective function.

The corresponding surrogate is formulated as

$$s(\boldsymbol{x}) = \sum_{k=1}^{M} \alpha_k v_k(\boldsymbol{x}), \tag{3.32}$$

where α_k are the model parameters and v_k are the basis functions. The model parameters α_k can be normally determined as a least-square solution to the linear system

$$\boldsymbol{F} = \boldsymbol{X}\boldsymbol{\alpha}, \tag{3.33}$$

where $\boldsymbol{\alpha}$ is an M-element vector containing the model parameters as

$$\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2 \ \ldots \ \alpha_M]^T, \tag{3.34}$$

and $\boldsymbol{F}$ is a q-element vector containing the objective function values defined as

$$\boldsymbol{F} = [f(\boldsymbol{x}^{(1)}) \ f(\boldsymbol{x}^{(2)}) \ \ldots \ f(\boldsymbol{x}^{(q)})]^T \tag{3.35}$$

$\boldsymbol{X}$ represents a $q \times M$ matrix containing the basis function values at the sample points. If $q \geq M$ and rank($\boldsymbol{X}$) = M, the model parameters α_k can be found in the form of a solution to the least-squares problem

$$\boldsymbol{\alpha} = \boldsymbol{X}^+ \boldsymbol{F} \tag{3.36}$$

where $\boldsymbol{X}^+$ is the pseudoinverse of $\boldsymbol{X}$ defined as

$$\boldsymbol{X}^+ = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \tag{3.37}$$

Note that the basis functions do not need to be necessarily polynomials. Depending on the application, they can be sinusoidal or exponential as well.

3.3.2.2 Radial Basis Function

Another commonly used technique for the construction of data-driven surrogates is radial basis function (RBF) interpolation. Therein, the RBF surrogate is generated using a linear combination of radially symmetric functions φ

$$s(\boldsymbol{x}) = \sum_{k=1}^{M} \theta_k \varphi(\| \boldsymbol{x} - \boldsymbol{w}^{(k)} \|), \tag{3.38}$$

where $\boldsymbol{w}^{(k)}$, $k = 1, \ldots, M$, are the basis function centers, and θ_k are the model parameters the values of which can be found using a least-square problem formulation similar to that of the previous section

$$\boldsymbol{\theta} = \boldsymbol{\varphi}^+ \boldsymbol{F} \tag{3.39}$$

where $\boldsymbol{\theta}$ is an M-element vector containing the model parameters

$$\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_M]^T, \tag{3.40}$$

and $\boldsymbol{F}$ is a q-element vector containing the objective function values

$$\boldsymbol{F} = [f(\boldsymbol{x}^{(1)}) \ f(\boldsymbol{x}^{(2)}) \ \dots \ f(\boldsymbol{x}^{(q)})]^T, \tag{3.41}$$

whereas $\boldsymbol{\phi}^+$ is the pseudoinverse of the $q \times M$ matrix $\boldsymbol{\phi} = [\phi_{jl}]_{j=1,\dots,q; \ l=1,\dots,M}$, defined as

$$\boldsymbol{\phi}^+ = (\boldsymbol{\phi}^T \boldsymbol{\phi})^{-1} \boldsymbol{\phi}^T, \tag{3.42}$$

with the corresponding entries ϕ_{jl} calculated using

$$\phi_{jl} = \varphi(\|\ \boldsymbol{x}^{(j)} - \boldsymbol{w}^{(l)}\ \|) \tag{3.43}$$

If $\boldsymbol{\phi}$ is a regular square matrix, i.e. $q = M$, and the centers of the basis functions coincide with and the training data points, the vector of the model parameters $\boldsymbol{\theta}$ can be calculated as

$$\boldsymbol{\theta} = \boldsymbol{\phi}^{-1} \boldsymbol{F} \tag{3.44}$$

A commonly used basis function is a Gaussian one, $\varphi(r) = \exp(-r^2/2\sigma^2)$, with σ being the scaling coefficient. Other basis functions include thin plate spline $\varphi(r) = r^2\ln(r)$, multiquadric $\varphi(r) = (1 + \sigma^2 r^2)^{1/2}$, inverse quadratic $\varphi(r) = (1 + \sigma^2 r^2)^{-1}$, or inverse multiquadric $\varphi(r) = (1 + \sigma^2 r^2)^{-1/2}$.

3.3.2.3 Kriging

Kriging is considered as one of the most widely used surrogate models in the automated design of the contemporary engineering systems. A basic kriging model is represented in the form of a following function:

$$f(\boldsymbol{x}) = \boldsymbol{c}(\boldsymbol{x})^T \boldsymbol{\lambda} + I(\boldsymbol{x}), \tag{3.45}$$

where $\boldsymbol{c}(\boldsymbol{x}) = [c_1(\boldsymbol{x}) \ c_2(\boldsymbol{x}) \ \dots \ c_M(\boldsymbol{x})]^T$ are known constant functions, and $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \dots \ \lambda_M]^T$ are the unknown model parameters; $I(\boldsymbol{x})$ is a realization of a normally distributed Gaussian random process with zero mean and variance σ^2. The formulation (3.45) consists of two terms: the first is a regression part $\boldsymbol{c}(\boldsymbol{x})^T\boldsymbol{\lambda}$, serving as a trend function for f, and the second term represents localized variations denoted as $I(\boldsymbol{x})$. The covariance matrix of these variations is defined as

$$Cov[I(\boldsymbol{x}^{(j)})I(\boldsymbol{x}^{(l)})] = \sigma^2 \boldsymbol{R}([R(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(l)})]), \tag{3.46}$$

where $\boldsymbol{R}$ is a $q \times q$ correlation matrix whose entries are the correlation functions between the pair of sampled data points $\boldsymbol{x}^{(j)}$ and $\boldsymbol{x}^{(l)}$

$$R_{jl} = R(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(l)}). \ R(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(l)}) \tag{3.47}$$

Gaussian function is the most commonly used correlation function. It is typically represented as

$$R(\boldsymbol{x}, \boldsymbol{y}) = \exp\left[-\sum_{i=1}^{m} \beta_i \ |\ x_i - y_i\ |^2\right], \tag{3.48}$$

where x_i and y_i are the ith components of the corresponding vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, and β_i are the correlation parameters to be found. The kriging predictor is formulated as

$$s(\boldsymbol{x}) = \boldsymbol{c}(\boldsymbol{x})^T \boldsymbol{\lambda} + \boldsymbol{r}^T(\boldsymbol{x})\boldsymbol{R}^{-1}(\boldsymbol{F} - \boldsymbol{C}\boldsymbol{\lambda}), \tag{3.49}$$

where $\boldsymbol{C}$ is a $q \times M$ matrix with $C_{jl} = c_l(\boldsymbol{x}^{(j)})$, $\boldsymbol{F} = [f(\boldsymbol{x}^{(1)})\ f(\boldsymbol{x}^{(2)})\ \dots\ f(\boldsymbol{x}^{(q)})]^T$, and $\boldsymbol{r}(\boldsymbol{x}) = [R(\boldsymbol{x}, \boldsymbol{x}^{(1)})$ $\dots\ R(\boldsymbol{x}, \boldsymbol{x}^{(q)})]^T$. The vector of unknown model parameters $\boldsymbol{\lambda} = (\boldsymbol{C}^T\boldsymbol{R}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}^T\boldsymbol{R}^{-1}\boldsymbol{F}$ is determined in model fitting procedure using maximum likelihood for λ_i.

An appealing feature specific to kriging is the availability of the model approximation error provided by the random process $I(\boldsymbol{x})$. This property helps improve the model accuracy by locating the highest model errors and accordingly updating the surrogate with additional data samples at the corresponding locations.

3.3.2.4 Artificial Neural Networks

Artificial neural networks (ANNs), also known as neural nets (NNs), constitute a wide and versatile research area with their main applications in categorization and classification. Nevertheless, they can be discussed as intricate nonlinear regression models in the context of this chapter.

The main idea behind ANN is the imitation of a biological neural network through a first-order approximation model. A neuron as the main component of ANN, gives rise to a nonlinear regression operation [106] as shown in Fig 3.5(a). The corresponding regression coefficients include w_1 through w_n; while β is bias value for the neuron, and T is a slope coefficient. Figure 3.5(b) demonstrates the most popular NN architecture consisting of a multi-layer feed-forward network.

The two critical stages in the construction of ANN include a selection of a proper NN architecture, and a subsequent network training. In the case of a multi-layer ANN as in Fig 3.5(b), the network flexibility, or in other words, the training data approximation capability, has a strong dependency on the configuration of the hidden layers, i.e., the number of the hidden layers and the corresponding neurons. In the network training stage, the neuron weights are applied values through solving a nonlinear least-squares regression problem. The error back-propagation procedure is a commonly used method to solve this specific regression problem. Depending on the size and complexity of the network architecture, global optimization methods may be required for the identification of the optimal ANN architecture. Too simple a network architecture may result in the training data approximation deficiencies, whereas an excess in degrees of freedom may lead to poor generalization capability.
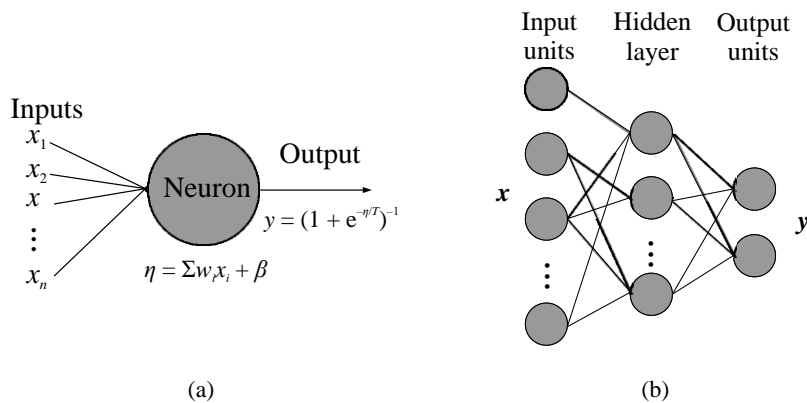

Fig. 3.5. Basic concepts of artificial neural networks: (a) structure of a neuron; (b) two-layer feed-forward ANN architecture [106]

3.3.3 Physics-Based Surrogate Modeling

In contrast to data-driven surrogates described in the previous section, physics-based models require problem-specific knowledge of the system at hand for their setup [106]. Therein, a physics-based description of the system behavior is embedded within a so-called low-fidelity surrogate. An iterative correction of the low-fidelity model is carried out using a limited number of high-fidelity system evaluations normally by means of either linear or nonlinear regression. This section starts by providing a description of low-fidelity models and how they are constructed. Subsequently, the selected physics-based surrogate modeling techniques including response correction models [106], and feature-based modeling [106] are discussed.

3.3.3.1 Low Fidelity Modeling

The low-fidelity model, also known as coarse computational model, is the essential element of any physics-based surrogate. It is actually a simplified version of the high-fidelity (a.k.a. fine) computational model intended to reduce the evaluation costs and therefore speed up the optimization process. While this is accomplished at the expense of lowering the model accuracy, a trade-off between the speed and accuracy needs to be assessed in order to guarantee the surrogate reliability. The low-fidelity model has to adequately represent the trends of the high-fidelity model with respect to design variable changes. Considering a specific performance figure of the system of interest being evaluated by both models, if certain variable changes make either an upward or downward trend in the high-fidelity model response, the low-fidelity model response should follow the same trend. Nevertheless, an offset value is allowed, i.e., the performance figures evaluated by the two models do not have to be the same. Low-fidelity models can be developed through a number of approaches including simplified physics (a.k.a. variable-fidelity physics model), coarse discretization (variable-resolution model), relaxed convergence criteria, or a combination of them. The simplified physics method involves a replacement of the high-fidelity model equations by a set of simplified ones, whereas the coarse discretization method utilizes the same high-fidelity model equations, however, with a coarser discretization of the computational mesh grid. The relaxed convergence criteria method (a.k.a. variable-accuracy model), as implied, deals with relaxing the solver convergence criteria.

3.3.3.2 Response Correction Models

Let $R_f(x)$ and $R_c(x)$ denote the high and low-fidelity EM-simulated responses of the system of interest, respectively. The response correction surrogate model $R_s(x)$ is constructed by applying a proper correction function to the low-fidelity response $R_c(x)$ as follows:

$$R_s(x) = C(R_c(x)) \qquad (3.50)$$

where $C: R^n \rightarrow R^n$ represents the response correction function. The surrogate is updated iteratively throughout the optimization process. Therefore, it is reformulated as $R_s^{(i)}(x) = C^{(i)}(R_c(x))$, with $C^{(i)}$ being the correction function at the iteration i. The minimum requirement for the correction based surrogates is zero-order consistency between the surrogate and the high-fidelity model at the current design $x^{(i)}$, defined as follows

$$R_s^{(i)}(x^{(i)}) = R_f(x^{(i)}) \tag{3.51}$$

However, an additional first-order consistency guarantees the convergence of the algorithm. The condition is formulated as

$$J[R_s^{(i)}(x^{(i)})] = J[R_f(x^{(i)})] \tag{3.52}$$

where $J[.]$ represents the model Jacobian.

A basic response correction scheme is defined as

$$C(R_c(x)) = R_c(x) + [R_f(x^{(i)}) - R_c(x^{(i)})] \tag{3.53}$$

where the aforementioned zero-order consistency is obviously satisfied.

3.3.3.3 Feature-Based Modeling

Development of the surrogate models for high-frequency structures involves handling of complex-valued system output vectors mainly concerning frequency characteristics. As mentioned in Section 2.2.1, the high-frequency system characteristics in the context of this thesis include reflection coefficient, power split ratio, axial ratio. The nonlinear nature of the system responses either as a function of the frequency or geometrical parameters, introduces challenges in the construction of the surrogates. Feature-based surrogates work around this problem by a meticulous selection of a set of specific points known as response features from the entire frequency characteristics. The response features are problem-dependent and may include the response minima, the response maxima, the center frequency, or points with specific response levels (e.g. –3 dB and –10 dB) all within the operating band of the specific structure. They exhibit a much weaker nonlinearity in their functional dependency on either frequency or structure parameters as compared to that of the entire frequency characteristics. Figure 3.6. illustrates the reflection coefficient $|S_{11}|$ of a bandpass filter [106] with the horizontal line as the minimax specifications. Design specifications include $|S_{11}(x,f)| \leq -10$ for $f \in [10.55 \ 11.45]$ GHz as well as $|S_{11}(x,f)| \geq -1$ for $f \leq 10.55$ GHz and $f \geq 11.45$ GHz. A set of feature points is considered as response maxima as well as –1 and –20 dB levels, all within the filter passband.
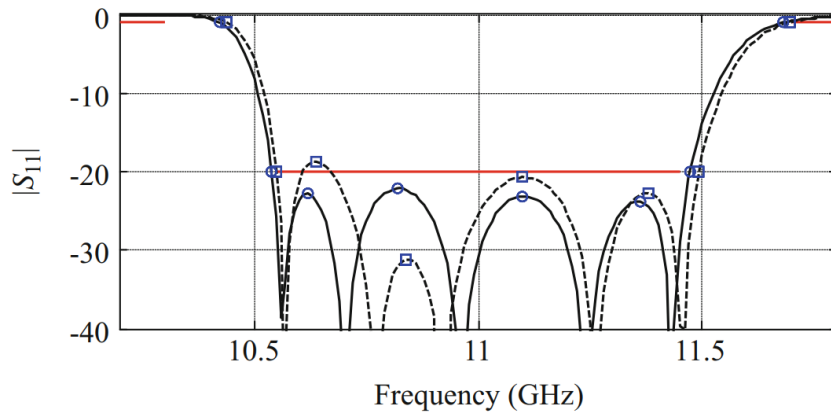

Fig. 3.6. Reflection coefficient of a bandpass filter [106]; (—): optimum design, (- - -): perturbed design. Feature points are represented by circles and squares corresponding to the response maxima, –1 and –20 dB levels, all within the filter passband.

Let $\boldsymbol{R}_f: X \rightarrow R^k$, $X \subseteq R^m$ be the high-fidelity EM-simulated response vector of the system of interest over a k-element frequency interval $[\omega_1, \omega_k]$, i.e., $\boldsymbol{R}_f(\boldsymbol{x}) = [R_f(\boldsymbol{x}, \omega_1), \ldots, R_f(\boldsymbol{x}, \omega_k)]^T$. The task is to construct a fast replacement model $\boldsymbol{R}_s$ that adequately represents $\boldsymbol{R}_f$ within X.

We denote by $X_T = \{ \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^M \} \subset X$ the training data set, meaning that the high-fidelity response vectors $\boldsymbol{R}_f(\boldsymbol{x}^i)$, $i = 1, \ldots, M$ at $\boldsymbol{x}^i$ are evaluated. The selected feature points are represented as the pairs including the frequencies and the corresponding response levels, that is $\boldsymbol{f}_n^p = [\omega_n^p, l_n^p]^T$, $n = 1, \ldots, N$, $p = 1, \ldots, M$; where ω_n^p and l_n^p are the frequency and response level components of $\boldsymbol{f}_n^p$, N is the number of feature points, and M is the number of the training data vectors. The feature-based surrogate model is constructed by approximating the properly selected set of feature points of the training data set rather than the entire frequency response. This inevitably leads to some loss of the information, yet the amount of information contained in the surrogate suffices to employ it for design purposes.

3.4   Surrogate-Based Optimization

Surrogate-based optimization techniques (SBO) [37] are intended to facilitate automated design optimization problems with costly computational models. Therein, the direct optimization of the original high-fidelity model is substituted by an approximate optimization problem involving a fast replacement surrogate. The SBO starts with an initial design $\boldsymbol{x}^{(0)}$ and subsequently obtains a series of approximate solutions $\boldsymbol{x}^{(i)}$, $i = 0, 1, \ldots$ to the original problem (2.9), using a reformulated version as

$$\boldsymbol{x}^{(i+1)} = \arg \min_{\boldsymbol{x} \in X} U_p(s^{(i)}(\boldsymbol{x})) \tag{3.54}$$

where $s^{(i)}(\boldsymbol{x})$ represents the surrogate model at $\boldsymbol{x}^{(i)}$. It is normally updated at each iteration i using the high-fidelity model responses.

The immediate benefit of SBO are the potential computational savings as compared to the conventional optimization methods. Given a fast surrogate, computational cost of solving (3.47) can be negligible. The only contributing factor to the overall cost of the optimization process is the number of the high-fidelity model evaluations required to update the surrogate at each iteration. Typical SBO procedures converge after a significantly smaller number of iterations when solving (2.9) as compared to majority of conventional optimization algorithms. This section provides a brief description of data-driven SBO techniques including response surface approximations (RSAs), sequential approximate optimization (SAO), and SBO with kriging surrogates.

3.4.1 Response Surfaces

Response surface approximation (RSA) [37] is a data-driven SBO technique operating within two major stages. Stage 1 is responsible for the construction of the initial surrogate, while Stage 2 implements a prediction-correction procedure in an iterative manner. The purpose of Stage 2 is to determine the promising designs within the search space so that the surrogate is updated using high-fidelity model evaluations at the suitable locations. The operational flow of the optimization process is as follows:

1. Sampling of the training data;
2. Construction of the initial surrogate using the sampled training data;
3. Identification of a candidate design by optimizing the surrogate model;

4. Evaluation of the high-fidelity model at the candidate solution;
5. Updating the surrogate by newly evaluated high-fidelity responses;
6. If the termination condition is not satisfied, go to 3.

The abovementioned procedure generates the new infill points through the optimization of the surrogate model. As we will see in Section 3.4.3, enhancing global accuracy of the surrogate model can be an important infill criterion.

3.4.2 Sequential Approximate Optimization

Sequential approximate optimization (SAO) is one of the most basic data-driven SBO techniques. The principal operation is based upon restricting the search to a small region of the parameter space and therefore facilitating the use of a low-order local approximation surrogate model. The surrogate is optimized and the search region is adjusted iteratively according to the newly predicted design in combination with a relocation strategy.

The optimization flow is as follows:

1. Set the iteration count $i = 0$;
2. Evaluate the high-fidelity response vector $\boldsymbol{R}_f$ at the current design $\boldsymbol{x}^{(i)}$;
3. Restrict the search space to a small region $r^{(i)}$ using a relocation strategy;
4. Perform sampling of $r^{(i)}$ based on a proper design-of-experiment (DOE);
5. Evaluate high-fidelity responses of the sampled designs obtained in Step 4;
6. Construct the surrogate $s^{(i)}$ using the data set acquired in Step 5;
7. Perform optimization of the surrogate $s^{(i)}$ within $r^{(i)}$ to obtain a new candidate solution $\boldsymbol{x}^{(i+1)}$;
8. If the termination condition is met, set $i = i + 1$ and go to 2.

The main advantage of SAO is the small number of training samples required at each iteration due to the simplicity of the surrogates. At the same time, it is only capable of performing local search routines.

A critical concept in SAO implementations is the relocation strategy referring to the adjustment of the restricted search region $r^{(i)}$ with respect to location and size. A typical relocation strategy utilizes the current design $\boldsymbol{x}^{(i)}$ as the center of the subregion $r^{(i)}$ with the previous design $\boldsymbol{x}^{(i-1)}$ as its corner point. Further size adjustment of the subregion $r^{(i)}$ may be based on standard TR-based rules [87].

3.4.3 SBO with Kriging Surrogates

A critical aspect of SBO techniques is an appropriate selection of infill (training) points [105]. A proper infill point not only reduces the objective function value, but also improves the global accuracy of the surrogate.

The most basic infill strategy applicable to local search routines, selects the optimum of the surrogate model as the infill point. The strategy leads to convergence capabilities of the SBO at least to a local minimum of the high-fidelity model. Nevertheless, the convergence is based upon two assumptions: (*i*) the SBO is implemented using TR-based framework, (*ii*) the surrogate model is first-order consistent. In order for the infill strategy to improve the global

accuracy, the surrogate model requires a feedback on its prediction power as well as on its generalization capability. As previously mentioned in Section 3.3.2.3, kriging surrogate models feature a random process that provides iterative model approximation errors. Therefore, they appear as appealing candidates in this context, as they provide the required aforementioned feedback. Popular infill criteria are established based on this particular feature of the kriging surrogate models, a brief description of which is as follows:

1. Maximizing the improvement expected for an untried candidate solution x;
2. Minimizing the surrogate prediction of the objective function $\hat{s}(x)$, assuming the surrogate model has a reasonable global accuracy;
3. Minimizing $LB(x) = \hat{s}(x) - Qi(x)$, i.e., the statistical lower bound; where $i^2(x)$ is the variance, and Q is a user-defined factor;
4. Maximizing the probability of improvement by identification of the locations with the highest likelihood to improve the objective function value;
5. Maximizing the mean square error through identification of the locations with the highest surrogate prediction of the mean square error.

Sampling of the new infill points according to the abovementioned infill criteria calls for a global optimization routine, in particular, population-based metaheuristics; a description of which has been provided in Section 3.2.3.

3.5 Thesis Methodology

This section summarizes the methodology adopted in this thesis work. After a comprehensive literature review in the area of high-frequency circuit miniaturization, our dissertation is focused specifically on EM-driven size reduction of high-frequency structures through constrained optimization.

The design task may be handled through the employment of local and global optimization techniques. Starting with the less complicated and lower cost technique, i.e. local optimization, we aim at addressing constraint satisfaction and size-reduction rate-related challenges of optimization problems with a single inequality constraint by developing an adaptive penalty function approach. Next, we extend the algorithm to a more flexible, sophisticated version, applicable to designs with multiple constraints. Subsequently, we develop a convergence-feasibility-based variable-fidelity EM model management scheme to accelerate the optimization process. Later on, a novel correction-based procedure for proper handling of equality constraints is developed. Finally, a new surrogate modeling technique is developed which enables cost-effective global miniaturization. Figure 3.7 illustrates a flow chart of the described methodology adopted in our thesis work.
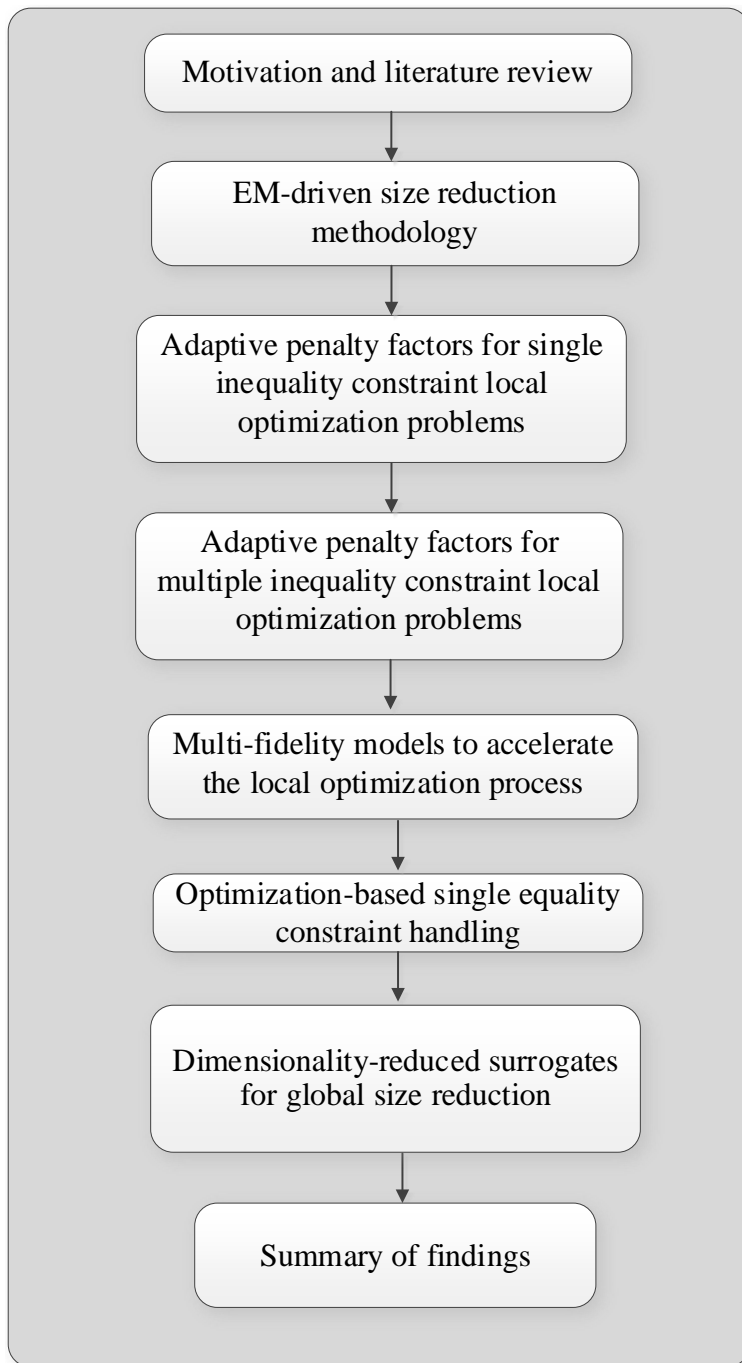
Figure 3.7: Flow chart of the collective methodology adopted in the thesis work.

Chapter 4

4 Paper # 1

Marzieh Mahrokh and Slawomir Koziel

Article

Optimization-Based Antenna Miniaturization Using Adaptively Adjusted Penalty Factors

Marzieh Mahrokh [1],* and Slawomir Koziel [1,2]

[1] Engineering Optimization and Modeling Center, Reykjavik University, 102 Reykjavik, Iceland; Koziel@ru.is
[2] Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, 80-233 Gdańsk, Poland
* Correspondence: marziehm@ru.is; Tel.: +354-599-6376

Abstract: The continuing trend for miniaturization of electronic devices necessitates size reduction of the comprising components and circuitry. Specifically, integrated circuit-antenna modules therein require compact radiators in applications such as 5G communications, implantable and on-body devices, or internet of things (IoT). The conflict between the demands for compact size and electrical and field performance can be mitigated by means of constrained numerical optimization. Evaluation of performance-related constraints requires expensive electromagnetic (EM) analysis of the system at hand; therefore, their explicit handling is inconvenient. A workaround is the penalty function approach where the primary objective (typically, antenna size) is complemented by additional terms quantifying possible constraint violations. The penalty coefficients that determine contributions of these terms are normally adjusted manually, which hinders precise control over antenna performance figures and often leads to inferior results in terms of achieved miniaturization rates. This paper proposes a novel algorithm featuring an automated adjustment of the penalty factors throughout the optimization process. Our methodology is validated using three broadband antenna structures. The obtained results demonstrate that the presented adaptive adjustment permits a precise control over the constraint violations while leading to better miniaturization rates as compared to manual penalty term setup.

Keywords: antenna miniaturization; compact antennas; EM-driven design; constrained optimization; penalty functions; constraint violations

1. Introduction

Accommodation along with the integration requirements of antennas with the circuit parts has rendered miniaturization a necessity in applications such as wireless communications, internet of things, or portable and on-body devices [1,2]. As the majority of antenna performance figures (reflection, gain, bandwidth, radiation efficiency, radiation pattern) are linked to the physical size [3], a miniaturization task is far from trivial. Miniaturization techniques based on utilization of high-permittivity substrates [3,4], have been successful in reducing the antenna size at the expense of degrading the bandwidth. Several other techniques based on geometrical modifications of the antenna structure have been proposed, including the use of meandered traces [1,5], introduction of corrugations in the ground plane and the radiator [6,7], or incorporation of L-shaped slits [1].

Although the aforementioned techniques provide degrees of freedom to work out a compromise between the antenna size and the electromagnetic (EM) performance, they also contribute to the complexity of the antenna design process, primarily due to an increase in the number of geometry parameters. The unintuitive interrelations between the antenna dimensions and the performance figures virtually eliminate the possibility of finding an optimum design using traditional tuning techniques, especially when multiple objectives are to be taken into account. The only way to produce the optimum design is through simultaneous numerical optimization of all geometrical parameters. Depending on the

antenna type, available initial design, and the design goals and constraints, this can be done with either global [8–10], quasi-global [11], or local search routines [12,13].

The reliability of the optimization process is strongly correlated with the accuracy of the computational model, which, most often, is based on full-wave electromagnetic (EM) analysis. Conventional optimization routines require numerous EM simulations, which entails significant computational costs. To alleviate this difficulty, a number of techniques have been proposed, including adjoint sensitivities [14–19], surrogate-based methods involving multi-fidelity simulation models [13,20–22], response surface approximations [23], several variations of space mapping (SM) [24] (e.g., aggressive space mapping [25], implicit SM [26]), feature-based optimization [27], but also machine learning methods [28,29], and surrogate-assisted versions of nature-inspired algorithms [30,31].

In the context of EM-driven antenna miniaturization, explicit reduction of the structure size (footprint area, volume) is the preferred approach because it enables direct control over the primary objective. Notwithstanding, the problem becomes a constrained task as electrical and field performance figures (impedance matching, gain, axial ratio) have to be controlled so that the appropriate levels thereof are ensured [32–34]. The principal inconvenience is that the majority of constraints are expensive to evaluate, i.e., require EM analysis of the antenna. A workaround is to handle the constraints in an implicit manner using the penalty function approach [34]. Therein, a satisfaction of the constraints is enforced by complementing the main objective with the contributions proportional to constraint violations (evaluated using suitable metrics).

The appropriate adjustment of the penalty coefficients is a non-trivial task. A coefficient that is too small, leads to underestimating the contributions of constraint violations to the objective function. Conversely, if the penalty coefficient is too large, the objective function becomes extremely steep in the vicinity of the feasible region boundary, which may cause a premature termination of the optimization process. A workaround was proposed in [35], where the acceptance threshold for maximum in-band reflection has been adaptively adjusted to facilitate exploration of the feasible space. An alternative constraint-oriented objective function has been adopted in [34] to identify the constrained optimum in the boundary of the feasible region [34], or to enable objective relaxation by switching between miniaturization and the electrical performance figures of interest [10]. However, in all cases, the efficacy of the optimization process relies on a manual selection of the penalty factors.

The discussion above indicates that the appropriate setup of the objective function, in particular, the penalty terms, is of paramount importance for the reliability and overall performance of EM-driven antenna miniaturization. This paper proposes a novel algorithm for explicit antenna size reduction, which features automated adjustment of the penalty factor values in the course of the optimization process. The adjustment process is focused on identifying the optimum penalty factor values based on current constraint violations and eliminates the need for manual, trial-and-error efforts. This, in turn, allows fulfillment of constraint satisfaction (up to the requested tolerance) while leading to improved miniaturization rates as compared to the conventional approach.

Our methodology is validated using three broadband antenna structures optimized for minimum size. Extensive benchmarking indicates that the adaptive adjustment of the penalty factors allows for a precise control over the constraint violations while leading to overall better results in terms of the achievable miniaturization rates (as compared to algorithms using fixed penalty terms).

2. Optimization-Based Antenna Miniaturization

This section recalls the formulation of EM-driven antenna miniaturization as a constrained numerical problem with explicit size reduction. Subsequently, we present the reference trust-region-based algorithm, as well as outline the proposed algorithm for automated adjustment of the penalty factors.

2.1. Problem Formulation

We will use $R(x)$ to denote the response of the EM simulation model of the antenna structure of interest. Here, x is a vector of the geometry parameters of the structure, which are to be adjusted in the course of the optimization process. The optimization problem at hand is to minimize the antenna size $A(x)$ while ensuring acceptable levels of performance figures pertinent to electrical and field characteristics of the structure, e.g., of the standard form.

$$s_j(x) \leq S_j, j = 1, \ldots, k \tag{1}$$

Note that $s_j(x)$ are evaluated based on the simulated antenna response $R(x)$. As mentioned before, the penalty function approach provides convenient means of handling the performance constraints, in which case the objective function can be defined as

$$U_A(R(x)) = A(x) + \beta_1 c_1(x)^2 + \ldots + \beta_k c_k(x)^2 \tag{2}$$

where k is the total number of constraints, $c_j(x)$ is the penalty function quantifying violation of the jth constraint (1), with β_j being the corresponding penalty factor. The penalty functions only contribute to (2) if the corresponding constraints are violated, i.e., $s_j(x) > S_j$. A typical definition of the penalty function would be the one measuring a relative violation of the constraint, i.e.,

$$c_j(x) = \max\left\{\frac{\gamma_j}{S_j}, 0\right\} \tag{3}$$

where γ_j is the constraint violation defined as

$$\gamma_j = s_j(x) - S_j \tag{4}$$

The design task is defined as a nonlinear minimization problem of the form.

$$x^* = \underset{x \in X}{\operatorname{argmin}} U_A(R(x)) \tag{5}$$

where X is the parameter space, typically determined by the lower and upper bounds for antenna geometry parameters x. It should be noted that without the penalty function approach, the design task would be subject to additional constraints (1). Whereas, when using (1)–(3), it becomes an unconstrained problem, apart from the aforementioned box constraints.

2.2. Trust-Region Gradient-Based Algorithm

The standard trust-region-based algorithm [33] is employed in this work as the core optimization engine. The procedure approximates x^* with a series $x^{(i)}$, $i = 0, 1, \ldots$, obtained by solving

$$x^{(i+1)} = \arg \min_{x; \, ||x-x^{(i)}|| \leq \delta} U_A(L^{(i)}(x)) \tag{6}$$

In (6), $L^{(i)}(x)$ is a linear (first-order Taylor) approximation of $R(x^{(i)})$. The candidate solution rendered by (6) is only accepted if $U_A(R(x^{(i+1)})) < U_A(R(x^{(i)}))$. The trust region radius δ is adaptively adjusted based on a gain ratio calculated as

$$\rho_A = \frac{U_A(R(x^{(i+1)})) - U_A(R(x^{(i)}))}{U_A(L^{(i)}(x^{(i+1)})) - U_A(L^{(i)}(x^{(i)}))} \tag{7}$$

The algorithm is terminated when the trust region radius is diminished below a certain user-defined limit δ_1, or convergence in the argument is achieved, i.e., $||x^{(i+1)} - x^{(i)}|| \leq \delta_x$.

3. Size Reduction with Adaptive Penalty Coefficients

Formulation (1) of the objective function offers an efficient way of handling the design constraints, especially those that are expensive to evaluate (i.e., require EM analysis of the antenna). Notwithstanding, an appropriate setup of penalty coefficients is a non-trivial matter and may affect the performance of the optimization process as elaborated on in Section 1.

This work proposes a novel algorithm incorporating adaptive adjustment of penalty coefficients in the course of the optimization run. The presented procedure eliminates the need for trial-and-error objective function setup as well as leads to improved antenna miniaturization rates, as demonstrated in Section 4. In the following, we discuss the penalty factor adjustment concept (Section 3.1), the adjustment procedure (Section 3.2), as well as the complete optimization algorithm (Section 3.3).

This section recalls the formulation of EM-driven antenna miniaturization as a constrained numerical.

3.1. Adaptively Adjusted Penalty Factors. The Concept

As mentioned before, an appropriate setup of penalty coefficients is of utmost importance for the performance of the optimization process. On the one hand, having the coefficients set at too low values will result in excessive violations of the constraints. On the other hand, the values that are too high make the optimization problem numerically challenging, especially in the context of size reduction, which requires exploration of the feasible region boundary.

The objective of the technique presented in this is to automate the process of setting up the penalty coefficients so that their specific values are based on currently detected constraint violations. We use the following prerequisites, which involve the current constraint violations as well as their possible improvements over the last consecutive iterations:

- If the parameter vector $x^{(i+1)}$ produced at the iteration i is feasible from the point of view of the jth constraint, the corresponding penalty coefficient β_j may be reduced;
- If $x^{(i+1)}$ is infeasible but the violation of the jth constraint was reduced to a sufficient extent w.r.t. the $(i–1)$th iteration, the coefficient β_j remains intact;
- If $x^{(i+1)}$ is infeasible and there was no improvement of the jth constraint violation or the improvement was insufficient, the coefficient β_j should be increased.

The above can be viewed as a set of rules (applied to each and every constraint), which are simple, yet allow us to relax the 'pressure' from the penalty terms when the algorithm operates in a feasible region and increase it if the discussed indicators show the lack of improvement in terms of reducing constraint violations. Furthermore, implementing these rules facilitates exploration of the feasible region boundary, which is where the minimum-size design is likely to reside. The notion of the aforementioned sufficient constraint violation improvement will be specified and discussed at length below, along with providing a rigorous formulation of the adjustment rules. To clarify the matter, Figure 1, conceptually illustrates the possible situations and the actions performed.

		Constraint violation improvement	
		Insufficient	Sufficient
Solution feasibility	Infeasible	Increase penalty coefficient	Keep penalty coefficient intact
	Feasible	Decrease penalty coefficient	Decrease penalty coefficient

Figure 1. The concept of adaptive adjustment of penalty coefficients. Shown are the four possible situations concerning constraint violation, and the actions undertaken.

3.2. Adaptively Adjusted Penalty Factors. The Procedure

In order to formalize the set of penalty factor adjustment rules considered in Section 3.1, we need to quantify the sufficient constraint violation improvement, which will be defined, for the *j*th constraint, as

$$\Delta_j = M\gamma_j \tag{8}$$

In (8), γ_j is the absolute violation, cf. (4), whereas $0 < M < 1$ is the improvement factor. We will elaborate on the selection of the value of M later in this section. Both the violation and its improvement are considered in relation to the new parameter vector $x^{(i+1)}$ produced in the *i*th iteration and the previous point $x^{(i)}$, therefore, we will use the superscript $i + 1$ to specify which iteration the above quantities are referring to. In particular, we have γ_j^{i+1} and Δ_j^{i+1} as the constraint violation at $x^{(i+1)}$ and the improvement from $x^{(i)}$ to $x^{(i+1)}$. Similarly, the penalty coefficient for the *j*th constraint at iteration *i* will be denoted as β_j^i.

We are now in a position to formulate the adjustment rules in a rigorous manner. These are described using the following pseudocode:

```
if γj^{i+1} ≤ 0
        βj^{i+1} = βj^i/m_decr;
    else
            if γj^i − γj^{i+1} > Δj^{i+1}
                βj^{i+1} = βj^i;
        else
                βj^{i+1} = βj^i m_incr;
        end
    end
```

The multiplication factors m_{decr} and m_{incr} determine the amount of penalty factor modifications. In our numerical experiments, they are set to $m_{decr} = 1.25$ and $m_{incr} = 5$, but their values are not critical. As discussed in Section 3.1, maintaining fixed penalty coefficients under sufficient constraint violation improvement (while the design is still infeasible) allows us to maintain stability of the optimization process, i.e., the values of β_j do not bounce back and forth (i.e., are immediately increased or decreased after the design crosses the feasible region boundary). This facilitates exploration of the boundary and leads to improved size reduction ratios.

Let us now consider the improvement factor M introduced in (8). Assuming that the parameter vector $x^{(i)}$ resides in the infeasible region, and sufficient constraint violation

improvement is observed for a few consecutive iterations, say, from i to $i + k$, the upper bound on the constraint violation at the iteration $i + k$ can be calculated as

$$\gamma_j^{i+k} \leq M\gamma_j^{i+k-1} \leq M^2\gamma_j^{i+k-2} \leq \ldots \leq M^k\gamma_j^i \qquad (9)$$

As $M < 1$, the constraint violation is reduced at a geometric rate, and the improvement is faster when M is closer to zero. At the same time, satisfaction of the sufficient improvement condition is more demanding for lower values of M. For $M = 0.5$, as selected in this work, we get a balance between the rate of approaching the feasible region boundary (e.g., the constraint violation is reduced to only about six percent of its original values after only four iterations), and the difficulty of satisfying the improvement condition (reduction of only half of the current violation has to be achieved per iteration).

3.3. Optimization Framework

This section provides a summary of the operation of the trust-region gradient-based algorithm incorporating the adaptive adjustment of penalty coefficients as presented in Section 3.2. Here is a summary of the control parameters of the algorithm:

- δ_x, δ_{TR}—termination thresholds (cf. Section 2.1);
- m_{inc}, m_{dec}—increase and decrease factors for the automated adjustment of penalty coefficients (cf. Section 3.2);
- M—a factor used to determine sufficient constraint violation improvement (cf. Section 3.2);
- β_j^{max}, β_j^{min}—maximum and minimum values of penalty coefficients; $j = 1, \ldots, k$;
- β_j^0—initial values of the penalty coefficients; $j = 1, \ldots, k$;

The algorithm operation has been presented in Figure 2. Additional clarification is provided in the form of a flow diagram in Figure 3. In the algorithm, Steps 1 and 2 are used to initialize the optimization procedure. In Steps 3 and 4, the antenna response and its sensitivity matrix are evaluated using EM analysis. The linear approximation model $L^{(i)}(x)$ of the antenna responses is constructed in Step 5, along with its corresponding objective function $U_A(L^{(i)}(x))$ (Step 6). The candidate design is produced in Step 7 by minimizing $U_A(L^{(i)}(x))$. It is validated in Steps 8 and 9, where the gain ratio is calculated for the purpose of either accepting or rejecting $x^{(i+1)}$ and computing constraint violation improvements (Step 10). The latter is then used in Step 11 to update the penalty coefficients. The termination condition for the procedure is convergence in argument or reducing the trust region size beyond the user-defined threshold (Step 12).

1. Set the iteration counter $i = 0$; provide $x^{(i)}$;

2. Set m_{inc}, m_{dec}, β_j^i, S_j, β_j^{max}, β_j^{min}, $j = 1, \ldots, k$;

3. Evaluate antenna response $R(x^{(i)})$;

4. Evaluate antenna sensitivities $J(x^{(i)})$ using finite differentiation;

5. Construct a linear model $L^{(i)}(x) = R(x^{(i)}) + J(x^{(i)}) \cdot (x - x^{(i)})$;

6. Construct objective function $U_A(L^{(i)}(x)) = A(x) + \beta_1^i . c_{L,1}(x)^2 + \ldots + \beta_k^i . c_{L,k}(x)^2$;

7. Solve $\quad x^{(i+1)} = \arg \min\limits_{x;\; \|x - x^{(i)}\| \leq \delta} U_A(L^{(i)}(x))$;

8. Evaluate antenna response $R(x^{(i)})$;

9. Compute gain ratio ρ_A and update TR size $\quad \rho_A = \dfrac{U_A(R(x^{(i+1)})) - U_A(R(x^{(i)}))}{U_A(L^{(i)}(x^{(i+1)})) - U_A(L^{(i)}(x^{(i)}))}$;

    if $\rho_A < 0$,

        Go to 7;

    else

        $\Delta_j^{i+1} = M(s_j(x^{(i)}) - S_j)$;

    end

10. **if** $\gamma_j^i - \gamma_j^{i+1} > \Delta_j^{i+1}$ && $\gamma_j^{i+1} > 0$

        $\beta_j^{i+2} = \beta_j^{i+1} m_{incr}$; $i = i + 1$;

    elseif $\quad \gamma_j^{i+1} \leq 0$

        $\beta_j^{i+1} = \beta_j^i / m_{decr}$; $i = i + 1$;

    else

        $\beta_j^{i+1} = \beta_j^i$; $i = i + 1$;

    end

11. **if** $||x^{(i+1)} - x^{(i)}|| \leq \delta_x$ OR TR $\leq \delta_{TR}$

        END;

    else

        Go to 4;

    end

Figure 2. The concept of adaptive adjustment of penalty coefficients. Shown are the four possible situations concerning constraint violation, and the actions undertaken.
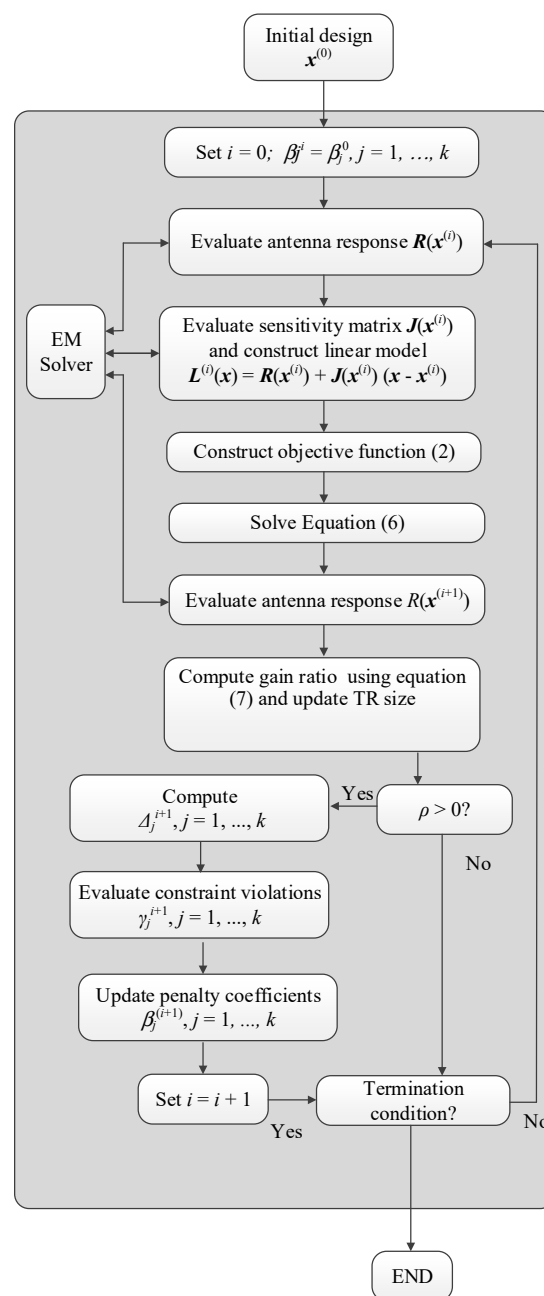
Figure 3. Operational flow of the proposed size-reduction algorithm with automated adjustment of penalty factors.

4. Verification Case Studies

This section provides the results of numerical experiments conducted to validate the proposed procedure for adaptive adjustment of penalty coefficients, introduced in Section 3. The verification studies involve three broadband antennas optimized for minimum size with the maximum acceptable in-band reflection considered as the only constraint. The results are compared with those obtained for manually set penalty terms, ranging from the relaxed to very tight constraint satisfaction conditions. All of the considered antenna structures have been previously described in the literature and experimentally validated therein [36–38]. Consequently, no experimental results are provided here.

The remaining part of this section is organized as follows. The experimental setup is described in Section 4.1, whereas the benchmark antennas are described along with

the numerical results in Section 4.2, Section 4.3, and Section 4.4, respectively. Section 4.5 provides a detailed discussion.

4.1. Experimental Setup

The objective is to optimize the considered antennas for minimum size, as determined by the substrate area $A(x)$. The optimization process is subject to a single constraint imposed on the antenna reflection coefficient $|S_{11}(x)|$.

More specifically, we have $s_1(x) \leq -10$, where $s_1(x)$ stands for the maximum value of $|S_{11}(x)|$ within the operating frequency range of the antennas, here, 3.1 GHz to 10.6 GHz. Correspondingly, a single penalty coefficient is used, denoted as β_1. The penalty function c is defined based on a relative violation of the constraint, cf. (3).

The performance of the algorithm is evaluated statistically through multiple runs initiated from random starting points. This allows for reducing a possible bias associated with a particular choice of the initial design. The performance figures include the antenna size averaged over ten independent runs, and the average value of constraint violation, both evaluated at the final design yielded by the optimization algorithm.

The proposed algorithm is compared to the standard TR algorithm executed for different (fixed) values of the penalty coefficients $\beta = 10^q$, $q = 2, 3, 4, 5, 6$. The values set for the control parameters are $\beta_j^{min} = 100$, $\beta_j^{max} = 10^6$, and $\beta_j^0 = 100$. Other parameters are set as in Section 3.2. The termination thresholds are set to $\delta_x = \delta_{TR} = 10^{-3}$.

4.2. Antenna I

Figure 4 shows the geometry of the first benchmark structure (Antenna I) [36]. It is a broadband monopole antenna operating within the UWB band (3.1 GHz to 10.6 GHz). The design parameters are $x = [L_0\ dR\ R\ r_{rel}\ dL\ dw\ L_g\ L_1\ R_1\ dr\ c_{rel}]^T$ (all dimensions in mm). The feed line width is fixed to $w_0 = 1.7$ mm. The antenna is implemented on RF-35 substrate with relative permittivity $\varepsilon_r = 3.5$, and thickness $h = 0.762$ mm. The computational model incorporating the SMA connector is simulated in the time-domain solver of CST Microwave Studio.
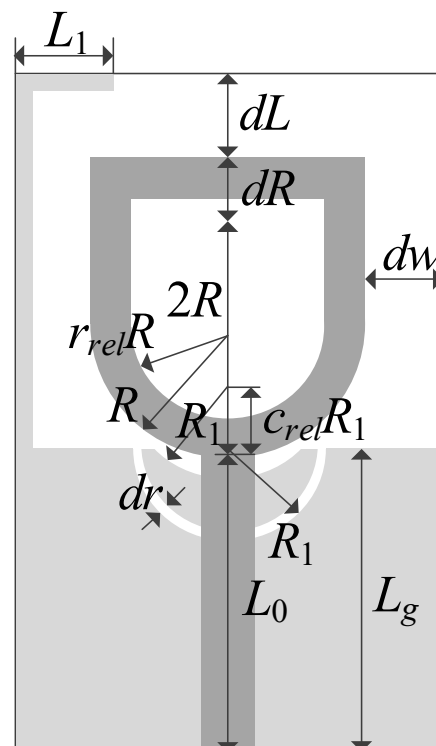
Figure 4. Geometry of Antenna I. The light gray shade is the ground plane on the backside of the structure.

Table 1 provides the data including the average footprint area, average constraint violation, as well as the standard deviations thereof; the latter is used to quantify the repeatability of solutions. Figure 5a illustrates the reflection coefficient of Antenna I at the initial and final designs obtained for a selected run of the proposed optimization algorithm, whereas Figure 5b shows the evolution of the penalty factor throughout the optimization process.

Table 1. Optimization results for antenna I.

Performance Figure	$\beta = 10^2$ (Fixed)	$\beta = 10^3$ (Fixed)	$\beta = 10^4$ (Fixed)	$\beta = 10^5$ (Fixed)	$\beta = 10^6$ (Fixed)	Adaptive β (This Work)
Antenna area [mm²] [1]	113.7	250.4	318.6	331.6	367.6	222.6
Std(A) [2]	9.07	24.0	60.0	63.4	51.9	49.6
Constraint violation γ [dB] [3]	8.4	1.2	0.14	0.10	0.05	0.08
Std(γ) [4]	0.53	0.5	0.1	0.14	0.11	0.06

[1] Average miniaturized antenna area for ten algorithm runs. [2] Standard deviation of the miniaturized antenna area for ten algorithm runs. [3] Average constraint violation for ten algorithm runs. [4] Standard deviation of the constraint violation γ, for ten algorithm runs.

(**a**)　　　　　　　(**b**)

Figure 5. Antenna I: representative run of the proposed algorithm with automated adjustment of the penalty factors: (**a**) reflection coefficient at the initial (- - -) and optimized design (—); (**b**) evolution of the penalty coefficient throughout the algorithm iterations.

4.3. Antenna II

The second verification case is a broadband rectangular-slot monopole antenna (Antenna II), shown in Figure 6 [37], operating in the UWB range. The design parameters of the antenna are $x = [L_g \ L_0 \ L_s \ W_s \ d \ dL \ d_s \ dW_s \ dW \ a \ b]^T$ (all in mm). The feed line width is $W_0 = 3$ mm. The structure is fabricated on FR4 substrate ($\varepsilon_r = 4.3$, $h = 1.55$ mm). The computational model incorporates the SMA connector, and it is simulated using the time-domain solver of CST Microwave Studio. Table 2 gathers the numerical results. Figure 7 shows the antenna responses and evolution of the penalty factor for a representative run of the proposed algorithm.
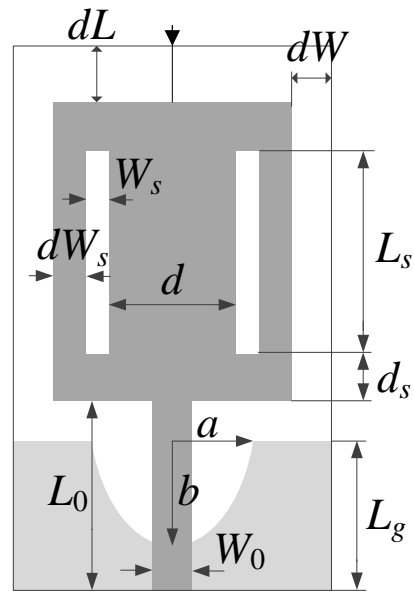
Figure 6. Geometry of Antenna II: the light gray shade is the ground plane on the backside of the structure.

Table 2. Optimization results for antenna II.

Performance Figure	$\beta = 10^2$ (Fixed)	$\beta = 10^3$ (Fixed)	$\beta = 10^4$ (Fixed)	$\beta = 10^5$ (Fixed)	$\beta = 10^6$ (Fixed)	Adaptive β (This Work)
Antenna area [mm^2] [1]	56.1	212.8	225.0	280.1	258.8	180.7
Std(A) [2]	3.8	14.3	25.1	47.4	29.6	11.1
Constraint violation γ [dB] [3]	8.6	1.0	0.15	0.05	0.00	0.17
Std(γ) [4]	0.60	0.4	0.10	0.07	0.01	0.23

[1] Average miniaturized antenna area for ten algorithm runs. [2] Standard deviation of the miniaturized antenna area for ten algorithm runs. [3] Average constraint violation for ten algorithm runs. [4] Standard deviation of the constraint violation γ, for ten algorithm runs.

(**a**)　　　　　　　　　　(**b**)
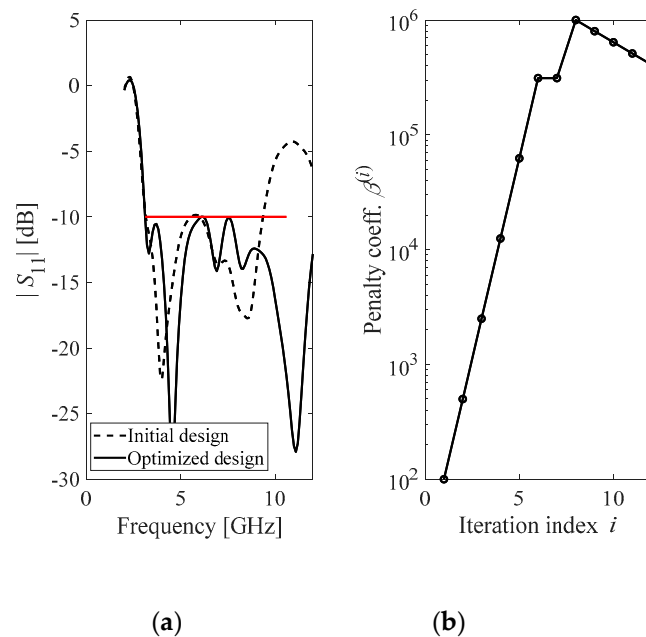
Figure 7. Antenna II: representative run of the proposed algorithm with automated adjustment of the penalty factors: (**a**) reflection coefficient at the initial (- - -) and optimized design (—); (**b**) evolution of the penalty coefficient throughout the algorithm iterations.

4.4. Antenna III

Figure 8 shows Antenna III, the third verification structure. It is also a UWB monopole antenna [38]. The design parameters are $x = [L_0\ g\ a\ l_1\ l_2\ w_1\ o]^T$, (all dimensions in mm). We also have $w_0 = 2o + a$, and $w_r = 1.7$ mm. The antenna is implemented on FR4 substrate ($\varepsilon_r = 4.3$, $h = 7.62$ mm). The computational model incorporates the SMA connector, and it is simulated in the time-domain solver of CST Microwave Studio.
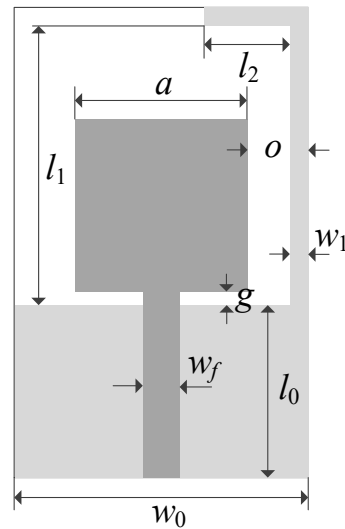
Figure 8. Geometry of Antenna III: the light gray shade is the ground plane on the backside of the structure.

Table 3 shows the numerical results, arranged the same way as for Antennas I and II. Antenna responses and evolution of the penalty coefficient for a selected algorithm run have been shown in Figure 9.
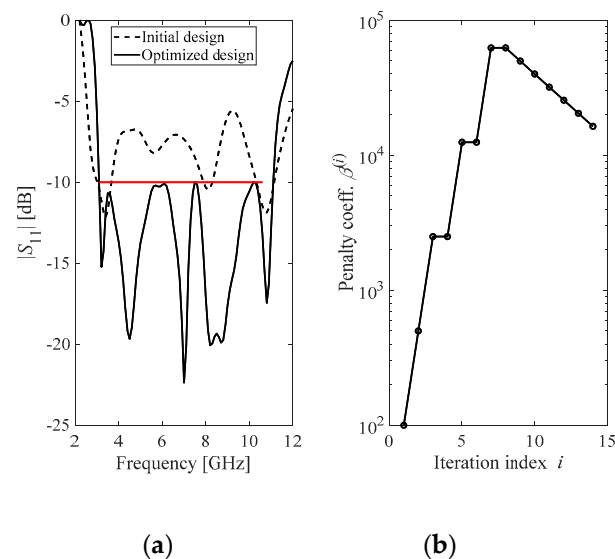

(**a**)   (**b**)
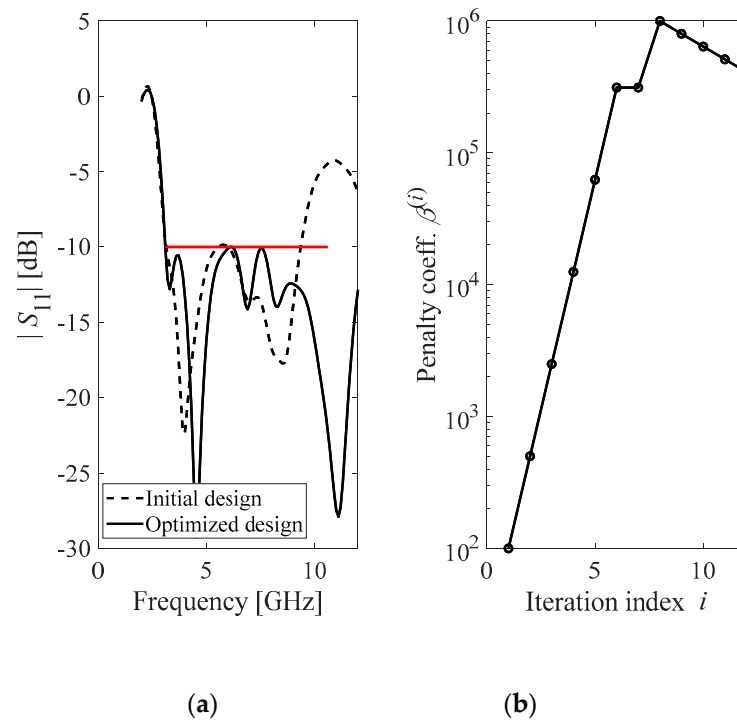
Figure 9. Antenna III: representative run of the proposed algorithm with automated adjustment of the penalty factors: (**a**) reflection coefficient at the initial (- - -) and optimized design (—); (**b**) evolution of the penalty coefficient throughout the algorithm iterations.

Table 3. Optimization results for antenna III.

Performance Figure	$\beta = 10^2$ (Fixed)	$\beta = 10^3$ (Fixed)	$\beta = 10^4$ (Fixed)	$\beta = 10^5$ (Fixed)	$\beta = 10^6$ (Fixed)	Adaptive β (This Work)
Antenna area [mm^2] [1]	305.4	318.1	317.7	318.8	320.9	304.43
Std(A) [2]	49.7	42.6	42.3	43.3	45.8	37.2
Constraint violation γ [dB] [3]	6.7	1.2	0.4	0.05	0.06	0.45
Std(γ) [4]	1.7	0.4	0.2	0.07	0.3	0.49

[1] Average miniaturized antenna area for ten algorithm runs. [2] Standard deviation of the miniaturized antenna area for ten algorithm runs. [3] Average constraint violation for ten algorithm runs. [4] Standard deviation of the constraint violation γ, for ten algorithm runs.

4.5. Discussion

The analysis of results provided in Section 4 allows for several conclusions regarding the importance of the automated adjustment of the penalty factors, as well as the performance-wise advantages of the proposed adaptive algorithm over the fixed-setup approach. To facilitate the interpretation, a graphical illustration of the data from Tables 1–3 has been provided in Figure 10. Therein, the average antenna footprint area along with the average constraint violation is shown versus the penalty coefficient for fixed-setup optimization runs. The horizontal lines represent the antenna area and constraint violation obtained for the automated adjustment procedure. The range of these lines is representative of the span in which the penalty coefficient varies throughout the optimization iterations. As it can be seen, not only the optimum value of the penalty coefficient, but also the span is problem-dependent and may not even reach the maximum set value in some cases. The following observations can be formulated:

- Although the optimum value of penalty coefficient in the fixed-setup optimization seems to be about $\beta = 10^4$ for Antenna I, between $\beta = 10^3$ and $\beta = 10^4$ for Antenna II, and between $\beta = 10^4$ and $\beta = 10^5$ for Antenna III, considering the achievable miniaturization rates along with sufficient constraint satisfaction, the optimum value of penalty coefficient is problem-dependent. The optimum values should be identified for particular iterations of the optimization process and they are generally dependent on the status of constraint violation.

- In both fixed and automated adjustment setups, using a penalty coefficient lower than the optimum value, results in significant constraint violation. As for the former, the violation can easily become as high as five times of the tolerance threshold or even more. Antennas I and II are representative examples of this design quality degradation.

- Automated adjustment of penalty coefficients seeks to improve the final design quality by the optimum value of penalty coefficients at the level of iterations of the optimization process. This, in turn, permits a better control of constraint violations along with better achievable miniaturization rates.

- The performance improvements are significant. For the corresponding levels of constraint violations, the procedure proposed in this work leads to antenna footprints that are smaller by 110, 44, and 13 mm^2 for Antenna I, II, and III, respectively.
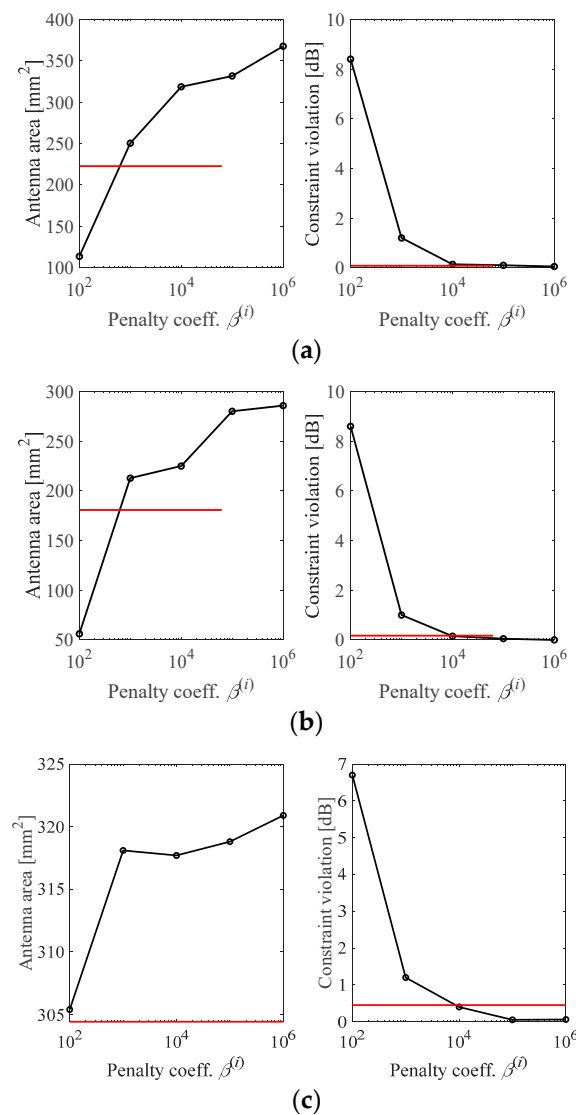
Figure 10. The average antenna footprint (left panel), and the average constraint violations (right panel) versus discrete values of the penalty coefficient utilized by the fixed-setup algorithm. The horizontal lines on the left show the average antenna footprint, whereas the ones on the right denote the average constraint violation for the designs obtained using the proposed automated adjustment algorithm. Defining the same constraint violation levels as the comparison criterion, the proposed algorithm outperforms the fixed-setup algorithm in terms of the achievable miniaturization rates, even under the optimistic scenario of having set beforehand the optimum value of β for the latter: (**a**) Antenna I, (**b**) Antenna II, (**c**) Antenna III.

5. Conclusions

This paper proposed a novel algorithm for optimization-based antenna miniaturization using local trust region gradient search routines. Our methodology incorporates the frameworks employing a penalty function approach for handling design constraints in an implicit manner. Therein, the appropriate adjustment of the objective function, specifically the penalty terms, has a strong correlation with the reliability as well as the efficacy of the optimization process, both in terms of the achievable miniaturization rate and constraint satisfaction. The proposed methodology effectively eliminates the need for trial-and-error efforts by automated adaptive adjustment of the penalty coefficients. The latter is based on a sufficiency of constraint violation improvement between consecutive iterations throughout the entire optimization process. The optimum values of penalty coefficients are identified for particular iterations and consequently enabling the optimization process to realize a

precise control over the constraint violations while leading to overall better results in terms of the achievable miniaturization rates as compared to the fixed adjustment setup.

The proposed procedure has been validated using three broadband antenna structures optimized for minimum size with constraints imposed on their maximum in-band reflection coefficient. Benchmarking the against the fixed penalty coefficient setups indicates superiority of automated adaptive adjustment setup in terms of achieving smaller size with a precise control over the constraint violations. The future work will involve a more comprehensive validation of the methodology in applications featuring multiple constraints, e.g., circularly polarized antennas, high-gain antennas, etc.

References

1. Le, T.T.; Yun, T.-Y. Miniaturization of a dual band wearable antenna for dual-band WBAN applications. *IEEE Antennas Wirel. Propag. Lett.* **2020**, *19*, 1452–1456. [CrossRef]
2. Agneessens, S.; Rogier, H. Compact half diamond dual-band textile HMSIW on-body antenna. *IEEE Trans. Antennas Propag.* **2014**, *62*, 2374–2381. [CrossRef]
3. Upadhyay, D.; Dwivedi, R.P. Antenna miniaturization techniques for wireless applications. In Proceedings of the 2014 Eleventh International Conference on Wireless and Optical Communication Networks (WOCN), Vijayawada, India, 11–13 September 2014; pp. 1–4.
4. Oraizi, H.; Rezai, B. Dual-banding and miniaturization of planar triangular monopole antenna by inductive and dielectric loadings. *IEEE Antennas Wirel. Propag. Lett.* **2013**, *12*, 1594–1597. [CrossRef]
5. Oh, J.; Sarabandi, K. A topology-based miniaturization of circularly polarized patch antennas. *IEEE Trans. Antennas Propag.* **2013**, *61*, 1422–1426. [CrossRef]
6. Abbosh, A.M. Miniaturized microstrip-fed tapered-slot antenna with altrawideband performance. *IEEE Antennas Wirel. Propag. Lett.* **2009**, *8*, 690–692. [CrossRef]
7. Abbosh, A.M. Miniaturization of planar ultrawideband antenna via corrugation. *IEEE Antennas Wirel. Propag. Lett.* **2008**, *7*, 685–688. [CrossRef]
8. Al-Azza, A.A.; Al-Jodah, A.A. Spider monkey optimization: A novel technique for antenna optimization. *IEEE Antennas Wirel. Propag. Lett.* **2016**, *15*, 1016–1019. [CrossRef]
9. Lalbakhsh, A.; Afzal, M.U.; Esselle, K.P. Multiobjective particle swarm optimization to design a time-delay equalizer metasurface for an electromagnetic band-gap resonator antenna. *IEEE Antennas Wirel. Propag. Lett.* **2017**, *16*, 912–915. [CrossRef]
10. Goudos, S.K.; Siakavara, K.; Samaras, T.; Vafiadis, E.E.; Sahalos, J.N. Self-adaptive differential evolution applied to real-valued antenna and microwave design problems. *IEEE Trans. Antennas Propag.* **2011**, *59*, 1286–1298. [CrossRef]
11. Tomasson, J.A.; Koziel, S.; Pietrenko Dabrowska, A. Quasi-global optimization of antenna structures using principal components and affine subspace-spanned surrogates. *IEEE Access* **2020**, *8*, 50078–50084. [CrossRef]
12. Koziel, S.; Pietrenko Dabrowska, A. Performance-based nested surrogate modeling of antenna input characteristics. *IEEE Trans. Antennas Propag.* **2019**, *67*, 2904–2912. [CrossRef]
13. Song, Y.; Cheng, Q.S.; Koziel, S. Multi-fidelity local surrogate model for computationally efficient microwave component design optimization. *Sensors* **2019**, *19*, 3023. [CrossRef]
14. Director, S.; Rohrer, R. The generalized adjoint network and network sensitivities. *IEEE Trans. Circuit Theory* **1969**, *16*, 318–323. [CrossRef]
15. Paronneau, O. *Optimal Shape Design for Elliptic Systems*; Springer: Berlin/Heidelberg, Germany, 1982; Volume 3, pp. 42–66.
16. Jameson, A. Aerodynamic design via control theory. *J. Sci. Comput.* **1988**, *3*, 233–260. [CrossRef]
17. El Sabbagh, M.A.; Bakr, M.H.; Nilolova, N.K. Sensitivity analysis of the scattering parameters of microwave filters using the adjoint network method. *Int. J. RF Microw. Comput. Aided Eng.* **2006**, *16*, 569–606. [CrossRef]

18. Papadimitriou, D.; Giannakoglou, K. Aerodynamic shape optimization using first and second order adjoint and direct approaches. *Arch. Comput. Methods Eng.* **2008**, *15*, 447–488. [CrossRef]

19. Toivann, J.I.; Mäkinen, R.A.E.; Järvenpää, S.; Ylä-Oijala, P.; Rahola, J. Electromagnetic sensitivity analysis and shape optimization using method of moments and automatic differentiation. *IEEE Trans. Antennas Propag.* **2009**, *57*, 168–175. [CrossRef]

20. Koziel, S.; Ogurtsov, S. Model management for cost-efficient surrogate-based optimization of antennas using variable-fidelity electromagnetic simulations. *IET Microw. Antennas Propag.* **2012**, *6*, 1643–1650. [CrossRef]

21. Koziel, S.; Ogurtsov, S. Low-fidelity model considerations for simulation-based optimization of miniaturized wideband antennas. *IET Microw. Antennas Propag.* **2012**, *12*, 1613–1619. [CrossRef]

22. Koziel, S.; Ogurtsov, S. Robust multi-fidelity simulation-driven optimization of microwave structures. In Proceedings of the 2010 IEEE MTT-S International Microwave Symposium, Anaheim, CA, USA, 23–28 May 2010; pp. 201–204.

23. De Villiers, D.I.L.; Couckuyt, I.; Dhaene, T. Multi-objective optimization of reflector antennas using Kriging and probability of improvement. In Proceedings of the IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, San Diego, CA, USA, 9–14 July 2017; pp. 985–986.

24. Bandler, J.W.; Biernacki, R.M.; Chen, S.H.; Grobelny, P.A.; Hemmers, R.H. Space mapping technique for electromagnetic optimization. *IEEE Trans. Microw. Theory Technol.* **1994**, *42*, 2536–2544. [CrossRef]

25. Ossorio, J.; Melgarejo, J.C.; Boria, V.E.; Guglielmi, M.; Bandler, J.W. On the alignment of low-fidelity and high-fidelity simulation spaces for the design of microwave waveguide filters. *IEEE Trans. Microw. Theory Technol.* **2018**, *66*, 5183–5196. [CrossRef]

26. Bandler, J.W.; Cheng, Q.S.; Georgieva, N.; Ismail, M.A. Implicit space mapping EM-based modeling and design exploiting preassigned parameters. In Proceedings of the 2002 IEEE MTT-S International Microwave Symposium Digest (Cat. No. 02CH37278), Seattle, WA, USA, 2–7 June 2002; Volume 2, pp. 713–716.

27. Koziel, S.; Ogurtsov, S. Fast surrogate-assisted simulation-driven optimization of add-drop resonators for integrated photonic circuits. *IET Microw. Ant. Propag.* **2015**, *9*, 672–675. [CrossRef]

28. Xiao, L.; Shao, W.; Ding, X.; Wang, B. Dynamic adjustment kernel extreme learning machine for microwave component design. *IEEE Trans. Microw. Theory Technol.* **2018**, *66*, 4452–4461. [CrossRef]

29. Zhang, C.; Zhu, Y.; Cheng, Q.; Fu, H.; Ma, J.; Zhang, Q. Extreme learning machine for the behavioral modeling of RF power amplifiers. In Proceedings of the 2017 IEEE MTT-S International Microwave Symposium (IMS), Honolulu, HI, USA, 4–9 June 2017; pp. 558–561.

30. Lim, D.K.; Woo, D.K.; Yeo, H.K.; Jung, S.Y.; Ro, J.S.; Jung, H.K. A novel surrogate-assisted multi-objective optimization algorithm for an electromagnetic machine design. *IEEE Trans. Magn.* **2015**, *51*, 1–4.

31. Toktas, A.; Ustun, D.; Tekbas, M. Multi-objective design of multi-layer radar absorber using surrogate-based optimization. *IEEE Trans. Microw. Theory Technol.* **2019**, *67*, 3318–3329. [CrossRef]

32. Koziel, S.; Pietrenko Dabrowska, A. Reduced-cost design closure of antennas by means of gradient search with restricted sensitivity update. *Metrol. Meas. Syst.* **2019**, *26*, 595–605.

33. Haq, M.A.; Koziel, S. Ground plane alterations for design of high-isolation compact wideband MIMO antenna. *IEEE Access* **2018**, *6*, 48978–48983. [CrossRef]

34. Johanesson, D.O.; Koziel, S. Feasible space boundary search for improved optimization-based miniaturization of antenna structures. *Microw. Antennas Propag.* **2018**, *12*, 1273–1278. [CrossRef]

35. Johanesson, D.O.; Koziel, S.; Bekasiewicz, A. EM-driven constrained miniaturization of antennas using adaptive in-band reflection acceptance threshold. *Int. J. Numer. Model. Electron. Netw. Devices Fields* **2018**, *32*, e2513. [CrossRef]

36. Alsath, M.G.N.; Kanagasabai, M. Compact UWB monopole antenna for automotive communications. *IEEE Trans. Antennas Propag.* **2015**, *63*, 4204–4208. [CrossRef]

37. Haq, M.A.; Koziel, S. Simulation-based optimization for rigorous assessment of ground plane modifications in compact UWB antenna design. *Int. J. RF Microw. Comput. Aided Eng.* **2018**, *28*, e21204. [CrossRef]

38. Koziel, S.; Bekasiewicz, A. Comprehensive comparison of compact UWB antenna performance by means of multi-objective optimization. *IEEE Trans. Antennas Propag.* **2017**, *65*, 3427–3436. [CrossRef]

4.1   Design Feasibility

This section elaborates on how feasibility is defined in each chapter, specifically Chapters 4 to 8. The aim is to clarify discrepancies between seemingly different feasibility definitions in different scenarios. To this end, an exemplary parameter space is shown in Fig. 4.1 with feasible and infeasible region indicated along with the boundary region marked in gray. The feasible region corresponds to a region in the parameter space where there is no violation of the constraints. In the case of the presence of multiple constraints, a specific region can be feasible with respect to the constraint, which is inactive, or in other words, there is no violation. At the same time, the same region can be infeasible with respect to the other constraints, for which the violations are present.

In Chapter 4, the developed algorithm is applied to benchmark structures with only a single constraint present. Therefore, the feasible region is simply defined as the specific region of the parameter space where the constraint is inactive.
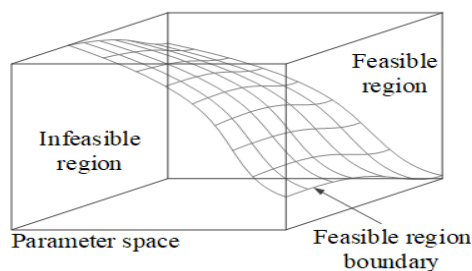
Figure 4.1:   An exemplary parameter space with indicated feasible and infeasible regions. The boundary is shown in gray.

Chapter 5 utilizes a similar definition of the feasibility. The only distinguishing factor is the presence of more than one constraint in the benchmark examples. In this case, the overall feasible region of interest is defined as the intersection of the individual feasible regions corresponding to each constraint. The resulting feasible region is the one where all the existing constraints are inactive, i.e., there is no violation of any of the constraints therein.

Similar definition of the feasible region is used in Chapter 6. However, a novel metric is defined that measures how the current design approaches or moves away from the feasible region in a relative manner. In other words, the metric provides a quantification of the feasibility status using an exponential function with the exponent accounting for aggregated constraint violation level. The exponent is formulated as a maximization problem where it takes the value of the maximum constraint violation among all the constraints. This quantification is used as a criterion for model fidelity selection. Further details are provided in the corresponding chapter.

In Chapters 7 and 8, the optimization problem is to handle a combination of equality and inequality constraints. The general description of the feasibility provided for Chapter 5 is applicable to Chapters 7 and 8 as well. The only difference is the specific geometry of the

feasible region corresponding to the equality constraints, which is a thin subset of the parameter space. Therefore, the intersection of the feasible regions corresponding to the inequality constraints, and the ones corresponding to the equality constraints, represents a thin subset of the design space.

4.2   Selection of Multiplication Factors for Penalty Factor Adjustment

This section provides an elaboration on how the values of the multiplication factors employed in our approach are chosen in the proposed algorithm concerning adaptive penalty factor adjustment.

The abovementioned multiplication factors m_{decr} and m_{incr} determine the amount of penalty factor modifications throughout the optimization process. In our experimental setups, the corresponding values are set to $m_{decr} = 1.25$ and $m_{decr} = 5$. However, the selection of the specific values can fall within a reasonable range such that a back and forth bouncing of β_i does not occur. For further elaboration on the algorithmic stability considerations of the proposed procedure, we provide a specific example with the values $m_{decr} = 5$ and $m_{decr} = 10$: this can be considered as an extreme selection of the values as the large multiplication factors will result in abrupt changes in penalty factor values, i.e. when the design is feasible, the penalty factor value decreases rapidly to a much lower value, resulting in the design to bounce deep back in the infeasible region in the next iteration. On the other hand, when the design is infeasible, there will be an abrupt large increase in the penalty factor value, resulting in the design to bounce forward in the feasible region in the next iteration. The back-and forth switching of the feasible and infeasible regions will result in instability of the optimization process. Therefore, the selection of the multiplication factors should be done in such a way that it avoids immediate increase or decrease of β_i after the design crosses the feasible region boundary. This facilitates exploration of the feasible region boundary and leads to improved size reduction ratios.

Chapter 5

5   Paper # 2

Marzieh Mahrokh and Slawomir Koziel

Explicit Size-Reduction of Circularly Polarized Antennas through Constrained Optimization with Penalty Factor Adaptation

Explicit Size-Reduction of Circularly Polarized Antennas Through Constrained Optimization With Penalty Factor Adjustment

MARZIEH MAHROKH[1] **AND SLAWOMIR KOZIEL**[1,2], **(Senior Member, IEEE)**

[1]Engineering Optimization and Modeling Center, Reykjavik University, 102 Reykjavik, Iceland
[2]Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, 80-233 Gdańsk, Poland

Corresponding author: Marzieh Mahrokh (marziehm@ru.is)

ABSTRACT Modern communication systems of high data capacity incorporate circular polarization (CP) as the preferred antenna radiation field configuration. In many applications, integration of the system circuitry with antennas imposes size limitations on CP radiators, which makes their development process a challenging endeavor. This can be mitigated by means of simulation-driven design, specifically, constrained numerical optimization. Majority of the performance-related constraints are expensive to evaluate, i.e. require full-wave electromagnetic (EM) analysis of the system. Their practical handling can be realized using a penalty function approach, where the primary objective (antenna size reduction) is complemented by contributions proportional to properly quantified constraint violations. The coefficients determining the contribution of the penalty terms are normally set up using designer's experience, which is unlikely to render their optimum values in terms of the achievable miniaturization rates as well as constraint satisfaction. This paper proposes a procedure for automated penalty factor adjustment in the course of the optimization process. Our methodology seeks for the most suitable coefficient levels based on the detected constraint violations and feasibility status of the design. It is validated using two CP antenna structures. The results demonstrate a possibility of a precise constraint control as well as superior miniaturization rates as compared to the manual penalty term setup.

INDEX TERMS Circular polarization antennas, compact antennas, constrained optimization, penalty functions, simulation-driven design.

I. INTRODUCTION

With the growing demands for reliable high-capacity data transfer, an increasing attention has been given to incorporation of CP antennas into modern communication systems. The orthogonal radiation field configuration can assure the reliability of these systems due to attractive features, including a reduction of polarization mismatch and multipath losses [1], as well as mitigation of the Faraday's effect [2]. The continuing trend towards miniaturization enforces CP antennas to be compatible with space constraints in applications such as Aerospace and Synthetic Aperture Radar (SAR) [3], Global Positioning System (GPS) [4],

picosatellites [5], 5G communication systems [6], or wearable and on-body devices.

While preservation of high CP purity along with satisfying other electrical and field performance requirements is already challenging, ensuring compact size is an additional contribution to the design complexity. Several miniaturization techniques based on topological modifications of the antenna structure have been proposed, including the use of slots and fractals [7], defected ground structure [8], fractal metasurfaces and fractal resonators [9], or mushrooms and reactive impedance surfaces (RIS) [10]. These techniques have been successful in working out a compromise between the compact size and performance figures of CP antenna. Notwithstanding, the evolution of antenna topology into complex multi-parameter geometries hinders the process of finding an optimum design, especially with conventional,

manual or trial-and-error efforts. A workaround is numerical optimization, which, depending on the nature of the design problem and available resources, may resort to either global [11]–[13], or local search [14], [15]. Full-wave EM analysis is most often used as the computational model, the accuracy of which determines the reliability of the optimization process. Yet, EM models tend to be expensive to evaluate. Cost-efficient optimization methods have been developed to mitigate this problem, including space mapping [16], incorporation of adjoint sensitivities [17]–[19], data-driven surrogate-based methods [15], [20]–[22], and machine learning approaches [23], [24].

EM-driven miniaturization is the most efficient when size reduction is explicitly handled as the primary objective. On the other hand, the need for ensuring the appropriate levels of electrical performance figures necessitates constrained optimization, with the constraints being expensive to evaluate. A convenient way of constraint control is the penalty function approach [25]. Therein, properly quantified constraint violations appear as additional terms complementing the main objective. The efficacy of this approach relies on the proper adjustment of the penalty factors. Setting these too large leads to an extreme steepness of the objective function in the vicinity of the feasible region boundary. Having them too small results in excessive constraint violations. A possible workaround is adaptive adjustment of the acceptance threshold for the maximum in-band reflection level [26]. Other approaches include feasible space boundary exploration procedure [25], or alternating the size-reduction- and constraint-improvement-oriented search steps [26]. However, in all these cases, the performance of the optimization process depends on a proper manual selection of the penalty factors.

In the context of constrained optimization using genetic algorithm, the aforementioned problem has been mitigated by incorporating an adaptive, tune-free penalty function [27], non-stationary penalty function [28], or a self-adaptive penalty function [29]. The abovementioned methods have been demonstrated successful in identifying feasible solutions without any manual tuning of the penalty function.

This paper proposes a novel procedure for explicit size-reduction of antenna structures featuring an automated penalty factor adjustment throughout the optimization process. The adjustment process employs monitoring of the feasibility status of the current design and the constraint violation levels. Our methodology is validated using two CP antenna structures miniaturized under reflection and axial ratio constraints. Extensive benchmarking demonstrates superior size reduction rates along with a possibility of precise control over the design constraints as compared to the manual penalty term setup.

II. EXPLICIT SIZE-REDUCTION THROUGH CONSTRAINED OPTIMIZATION

This section recalls a formulation of EM-driven antenna size reduction problem, as well as outlines the standard trust-region-based algorithm employed as the main optimization engine.

A. PROBLEM FORMULATION

We denote by $\boldsymbol{R}(\boldsymbol{x})$ the response of the EM simulation antenna model, where $\boldsymbol{x}$ is a vector of the geometry parameters. The task is to minimize the antenna size $A(\boldsymbol{x})$, subject to performance-related constraints of the form

$$s_j(\boldsymbol{x}) \leq S_j, \quad j = 1, \ldots, k \quad (1)$$

The constraints are handled implicitly, using the penalty function approach. The objective function takes the form of

$$U_A(\boldsymbol{R}(\boldsymbol{x})) = A(\boldsymbol{x}) + \sum_{j=1}^{k} \beta_j c_j(\boldsymbol{x})^2 \quad (2)$$

The penalty function c_j quantifies relative violation of the jth constraint as $c_j(\boldsymbol{x}) = \max\{\zeta_j/S_j, 0\}$, with the absolute violation defined as

$$\zeta_j = s_j(\boldsymbol{x}) - S_j \quad (3)$$

The penalty coefficients β_j determine the contribution of the c_j-measured violation to (2).

The design problem is formulated as

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x} \in X} U_A(\boldsymbol{R}(\boldsymbol{x})) \quad (4)$$

B. TRUST-REGION GRADIENT BASED ALGORITHM

Our core optimization procedure utilizes the standard trust-region gradient-based algorithm [30], therein, a series of candidate solutions to (4) are obtained as

$$\boldsymbol{x}^{(i+1)} = \arg \min_{\boldsymbol{x}; \, ||\boldsymbol{x} - \boldsymbol{x}^{(i)}|| \leq \delta} U_A(\boldsymbol{L}^{(i)}(\boldsymbol{x})), \quad i = 0, 1, \ldots \quad (5)$$

where $\boldsymbol{L}^{(i)}(\boldsymbol{x})$ is a first-order Taylor approximation of $\boldsymbol{R}(\boldsymbol{x})$ at $\boldsymbol{x}^{(i)}$, constructed using the antenna response sensitivities estimated using finite differentiation. The vector $\boldsymbol{x}^{(i+1)}$ is accepted if $U_A(\boldsymbol{R}(\boldsymbol{x}^{(i+1)})) < U_A(\boldsymbol{R}(\boldsymbol{x}^{(i)}))$. The standard TR-based rules [32] are employed to adjust the search radius δ upon each iteration.

III. AUTOMATED PENALTY FACTOR ADJUSTMENT

Formulation (2) facilitates handling of performance-related constraints yet the efficacy of this approach relies upon appropriate adjustment of the penalty coefficients (cf. Section I). This paper proposes a procedure for automated penalty factor adjustment, which eliminates the need for the trial-and-error, or experience-based penalty term setup, and leads to a more precise control of constraint violations as well as improved size reduction rates, as demonstrated in Section IV. This section discusses the underlying concept, the adjustment procedure, and the overall optimization algorithm.

A. AUTOMATED PENALTY FACTOR ADJUSTMENT

The setup of the penalty terms is instrumental in achieving top performance of the optimization process. Having the penalty coefficients too small results in excessive constraint

violations, whereas keeping them too large leads to numerical challenges in exploring the feasible region boundary.

Our technique aims at automating the penalty term setup using a set of rules, derived from current constraint violations along with a notion of sufficient improvement in successive iterations (i.e., between $x^{(i)}$, and $x^{(i+1)}$):

- If $x^{(i+1)}$ is feasible w.r.t. the jth constraint, reduce β_j;
- If $x^{(i+1)}$ is infeasible but there is sufficient improvement of the jth constraint violation w.r.t that of $x^{(i)}$, keep β_j intact;
- If $x^{(i+1)}$ is infeasible and there is either no improvement or insufficient improvement of the jth constraint, increase β_j;

Sufficient constraint violation improvement for the jth constraint, is defined as $\Delta_j = M\zeta_j$, where M is the improvement factor elaborated on below. In the following descriptions, ζ_j^{i+1} is the jth constraint violation at $x^{(i+1)}$ calculated as in (3), whereas Δ_j^{i+1} is defined for the last two consecutive vectors $x^{(i+1)}$ and $x^{(i)}$. Similarly, β_j^i stands for β_j at iteration i.

Rigorous formulation of the adjustment rules is provided in Step 4 of the pseudocode in Fig. 1. Therein, m_{incr} and m_{decr} are the modification factors for penalty coefficient modifications (here, set to $m_{incr} = 5$ and $m_{decr} = 1.25$). Keeping the penalty coefficient unchanged upon detecting sufficient constraint violation improvement allows for improving stability of the optimization process, i.e., avoiding over-multiplication of β_j, which would otherwise bounce back and forth throughout the optimization process.

1. Set $i = 0$; provide $x^{(0)}$;
2. Solve $x^{(i+1)} = \arg \min\limits_{x; \|x-x^{(i)}\| \le \delta} U_A(L^{(i)}(x))$;
3. Compute gain ratio ρ_A and update TR size

$$\rho_A = \frac{U_A(R(x^{(i+1)})) - U_A(R(x^{(i)}))}{U_A(L^{(i)}(x^{(i+1)})) - U_A(L^{(i)}(x^{(i)}))} ;$$

   if $\rho_A < 0$,
      Go to 3;
   else
      $\Delta_j^{i+1} = M(s_j(x^{(i)}) - S_j)$;
   end
4. **if** $\zeta_j^i - \zeta_j^{i+1} > \Delta_j^{i+1}$ && $\zeta_j^{i+1} > 0$
      $\beta_j^{i+2} = \beta_j^{i+1} m_{incr}$; $i = i + 1$;
   elseif $\zeta_j^{i+1} \le 0$
      $\beta_j^{i+1} = \beta_j^i / m_{decr}$; $i = i + 1$;
   else
      $\beta_j^{i+1} = \beta_j^i$; $i = i + 1$;
   end
5. **if** $\|x^{(i+1)} - x^{(i)}\| \le \delta_x$ OR TR $\le \delta_{TR}$
      END;
   else
      Go to 2;
   end
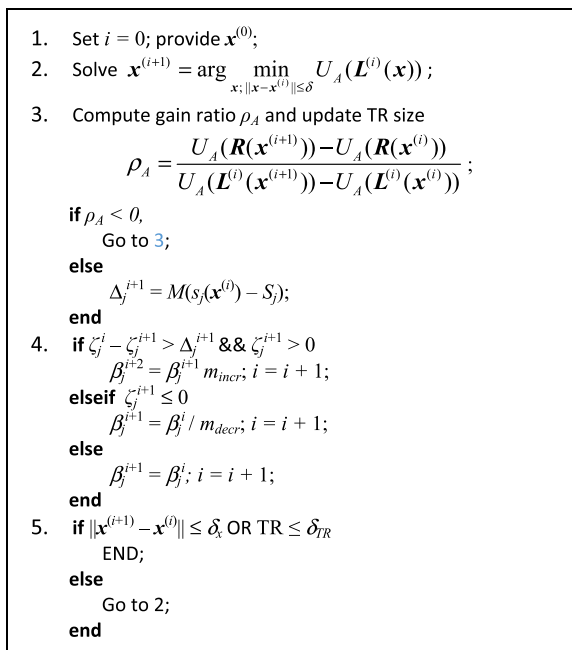
FIGURE 1. Operation of the proposed antenna size reduction algorithm with automated penalty factor adjustment.

The aforementioned improvement factor M is selected as follows. Let us assume that the vector $x^{(i)}$ is infeasible, and a sufficient constraint violation improvement is observed for n consecutive iterations, from i to $i+n$. As $M < 1$, this results in
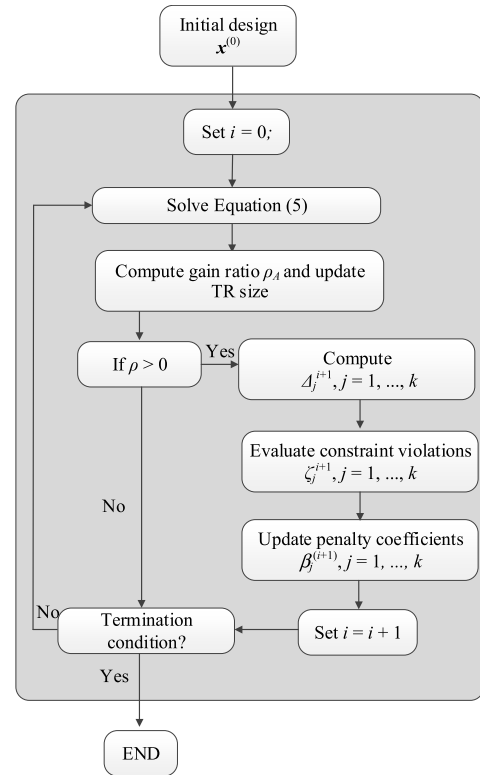
FIGURE 2. Operating flow of the proposed antenna size reduction algorithm with automated penalty factor adjustment.

a geometrical decrease of constraint violation, for which the upper bound at the iteration $i + n$ can be calculated as $\zeta_j^{i+n} \le M\zeta_j^{i+n-1} \le M\zeta_j^{i+n-2} \le \cdots \le M\zeta_j^i$. The improvement rate becomes faster as M gets smaller. On the other hand, the fulfillment of the sufficient improvement becomes more demanding. A value of $M = 0.5$ is chosen as a compromise.

B. OPTIMIZATION FRAMEWORK

The operation flow of the complete optimization algorithm has been shown in Fig. 1. The control parameters δ_x and δ_{TR} are the termination thresholds. Step 1 of the algorithm initializes the optimization procedure. Step 2 produces the candidate design by minimizing $U_A(L(x^{(i)}))$. Step 3 calculates the gain ratio, used to decide about the acceptance of rejection of $x^{(i+1)}$. Subsequently, constraint violation improvements are computed, which are used to update the penalty coefficients in Step 4.

IV. DEMONSTRATION CASE STUDIES

This section provides numerical validation of the automated adjustment procedure introduced in Section III. The verification case studies include two CP antennas optimized for minimum size with the constraints imposed on their axial ratio and reflection responses. The obtained results are compared to those produced with the fixed penalty coefficient setups ranging from the very relaxed to tight conditions regarding constraint satisfaction. The benchmark structures, experimental

setup, numerical results and their detailed discussion, are provided in Sections IV. *A* through IV. *C*, respectively.

A. VERIFICATION EXAMPLES AND EXPERIMENTAL SETUP

Figure 3 shows the geometries of the two benchmark structures (Antenna I [31], and Antenna II [32]) employed for verification purposes. Antenna I is a stacked microstrip patch structure supposed to be optimized within the frequency band 5.36 GHz to 5.9 GHz, whereas, Antenna II is a circular patch structure with annular and rectangular slots, to be optimized within a frequency band from 8.1 GHz to 8.3 GHz. Table 1 provides the details of the substrate materials, designable variables, as well as the initial design vectors of both antennas. The computational models are simulated using the time domain solver of CST Microwave Studio. The initial designs have been obtained by an auxiliary optimization process so as to provide a reasonable margin for both the axial ratio and reflection coefficient constraints, thereby creating room for size reduction.
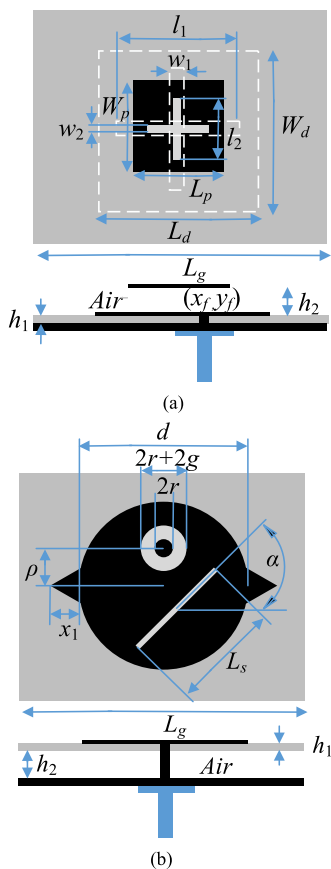
FIGURE 3. Geometries of the benchmark CP antennas: (a) Antenna I, (b) Antenna II.

The goal is to optimize the considered CP antennas for minimum size, defined as the substrate area $A(\boldsymbol{x})$. The constraints are imposed on the axial ratio $AR(\boldsymbol{x})$, and the reflection coefficient $|S_{11}(\boldsymbol{x})|$ of the antennas. In particular, we have $s_{AR}(\boldsymbol{x}) \leq 3$ and $s_{S11}(\boldsymbol{x}) \leq -10$, where $s_{AR}(\boldsymbol{x})$ and $s_{S11}(\boldsymbol{x})$ stand for the maximum value of $AR(\boldsymbol{x})$ and $|S_{11}(\boldsymbol{x})|$ respectively.

TABLE 1. Benchmark antenna structures.

	Antenna I [31]	Antenna II [32]
Substrate I	Arlon ($\varepsilon_r = 2.2$, $h = 1.575$ mm)	Arlon ($\varepsilon_r = 2.5$, $h = 3.8$ mm)
Substrate II	Air ($\varepsilon_r = 1$, $h = 3.8$ mm)	Air ($\varepsilon_r = 1.08$, $h = 2$ mm)
Design variables [mm]	$\boldsymbol{x} = [x_f\, y_f\, l_1\, l_2\, W_p\, W_d\, L_p$ $L_d\, w_2\, w_1\, L_g]$	$\boldsymbol{x} = [r\, g\, L_g\, d\, \rho\, L_s\, \alpha\, x_1]$
Initial design [mm]	$\boldsymbol{x} = [4.16\ 3.09\ 8.26$ $12.08\ 17.23\ 12.93\ 17.7$ $15.96\ 1.15\ 0.89\ 26.04]$	$\boldsymbol{x} = [1.58\ 0.48\ 21.7$ $12.46\ 3.4\ 9.4\ 52.4$ $1.52]$

Correspondingly, the two penalty coefficients are defined as β_{AR} and β_{S11}. Note that descriptive subscripts AR and $S11$ are used here rather than numerical ones (cf. Section III) to allow for a better clarity.

The proposed procedure is compared with the standard trust-region-based algorithm executed with fixed values of both of penalty coefficients i.e. $\beta_{AR} = 10^y$, $y = 1, 2, 3, 4$, and $\beta_{S11} = 10^z$, $z = 2, 3, 4, 5$. The values set for the termination thresholds are $\delta_x = \delta_{TR} = 10^{-3}$.

B. RESULTS

Figures 4 and 5 show the axial ratio and reflection coefficient responses along with the evolution of their corresponding penalty factors throughout the optimization process for Antennas I and II, respectively. Table 2 provides the optimization results of both the fixed and the automated adjustment setups for the two antennas. The data includes antenna size along with constraint violations of the axial ratio and the reflection coefficient, denoted by ζ_{AR} and ζ_{S11}, respectively. The final optimized design vectors are $\boldsymbol{x} = [2.96\ 3.16\ 8.74\ 14.10\ 16.34\ 13.32\ 16.15\ 15.80\ 1.02\ 1.00\ 24.28]$ (mm) and $\boldsymbol{x} = [1.77\ 0.66\ 19.32\ 12.20\ 2.97\ 9.28\ 52.56\ 1.41]$ (mm) for Antenna I and Antenna II, respectively. Further discussion of the results can be found Section IV. *C*.

C. DISCUSSION

The analysis of the results reported in Table 2 allows for drawing several conclusions regarding the importance of the automated penalty factor adjustment, as well as the performance superiority of the proposed automated procedure over the conventional approach.

The major observations are as follows:

- The optimum values of the penalty coefficients are problem dependent, therefore, finding the appropriate setup beforehand is a matter of a guess work. Clearly, this affects the performance of the optimization process and increases its computational cost (e.g., if the initially adopted setup turns out to be sub-optimal).
- Using a penalty factor higher than the optimum value results in degradation of the performance in terms of the
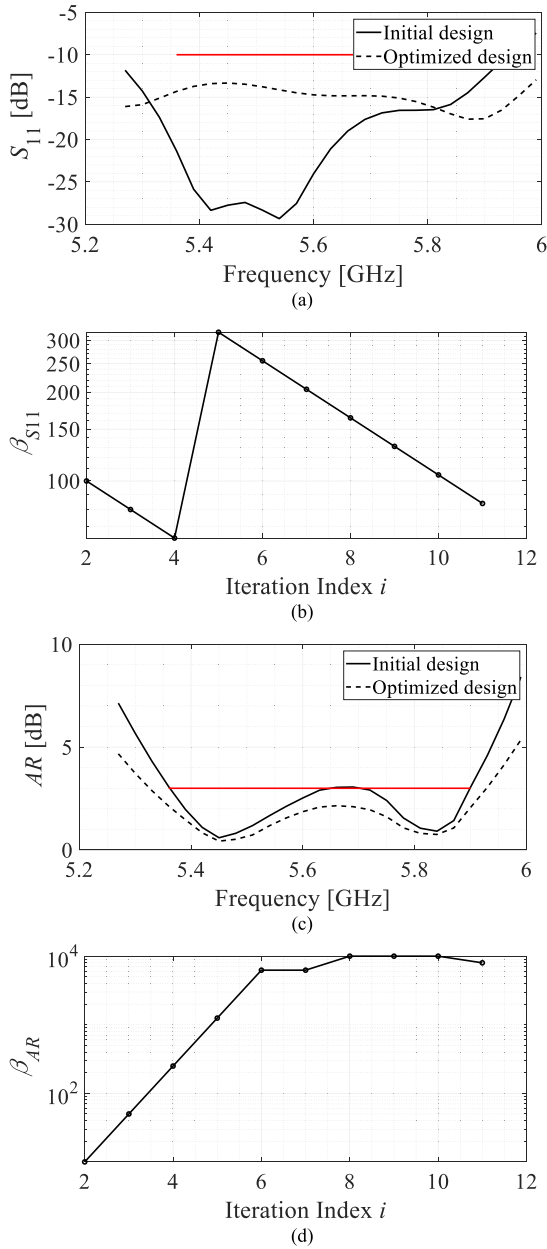
FIGURE 4. Antenna I: optimization results using the proposed automated penalty factor adjustment procedure: (a) reflection coefficient at the initial (- - -) and the optimized design (—); (b) axial ratio at the initial (- - -) and the optimized design (—); (c) evolution of the reflection coefficient penalty factor throughout the iterations of the optimization process; (d) evolution of the axial ratio penalty factor throughout the iterations of the optimization process.

FIGURE 5. Antenna II: optimization results using the proposed automated penalty factor adjustment procedure: (a) reflection coefficient at the initial (- - -) and the optimized design (—); (b) axial ratio at the initial (- -) and the optimized design (—); (c) evolution of the reflection coefficient penalty factor throughout the iterations of the optimization process; (d) evolution of the axial ratio penalty factor throughout the iterations of the optimization process.

achievable size reduction rates, whereas too low values, leads to significant constraint violation.

- The (fixed) penalty coefficient setup, which is optimum from the point of view of constraint violation, is still inferior in terms of achievable antenna size as compared to the proposed adaptive approach.

- Automated penalty factor adjustment allows to improve the quality of the final design by identifying the optimum values of the penalty coefficients for every iteration

throughout the optimization process. The history of the iteration-wise penalty factor adjustment for the reflection constraint, ζ_{S11}, of Antenna I is illustrated in Fig. 4(b). It starts with the minimum value of β_{S11} and continues with a decreasing trend for two subsequent iterations, i.e., as long as there is no constraint violation. The sudden increase of β_{S11} that can be observed between the fourth and the fifth iteration, is representative of constraint violation of the design obtained at the

TABLE 2. Optimization results for antennas I and II.

Penalty Factor Setup $[\beta_{AR}, \beta_{S11}]$		Antenna I			Antenna II		
		Area A [mm²]	Constraint violation ζ_{AR} [dB]	Constraint violation ζ_{S11} [dB]	Area A [mm²]	Constraint violation ζ_{AR} [dB]	Constraint violation ζ_{S11} [dB]
Reference algorithms	$[10^1, 10^2]$	313.26	5.13	0.91	374.38	2.77	0
	$[10^1, 10^3]$	313.26	5.13	0.91	323.37	2.24	0.2
	$[10^1, 10^4]$	313.26	5.13	0.91	334.37	2.69	0.01
	$[10^1, 10^5]$	313.26	5.13	0.91	340.24	2.15	0.01
	$[10^2, 10^2]$	587.9	0.36	0	309.99	0.05	3.8
	$[10^2, 10^3]$	587.9	0.36	0	361.73	0.33	0
	$[10^2, 10^4]$	587.9	0.36	0	359.71	0.27	0
	$[10^2, 10^5]$	587.9	0.36	0	356.18	0.20	0
	$[10^3, 10^2]$	607.39	0.02	0	404.58	0.07	0
	$[10^3, 10^3]$	607.39	0.02	0	421.75	0.05	0
	$[10^3, 10^4]$	607.39	0.02	0	421.75	0.05	0
	$[10^3, 10^5]$	607.39	0.02	0	421.75	0.05	0
	$[10^4, 10^2]$	663.73	0.008	0	455.41	0.03	0
	$[10^4, 10^3]$	663.73	0.008	0	455.41	0.03	0
	$[10^4, 10^4]$	663.73	0.008	0	455.41	0.03	0
	$[10^4, 10^5]$	663.73	0.008	0	455.41	0.03	0
Adaptive β (This work)		589.78	0.07	0	373.36	0	0

fourth iteration. The following decreasing trend up to the last iteration indicates a lack of constraint violation for the corresponding iterations. Similar trends can be observed for axial-ratio-related penalty factor, as well as for Antenna II. The current values of both β_{S11} and β_{AR} are set to either reduce constraint violation or to maintain the solution in the vicinity of the feasible region boundary.

In general, the described automated adjustment procedure, in turn, enables improved size reduction rates along with a better control over constraint violations.

V. CONCLUSION

This paper proposed a novel methodology for optimization-based antenna size reduction using local trust-region-based search routines. Our procedure can be incorporated into frameworks involving a penalty function approach for implicit handling of design constraints. Therein, the proper adjustment of the penalty factors strongly correlates with the efficacy as well as the reliability of the optimization process, both in terms of constraint satisfaction and the achievable size reduction rates, yet it is difficult to be identified beforehand. The proposed procedure virtually eliminates the need for

manual, or guess-work efforts by an automated penalty factor adjustment. The latter is conducted based sufficient constraint violation in successive iterations and consequently allowing for better size reduction rates while leading to a precise control over the design constraints as compared to the fixed penalty coefficient setup.

The proposed methodology has been validated using two CP antenna structures optimized for minimum size, with the constraints imposed on their axial ratio and reflection responses. Benchmarking against fixed penalty factor setups indicates superior performance of the automated adjustment in terms of the achievable size reduction rates and a precise control of the design constraints.

VI. ACKNOWLEDGMENT

REFERENCES

[1] Y.-H. Yang, B.-H. Sun, and J.-L. Guo, "A low-cost, single-layer, dual circularly polarized antenna for millimeter-wave applications," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 4, pp. 651–655, Apr. 2019.

[2] W. Wang, C. Chen, S. Wang, and W. Wu, "Circularly polarized patch antenna with filtering performance using polarization isolation and dispersive delay line," *IEEE Antennas Wireless Propag. Lett.*, vol. 19, no. 8,

pp. 1457–1461, Aug. 2020.

[3] A. Akbarpour and S. Chamaani, "Ultrawideband circularly polarized antenna for near-field SAR imaging applications," *IEEE Trans. Antennas Propag.*, vol. 68, no. 6, pp. 4218–4228, Jun. 2020.

[4] Z.-P. Zhong, X. Zhang, J.-J. Liang, C.-Z. Han, M.-L. Fan, G.-L. Huang, W. Xu, and T. Yuan, "A compact dual-band circularly polarized antenna with wide axial-ratio beamwidth for vehicle GPS satellite navigation application," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8683–8692, Sep. 2019.

[5] A. H. Lokman, P. J. Soh, S. N. Azemi, M. F. Jamlos, A. A. Al-Hadi, S. Chalermwisutkul, and P. Akkaraekthalin, "Compact circularly polarized S-band antenna for pico-satellites," in *Proc. Int. Symp. Antennas Propag. (ISAP)*, Oct. 2017, pp. 1–2.

[6] Y. Cheng and Y. Dong, "Wideband circularly polarized split patch antenna loaded with suspended rods," *IEEE Antennas Wireless Propag. Lett.*, vol. 20, no. 2, pp. 229–233, Feb. 2021.

[7] T. Mondal, S. Maity, R. Ghatak, and S. R. B. Chaudhuri, "Compact circularly polarized wide-beamwidth fern-fractal-shaped microstrip antenna for vehicular communication," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5126–5134, Jun. 2018.

[8] K. Wei, J. Y. Li, L. Wang, R. Xu, and Z. J. Xing, "A new technique to design circularly polarized microstrip antenna by fractal defected ground structure," *IEEE Trans. Antennas Propag.*, vol. 65, no. 7, pp. 3721–3725, Jul. 2017.

[9] T. Cai, G.-M. Wang, X.-F. Zhang, and J.-P. Shi, "Low-profile compact circularly-polarized antenna based on fractal metasurface and fractal resonator," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 1072–1076, 2015.

[10] M. Ameen and R. K. Chaudhary, "Dual-layer and dual-polarized metamaterial inspired antenna using circular-complementary split ring resonator mushroom and metasurface for wireless applications," *Int. J. Electron. Commun.*, vol. 113, pp. 1–18, Jan. 2020.

[11] B. Liu, H. Yang, and M. J. Lancaster, "Global optimization of microwave filters based on a surrogate model-assisted evolutionary algorithm," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 6, pp. 1976–1985, Jun. 2017.

[12] A. Lalbakhsh, M. U. Afzal, and K. P. Esselle, "Multiobjective particle swarm optimization to design a time-delay equalizer metasurface for an electromagnetic band-gap resonator antenna," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 912–915, 2016.

[13] J. A. Tomasson, S. Koziel, and A. Pietrenko-Dabrowska, "Quasi-global optimization of antenna structures using principal components and affine subspace-spanned surrogates," *IEEE Access*, vol. 8, pp. 50078–50084, 2020.

[14] S. Koziel and A. Pietrenko-Dabrowska, "Performance-based nested surrogate modeling of antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 67, no. 5, pp. 2904–2912, May 2019.

[15] Y. Song, Q. S. Cheng, and S. Koziel, "Multi-fidelity local surrogate model for computationally efficient microwave component design optimization," *Sensors*, vol. 19, no. 13, p. 3023, Jul. 2019.

[16] I. A. Baratta, C. B. de Andrade, R. R. de Assis, and E. J. Silva, "Infinitesimal dipole model using space mapping optimization for antenna placement," *IEEE Antennas Wireless Propag. Lett.*, vol. 17, no. 1, pp. 17–20, Jan. 2018.

[17] F. Feng, W. Na, W. Liu, S. Yan, L. Zhu, and Q.-J. Zhang, "Parallel gradient-based EM optimization for microwave components using Adjoint-sensitivity-based neuro-transfer function surrogate," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 3606–3620, Sep. 2020.

[18] S. Koziel and A. Pietrenko-Dabrowska, "Accelerated gradient-based optimization of antenna structures using multi-fidelity simulations and convergence-based model management scheme," *IEEE Trans. Antennas Propag.*, early access, Jun. 2, 2021, doi: 10.1109/TAP.2021.3083742.

[19] E. Hassan, D. Noreland, R. Augustine, E. Wadbro, and M. Berggren, "Topology optimization of planar antennas for wideband near-field coupling," *IEEE Trans. Antennas Propag.*, vol. 63, no. 9, pp. 4208–4213, Sep. 2015.

[20] F. Feng, J. Zhang, W. Zhang, Z. Zhao, J. Jin, and Q.-J. Zhang, "Coarse- and fine-mesh space mapping for EM optimization incorporating mesh deformation," *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 8, pp. 510–512, Aug. 2019.

[21] F. Feng, C. Zhang, W. Na, J. Zhang, W. Zhang, and Q. J. Zhang, "Adaptive feature zero assisted surrogate-based EM optimization for microwave filter design," *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 1, pp. 2–4, Jan. 2019.

[22] J. Zhang, F. Feng, J. Jin, W. Zhang, Z. Zhao, and Q.-J. Zhang, "Adaptively weighted yield-driven EM optimization incorporating neurotransfer function surrogate with applications to microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 1, pp. 518–528, Jan. 2021.

[23] L.-Y. Xiao, W. Shao, X. Ding, and B.-Z. Wang, "Dynamic adjustment kernel extreme learning machine for microwave component design," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 10, pp. 4452–4461, Oct. 2018.

[24] J. Gao, Y. Tian, and X. Chen, "Antenna optimization based on co-training algorithm of Gaussian process and support vector machine," *IEEE Access*, vol. 8, pp. 211380–211390, 2020.

[25] D. O. Johanesson and S. Koziel, "Feasible space boundary search for improved optimisation-based miniaturisation of antenna structures," *IET Microw., Antennas Propag.*, vol. 12, no. 8, pp. 1273–1278, Jul. 2018.

[26] D. O. Johanesson, S. Koziel, and A. Bekasiewicz, "EM-driven constrained miniaturization of antennas using adaptive in-band reflection acceptance threshold," *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 32, no. 2, Mar. 2019, Art. no. e2513.

[27] B. Tessema and G. G. Yen, "An adaptive penalty formulation for constrained evolutionary optimization," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 3, pp. 565–578, May 2009.

[28] J. A. Joines and C. R. Houck, "On the use of non-stationary penalty functions to solve nonlinear constrained optimization problems with GA's," in *Proc. 1st IEEE Conf. Evol. Comput. IEEE World Congr. Comput. Intell.*, Jun. 1994, pp. 579–584.

[29] B. Tessema and G. G. Yen, "A self adaptive penalty function based algorithm for constrained optimization," in *Proc. IEEE Int. Conf. Evol. Comput.*, Jul. 2006, pp. 246–253.

[30] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust Region Methods, MPS-SIAM Series on Optimization*. Philadelphia, PA, USA: SIAM, 2000.

[31] S. A. Malekabadi, A. R. Attari, and M. M. Mirsalehi, "Compact and broadband circular polarized microstrip antenna with wideband axial-ratio bandwidth," in *Proc. Int. Symp. Telecommun.*, Aug. 2008, pp. 106–109.

[32] B. P. Kumar, C. Kumar, and D. Guha, "A new design approach to improve the circular polarization characteristics of a microstrip antenna," in *Proc. IEEE Indian Conf. Antennas Propogation (InCAP)*, Dec. 2018, pp. 1–2.
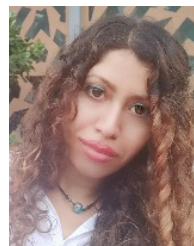
MARZIEH MAHROKH received the M.Sc. degree in telecommunications engineering from Shahid Beheshti University, Tehran, Iran, in 2016. She is currently pursuing the Ph.D. degree in electrical engineering with Reykjavik University, Reykjavik, Iceland. From 2016 to 2018, she was a Microwave Design Engineer with the Research and Development Department, Electronic Research & Production Company, TAKTA, Tehran. From 2018 to 2020, she was a Research Assistant with the School of Microelectronics, Southern University of Science and Technology, Shenzhen, China. Her research interests include simulation-driven design, multi-objective optimization, and miniaturization of microwave/RF components.

SLAWOMIR KOZIEL (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electronic engineering from Gdańsk University of Technology, Poland, in 1995 and 2000, respectively, the M.Sc. degrees in theoretical physics and in mathematics, in 2000 and 2002, and the Ph.D. degree in mathematics from the University of Gdańsk, Poland, in 2003. He is currently a Professor with the Department of Engineering, Reykjavik University, Iceland. His research interests include CAD and modeling of microwave and antenna structures, simulation-driven design, surrogate-based optimization, space mapping, circuit theory, analog signal processing, evolutionary computation, and numerical analysis.

Chapter 6

6   Paper # 3

Marzieh Mahrokh and Slawomir Koziel

Improved-Efficacy EM-Based Antenna Miniaturization by Multi-Fidelity Simulations and Objective Function Adaptation

Improved-Efficacy EM-Based Antenna Miniaturization by Multi-Fidelity Simulations and Objective Function Adaptation

Marzieh Mahrokh [1,*] and **Slawomir Koziel** [1,2]

1   Engineering Optimization and Modeling Center, Reykjavik University, 102 Reykjavik, Iceland; koziel@ru.is
2   Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology,
     80-233 Gdańsk, Poland
*   Correspondence: marziehm@ru.is; Tel.: +354-599-6376

Abstract: The growing demand for the integration of surface mount design (SMD) antennas into miniaturized electronic devices has imposed increasing limitations on the structure dimensions. Examples include embedded antennas in applications such as on-board devices, picosatellites, 5G communications, or implantable and wearable devices. The demands for size reduction while ensuring a satisfactory level of electrical and field performance can be managed through constrained numerical optimization. The reliability of optimization-based size reduction requires utilization of full-wave electromagnetic (EM) analysis, which entails significant computational costs. This can be alleviated by incorporating surrogate modeling techniques, adjoint sensitivities, or the employment of sparse sensitivity updates. An alternative is the incorporation of multi-fidelity simulation models, normally limited to two levels, low and high resolution. This paper proposes a novel algorithm for accelerated antenna miniaturization, featuring a continuous adjustment of the simulation model fidelity in the course of the optimization process. The model resolution is determined by factors related to violation of the design constraints as well as the convergence status of the algorithm. The algorithm utilizes the lowest-fidelity model for the early stages of the optimization process; it is gradually refined towards the highest-fidelity model upon approaching convergence, and the constraint violations improve towards the preset tolerance threshold. At the same time, a penalty function approach with adaptively adjusted coefficients is applied to enable the precise control of constraints, and to increase the achievable miniaturization rates. The presented procedure has been validated using five microstrip antennas, including three broadband, and two circularly polarized structures. The obtained results corroborate the relevance of the implemented mechanisms from the point of view of improving the average computational efficiency of the optimization process by 43% as compared to the single-fidelity adaptive penalty function approach. Furthermore, the presented methodology demonstrates a performance that is equivalent or even superior to its single-fidelity counterpart in terms of an average constraint violation of 0.01 dB (compared to 0.03 dB for the reference), and an average size reduction of 25% as compared to 25.6%.

Keywords: antenna miniaturization; surface mount design (SMD); constrained optimization; EM-driven design; multi-fidelity simulations; penalty coefficients

1. Introduction

The emerging trends in integrated wireless communication technology require the integration of surface mount design (SMD) antennas with other on-chip system components. This, in turn, imposes miniaturization requirements in applications such internet of things (IoT), portable and implantable devices [1,2], or 5G communication systems. Several antenna size-reduction techniques involving topological alterations of the basic geometries have been proposed, including the use of corrugations in the radiator and the ground plane [3,4], the introduction of meandering slits and fractals [5], or incorporation of slots and slits [6].

The abovementioned techniques offer degrees of freedom to facilitate reaching a compromise between the compact size and the electromagnetic (EM) performance. Notwithstanding, as antenna topology evolves into more complex geometries due to topological modifications of the structure, manual or trial-and-error efforts fall short of identifying the optimum design. This shortcoming is more pronounced when multiple objectives need to be handled.

Fulfilling the stringent demands concerning the electrical and field performance of the system, along with miniaturization of the comprising structures, can be handled through constrained numerical optimization. Depending on the antenna type, specifications, and the available design database, this can be accomplished using local [7,8], quasi-global [9], or global search routines [10–12]. Maintaining the reliability of optimization-based antenna miniaturization requires an accurate computational model, which is most often based on full-wave EM analysis. The bottleneck is high computational cost of EM models. Numerous evaluations are required by the optimization routines, especially for complex geometries, thus this cost may become prohibitive, even in the less challenging case of local optimization, e.g., those realized using pattern search [13], or gradient-based [14,15] algorithms.

Addressing the aforementioned issues of optimization-based antenna size reduction necessitates the development of CPU-efficient numerical techniques. Based on a comprehensive analysis of the available literature, the methods used to mitigate the computational burden of EM-driven design can be categorized into two groups. The first is strictly oriented toward algorithmic improvements that primarily target faster evaluation of antenna response gradients. These include utilization of adjoint sensitivities [16–21], or the employment of sparse sensitivity updates [22]. The second group involves the utilization of surrogate modeling techniques including both data-driven [23] and physics-based models [24].

Data-driven surrogates either replace the high-cost EM simulations altogether upon initial construction, or gradually develop and train a statistical model of the system using sequential sampling techniques throughout the optimization process [25]. Examples of the modeling techniques include kriging [26], artificial neural networks [27], support vector regression [28], or fuzzy systems [29]. Unfortunately, the application of data-driven surrogates is impeded by a typically considerable nonlinearity of high-frequency system responses (sharp resonances [30]), and additionally, by the curse of dimensionality.

Physics-based surrogates typically embed problem-specific knowledge of the system at hand in an underlying low-fidelity model (equivalent networks [31], coarse-discretization EM simulations [32]). A few popular modeling techniques include feature-based optimization [33], response correction methods [34], or space mapping [35].

While global optimization routines are important in application areas such as synthesis of array antennas [36], beam-shaping and beam-steering [37], local optimization routines are employed in the majority of scenarios such as design closure (final parameter tuning), or optimization-based antenna miniaturization. This is due to the availability of reasonably good initial designs, obtained as a result of the early stages of topological developments in the antenna design process. As previously mentioned, gradient-based search routines can be greatly expedited using adjoint sensitivities. However, the availability of this technology in commercial software packages is limited. Variable-fidelity techniques including response correction methods [38,39], or space mapping [40,41], and variations thereof [42,43], can also be employed to improve the cost efficiency. Notwithstanding, the efficacy of these techniques relies on the meticulous selection of the model-fidelity and response-type dependent correction techniques [44]. Other alternatives include the employment of restricted sensitivity updates [45], incorporation of updating formulas (e.g., Broyden [9]), also in conjunction with response feature techniques [46]. These methods offer up to forty [46] or even sixty percent acceleration [47] without minor degradation of the design quality compared to that of the reference algorithms. Further benefits in terms of accelerating EM-based optimization processes can be obtained through the incorporation of multi-fidelity

simulation models. They can be either incorporated into the above-mentioned sensitivity-based speedup mechanisms to boost the acceleration, or applied in the different context of expediting antenna miniaturization procedures while maintaining precise constraint control and the efficacy of the miniaturization process, which is the main focus of this paper. Nevertheless, proper management of the model fidelity is far from trivial [48].

Another issue is the efficacy of the optimization-based antenna size reduction in terms of achievable miniaturization rates. Efficient miniaturization requires explicit treatment of the antenna size as the primary objective. At the same time, satisfaction of the design constraints necessitates their appropriate handling, which is realized implicitly by means of the penalty function approach [49]. The formulation of the penalty function includes setting up the values of penalty coefficients that determine the contributions of the constraint violations to the main objective. Optimum determination thereof is a challenging task. Excessively high or low values may result in low efficacy in terms of miniaturization rates or constraint violation control. Further details of the formulation will be discussed in Section 2.1. A workaround is the adaptive adjustment of the penalty factors [50,51], which meticulously identifies the optimum setup based on the level of the constraint violations throughout the optimization process. This approach is adopted in this work as the constraint control mechanism. This paper proposes a novel procedure for accelerated miniaturization of antenna structures, incorporating variable-fidelity EM models with a continuous adjustment of the model fidelity throughout the optimization process. The model resolution is controlled by factors related to violation of the design constraints as well as the convergence status of the algorithm. The procedure utilizes the lowest-fidelity model in the early stages of the optimization process, which permits a cost-efficient exploration of the design space. The reliability is ensured by a gradual refinement of the model resolution towards the highest-fidelity model as the constraint violations are reduced beyond the preset tolerance threshold, and the optimization process approaches convergence. These mechanisms are supplemented by a penalty function approach with adaptive penalty coefficient adjustment to enable the precise control of constraints, and to achieve better miniaturization rates. The presented procedure addresses parameter tuning of the existing designs with fixed topologies, meaning the adjustment of geometry variables (antenna sizing) without changing the basic topology. The entire antenna design process, especially the development of antenna geometry, is outside the scope of this work. Our methodology was validated through miniaturization of five microstrip antennas, including two circular polarization (CP) ones, and three broadband structures. An average computational speedup of over forty percent was achieved across the benchmark set as compared to the reference algorithm, while ensuring the precise control of the constraints and improved miniaturization rates.

The originality and the technical contributions of this paper include (i) the development of a multi-fidelity model management scheme based on the design constraint violations and the convergence status of the optimization process, which permits a reliable low-cost optimization-based miniaturization of antenna structures, (ii) integration of the local gradient-based search with the multi-fidelity model management scheme as well as a penalty function approach with an adaptive penalty coefficient adjustment, and (iii) demonstration of a significant speedup in the miniaturization process that can be attained using the presented framework along with precise control over the constraint violations. The main feature that distinguishes this work from previous attempts to utilize multi-fidelity models, especially in the context of EM-driven miniaturization, is a continuous adjustment of the model fidelity based on both the convergence status and constraint violation levels. Operating over a continuous spectrum of model resolution avoids the experience-based model fidelity setup. The main advantages of this approach include a considerable speedup of the miniaturization process along with precise constraint control, along with improved quality of miniaturization as compared to the single-fidelity adaptive penalty function approach. According to the authors' knowledge, the presented methodology is the first rigorous approach to simulation-based size reduction that integrates model resolution

and constraint management schemes into a single algorithmic framework. The remaining part of the paper is organized as follows. Section 2 discusses the details of EM-based antenna miniaturization and the underlying mechanisms, which include the penalty function approach (Section 2.1), trust-region algorithm (Section 2.2), adaptive penalty coefficient approach (Section 2.3), variable-fidelity model management (Section 2.4), and finally, the proposed constraint–convergence-based procedure is discussed in Sections 2.5 and 2.6. Section 3 provides the numerical validation of the proposed optimization frameworks, including a description of the benchmark antenna structures, the experimental setup, results and their discussion. Section 4 concludes the paper.

2. Accelerated Antenna Miniaturization by Model Fidelity and Constraint Management

This section introduces the proposed procedure for accelerated miniaturization of antenna structures, involving multi-fidelity simulation models and penalty functions with adaptive coefficient adjustment. The methodology presented here is based on two earlier works: accelerating EM-driven design through multi-fidelity model simulations [52] and the adaptive penalty function approach [50]. The incorporation thereof leads to results that were unattainable by any other method currently available. Utilization of the multi-fidelity models is widespread in high-frequency CAD (as elaborated on in Section 1). However, it has never been used in the context of constrained size reduction. In other words, we exploit this specific algorithmic tool and incorporate it into our algorithm to achieve further computational benefits. Additionally, the multi-fidelity scheme is adopted for our particular setup where the fidelity level is controlled based on the factors related to the feasibility status as well as the convergence status of the algorithm. The adjustment of the model fidelity based on the feasibility status is a novel concept and is used for the first time. This particular setup is different than the one in the prior work.

The optimization engine is a trust-region-based algorithm. A continuous adjustment of the EM analysis fidelity is realized by altering the model resolution based on factors related to the constraint violations and the convergence status of the algorithm. Utilization of the lowest-fidelity model in the initial stages of the optimization process enables a fast exploration of the parameter space. Reliability is ensured by gradually refining the model resolution towards the final stages of the optimization process. This section starts by recalling a formulation of the EM-based antenna miniaturization task using fixed penalty coefficients (Section 2.1). Antenna miniaturization with adaptive penalty coefficients is outlined in Section 2.2. Section 2.3 discusses the standard trust-region gradient-based procedure as the main optimization engine. Section 2.4 elaborates on multi-fidelity EM models, whereas Sections 2.5 and 2.6 formulate the operating flow for the complete size reduction procedure.

2.1. EM-Based Antenna Miniaturization with Penalty Functions

We will use $R(x)$ to designate the response of the EM simulation model of the antenna structure of interest. Here, x denotes a vector of geometry parameters to be adjusted throughout the optimization process. The miniaturization problem at hand is to minimize the antenna size $A(x)$, subject to constraints related to the electrical and field performance. The constraints are defined as

$$s_j(x) \leq S_j, j = 1, \dots, k \tag{1}$$

where $s_j(x)$ is a scalar function representing a given figure of interest (e.g., maximum in-band reflection over a frequency range of interest), whereas S_j is a user-defined acceptance threshold.

Evaluation of $s_j(x)$ is computationally expensive as it requires EM simulation of the antenna. The penalty function approach [49] facilitates constraint handling by turning the problem into an unconstrained one. This is achieved by supplementing the objective

function U_A (here, corresponding to antenna size minimization) with a linear combination of penalty functions c_j ($j = 1, \ldots, k$) quantifying constraint violations. We have

$$U_A(\boldsymbol{R}(\boldsymbol{x})) = A(\boldsymbol{x}) + \beta_1 c_1(\boldsymbol{x})^2 + \ldots + \beta_k c_k(\boldsymbol{x})^2 \tag{2}$$

In this work, relative penalty functions are employed

$$c_j(\boldsymbol{x}) = \max\{\zeta_j / S_j, 0\} \tag{3}$$

$$\zeta_j = s_j(\boldsymbol{x}) - S_j \tag{4}$$

stands for absolute violations. Note that (3) ensures that the contribution of c_j is non-zero only if violation of the jth constraint occurs. The corresponding penalty coefficient $\vdots \hat{A}_j$ determines the proportion of the aforementioned contribution to (2).

The size reduction task is formulated as

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in X}{\arg\min} U_A(\boldsymbol{R}(\boldsymbol{x})) \tag{5}$$

The solution to problem (5) is subject to constraints (1). The parameter space X is determined by the lower and upper bounds for geometry parameters (entries of vector $\boldsymbol{x}$).

2.2. Trust-Region Gradient-Based Algorithm

The optimization framework proposed in this paper builds on the standard trust-region (TR) gradient-based algorithm [53]. It solves problem (4) iteratively by generating approximations $\boldsymbol{x}^{(i)}$, $i = 0, 1, \ldots$, to the optimum solution $\boldsymbol{x}^*$ through constrained optimization of a first-order Taylor approximation model $\boldsymbol{L}^{(i)}$ of the antenna responses $\boldsymbol{R}(\boldsymbol{x})$. We have

$$\boldsymbol{x}^{(i+1)} = \arg \min_{\boldsymbol{x}; \boldsymbol{x}^{(i)} - \boldsymbol{\delta} \leq \boldsymbol{x} \leq \boldsymbol{x}^{(i)} + \boldsymbol{\delta}} U_A(\boldsymbol{L}^{(i)}(\boldsymbol{x})), \ i = 0, 1, \ldots \tag{6}$$

Note that the solution of (5) is constrained to the interval $[\boldsymbol{x}^{(i+1)} - \boldsymbol{\delta}^{(i)}, \boldsymbol{x}^{(i+1)} + \boldsymbol{\delta}^{(i)}]$, referred to as the trust region. This arrangement accounts for different ranges of the geometry parameters. The new design $\boldsymbol{x}^{(i+1)}$ is only accepted if $U_A(\boldsymbol{R}(\boldsymbol{x}^{(i+1)})) < U_A(\boldsymbol{R}(\boldsymbol{x}^{(i)}))$, i.e., the objective function is improved. Otherwise, the TR size vector $\boldsymbol{\delta}^{(i)}$ is reduced [53] and the current iteration is repeated. The procedure is terminated if $||\boldsymbol{\delta}^{(i)}||$ is reduced below a preset limit δ_{TR}, or if $||\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}|| \leq \delta_{arg}$ (convergence in argument).

2.3. EM-Based Antenna Miniaturization and Adaptive Penalty Coefficients

As discussed in Section 2.1, the penalty function approach offers a convenient way of handling constraints. At the same time, the efficacy of the optimization process relies upon a proper adjustment of the penalty factors β_j. A workaround is the adaptive adjustment of the penalty factors [49,50], which eliminates the costly stage of trial-and-error-based penalty term setup. Additionally, adaptive adjustment allows for the precise control of the constraints, which leads to improved size-reduction rates [49]. This technique is incorporated into the optimization framework proposed in this work. The formulation of adaptive penalty factor adjustment is based upon the concepts of constraint violation as in (3), as well as sufficient constraint violation improvement defined as

$$\Delta_j = M\zeta_j \tag{7}$$

where ζ_j is the absolute constraint violation, cf. (3), whereas $0 \leq M \leq 1$ is the improvement factor; here, it is set to 0.5 as recommended in [49].

The adjustment of the penalty factor β_j is governed by the following rules:

- if $\boldsymbol{x}^{(i+1)}$ produced in the ith iteration of (5) is infeasible from the point of view of the jth constraint but constraint violation is improved by at least Δ_j w.r.t $\boldsymbol{x}^{(i)}$, $\vdots \hat{A}_j$ is kept intact;

- if $x^{(i+1)}$ is feasible w.r.t. the jth constraint, β_j is reduced;
- if $x^{(i+1)}$ is infeasible w.r.t the jth constraint and there is either insufficient improvement or no improvement in the constraint violation, β_j is increased.

The quantification details concerning penalty factor decrease/increase can be found in [49].

2.4. Multi-Fidelity EM Simulation Models

Section 2.3 addressed the issue of efficient handling of design constraints, which is of primary importance for ensuring the reliability aspects of the optimization process. Another issue is its computational cost. In this paper, in order to expedite size reduction while retaining generality, we employed variable resolution EM simulations. Reducing the fidelity level of the structure results in a faster analysis at the expense of certain accuracy loss. Because they share the same underlying physics, coarse discretization models are normally well correlated with their high-fidelity counterparts, which has been widely explored in the literature to speed up optimization processes [32]. An alternative is utilization of simplified-physics representations, e.g., equivalent networks [31] or even analytical models. Yet, to specifically address antennas, reducing the resolution of EM analysis is the only universal option [32]. Regardless of the low-fidelity model origin, only two resolution levels are typically used [54]. The low-fidelity model is typically refined by using an appropriate response-type dependent correction technique [44], and it is employed as a substitute to the high-fidelity model. Popular approaches of this class include response correction methods [34] or space mapping [35]. Low-fidelity models can also offer a cost-efficient initial exploration of the design space within variable-fidelity modeling techniques (co-kriging [55]) or machine learning frameworks [56]. Notwithstanding, the reliability and the efficacy of these techniques rely on the meticulous selection of the model-fidelity [44].

Figure 1 illustrates the geometry of a circular polarization antenna along with its reflection and axial ratio responses for various fidelity levels. The dimensions of the antenna are as given in Table 1. The model resolution is adjusted using a lines per wavelength (LPW) parameter in the mesh properties setup, thus controlling the discretization density of the structure in the Time Domain Solver of CST Microwave Studio. The LPW parameter is adjusted according to the strategy described in Section 2.4, and the evaluation time corresponding to each LPW is accounted for while computing the overall CPU cost of the optimization process. The required range of LPWs is defined in the code as discrete values. They are communicated to CST through a MATLAB-CST socket one at a time. The socket actually connects the code to the CST environment. Figure 1b shows a relationship between model fidelity and the simulation time.

Note that relaxing the accuracy criteria to the lowest usable level reduces the antenna evaluation time by a factor of about three. Although the accuracy is compromised when reducing model resolution, the lower-fidelity model still preserves the overall response shape. Therefore, it can be successfully used for initial design space exploration in the antenna miniaturization process.

| (a) | (b) |

Figure 1. *Cont.*

(**c**)

(**d**)
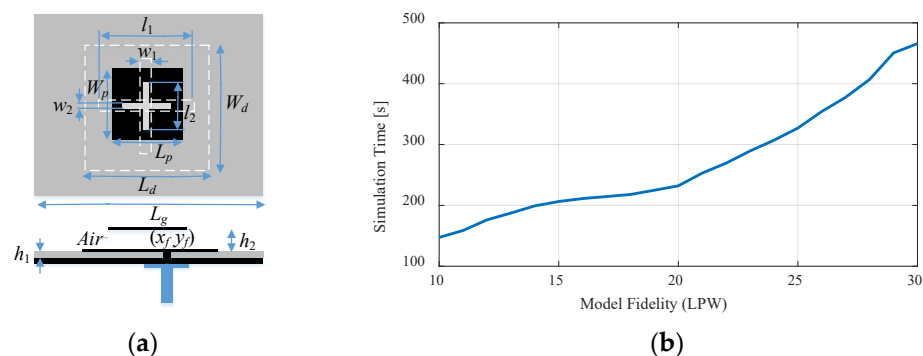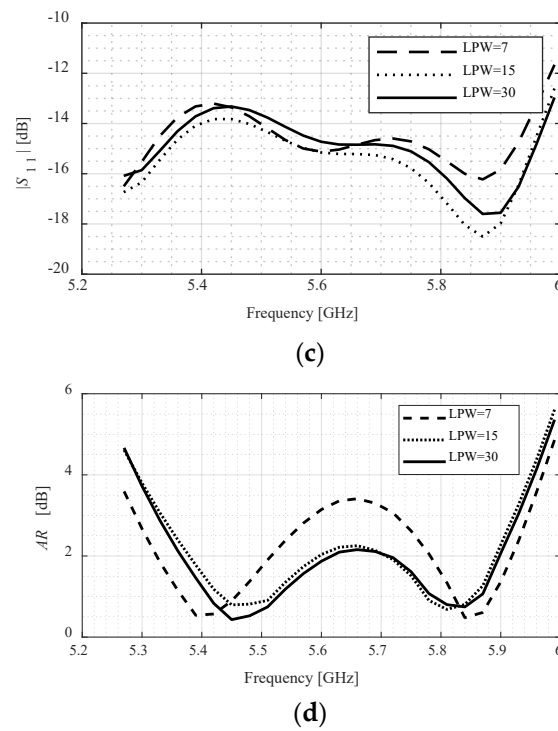
Figure 1. Circularly-polarized patch antenna; EM responses for various simulation model fidelities: (**a**) antenna structure, (**b**) the increasing trend of the simulation time as proceeding towards higher fidelities, (**c**) reflection responses for various fidelity levels, (**d**) axial ratio responses for various fidelity levels.

The lowest practical LPW denoted by F_{min} can be established by inspecting the antenna responses for various fidelity levels. Similarly, the highest LPW (denoted by F_{max}) corresponds to the high-fidelity model, which renders the system characteristics with sufficient accuracy. In this work, the objective is to accelerate the size reduction process by continuous control of the model fidelity within the range $F_{min} \leq F \leq F_{max}$. The following section provides a description of a set of prerequisites based on which the model-fidelity adjustment scheme was developed.

Table 1. Benchmark antenna structures.

	Antenna I [57]	**Antenna II [58]**	**Antenna III [59]**	**Antenna IV [60]**	**Antenna V [61]**
Substrate I	RF-35 (ε_r = 3.5 h = 0.762 mm)	RF-35 (ε_r = 3.5 h = 0.762 mm)	FR4 (ε_r = 4.3 h = 1.55 mm)	Arlon AD250 (ε_r = 2.5 h = 3.8 mm)	Arlon (ε_r = 2.2 h = 1.575 mm)
Substrate II	–	–	–	Air (ε_r = 1.08, h = 2 mm)	Air (ε_r = 1, h = 3.8 mm)
Designable parameters (mm)	$x = [L_0\ g\ a\ l_1\ l_2\ w_1\ o]^T$	$x = [L_0\ dR\ R\ r_{rel}\ dL\ dw\ L_g$ $L_1\ R_1\ dr\ c_{rel}]^T$	$x = [L_g\ L_0\ L_s\ W_s\ d\ dL\ d_s$ $dW_s\ dW\ a\ b]^T$	$x = [r\ g\ L_g\ d\ \rho\ L_s\ \alpha\ x_1]$	$x = [x_f\ y_f\ l_1\ l_2\ W_p\ W_d\ L_p$ $L_d\ w_2\ w_1\ L_g]$
Other parameters (dB)	$w_0 = 2o + a, w_f = 1.7$	$w_0 = 1.7$	$W_0 = 3$	–	–
Target operating bandwidth	3.1 GHz to 10.6 GHz	3.1 GHz to 10.6 GHz	3.1 GHz to 10.6 GHz	8.1 GHz to 8.3 GHz	5.36 GHz to 5.9 GHz
Design constraints	$\|S_{11}\| \leq -10$ dB	$\|S_{11}\| \leq -10$ dB	$\|S_{11}\| \leq -10$ dB	$\|S_{11}\| \leq -10$ dB, AR ≤ 3 dB	$\|S_{11}\| \leq -10$ dB, AR ≤ 3 dB
Initial design (mm)	$x = [20.23\ 18.62\ 9.23\ 6.67$ $5.64\ 3.84\ 2.29]$	$x = [8.74\ 0.66\ 4.59\ 0.75$ $4.75\ 1.84\ 10.00\ 5.94\ 3.67$ $0.49\ 0.79]$	$x = [8.53\ 12.35\ 9.68\ 0.33$ $3.90\ 1.72\ 1.04\ 1.48\ 1.95$ $0.37\ 0.57]$	$x = [1.58\ 0.48\ 21.7\ 12.46$ $3.40\ 9.40\ 52.40\ 1.52]$	$x = [4.16\ 3.09\ 8.26\ 12.08$ $17.23\ 12.93\ 17.70\ 15.96$ $1.15\ 0.89\ 26.04]$

2.5. Constraint–Convergence-Based Model Management

The model management scheme employed in this work seeks to control the fidelity level of the EM simulation model throughout the optimization process. The trust-region procedure of Section 2.2 is employed as the main optimization engine. The fidelity level, represented here using a parameter F, is adjusted within the range $F_{min} \leq F \leq F_{max}$ as

defined in Section 2.4. The decision concerning the value of F is based upon a set of prerequisites as follows (see also Figure 2):

- Fidelity level is set to the lowest value $F_{\min}$ in the early stages of the optimization process (away from convergence). The decision is made regardless of the feasibility status of the solution. This permits a cost-efficient initial search within the design space;
- Fidelity is set to the highest value $F_{\max}$ upon convergence. This allows to ensure reliability of the final solution;
- Fidelity selection in the transition phase, either from infeasible to feasible, or approaching convergence, is based upon both the feasibility status of the solution (to be formulated later), and the convergence status of the procedure;
- The fidelity is selected from a continuous range of F-values, which improves the stability of the procedure. In particular, it allows for a smooth transition between model fidelities throughout the optimization process.

Figure 2. Conceptual illustration of the constraint–convergence-based model-fidelity adjustment. The six possible situations concerning feasibility status and convergence status, as well as their corresponding actions are shown.

The feasibility status is quantified using an exponential function $e^{-\tau^{(i)}}$, with $\tau^{(i)}$ being the aggregated constraint violation at the ith iteration, defined as

$$\tau^{(i)} = \max\{j = 1, \dots, k: \tau_j^{(i)}\}, \tag{8}$$

where

$$\tau_j^{(i)} = \zeta_j / \tau_{cj} \tag{9}$$

The normalization factors τ_{cj} are selected to have $\tau_j^{(i)}$ equal to unity when constraint violation reaches a user-defined threshold, which is constraint-specific. In our numerical experiments, this threshold is set to 2dB for reflection-related constraint and 1 dB for axial-ratio-related one, cf. Section 3.

The convergence status is quantified based on two convergence criteria, namely, the distance between consecutive vectors $||x^{(i+1)} - x^{(i)}|| \leq \delta_{arg}$, and the difference between consecutive objective function values $|U_A(R(x^{(i+1)})) - U_A(R(x^{(i)}))| \leq \delta_{obj}$, where δ_{arg} and δ_{obj} are the respective termination thresholds. We define

$$Q^{(i)}(\delta_{arg}, \delta_{obj}) = \max\left\{C_{arg}, C_{obj}\right\} \tag{10}$$

where C_{arg} and C_{obj} represent relative convergence factors, computed as

$$C_{arg} = \frac{\delta_{arg}}{||x^{(i+1)} - x^{(i)}||} \tag{11}$$

and

$$C_{obj} = \frac{\delta_{obj}}{\left|U_A(\boldsymbol{R}(\boldsymbol{x}^{(i+1)})) - U_A(\boldsymbol{R}(\boldsymbol{x}^{(i)}))\right|} \tag{12}$$

The model fidelity adjustment rules have been implemented in the form of the updating formula

$$F^{(i+1)} = \begin{cases} F_{\min} \text{ if } Q^{(i)}(\delta_{arg}, \delta_{obj}) \le \delta_{th} \\ \max\left\{F^{(i)}, F_{\min} + (F_{\max} - F_{\min})e^{-\tau^{(i)}}Q_L^{(i)}\right\} \text{ otherwise} \end{cases} \tag{13}$$

where $F^{(i)}$ is model fidelity at the ith iteration of the optimization process, and δth is a threshold for initiating an increase in model fidelity. In (13), $Q_L^{(i)}$ is the aggregated convergence status defined as

$$Q_L^{(i)} = \left[1 - \frac{\log(Q^{(i)}(\delta_{arg}, \delta_{obj}))}{\log \delta_{th}}\right] \tag{14}$$

Note that $Q_L^{(i)}$ changes between zero (for $Q^{(i)}$ equal to δth) and one (upon algorithm convergence, i.e., when $Q^{(i)}$ reaches a unity), which enables model fidelity adjustment between $F_{\min}$ and $F_{\max}$. Notwithstanding, formulation (14) does not account for an unexpected termination of the algorithm without reaching the high-fidelity discretization level. This may occur due to a reduction in the TR search radius below the termination thresholds. A workaround is an additional termination mechanism as suggested in [51]. Therein, the TR search radius $\delta^{(i)}$ is forcefully increased upon the convergence of the algorithm to

$$\delta^{(i+1)} = \frac{M_\delta \delta^{(i)} \delta_{arg}}{||\delta^{(i)}||} \tag{15}$$

whereas the fidelity level is set to the highest value, $F^{(i+1)} = F_{\max}$. This allows for finalizing the search process at the high-fidelity level, thereby ensuring reliability.

Figure 3 visualizes the dependence between the simulation model fidelity and the convergence status Q, as well as the aggregated constraint violation τ. The said dependence is monotonic with respect to both control factors, except from the early stages of the algorithm and when close to convergence.
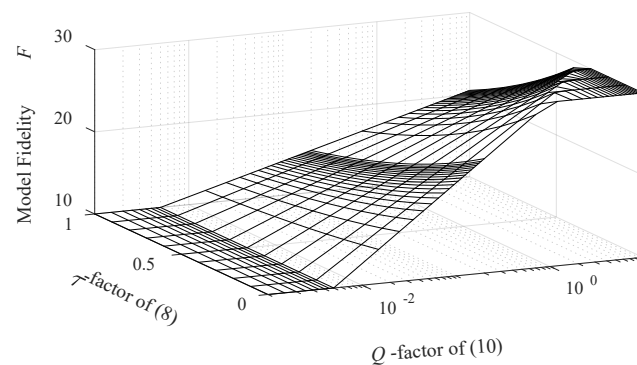
Figure 3. The dependence between the model fidelity and the convergence status Q, as well as aggregated constraint violation τ, cf. (13) and (14). Except from the initial stages of the optimization process and when approaching convergence, this dependence is a monotonic function of both control factors. The surface plot is created for exemplary model resolution levels $F_{\min} = 10$, $F_{\max} = 30$, assuming δth = 5×10^{-3}.

2.6. Proposed Miniaturization Procedure

The proposed expedited miniaturization algorithm combines the constraint–convergence-based model adjustment process of Section 2.5 and the adaptive penalty factor adjustment of Section 2.3.

The control parameters include:

- δ_{arg}, δ_{obj}, δ_{TR}—termination thresholds (cf. Sections 2.2 and 2.5);
- δth—a threshold used to initiate an increase in the model fidelity (cf. Section 2.5);
- M—sufficient constraint violation improvement factor (cf. Section 3.3);
- M_δ—a multiplication factor used to increase the TR search radius in (15) (upon convergence);
- τ_{cj}—constraint violation normalization factors in (8).

The termination thresholds are set to $\delta_{arg} = \delta_{obj} = \delta_{TR} = 10^{-3}$, which corresponds to a typically expected optimization process resolution, whereas the TR search region multiplication factor is selected as $M_\delta = 10$, as suggested in [51]. The values for other control parameters are as mentioned in the previous sections. Additionally, for any given antenna structure under design, a grid convergence study is conducted to determine the values of the lowest-fidelity model $F_{\min}$, and the high-fidelity model $F_{\max}$.

As mentioned earlier, the threshold value δth is used discriminate whether the optimization process is away from the convergence or in the transition phase to convergence, which corresponds to an increase in model fidelity.

Figure 4 provides a pseudocode of the algorithm. Steps 1 and 2 set the required values for the control parameters of the algorithm to initialize the miniaturization process. In Step 3, antenna response $R(x^{(i)})$ is acquired at the current fidelity level $F^{(i)}$. Step 4 uses finite differentiation to evaluate antenna sensitivity matrix $J(x^{(i)})$ at $F^{(i)}$. In Step 5, a linear approximation model of antenna responses $L^{(i)}(x)$ at the current design vector $x^{(i)}$ is identified, whereas in Step 6, the objective function $U_A(L^{(i)}(x))$ is constructed based on the linear approximation of antenna responses. The candidate design is found in Step 7 by minimizing $U_A(L^{(i)}(x))$. The antenna responses and the objective function at the new design $x^{(i+1)}$, $R(x^{(i+1)})$ and $U_A(x^{(i+1)})$ are evaluated in Steps 8 and 9, respectively. The TR search radius is updated in Step 10 (cf. [53]), whereas adaptive adjustment of the penalty coefficients is conducted in Step 11 [51]. Steps 12 determines whether to accept or reject $x^{(i+1)}$. If it is accepted, the new model-fidelity $F^{(i+1)}$ is computed using (13). Subsequently, the termination conditions are checked in Step 13. If the algorithm converges while the model fidelity has not reached its maximum value, we set $F^{(i)} = F_{\max}$, and the TR search radius is increased using (15). For additional explanation, Figure 5 provides a block-diagram illustration of the algorithm operation.

The algorithm is implemented in the MATLAB programming environment. The antenna models are simulated in the time-domain solver of CST Microwave Studio. A Visual-Basic-based MATLAB-CST socket is used to connect the algorithm to CST. The new design vector produced by the algorithm is submitted to CST through this socket. The antenna model is updated accordingly, and the responses are evaluated by the CST solver. Subsequently, the aforementioned responses are post-processed in MATLAB. This procedure continues until the convergence of the algorithm.

1. Set the iteration counter $i = 0$; $F^{(i)} = F_{min}$;　provide $x^{(i)}$;

2. Set　F_{min}, F_{max}, ∂th, S_j, τ_{cj}, $j = 1, \ldots, k$;

3. Evaluate antenna response $R(x^{(i)})$ at the fidelity level $F^{(i)}$;

4. Evaluate antenna sensitivities $J(x^{(i)})$ at the fidelity level $F^{(i)}$ using finite differentiation;

5. Construct a linear model

$$L^{(i)}(x) = R(x^{(i)}) + J(x^{(i)}) . (x - x^{(i)});$$

6. Construct objective function

$$U_A(L^{(i)}(x)) = A(x) + \beta_1^i . c_{L,1}(x)^2 + \ldots + \beta_k^i . c_{L,k}(x)^2;$$

7. Solve $\quad x^{(i+1)} = \arg \min_{x; \, x^{(i)} - \delta \leq x \leq x^{(i)} + \delta} U_A(L^{(i)}(x))$;

8. Evaluate antenna response $R(x^{(i+1)})$;

9. Evaluate objective function $U_A(x^{(i+1)})$;

10. Update TR size [53];

11. Update β_j^{i+1} [50];

12. **if**　$U_A(x^{(i+1)}) < U_A(x^{(i)})$

Compute $F^{(i+1)}$ using (13);

$i = i + 1$;

end

13. **if** $||x^{(i+1)} - x^{(i)}|| \leq \delta_{arg}$ OR $|U_A(R(x^{(i+1)})) - U_A(R(x^{(i)}))| \leq \quad \delta_{obj}$ OR $\quad ||\delta^{(i)}|| \leq \delta_{TR}$

if $F^{(i)} < F_{max}$

Set $F^{(i)} = F_{max}$;

Update $\quad \delta^{(i+1)} = \dfrac{M_\delta \delta^{(i)} \delta_{arg}}{|| \delta^{(i)} ||}$;

Go to 3;

else

Go to 14;

end

else

Go to 3;

end

14. **END**

Figure 4. Operation of the proposed optimization algorithm incorporating adaptive adjustment of penalty coefficients.
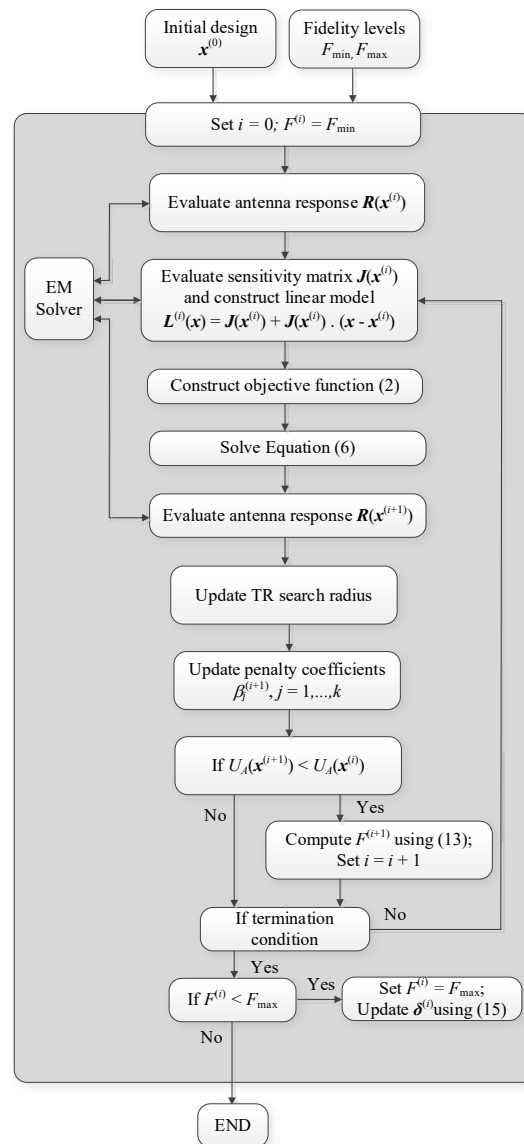
Figure 5. Operational flow of the proposed miniaturization procedure incorporating constraint–convergence-based model-fidelity adjustment scheme.

3. Verification Examples

This section provides numerical results of the proposed procedure conducted on five verification case studies. These include five benchmark structures optimized for minimum size. The benchmark structures include three broadband antennas with a single constraint defined as maximum in-band reflection coefficient. The other two structures are CP antennas with two constraints, maximum in-band reflection coefficient, and axial ratio response. The optimization results of the proposed miniaturization procedure with constraint-convergence-based model management scheme is compared to those of the benchmark algorithm incorporating a single-fidelity adaptive penalty-factor procedure.

The remaining part of this section is organized as follows. Section 3.1 provides a description of the benchmark antenna structures. Section 3.2 includes the experimental setup of the algorithm. Numerical results and their discussion are discussed in Sections 3.3 and 3.4, respectively.

3.1. Benchmark Antenna Structures

Figure 6 shows the verification antenna structures considered in this work. These include:

- A monopole antenna with L-shaped ground plane stub [57], Antenna I;
- A monopole antenna with a radiator slot and modified ground plane [58], Antenna II;
- A monopole antenna with two radiator slots and elliptical ground plane slit [59], Antenna III;
- A stacked circular polarization antenna with circular and annular slots [60], Antenna IV;
- A stacked circular polarization antenna with a cross-shaped radiator slot [61], Antenna V.
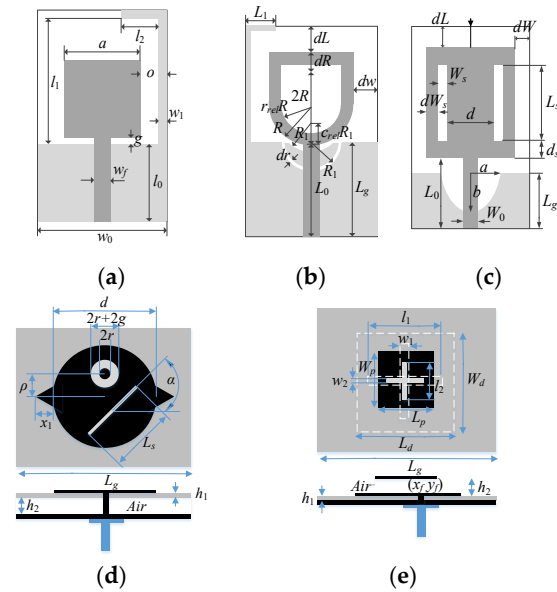
Figure 6. Topologies of the benchmark antennas: (**a**) Antenna I, (**b**) Antenna II, (**c**) Antenna III, (**d**) Antenna IV, (**e**) Antenna V.

The relevant data for all structures is shown in Table 1. This includes the target operating bandwidths as well as design constraints. The reflection constraints are set to $|S_{11}| \leq -10$ dB for all considered examples, which is a standard threshold commonly used in antenna design tasks. However, the presented method is generic and allows for setting up any limit as required by the user (cf. S_j in Equation (1)). Table 2 provides information about the simulation times associated with the lowest-fidelity model $F_{\min}$, and the high-fidelity model $F_{\max}$ for Antennas I–V. As mentioned before, these levels were decided upon through grid convergence studies (cf. Section 2.6).

Table 2. Antenna simulation time vs. model fidelity.

Benchmark Antenna Structure	Model Fidelity $[F_{\min} \ F_{\max}]$	Simulation Time $[T_{F\min} \ T_{F\max}]$ [s]
Antenna I	[11 30]	[145 466]
Antenna II	[12 20]	[49 124]
Antenna III	[10 24]	[31 164]
Antenna IV	[11 20]	[38 219]
Antenna V	[11 22]	[82 236]

3.2. Experimental Setup

The performance of the proposed procedure was evaluated by carrying out size reduction under the constraints listed in Table 1. The starting points were selected to be feasible for all antennas. This demonstrates that in all cases, there is a margin for size reduction as minimum-size designs are always allocated at the feasible region boundary.

As the primary contribution of this work is the incorporation of variable-fidelity models, the benchmark procedure is the algorithm employing the adaptively adjusted penalty coefficient scheme using a single (high-fidelity) EM model. The performance figures of interest include the antenna size, the levels of constraint violations, and the CPU cost of the optimization process. The latter is expressed in (i) absolute CPU time in hours, (ii) relative cost expressed in terms of the number of equivalent high-fidelity model evaluations, and (iii) relative computational savings enabled by the proposed algorithm as compared to the benchmark algorithm.

3.3. Result

Table 3 provides a comparison of the performance of the proposed and the benchmark procedure. As mentioned before, the performance figures include antenna size, constraint violation levels, and the computational cost of the optimization process. Figures 7–11 show—for Antennas I through V—the reflection responses at the initial and the optimized designs, along with the evolution of the model fidelity in the course of the algorithm run.

Table 3. Numerical results for Antennas I–V.

Performance Figures		Antenna I		Antenna II		Antenna III		Antenna IV		Antenna V	
		Adaptive β	This Work [4]	Adaptive β	This Work	Adaptive β	This Work	Adaptive β	This Work	Adaptive β	This Work
Area A (mm^2)		293	284	207	215	176	177	590	615	372.7	368
Constraint violation ζ_{S11} [1] (dB)		0.08	0.04	0.02	0	0.06	0	0	0	0	0.02
Constraint violation ζ_{AR} [2] (dB)		–	–	–	–	–	–	0.07	0.01	0	0
CPU Time	Absolute (h)	6.5	4.7	6.8	3.2	12.3	7.2	13.9	6.6	8.8	4.7
	Relative to R_f [3]	144	104	150	70	2.4	119	108	51	135	72
	Saving (%)	–	28	–	53	–	33	–	53	–	47

[1] Reflection coefficient constraint violation. [2] Axial ratio constraint violation. [3] High fidelity model. [4] Adaptive penalty function approach incorporating variable-fidelity model management.

(a)

(b)

Figure 7. Numerical results obtained for Antenna I using the proposed algorithm incorporating multi-fidelity model management and adaptive penalty coefficients: (**a**) reflection responses, (**b**) evolution of the model fidelity. The horizontal line represents the design specifications.

Figure 8. Numerical results obtained for Antenna II using the proposed algorithm incorporating multi-fidelity model management and adaptive penalty coefficients: (**a**) reflection responses, (**b**) evolution of the model fidelity. The horizontal line represents the design specifications.

Figure 9. Numerical results obtained for Antenna III using the proposed algorithm incorporating multi-fidelity model management and adaptive penalty coefficients: (**a**) reflection responses, (**b**) evolution of the model fidelity. The horizontal line represents the design specifications.

(**a**)

(**b**)

(**c**)

Figure 10. Numerical results obtained for Antenna IV using the proposed algorithm incorporating multi-fidelity model management and adaptive penalty coefficients: (**a**) reflection responses, (**b**) axial ratio responses, (**c**) evolution of the model fidelity. The horizontal lines represent the design specifications.

(**a**)
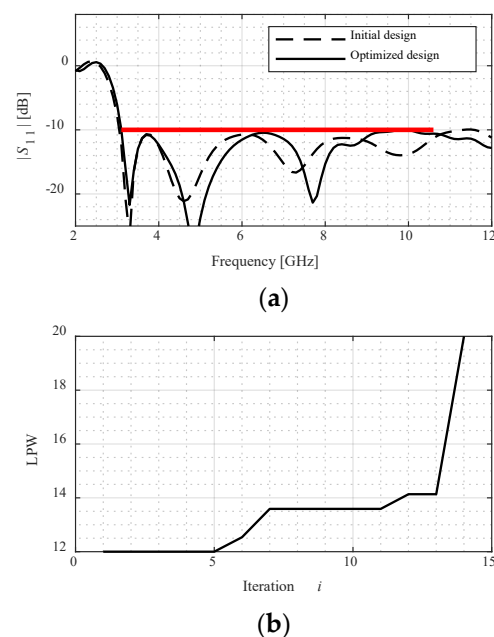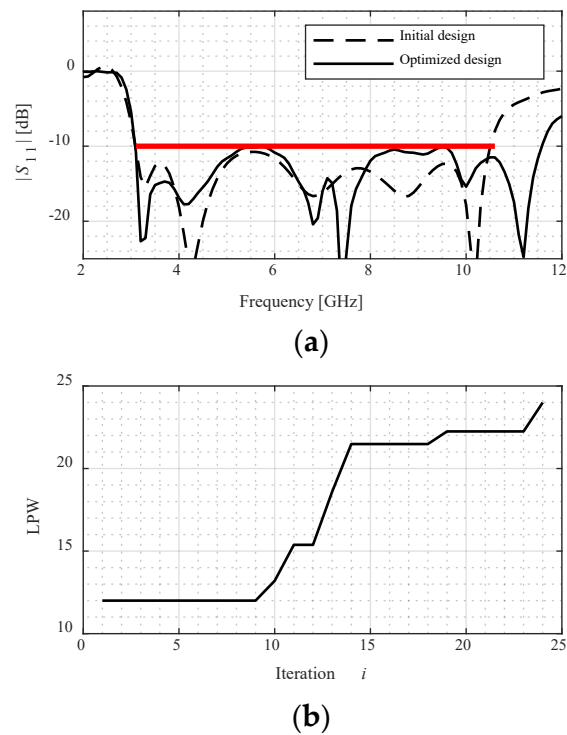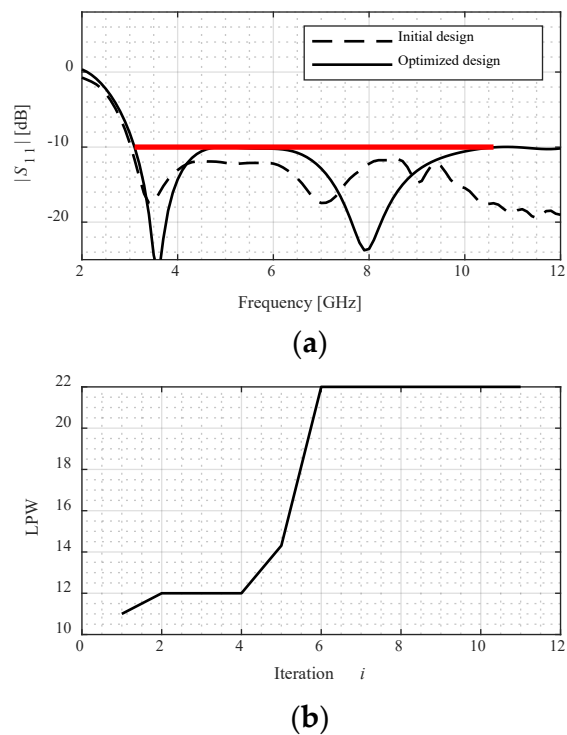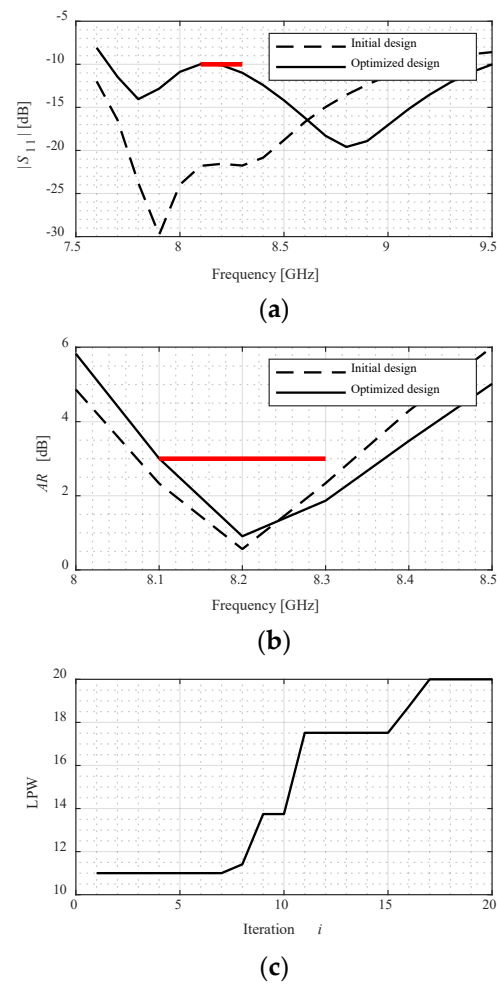

(**b**)

Figure 11. *Cont.*

(c)
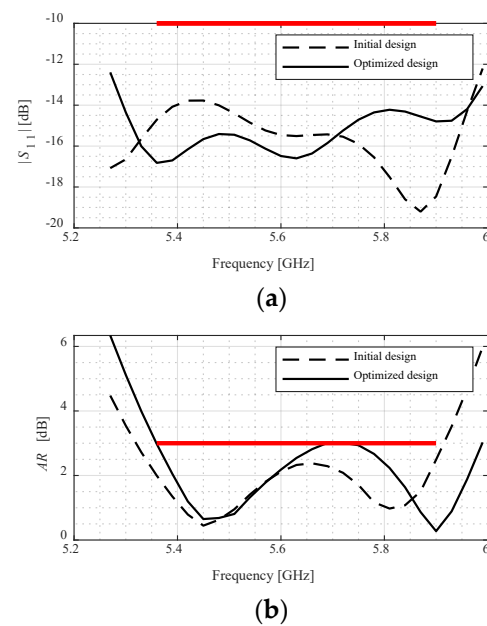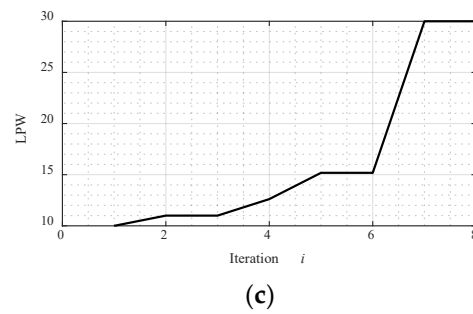
Figure 11. Numerical results obtained for Antenna V using the proposed algorithm incorporating multi-fidelity model management and adaptive penalty coefficients: (**a**) reflection responses, (**b**) axial ratio responses, (**c**) evolution of the model fidelity. The horizontal lines represent the design specifications.

3.4. Discussion

The numerical results lead to several conclusions concerning the efficacy of the proposed miniaturization algorithm incorporating the variable-fidelity EM model management scheme, and the adaptive penalty function approach. These can be briefly characterized as follows:

- The proposed variable-fidelity procedure allows for a considerable acceleration of the miniaturization process as compared to the single-fidelity adaptive penalty function approach, by about 28 to 53 percent and by 43 percent on average.
- The designs rendered by the proposed procedure are of a quality comparable to that produced by the single-fidelity procedure, both in terms of constraint satisfaction and achievable size reduction rates. For Antennas I–V, all constraint violations are kept at the same level of 0.0 dB, whereas the achieved antenna footprint area is smaller by 9 mm^2, larger by 8 mm^2, larger by 1 mm^2, degraded by 25 mm^2, and smaller by about 5 mm^2, respectively. In practical terms, these differences can be considered minor.
- The reliability of the miniaturization process is ensured by conducting the final iterations of the optimization process at the level of a high-fidelity model, which gives an accurate account of antenna characteristics. This can be observed in the plots showing the evolution of the model fidelity, as included in Figures 7–11. As can be observed in Figure 7, the initial design is allocated in the feasible region to demonstrate the margin for size reduction. It can be seen that the design moves towards the boundary of the feasible region as the final allocation of the optimized design, at which the reflection constraint is active. Figure 7b shows the evolution of the model fidelity across the iterations of the optimization process. It is set to the lowest value in the first few iterations, as the algorithm is away from convergence. There is a gradual increase in the model fidelity between iterations 5 and 13, corresponding to the transition phase either from infeasible to feasible, or due to approaching convergence. This increase is based on the feasibility and the convergence status of the optimization process as formulated in (13). The model fidelity is set to the highest level at the last iteration of the optimization process, when approaching convergence.

In brief, the proposed variable-fidelity miniaturization procedure demonstrably yields significant CPU time savings, while producing designs that are of comparable quality to those obtained with the single-fidelity adaptive penalty function approach. Further acceleration can be realized by the incorporation of sparse sensitivity updates [22], which will be a part of our future work.

4. Conclusions

This paper proposes a novel procedure for expedited EM-driven miniaturization of antenna structures. Our methodology incorporates a constraint-violation and convergence-based model management, combined with the adaptive penalty function procedure. The

model fidelity is continuously adjusted using the control factors related to violation of the design constraints as well as the convergence status of the algorithm. The proposed procedure enables a cost-efficient exploration of the design space by utilizing the lowest-fidelity level at the early stages of the optimization process. The model-fidelity is gradually refined towards the highest-fidelity level in the final stages of the optimization process. Comprehensive numerical verification involving five microstrip antenna structures demonstrated the advantages of the presented approach, primarily a considerable speedup of the optimization process. The average computational savings are as high as 43 percent with respect to the benchmark algorithm. At the same time, no quality degradation was observed. Our future work will focus on the development and incorporation of further acceleration mechanisms, including sparse sensitivity updates, as well as dimensionality reduction methods.

Author Contributions: Conceptualization, M.M. and S.K.; methodology, M.M. and S.K.; software, M.M.; validation, M.M.; formal analysis, S.K.; investigation, M.M.; resources, S.K.; data curation, M.M.; writing—original draft preparation, M.M. and S.K.; writing—review and editing, M.M. and S.K.; visualization, M.M.; supervision, S.K.; project administration, S.K.; funding acquisition, S.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Le, T.T.; Yun, T.-Y. Miniaturization of a dual band wearable antenna for dual-band WBAN applications. *IEEE Trans. Antennas Propag.* **2020**, *19*, 1452–1456. [CrossRef]
2. Agneessens, S.; Rogier, H. Compact half diamond dual-band textile HMSIW on-body antenna. *IEEE Trans. Antennas Propag.* **2014**, *62*, 2374–2381. [CrossRef]
3. Abbosh, A.M. Miniaturized microstrip-fed tapered-slot antenna with altrawideband performance. *IEEE Antennas Wirel. Propag. Lett.* **2009**, *8*, 690–692. [CrossRef]
4. Abbosh, A.M. Miniaturization of planar ultrawideband antenna via corrugation. *IEEE Antennas Wirel. Propag. Lett.* **2008**, *7*, 685–688. [CrossRef]
5. Arif, A.; Zubair, M.; Ali, M.; Khan, M.U.; Mehmood, M.Q. A compact, low-profile fractal antenna for wearable on-bode WBAN applications. *IEEE Antennas Wirel. Propag. Lett.* **2019**, *18*, 981–985. [CrossRef]
6. Yand, Y.; Zhao, Z.; Ding, X.; Nie, Z.; Liu, Q. Compact UWB slot antenna utilizing travelling-wave mode based on slitline transitions. *IEEE Trans. Antennas Propag.* **2019**, *67*, 140–150.
7. Dong, J.; Qin, W.; Wang, M. Fast multi-objective optimization of multi-parameter antenna structures based on improved BPNN surrogate model. *IEEE Access* **2019**, *7*, 77692–77701. [CrossRef]
8. Lee, K.; Sung, H.; Park, E.; Lee, I. Joint optimization for one and two-way MIMO AF multiple relay systems. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 3671–3681. [CrossRef]
9. Tomasson, J.A.; Koziel, S.; Pietrenko Dabrowska, A. Quasi-global optimization of antenna structures using principal components and affine subspace-spanned surrogates. *IEEE Access* **2020**, *8*, 50078–50084. [CrossRef]
10. Al-Azza, A.A.; Al-Jodah, A.A. Spider monkey optimization: A novel technique for antenna optimization. *IEEE Antennas Wirel. Propag. Lett.* **2016**, *15*, 1016–1019. [CrossRef]
11. Lalbakhsh, A.; Afzal, M.U.; Esselle, K.P. Multiobjective particle swarm optimization to design a time-delay equalizer metasurface for an electromagnetic band-gap resonator antenna. *IEEE Antennas Wirel. Propag. Lett.* **2017**, *16*, 912–915. [CrossRef]
12. Goudos, S.K.; Siakavara, K.; Samaras, T.; Vafiadis, E.E.; Sahalos, J.N. Self-adaptive differential evolution applied to real-valued antenna and microwave design problems. *IEEE Trans. Antennas Propag.* **2011**, *59*, 1286–1298. [CrossRef]
13. Koziel, S. Computationally efficient multi-fidelity multi-grid design optimization of microwave structures. *Appl. Comp. Electromagn. Soc. J.* **2010**, *25*, 578–586.

14. Ohira, M.; Miura, A.; Taromaru, M.; Ueba, M. Efficient gain optimization techniques for azimuth beam/null steering of inverted-F multiport parasitic array radiator (MuPAR) antenna. *IEEE Trans. Ant. Propag.* **2012**, *60*, 1352–1361. [CrossRef]

15. Wang, J.; Yang, X.S.; Wang, B.Z. Efficient gradient-based optimization of pixel antenna with large-scale connections. *IET Microw. Ant. Propag.* **2018**, *12*, 385–389. [CrossRef]

16. Kalantari, L.S.; Bakr, M.H. Wideband cloaking of objects with arbitrary shapes exploiting adjoint sensitivities. *IEEE Trans. Antennas Propag.* **2016**, *64*, 1963–1968. [CrossRef]

17. Paronneau, O. *Optimal Shape Design for Elliptic Systems*; Springer: Berlin/Heidelberg, Germany, 1982; Volume 3, pp. 42–66.

18. Jameson, A. Aerodynamic design via control theory. *J. Sci. Comput.* **1988**, *3*, 233–260. [CrossRef]

19. El Sabbagh, M.A.; Bakr, M.H.; Nilolova, N.K. Sensitivity analysis of the scattering parameters of microwave filters using the adjoint network method. *Int. J. RF Microw. CAE* **2006**, *16*, 569–606. [CrossRef]

20. Papadimitriou, D.; Giannakoglou, K. Aerodynamic shape optimization using first and second order adjoint and direct approaches. *Arch. Comput. Methods Eng.* **2008**, *15*, 447–488. [CrossRef]

21. Toivann, J.I.; Mäkinen, R.A.E.; Järvenpää, S.; Ylä-Oijala, P.; Rahola, J. Electromagnetic sensitivity analysis and shape optimization using method of moments and automatic differentiation. *IEEE Trans. Antennas Propag.* **2009**, *57*, 168–175. [CrossRef]

22. Director, S.; Rohrer, R. The generalized adjoint network and network sensitivities. *IEEE Trans. Circuit Theory* **1969**, *16*, 318–323. [CrossRef]

23. Easum, J.A.; Nagar, J.; Werner, P.L.; Werner, D.H. Efficient multi-objective antenna optimization with tolerance analysis through the use of surrogate models. *IEEE Trans. Ant. Propag.* **2018**, *66*, 6706–6715. [CrossRef]

24. Sarkar, T.K.; Chen, H.; Palma, M.S.; Zhu, M. Lessons learned using a physics based macro model for analysis of radio wave propagation in wireless transmission. *IEEE Trans. Antennas Propag.* **2019**, *67*, 2150–2157. [CrossRef]

25. Alzahed, A.M.; Mikki, S.M.; Antar, Y.M.M. Nonlinear mutual coupling compensation operator design using a novel electromagnetic machine learning paradigm. *IEEE Ant. Wirel. Propag. Lett* **2019**, *18*, 861–865. [CrossRef]

26. Hassan, A.K.S.O.; Etman, A.S.; Soliman, E.A. Optimization of a novel nano antenna with two radiation modes using kriging surrogate models. *IEEE Photonics J.* **2018**, *10*, 4800807. [CrossRef]

27. Rawat, A.; Yadav, R.N.; Shrivastava, S.C. Neural network applications in smart antenna arrays: A review. *AEU Int. J. Elec. Commun.* **2012**, *66*, 903–912. [CrossRef]

28. Cai, J.; King, J.; Yu, C.; Liu, J.; Sun, L. Support vector regression-based behavioral modeling technique for RF power transistors. *IEEE Microw. Wirel. Compon. Lett.* **2018**, *28*, 428–430. [CrossRef]

29. Van Der Herten, J.; Couckuyt, I.; Deschrijver, D.; Dhaene, T. A fuzzy hybrid sequential design strategy for global surrogate modeling of high-dimensional computer experiments. *SIAM J. Sci. Comput.* **2015**, *32*, A1020–A1039.

30. Rayas-Sánchez, J.E.; Chávez-Hurtado, J.L.; Brito-Brito, Z. Optimization of full-wave EM models by low-order low-dimension polynomial surrogate functionals. *Int. J. Numer. Model. Electron. Netw. Devices Fields* **2015**, *30*, e2094. [CrossRef]

31. Bandler, J.W.; Cheng, Q.S.; Dakroury, S.A.; Mohamed, A.S.; Bakr, M.H.; Madsen, K.; Sondergaard, J. Space mapping: The state of the art. *IEEE Trans. Microw. Theory Tech.* **2004**, *52*, 337–361. [CrossRef]

32. Robinson, T.D.; Eldred, M.S.; Willcox, K.E.; Haimes, R. Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. *AIAA J.* **2008**, *46*, 2814–2822. [CrossRef]

33. Echeverria, D.; Hemker, P.W. Space mapping and defect correction. *Comput. Methods Appl. Math.* **2005**, *5*, 107–136. [CrossRef]

34. Manchec, A.; Quendo, C.; Favennec, J.-F.; Rius, E.; Person, C. Synthesis of capacitive-coupled dual-behavior resonator (CCDBR) filters. *IEEE Trans. Microw. Theory Tech.* **2006**, *54*, 2346–2355. [CrossRef]

35. Feng, F.; Zhang, J.; Zhang, W.; Zhao, Z.; Jin, J.; Zhang, Q.J. Coarse-and fine-mesh space mapping for EM optimization incorporating mesh deformation. *IEEE Microw. Wirel. Compon. Lett.* **2019**, *29*, 510–512. [CrossRef]

36. Zhou, R.; Sun, J.; Wei, S.; Wang, J. Synthesis of conformal array antenna for hypersonic platform SAR using modified particle swarm optimization. *IET Radar Sonar Navigat.* **2017**, *11*, 1235–1242. [CrossRef]

37. Greda, L.A.; Winterstein, A.; Lemes, D.L.; Heckler, M.V.T. Beamsteering and beamshaping using a linear antenna array based on particle swarm optimization. *IEEE Access* **2019**, *7*, 11562–141573. [CrossRef]

38. Aoad, M.; Simsek, M.; Aydin, Z. Development of knowledge based response correction for a reconfigurable N-shaped microstrip antenna design. In Proceedings of the 2015 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), Ottawa, ON, Canada, 11–14 August 2015; pp. 1–3.

39. Koziel, S.; Ogurtsov, S.; Cheng, Q.S.; Bandler, J.W. Rapid EM-based microwave design optimization exploiting shape-preserving response prediction and adjoint sensitivities. *IET Microw. Ant. Propag.* **2014**, *8*, 775–781. [CrossRef]

40. Rayas-Sánchez, J.E. Power in simplicity with ASM: Tracing the aggressive space mapping algorithm over two decades of development and engineering applications. *IEEE Microw. Mag* **2016**, *17*, 64–76. [CrossRef]

41. Cervantes-González, J.C.; Rayas-Sánchez, J.E.; López, C.A.; Camacho-Pérez, J.R.; Brito-Brito, Z.; Chávez-Hurtado, J.L. Space mapping optimization of handset antennas considering EM effects of mobile phone components and human body. *Int. J. RF Microw. CAE* **2016**, *26*, 121–128. [CrossRef]

42. De Villiers, D.I.L.; Couckuyt, I.; Dhaene, T. Multi-objective optimization of reflector antennas using Kriging and probability of improvement. In Proceedings of the 2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, San Diego, CA, USA, 9–14 July 2017; pp. 985–986.

43. Bandler, J.W.; Cheng, Q.S.; Georgieva, N.; Ismail, M.A. Implicit space mapping EM-based modeling and design exploiting preassigned parameters. In Proceedings of the 2002 IEEE MTT-S International Microwave Symposium Digest (Cat. No.02CH37278), Seattle, WA, USA, 2–7 June 2002; pp. 713–716.

44. Aoad, A.; Simsek, M.; Aydin, Z. Knowledge based response correction method for design of reconfigurable N-shaped microstrip patch antenna using inverse ANNs. *Int. J. Numer. Model Electron. Netw. Devices Fields* **2015**, *30*, e2129. [CrossRef]

45. Diehl, M.; Walther, A.; Georg Bock, H.; Kostina, E. An adjoint-based SQP algorithm with quasi-Newton Jacobin updates for iequality constrained optimization. *Opt. Methods Softw.* **2009**, *25*, 531–552. [CrossRef]

46. Koziel, S.; Pietrenko-Dabrowska, A. Expedited feature-based quasi-global optimization of multi-band antenna input characteristics with jacobian variability tracking. *IEEE Access* **2020**, *8*, 83907–83915. [CrossRef]

47. Koziel, S.; Pietrenko-Dabrowska, A. Reduced-cost electromagnetic-driven optimization of antenna structures by means of trust-region gradient-search with sparse Jacobian updates. *IET Microw. Ant. Propag.* **2019**, *13*, 1646–1652. [CrossRef]

48. Koziel, S.; Ogurtsov, S. Multi-level design optimization of microwave structures with automated model fidelity adjustment. In Proceedings of the 2013 IEEE MTT-S International Microwave Symposium Digest (MTT), Seattle, WA, USA, 2–7 June 2013.

49. Koziel, S. Objective relaxation algorithm for reliable simulation-driven size reduction of antenna structure. *IEEE Ant. Wirel. Propag. Lett.* **2017**, *16*, 1949–1952. [CrossRef]

50. Mahrokh, M.; Koziel, S. Optimization-based antenna miniaturization using adaptively adjusted penalty factors. *Electronics* **2021**, *10*, 1751. [CrossRef]

51. Mahrokh, M.; Koziel, S. Explicit size-reduction of circularly polarized antennas through constrained optimization with penalty factor adjustment. *IEEE Access* **2021**, *9*, 132390–132396. [CrossRef]

52. Koziel, S.; Pietrenko-Dabrowska, A. Accelerated gradient-based optimization of antenna structures using multi-fidelity simulations and convergence-based model management scheme. *IEEE Trans. Ant. Propag.* **2021**, *69*, 8778–8789. [CrossRef]

53. Conn, A.R.; Gould, N.I.M.; Toint, P.L. *Trust Region Methods*; MPS-SIAM Series on Optimization; SIAM: Philadelphia, PA, USA, 2000.

54. Koziel, S.; Ogurtsov, S. Model management for cost-efficient surrogate-based optimization of antennas using variable-fidelity electromagnetic simulations. *IET Microw. Ant. Propag.* **2012**, *6*, 1643–1650. [CrossRef]

55. Sendrea, R.E.; Zekios, C.L.; Georgakopoulos, S.V. A multi-fidelity surrogate optimization method based on analytical models. In Proceedings of the 2021 IEEE MTT-S International Microwave Symposium (IMS), Atlanta, GA, USA, 7–25 June 2021; pp. 70–73.

56. Xiao, L.; Shao, W.; Ding, X.; Wang, B. Dynamic adjustment kernel extreme learning machine for microwave component design. *IEEE Trans. Microw. Theory Tech.* **2018**, *66*, 4452–4461. [CrossRef]

57. Koziel, S.; Bekasiewicz, A. Comprehensive comparison of compact UWB antenna performance by means of multi-objective optimization. *IEEE Trans. Ant. Propag.* **2017**, *65*, 3427–3436. [CrossRef]

58. Alsath, M.G.N.; Kanagasabai, M. Compact UWB monopole antenna for automotive communications. *IEEE Trans. Antennas Propag.* **2015**, *63*, 4204–4208. [CrossRef]

59. Haq, M.A.; Koziel, S. Simulation-based optimization for rigorous assessment of ground plane modifications in compact UWB antenna design. *Int. J. RF Microw. Comput. Aided Eng.* **2018**, *28*, e21204. [CrossRef]

60. Kumar, B.P.; Kumar, C.; Guha, D. A new design approach to improve the circular polarization characteristics of a microstrip antenna. In Proceedings of the 2018 IEEE Indian Conference on Antennas and Propogation (InCAP), Hyderabad, India, 16–19 December 2018; pp. 1–2.

61. Malekabadi, S.A.; Attari, A.R.; Mirsalehi, M.M. Compact broadband circular polarized microstrip antenna with wideband axial-ratio bandwidth. In Proceedings of the 2008 International Symposium on Telecommunications, Tehran, Iran, 27–28 August 2008; pp. 106–109.

6.1 Chu's Limit and Antenna Miniaturization

This section provides an elaboration on how Chu's lower bound [107] imposes a fundamental limit on antenna miniaturization, as well as a comparison of the quality factor, Q, of the optimized antennas with Chu's limit.

Chu's limit defines a lower bound for the antenna quality factor formulated as

$$Q > \frac{1}{(kr)^3} + \frac{1}{kr} \tag{6.1}$$

where r is the radius of the assumed virtual sphere enclosing the antenna structure, $k = 2\pi/\lambda$, and λ is the wavelength.

The aim is to provide a comparison of the initial antenna quality factor, $Q_{initial}$, the miniaturized antenna quality factor, $Q_{miniaturized}$, and Chu's lower bound, Q_{Chu}. The values of $Q_{initial}$ and $Q_{miniaturized}$ are obtained using CST based on the corresponding antenna dimensions, while Q_{Chu} is obtained analytically using equation (6.1) based on the dimensions of the initial antenna.

Figures 6.1 to 6.5 illustrate values of $Q_{initial}$, $Q_{miniaturized}$, and Q_{Chu} for the first ten propagation modes corresponding to the five microstrip antennas optimized for minimum size, including Example I: UWB monopole antenna, Example II: UWB rectangular monopole antenna, Example III: UWB rectangular-slot monopole antenna, Example IV: stacked patch CP antenna, Example V: circular patch CP antenna with annular and rectangular slots. It is noteworthy to say that $Q_{initial}$, $Q_{miniaturized}$ are scaled down by a factor of 10^3 for illustration purposes.
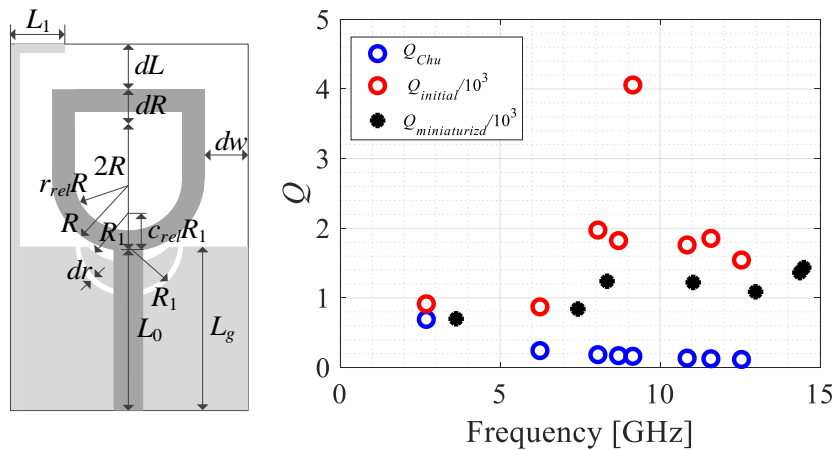
Figure 6.1:   A comparison of $Q_{initial}$, $Q_{miniaturized}$, and Q_{Chu} for Example I: UWB monopole antenna.
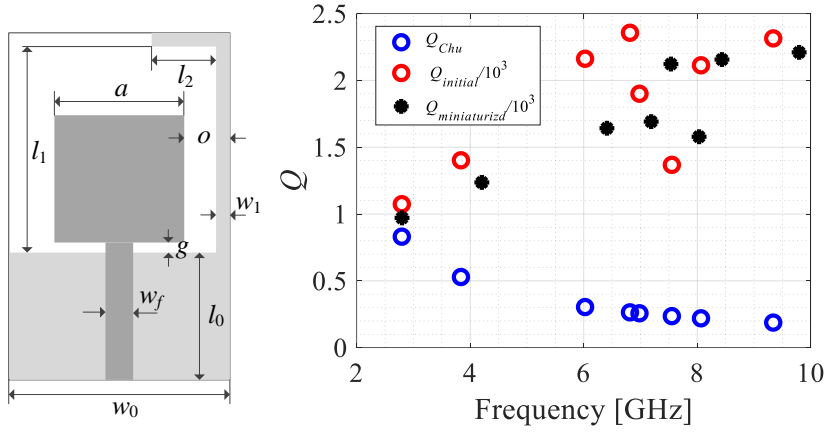
Figure 6.2: A comparison of $Q_{initial}$, $Q_{miniaturized}$, and Q_{Chu} for Example II: UWB rectangular monopole antenna.
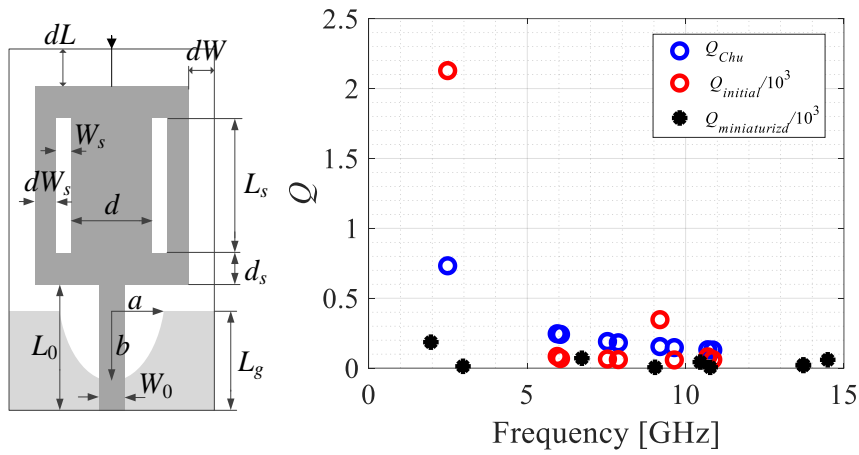
Figure 6.3: A comparison of $Q_{initial}$, $Q_{miniaturized}$, and Q_{Chu} for Example II: UWB rectangular-slot monopole antenna.
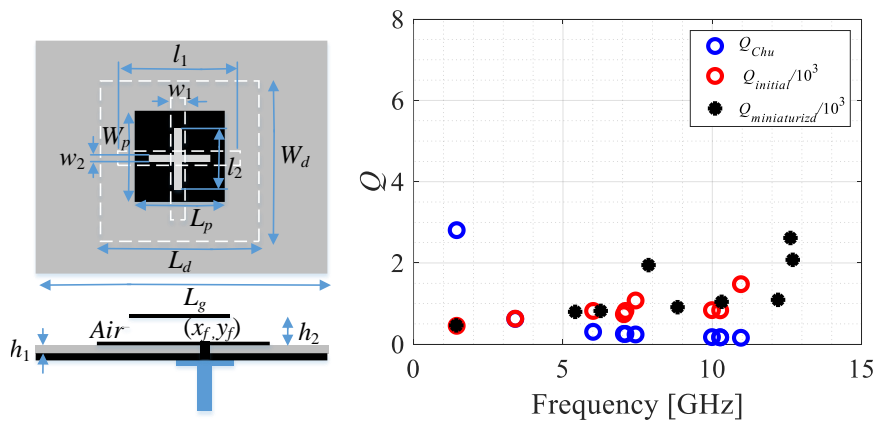
Figure 6.4: A comparison of $Q_{initial}$, $Q_{miniaturized}$, and Q_{Chu} for Example II: stacked patch CP antenna.
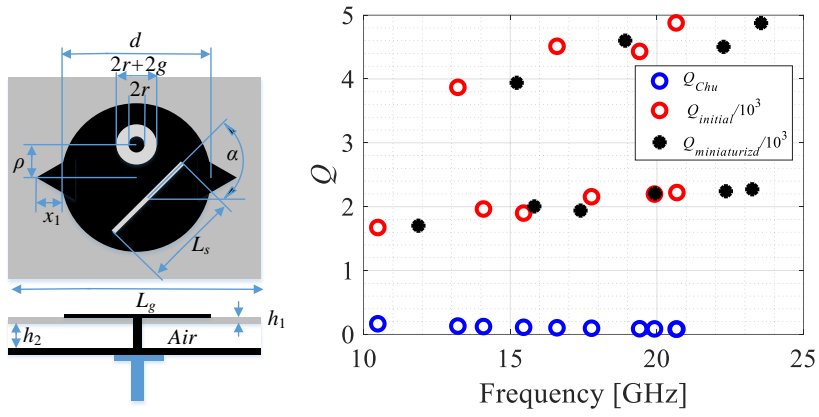
Figure 6.5: A comparison of $Q_{initial}$, $Q_{miniaturized}$, and Q_{Chu} for Example II: circular patch CP antenna with annular and rectangular slots.

Chapter 7

7 Paper # 4

Slawomir Koziel, Anna Pietrenko-Dabrowska, and Marzieh Mahrokh

On Decision-Making Strategies for Improved-Reliability Size Reduction of Microwave Passives: Intermittent Correction of Equality Constraints and Adaptive Handling of Inequality Constraints
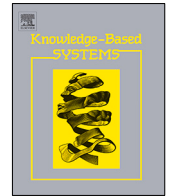
On decision-making strategies for improved-reliability size reduction of microwave passives: Intermittent correction of equality constraints and adaptive handling of inequality constraints

Slawomir Koziel [a,b], Anna Pietrenko-Dabrowska [b,*], Marzieh Mahrokh [a]

[a] *Engineering Optimization & Modeling Center, Reykjavik University, 102 Reykjavik, Iceland*
[b] *Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland*

ARTICLE INFO

ABSTRACT

Design optimization of passive microwave components is an intricate process, especially if the primary objective is a reduction of the physical size of the structure. The latter has become an important design consideration for a growing number of modern applications (mobile communications, wearable/implantable devices, internet of things), where miniaturization is imperative due to a limited space allocated for the electronic circuitry. Optimization-based size reduction is a heavily constrained task, with several acceptance thresholds imposed on electrical characteristics of the system. The challenges are pronounced whenever equality constraints are involved (e.g., related to power split ratio requirements), in which case the feasible space is a thin set, thereby difficult to be explored throughout the optimization process. This feature makes conventional methods, such as penalty function approaches or algorithms with explicit constraint handling, of limited reliability. In this paper, we introduce a novel technique for reliable control of equality constraints in simulation-driven size reduction of microwave components. Our methodology involves an intermittent optimization-based correction of equality constraints. This is essentially a knowledge-based decision-making strategy implemented as a supplementary optimization stage, and launched before each iteration of the core algorithm. Constraint violation is reduced without being detrimental to the remaining figures of merit, in particular, the circuit size and inequality constraints. Meanwhile, inequality constraints are handled using a penalty function approach with adaptive adjustment of penalty coefficients. The proposed technique facilitates exploration of the feasible space, and allows for achieving reduced miniaturization rates in comparison to the benchmark methods, while ensuring a reliable control of the design constraints. These advantages have been demonstrated using four microstrip couplers, with consistent results obtained for all considered circuits.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Compact size has become an important prerequisite in the design of contemporary high-frequency components, including passive microwave devices. The underlying reason is a limited physical space allocated for electronic circuitry in an increased number of application areas, some of which include internet of things [1], mobile communications [2], autonomous vehicles [3], energy harvesting [4], wearable [5] and implantable devices [6], as well as medical imaging [7]. Miniaturization of passive components is hindered by the fact that their dimensions are related to the guided wavelength, therefore, downscaling

requires the employment of dedicated techniques, e.g., transmission line (TL) meandering [8,9], multi-layer implementations [10], employment of the slow-wave phenomenon [11] (e.g., compact microwave resonant cells, CMRCs [12]), or the application of high-permittivity substrates [13]. Other methods include geometrical alterations (stubs [14], slots [15], defected ground structures [16], substrate-integrated waveguides [17], shorting pins [18], etc.). The circuits designed using these methods are geometrically more complex, and often feature dense layouts. For reliability reasons, they have to be evaluated using full-wave electromagnetic (EM) analysis (e.g., to account for EM cross-coupling effects) rather than equivalent network models.

Achieving the best possible performance of a circuit requires careful tuning of its geometry parameters. Modern structures, especially those developed using the techniques described in the previous paragraph, are described by larger numbers of parameters than conventional ones (e.g., typically, CMRC features

---

four to six parameters as compared to two for a conventional TL). In the case of compact circuits, the primary objective is to minimize the circuit size, whereas maintaining appropriate levels of electrical parameters (centre frequency, bandwidth, power split ratio, matching, port isolation, phase response) become design constraints. Parameter tuning under these circumstances is an intricate task that requires the employment of rigorous numerical optimization. Unfortunately, EM-driven optimization entails considerably computational costs. Local algorithms (both gradient-based [19] and derivative-free [20]) may require tens or even hundreds of EM analyses depending on the problem size. Global optimization (e.g., [21–23]) is nowadays most often performed using nature-inspired algorithms, e.g., [24–26]. Globalized search, but also multi-objective design [27–30], as well as statistical design (e.g., yield optimization [31,32]) is typically associated with significantly higher CPU costs.

Needless to say, improving computational efficiency of the optimization procedures has attracted a great deal of attention. Among many techniques developed to alleviate the cost-related difficulties, one can mention adjoint sensitivities [33,34], restricted Jacobian updating strategies [35–37], mesh deformation methods [38], as well as utilization of surrogate modelling methods [39–52]. The major advantage of surrogate-based optimization is that shifting the computational burden onto fast replacement models may lead to a remarkable reduction of the running costs of the algorithm in the way unattainable for other methods [53]. However, surrogate modelling faces difficulties on its own. For example, data-driven models (kriging [54], radial basis functions [55], neural networks [56], Gaussian process regression [57]) are greatly affected by the curse of dimensionality. Combining them with machine learning methods [58,59] and sequential sampling procedures [60], mitigates these issues to some extent, yet, for systems featuring highly-nonlinear responses (e.g., narrow- and multi-band structures [61]), only a few parameters can be handled in an efficient manner. Physics-based surrogates (space mapping [62], response correction [63–65]) are more immune to dimensionality problems but are also less generic, and more dependent on an appropriate selection and setup of lower-fidelity representations [66].

In the context of simulation-based circuit miniaturization, CPU expenses incurred by EM-driven optimization constitute only one of the practical problems. Another one is the enforcement of constraints imposed on electrical performance figures. Size reduction normally leads to a degradation thereof (e.g., reduction of the circuit bandwidth), meaning that the constraints are active at the minimum-size designs [67]. Exploration of the feasible region boundary is a challenging task [68]. At the same time, evaluation of the constraints is costly (as it necessitates executing EM analysis of the circuit), which aggravates the problem. A commonly used workaround is an implicit constraint handling using a penalty function approach [69], with the principal objective (size reduction) being augmented by additional terms proportional to violations of the conditions imposed on the S-parameters [70]. This approach simplifies the problem by turning it into a formally unconstrained one, but the reliability of the optimization process is contingent upon appropriate selection of the penalty coefficients, which is non-trivial [71]. Recently, algorithms with the adaptive adjustment of penalty factors have been proposed, and demonstrated superior over the manual setup [71,72]. A method with explicit constraint handling was also suggested, leading to comparable benefits [73].

In the case of a size reduction of microwave passive components, it is the equality constraints that are particularly troublesome. Representative examples are coupling structures with the performance requirements formulated for the power split ratio. The feasible space for equality constraints is a "thin" (measure zero) set, which is difficult to be explored because any deviation from the required power division level results in the design being infeasible. This is in contrast to inequality constraints, where the feasible region is of full dimensionality. Consequently, the operation of the optimization process is more sensitive to the treatment of equality-rather than inequality-type of performance conditions. This work proposes a novel approach to constrained miniaturization of microwave components, which involves dedicated decision-making strategies concerning constraint handling. In particular, we develop an intermittent optimization-based correction of equality constraints, where a separate optimization sub-problem is executed after each iteration of the algorithm, aiming at reduction of the equality constraint violation before proceeding with the core size-reduction procedure. The correction is arranged not to degrade the primary objective (circuit size) nor the inequality constraints. The inequality constraints are controlled by means of a penalty function approach with adaptive penalty coefficients, similarly as in [71]. As demonstrated using several microstrip couplers, the proposed technique offers a consistent control over both equality and inequality constraints, and results in improved miniaturization rates with regard to a penalty function method with manual coefficient setup.

The originality and the technical contribution of this work can be summarized as follows: (i) the development of the concept of optimization-based equality constraints control, (ii) the development of a management scheme with optimization-based correction of equality constraints and adaptive handling of inequality constraints for reliability-enhanced EM-driven size reduction of microwave passives, (iii) corroborating a possibility of a precise correction of potential violations of the equality constraints during the optimization run, (iv) demonstrating the improvement of the performance and reliability of the size-reduction process, which manifests itself in yielding smaller footprints of microwave circuits while maintaining the target values of electrical performance figures. To the best knowledge of the authors, none of these have been reported so far in the literature.

2. Size reduction of microwave passives with correction-based handling of equality constraints

This section outlines the optimization-based size reduction procedure proposed in the work. We begin by recapitulating the formulation of the EM-driven miniaturization problem with the emphasis on design constraints (Section 2.1), then, a brief characterization of a penalty function approach is given (Section 2.2). An outline of the adaptive adjustment of penalty coefficients, utilized here to handle inequality constraints, is provided in Section 2.3. In Section 2.4, we introduce the optimization-based correction scheme, developed to control equality constraint violations, whereas Section 2.5 summarizes the entire framework.

2.1. EM-driven miniaturization task

The size reduction task is formulated here as a nonlinear minimization problem of the form

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} A(\boldsymbol{x}) \tag{1}$$

where $A(\boldsymbol{x})$ is the circuit size (e.g., footprint area), and $\boldsymbol{x}$ is the vector of circuit (typically, geometry) parameters, $\boldsymbol{x} = [x_1 \ldots x_n]^T$. The problem (1) is subject to constraints, which can be of inequality

$$g_k(\boldsymbol{x}) \le 0, k = 1, \ldots, n_g \tag{2}$$

or equality type

$$h_k(\boldsymbol{x}) = 0, k = 1, \ldots, n_h \tag{3}$$

Table 1
Example constraints in size-reduction of microwave components and possible formulations of penalty functions.

Constraint	Type	Analytical description[a]	Penalty function
Input matching $|S_{11}|$ not exceeding −20 dB over the operating bandwidth $[f_1 \, f_2]$	Inequality	$|S_{11}(\boldsymbol{x}, f)| \leq -20$ dB for $f \in [f_1 \, f_2]$	$c(\boldsymbol{x}) = \left[\dfrac{\max\{\max f_1 \leq f \leq f_2 : |S_{11}(\boldsymbol{x}, f)| + 20, \, 0\}}{20} \right]^2$
Port isolation $|S_{41}|$ not exceeding −20 dB over the operating bandwidth $[f_1 \, f_2]$	Inequality	$|S_{41}(\boldsymbol{x}, f)| \leq -20$ dB for $f \in [f_1 \, f_2]$	$c(\boldsymbol{x}) = \left[\dfrac{\max\{\max f_1 \leq f \leq f_2 : |S_{41}(\boldsymbol{x}, f)| + 20, \, 0\}}{20} \right]^2$
In-band transmission ripple not exceeding 0.2 dB over the operating bandwidth $[f_1 \, f_2]$	Inequality	$|S_{21}(\boldsymbol{x}, f)| \geq -0.2$ dB for $f \in [f_1 \, f_2]$	$c(\boldsymbol{x}) = \left[\dfrac{\max\{-\min f_1 \leq f \leq f_2 : |S_{21}(\boldsymbol{x}, f)| - 0.2, \, 0\}}{0.2} \right]^2$
Power split ratio between output ports 2 and 3 equal to K_P at the centre frequency f_0	Equality	$|S_{31}(\boldsymbol{x}, f)| - |S_{21}(\boldsymbol{x}, f)| = K_P$ at $f = f_0$;	$c(\boldsymbol{x}) = [|S_{31}(\boldsymbol{x}, f_0)| - |S_{21}(\boldsymbol{x}, f_0)| - K_P]^2$
Phase difference between output ports 2 and 3 equal to 90° at the centre frequency f_0	Equality	$\angle S_{31}(\boldsymbol{x}, f) - \angle S_{21}(\boldsymbol{x}, f) = 90°$ at $f = f_0$;	$c(\boldsymbol{x}) = [\angle S_{31}(\boldsymbol{x}, f_0) - \angle S_{21}(\boldsymbol{x}, f_0) - 90°]^2$

[a]The symbol $|S_{jk}(\boldsymbol{x}, f)|$ stands for the modulus of the S-parameter S_{jk} at the design $\boldsymbol{x}$, and frequency f.

Table 1 provides a few examples of constraints pertinent to microwave passive components. In Section 3, we will consider a number of specific microwave coupler examples, where the equality constraint is formulated for the power split ratio of the circuit (cf. the second-but-last row of Table 1).

2.2. Constraint handling using penalty functions

In this work, the primary way of handling constraints is the penalty function approach [70]. It allows for reformulating the design problem (1) into a formally unconstrained task by supplementing the main objective with a linear combination of appropriately quantified constraint violations. The reformulated problem is

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} U_P(\boldsymbol{x}) \tag{4}$$

where cost function U_P is defined as

$$U_P(\boldsymbol{x}) = U(\boldsymbol{x}) + \sum_{k=1}^{n_g + n_h} \beta_k c_k(\boldsymbol{x}) \tag{5}$$

The penalty functions $c_k(\boldsymbol{x})$ quantify violations of the requirements imposed on particular circuit characteristics, whereas β_k are the penalty coefficients. Typically, the penalty functions are defined to assess relative violation of the constraints, e.g., with respect to the assumed acceptance threshold for the inequality constraints. Table 1 provides a few possible formulations of penalty functions.

One of practical issues of the above approach is the adjustment of penalty coefficients, which may have a profound effect on the efficacy of the size reduction process. In particular, the penalty terms should be sufficiently large to enforce satisfaction of constraints within acceptable tolerance. Yet, they cannot be excessively large, in which case exploration of the feasible region boundary becomes problematic, as the objective function landscape becomes extremely nonlinear at and around the feasible region boundary [68]. A manual adjustment of β_k usually leads to sub-optimal results [71]. Recently, adaptive adjustment of penalty coefficients has been proposed [71,72] to overcome this issue. One of these schemes, [71], will be employed in this work to control inequality constraints.

It should be noted that formulation (4), (5) has a loose resemblance to the weighted-sum approach, which is commonly used at the presence of multiple objectives [74]. According to this popular method, a scalar cost function is composed as a linear combination of the individual goals, with the weighting factors representing the user preferences for those goals (i.e., more important goals are distinguished by assigning higher weight values). The approach is often used in multi-criterial design, where the weights are swept in order to yield a family of solutions representing distinct Pareto-optimal parameter vectors [75]. Using the above terminology, the primary objective (footprint area of the optimized circuit) is always associated with the weighting factor of unity, whereas the remaining 'goal' (here, constraints), are associated with the weights equal to β_k, $k = 1, \ldots, n_g + n_h$.

Notwithstanding, what we really have here is a single objective, which is to be minimized (with the optimum value unknown beforehand), and one or several constraints, which are not to be optimized, but merely set at their boundary values, which is zero. As soon as the constraint value reaches zero (whether it is inequality or equality one), no further changes are intended, as the design entered the feasible region or is allocated on its boundary. Furthermore, the typical values of the penalty coefficients are much larger than unity because the objective is to enforce constraints to be close to zero. More specifically, in Section 3, combinations of penalty factors are considered, which take the values between ten and ten thousand.

2.3. Adaptive penalty factors for inequality constraint control

In this work, we employ the adaptive adjustment of penalty factors proposed in [71] to control the inequality constraints. These are generally easier to handle as compared to equality constraints. As mentioned earlier, formulation (5) is convenient because it presents a formally unconstrained task. However, the penalty coefficients have to be set up properly: the values too small may result in excessive constraint violations, whereas coefficients being too large may incur numerical problems associated with high nonlinearity of the merit function in the neighbourhood of the feasible region boundary. Here, we exploit the penalty term management scheme proposed in [71], in which the values of β_k are automatically adjusted depending on current constraint violation but also on the algorithm convergence status.

The main components of the management scheme of [71] have been provided in Table 2. Note that the optimization algorithm is assumed to be an iterative procedure that produces a series $\boldsymbol{x}^{(i)}$, $i = 0, 1, \ldots$, of approximations to the problem (4), (5). The adjustment of the penalty coefficients is based on the relation between the current constraint violation tolerance threshold $D_{k.t}^{(i+1)}$ and the actual violation $D_k(\boldsymbol{x}^{(i)})$ at the current iteration point $\boldsymbol{x}^{(i)}$. The details of the adjustment procedure can be found in Fig. 1.

Table 2
Main components of the penalty coefficient management scheme of [71].

Component	Analytical description	Comments
Termination conditions of the optimization algorithm	$\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}\| < \varepsilon_x$ OR $\|U_P(\boldsymbol{x}^{(i+1)}) - U_P(\boldsymbol{x}^{(i)})\| < \varepsilon_U$	Convergence in argument OR convergence in objective function value; typically, $\varepsilon_x = \varepsilon_U = 10^{-3}$
Violation of the kth constraint at the design $\boldsymbol{x}$	$D_k(\boldsymbol{x}) = \begin{cases} 0 \text{ if } c_k(\boldsymbol{x}) \leq 0 \\ c_k(\boldsymbol{x}) \text{ otherwise} \end{cases}$	Returns non-zero value only in the case of a violation
Tolerance for kth constraint violation	$D_{k.max}$	User-defined value
Tolerance multiplication factor	$M_k > 1$	User-defined, determines the maximum tolerance $M_k D_{k.max}$ for the kth constraint violation that may occur during the optimization process
Convergence metric	$C^{(i)}(\varepsilon_x, \varepsilon_U) = \max\left\{ \dfrac{\varepsilon_x}{\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}\|}, \dfrac{\varepsilon_U}{\|U_P(\boldsymbol{x}^{(i+1)}) - U_P(\boldsymbol{x}^{(i)})\|} \right\}$	Estimates convergence stage of the optimization process w.r.t. the assumed termination criteria
Current violation tolerance of the kth constraint for the (i+1)th iteration of the optimization algorithm	$D_{k.t}^{(i+1)} = \begin{cases} M_k D_{k.max} \text{ if } C^{(i)}(\varepsilon_x, \varepsilon_U) \leq M_C \\ \max\left\{ \begin{matrix} D_{k.max}, M_k D_{k.max} + \\ D_{k.max}(1 - M_k)\left[1 - \dfrac{\log(C^{(i)}(\varepsilon_x, \varepsilon_U))}{\log M_C}\right] \end{matrix} \right\} \end{cases}$	$D_{k.t}^{(i+1)} = M_k D_{k.max}$ if $C^{(i)}(\varepsilon_x, \varepsilon_U) \leq M_C$, where (user-defined) M_C determines the initialization of tolerance adjustment (here, we use $M_C = 10^{-2}$); Upon convergence, $D_{k.t}^{(i+1)} \rightarrow$ to $D_{k.max}$.

Procedure for calculating the kth penalty coefficient $\beta_k^{(i+1)}$ for the $(i+1)$th iteration:

if $D_k(\boldsymbol{x}^{(i)}) < D_{k.t}^{(i+1)} / M_D$

$\quad \beta_k^{(i+1)} = \max\{\beta_k^{(i)} / m_{dec}, \beta_{k.min}\}$

elseif $D_k(\boldsymbol{x}^{(i)}) > D_{k.t}^{(i+1)}$

$\quad \beta_k^{(i+1)} = \min\{\beta_k^{(i)} m_{inc}, \beta_{k.max}\}$

else

$\quad \beta_k^{(i+1)} = \beta_k^{(i)}$

end

The meaning of control parameters [67]:
- M_D – multiplication factor determining the threshold for penalty coefficient reduction;
- m_{dec}, m_{inc} – decrease and increase factors for penalty coefficient adjustment, respectively;
- $\beta_{k.min}$, $\beta_{k.max}$ – minimum and maximum penalty coefficient values, respectively.

Default values: $M_D = 2$, $m_{dec} = 2$, $m_{inc} = 3$, $\beta_{k.min} = 1$, and $\beta_{k.max} = 10^6$.

Fig. 1. Adaptive adjustment of penalty coefficients. The explanation of terms can be found in Table 2.

2.4. Optimization-based equality constraint correction scheme

The main contribution of this work is an optimization-based correction scheme developed to facilitate reliable handling of equality constraints. As discussed earlier, miniaturization under equality constraints is challenging primarily due to the fact that the feasible region for such constraints is a thin set, i.e., its dimensionality is normally lower than the dimensionality of the parameter space. Consequently, exploration of this region aiming at identification of the minimum size design is difficult. In particular, when using a penalty function approach, any design relocation from the feasible region leads to the increase of the objective function, which is in contrast to inequality constraints, where the feasible space interior is non-empty, and the design may stay feasible for a significantly larger number of parameter vectors, as shown in Fig. 2.

In this work, the equality constraints are corrected after each iteration of the optimization process, using an auxiliary search step, as described below. Let $\boldsymbol{x}^{(i+1)}$ be the new design produced in iteration i. Let $\boldsymbol{S}(\boldsymbol{x})$ denote the EM-simulated circuit responses, and $\boldsymbol{J}_S(\boldsymbol{x})$ stand for the Jacobian matrix of $\boldsymbol{S}(\boldsymbol{x})$. Assuming that

the core optimization procedure is a gradient-based routine, the matrix $\boldsymbol{J}_S(\boldsymbol{x}^{(i)})$ is known at the design $\boldsymbol{x}^{(i)}$ (cf. Section 2.5). Consider the linear expansion model

$$L_A^{(i)}(\boldsymbol{x}) = \boldsymbol{S}(\boldsymbol{x}^{(i+1)}) + \boldsymbol{J}_S(\boldsymbol{x}^{(i)}) \cdot (\boldsymbol{x} - \boldsymbol{x}^{(i+1)}) \quad (6)$$

which is established at $\boldsymbol{x}^{(i+1)}$ but using the sensitivity data at $\boldsymbol{x}^{(i)}$. The latter is to reduce the cost of establishing $L_A^{(i)}$ by re-using the Jacobian $\boldsymbol{J}_S(\boldsymbol{x}^{(i)})$ (the subscript A is to indicate the mentioned approximation). Note that upon convergence of the optimization process $\boldsymbol{J}_S(\boldsymbol{x}^{(i)})$ will converge to $\boldsymbol{J}_S(\boldsymbol{x}^{(i+1)})$, so that the model $L_A^{(i)}$ becomes increasingly closer to the true Taylor expansion at $\boldsymbol{x}^{(i+1)}$.

Consider the following minimization sub-problem

$$\boldsymbol{x}_{corr}(M) = \arg \min_{\boldsymbol{x}, \|\boldsymbol{x} - \boldsymbol{x}^{(i+1)}\| \leq M} \max\left\{ h_1(L_A^{(i+1)}(\boldsymbol{x})), \ldots, h_{n_h}(L_A^{(i+1)}(\boldsymbol{x})) \right\} \quad (7)$$

with constraints

$$A(\boldsymbol{x}) \leq A(\boldsymbol{x}^{(i+1)}) \quad (8)$$

Fig. 2. Design relocation from feasible region boundary for inequality and equality constraints: (a) inequality constraint: the designs produced from the current design $\boldsymbol{x}^{(i)}$ may be located in either feasible or infeasible region, (b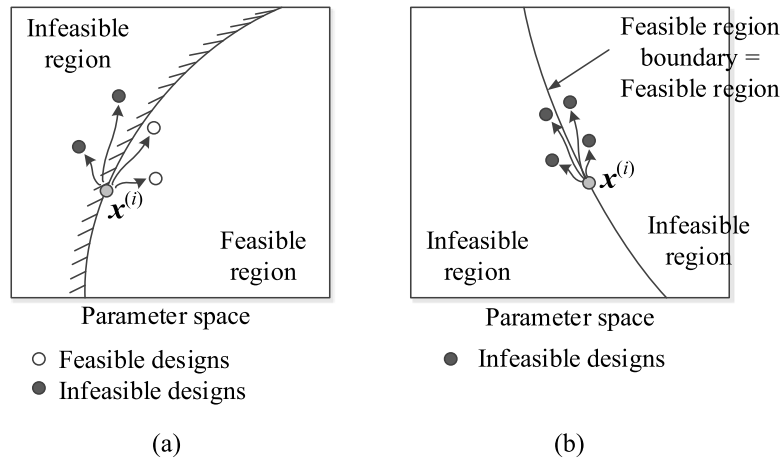) equality constraint: essentially all designs produced from $\boldsymbol{x}^{(i)}$ are infeasible due to feasible region being a thin set. Consequently, exploration of the feasible space is more difficult for the equality constraints than the inequality ones.

$$g_j(\boldsymbol{x}) \le g_j(\boldsymbol{x}^{(i+1)}), j = 1, \ldots, n_g \tag{9}$$

Here, the notation $h_k(L_{A^{(i+1)}}(\boldsymbol{x}))$ has been introduced to emphasize the fact that the constraints are evaluated based on the linear model $L_{A^{(i+1)}}$ rather than on EM simulation results. The corrected design is rendered as

$$\boldsymbol{x}_c^{(i+1)} = \arg \min_M \left\{ \max \left\{ h_1\left(L_A^{(i+1)}(\boldsymbol{x}_{corr}(M))\right), \ldots, h_{n_h} \right. \right.$$
$$\left. \left. \left(L_A^{(i+1)}(\boldsymbol{x}_{corr}(M))\right) \right\} \le \varepsilon_h \right\} \tag{10}$$

where, for any given M, $\boldsymbol{x}_{corr}(M)$ is obtained using (7)–(9), whereas ε_h is a small positive number (here, $\varepsilon_h = 10^{-3}$). Also, $M \le M_{max}$, where the upper bound $M_{\max}$ (here, set to $M_{\max} = 0.1$) is set up in order to limit the design relocation $| |\boldsymbol{x}^{(i+1)} - \boldsymbol{x}_c^{(i+1)}|$ due to the underlying approximations (i.e., the usage of the linear model (6) rather than EM simulation).

In plain words, the corrected design is the one that reduces the violation of equality constraints (in the minimax sense) as much as possible, without increasing the circuit size, and without degrading inequality constraints. At the same time, for the sake of reliability, the design relocation is limited by reducing the search radius M as much as possible. In effect, the correction procedure is a knowledge-based decision-making strategy, which fosters design relocation towards the feasible region while accounting for the remaining figures of merit (here, the primary objective and the inequality constraints).

Formally (cf. (10)) identification of the corrected design is handled as optimization of the parameter M, i.e., seeking for the smallest possible M, for which equality constraints at $\boldsymbol{x}_{corr}(M)$ are below ε_h. If satisfying the condition

$$\max \left\{ h_1\left(L_A^{(i+1)}(\boldsymbol{x}_{corr}(M))\right), \ldots, h_{n_h}\left(L_A^{(i+1)}(\boldsymbol{x}_{corr}(M))\right) \right\} \le \varepsilon_h \tag{11}$$

is not possible for $M \le M_{\max}$, the corrected design is assigned to be $\boldsymbol{x}_c^{(i+1)} = \boldsymbol{x}_{corr}(M_{\max})$, regardless of the equality constraint values therein. It should be noted that when the algorithm converges, the expected relocation of the iteration points $||\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}||$ becomes smaller, which improves the accuracy of the linear model $L_{A^{(i)}}$, along with the likelihood of satisfying (11) for $M < M_{\max}$. The proposed optimization-based correction scheme for efficient control of equality constraints is shown graphically in Fig. 3.

2.5. EM-driven size reduction. Algorithm summary

This section summarizes the entire size reduction procedure with optimization-based correction of equality constraints introduced in Section 2.4. The underlying optimization algorithm is a trust-region (TR) gradient-based search with numerical derivatives [76]. The TR algorithm yields approximations $\boldsymbol{x}^{(i)}$, $i = 0$, 1, …, to the solution of the problem (4), (5), with $\boldsymbol{x}^{(0)}$ being the starting point. We have

$$\boldsymbol{x}^{(i+1)} = \arg \min_{\boldsymbol{x};\ ||\boldsymbol{x}-\boldsymbol{x}^{(i)}|| \le d^{(i)}} U_L^{(i)}(\boldsymbol{x}) \tag{12}$$

The objective function $U_L^{(i)}$ is defined as similarly as the function (5)

$$U_L^{(i)}(\boldsymbol{x}) = A(\boldsymbol{x}) + \sum_{k=1}^{n_g} \beta_k^{(i)} c_{L.k}(\boldsymbol{x}) + \sum_{k=1}^{n_h} \beta_{n_g+k} c_{L.n_g+k}(\boldsymbol{x}) \tag{13}$$

In (13), the penalty functions $c_{L.k}$, corresponding to both inequality and equality constraints, are not evaluated from the EM-simulated circuit responses $\boldsymbol{S}(\boldsymbol{x})$ but using the linear model $L^{(i)}(\boldsymbol{x})$, defined as

$$L^{(i)}(\boldsymbol{x}) = \boldsymbol{S}(\boldsymbol{x}^{(i)}) + \boldsymbol{J}_S(\boldsymbol{x}^{(i)}) \cdot (\boldsymbol{x} - \boldsymbol{x}^{(i)}) \tag{14}$$

where $\boldsymbol{J}_S(\boldsymbol{x}^{(i)})$ is a Jacobian matrix of $\boldsymbol{S}$ at $\boldsymbol{x}^{(i)}$, computed by means of finite differentiation.

The penalty terms $\beta_k^{(i)}$ for the inequality constraints are adaptively adjusted using the procedure of Section 2.3. The penalty coefficients β_{ng+k} are fixed at low values (e.g., $\beta_{ng+k} = 10$, $k = 1$, …, n_h), because the enforcement of the equality constraints is mainly realized using the correction procedure of Section 2.4. The penalty terms are employed only to avoid excessive violations that might occur especially at the initial steps of the optimization process (the scheme of Section 2.4 is intended for correcting smaller violations).

The trust region size $d^{(i)}$ is re-adjusted before each algorithm iteration (12) based on the gain ratio $r = [U_L^{(i)}(\boldsymbol{x}^{(i+1)}) - U_L^{(i)}(\boldsymbol{x}^{(i)})]/[U_P(\boldsymbol{x}^{(i+1)}) - U_P(\boldsymbol{x}^{(i)})]$ (EM-evaluated versus model-predicted merit function improvement). The standard TR rules are applied [76]. Upon producing the candidate design $\boldsymbol{x}^{(i+1)}$, the correction procedure of Section 2.4 is launched to yield the corrected design $\boldsymbol{x}_c^{(i+1)}$. If $U_P(\boldsymbol{x}_c^{(i+1)}) < U_P(\boldsymbol{x}^{(i+1)})$, then $\boldsymbol{x}_c^{(i+1)}$ replaces $\boldsymbol{x}^{(i+1)}$. The new design is only accepted if $U_P(\boldsymbol{x}^{(i+1)}) < U_P(\boldsymbol{x}^{(i)})$. If $r < 0$, $d^{(i)}$ is shrunk, and the iteration is executed again. Also, in the TR framework, the termination condition of Table 2 (the first
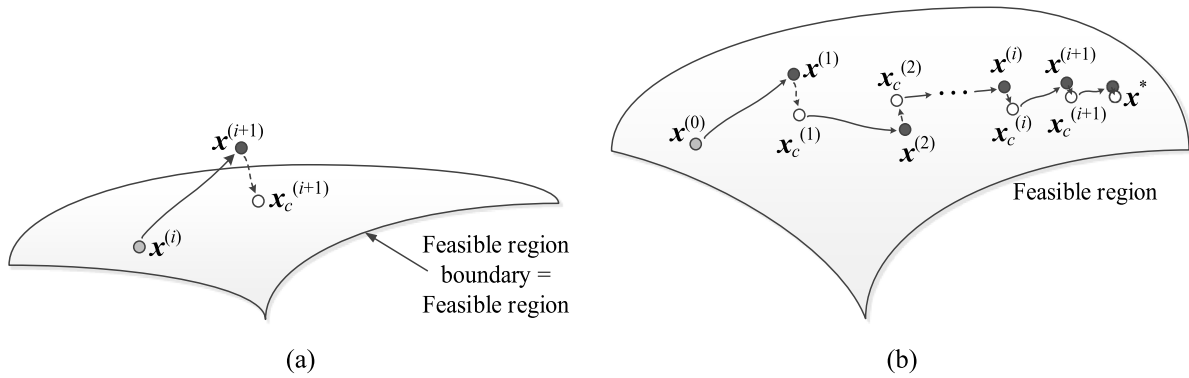
Fig. 3. Optimization-based correction scheme for efficient control of equality constraints: (a) the concept: design $\boldsymbol{x}^{(i+1)}$ obtained in the ith iteration of the optimization algorithm, starting from $\boldsymbol{x}^{(i)}$, is infeasible; it is brought closer to a feasible space using the procedure (7)–(10), without neither degrading inequality constraints nor the circuit area; the corrected design $\boldsymbol{x}_c^{(i+1)}$ features reduced equality constraint values $h_k(\boldsymbol{x}_c^{(i+1)}) < h_k(\boldsymbol{x}^{(i+1)})$, $k = 1, \ldots, n_h$, as well as $A(\boldsymbol{x}_c^{(i+1)}) \leq A(\boldsymbol{x}^{(i+1)})$, and $g_j(\boldsymbol{x}_c^{(i+1)}) \leq g_j(\boldsymbol{x}^{(i+1)})$, $j = 1, \ldots, n_g$, (b) graphical illustration of the intermittent equality constraint correction throughout the optimization run.

Table 3
Microwave structures used for verification of the proposed size reduction procedure.

	Case study			
	Circuit I	Circuit II	Circuit III	Circuit IV
Substrate	AD300 ($\varepsilon_r = 2.97$, $h = 0.76$ mm)	RO4003 ($\varepsilon_r = 3.38$, $h = 0.762$ mm)	FR4 ($\varepsilon_r = 4.4$, $h = 1.0$ mm)	FR4 ($\varepsilon_r = 4.4$, $h = 1.0$ mm)
Design parameters	$\boldsymbol{x} = [g\ l_{1r}\ l_a\ l_b\ w_1\ w_{2r}\ w_{3r}\ w_{4r}\ w_a\ w_b]^T$	$\boldsymbol{x} = [l_1\ l_2\ l_3\ d\ w\ w_1]^T$	$\boldsymbol{x} = [G\ g_1\ g_2\ g_3\ w_1\ w_3\ L_1\ L_2]^T$	$\boldsymbol{x} = [W\ w_{1\_r}\ w_{2\_r}\ w_3\ w_4\ L\ L_{1\_r}\ L_{2\_r}\ L_3\ L_4\ L_{5\_r}\ s]^T$
Other parameters	$L = 2dL + L_s$, $L_s = 4\ w_1+ 4g + s + l_a + l_b$, $W = 2dL + W_s$, $W_s = 4w_1+ 4g + s + 2w_a$, $l_1 = l_b\ l_{1r}$, $w_2 = w_a\ w_{2r}$, $w_3 = w_{3r} w_a$, and $w_4 = w_{4r}w_a$, $w_c = 1.9$ mm	$d_1 = d + |w - w_1|$, $d = 1.0$, $w_0 = 1.7$, and $l_0 = 15$ mm	$L = 4w_1 + 10w_3 + 15g_3 + 2L_2$, $W = 4w_3 + 2L_1 + G + 2g_1 + 2g_3$	–
Operating parameters (design scenario I)	$f_0 = 1.5$ GHz $F = [1.45\ 1.55]$ GHz	$f_0 = 1.0$ GHz $F = [0.9\ 1.1]$ GHz	$f_0 = 1.0$ GHz $F = [1.0]$ GHz	$f_0 = 2.0$ GHz $F = [2.0]$ GHz
Operating parameters (design scenario II)	$f_0 = 1.5$ GHz $F = [1.47\ 1.53]$ GHz	$f_0 = 1.0$ GHz $F = [0.95\ 1.05]$ GHz	N/A	N/A
Inequality constraint	$|S_{11}| \leq -20$ dB and $|S_{41}| \leq -20$ dB over the bandwidth F	$|S_{11}| \leq -20$ dB and $|S_{41}| \leq -20$ dB over the bandwidth F	$|S_{11}| \leq -20$ dB and $|S_{41}| \leq -20$ dB at f_0	$|S_{11}| \leq -20$ dB and $|S_{41}| \leq -20$ dB at f_0
Equality constraint	$|S_{31}| - |S_{21}| = 0$ at f_0	$|S_{31}| - |S_{21}| = 0$ at f_0	$|S_{31}| - |S_{21}| = 0$ at f_0	$|S_{31}| - |S_{21}| = 3$ dB at f_0
Initial design	$\boldsymbol{x}^{(0)} = [0.45\ 0.69\ 6.25\ 10.32\ 0.96\ 0.39\ 0.14\ 0.57\ 4.62\ 0.60]^T$	$\boldsymbol{x}^{(0)} = [5.27\ 13.33\ 21.51\ 0.96\ 0.89\ 0.90]^T$	$\boldsymbol{x}^{(0)} = [1.0\ 1.0\ 0.6\ 0.25\ 2.4\ 0.25\ 9.0\ 3.75]^T$	$\boldsymbol{x}^{(0)} = [15.0\ 0.63\ 0.93\ 3.45\ 3.0\ 12.4\ 0.42\ 0.81\ 1.50\ 1.0\ 0.9\ 0.5]^T$

row thereof) is supplemented by the condition $d^{(i)} < \varepsilon_x$, which is necessary in order to stop the procedure upon sufficient reduction of the TR size even if the latest iteration was unsuccessful.

The operating flow of the entire algorithm can be summarized in the form of the pseudocode as follows:

Fig. 4 shows the flow diagram of the proposed optimization framework. The initial values of penalty coefficients are not critical, typically set anywhere between 10^2 and 10^3 (will be automatically adjusted during the optimization procedure).

3. Demonstration examples

In this section, we validate the size reduction framework introduced in Section 2, and compare it to optimization-based miniaturization involving a penalty function approach. The latter is executed in several setups different in terms of the assumed values of penalty coefficient. The numerical experiments are based on four compact couplers. The major performance figures utilized for the evaluation of the algorithm efficacy are the final circuit footprint area, as well as violation of inequality and equality constraints.

3.1. Case studies

Fig. 5 shows the geometries of four miniaturized microstrip couplers, Circuit I through IV, utilized as case studies. In all cases, the EM models are implemented and simulated in CST Microwave Studio, using the time-domain solver. Table 3 contains relevant data concerning all circuits, including the substrate information, design variables, target operating frequencies, etc. In all cases, the primary objective is a reduction of the circuit footprint $A(\boldsymbol{x})$, under the two constraints:

- Inequality constraint for matching and port isolation, $g_1(\boldsymbol{x}) = \max\{f \in F : \max\{|S_{11}(\boldsymbol{x}, f)|, |S_{41}(\boldsymbol{x}, f)|\}\} + 20$ dB, which corresponds to a condition that both $|S_{11}(\boldsymbol{x}, f)|$ and $|S_{41}(\boldsymbol{x}, f)|$ should be not higher than -20 dB over the operating band F;

Algorithm 1: Size reduction with equality constraint correction

---

Input: $x^{(0)}$ – initial design; $g_k(x)$ – inequality constraints; $h_k(x)$ – equality constraints;

1   Set iteration index $i = 0$;

2   Set up penalty coefficients $\beta_k^{(i)} = \beta_{k.0}$ $k = 1, \ldots, n_g$, and $\beta_{ng+k} = \beta_{ng+k.0}$ $k = 1, \ldots, n_h$;

3   Evaluate circuit responses $S(x^{(i)})$;

4   Evaluate Jacobian $J_S(x^{(i)})$;

5   Construct linear model $L^{(i)}(x)$, cf. (14);

6   Construct merit function $U_L^{(i)}(x) = A(x) + \sum_{k=1}^{n_g} \beta_k^{(i)} c_{L.k}(x) + \sum_{k=1}^{n_h} \beta_{n_g+k} c_{L.n_g+k}(x)$ (cf. 13);

7   Find the candidate design, cf. (12) $x^{(i+1)} = \arg \min\limits_{x;\, \|x-x^{(i)}\| \le d^{(i)}} U_L^{(i)}(x)$;

8   Evaluate circuit responses $S(x^{(i+1)})$;

9   Construct linear model $L_A^{(i)}(x)$, cf. (6);

10   Find corrected design, cf. (10):

$$x_c^{(i+1)} = \arg \min_M \left\{ \max \left\{ h_1\left(L_A^{(i+1)}(x_{corr}(M))\right), \ldots, h_{n_h}\left(L_A^{(i+1)}(x_{corr}(M))\right) \right\} \le \varepsilon_h \right\} ;$$

11   Evaluate circuit responses $S(x_c^{(i+1)})$;

12   **if** $U_P(x_c^{(i+1)}) < U_P(x^{(i+1)})$ **then** set $x^{(i+1)} = x_c^{(i+1)}$; **end**

13   Compute gain ratio $r = [U_L^{(i)}(x^{(i+1)}) - U_L^{(i)}(x^{(i)})]/[U_P(x^{(i+1)}) - U_P(x^{(i)})]$;

14   Update TR size $d^{(i)}$;

15   **if** $U_P(x^{(i+1)}) < U_P(x^{(i)})$

16   Compute $D_{k.t}^{(i+1)}$, for inequality constraints $k = 1, \ldots, n_g$, cf. Table 2;

17   Evaluate constraint violations $D_k(x^{(i+1)})$, $k = 1, \ldots n_c$;

18   **if** $D_k(x^{(i)}) < D_{k.t}^{(i+1)} / M_D$

19             $\beta_k^{(i+1)} = \max\{\beta_k^{(i)} / m_{dec}, \beta_{k.\min}\}$

20   **elseif** $D_k(x^{(i)}) > D_{k.t}^{(i+1)}$

21             $\beta_k^{(i+1)} = \min\{\beta_k^{(i)} m_{inc}, \beta_{k.\max}\}$

22   **else**

23             $\beta_k^{(i+1)} = \beta_k^{(i)}$

24   **end**

25   Set $i = i + 1$;

26   **end**

27   **if** $\|x^{(i+1)} - x^{(i)}\| < \varepsilon_x$ OR $|U_P(x^{(i+1)}) - U_P(x^{(i)})| < \varepsilon_U$ OR $d^{(i)} < \varepsilon_x$;

28   **goto** 32;

29   **else**

30   **goto** 3;

31   **end**
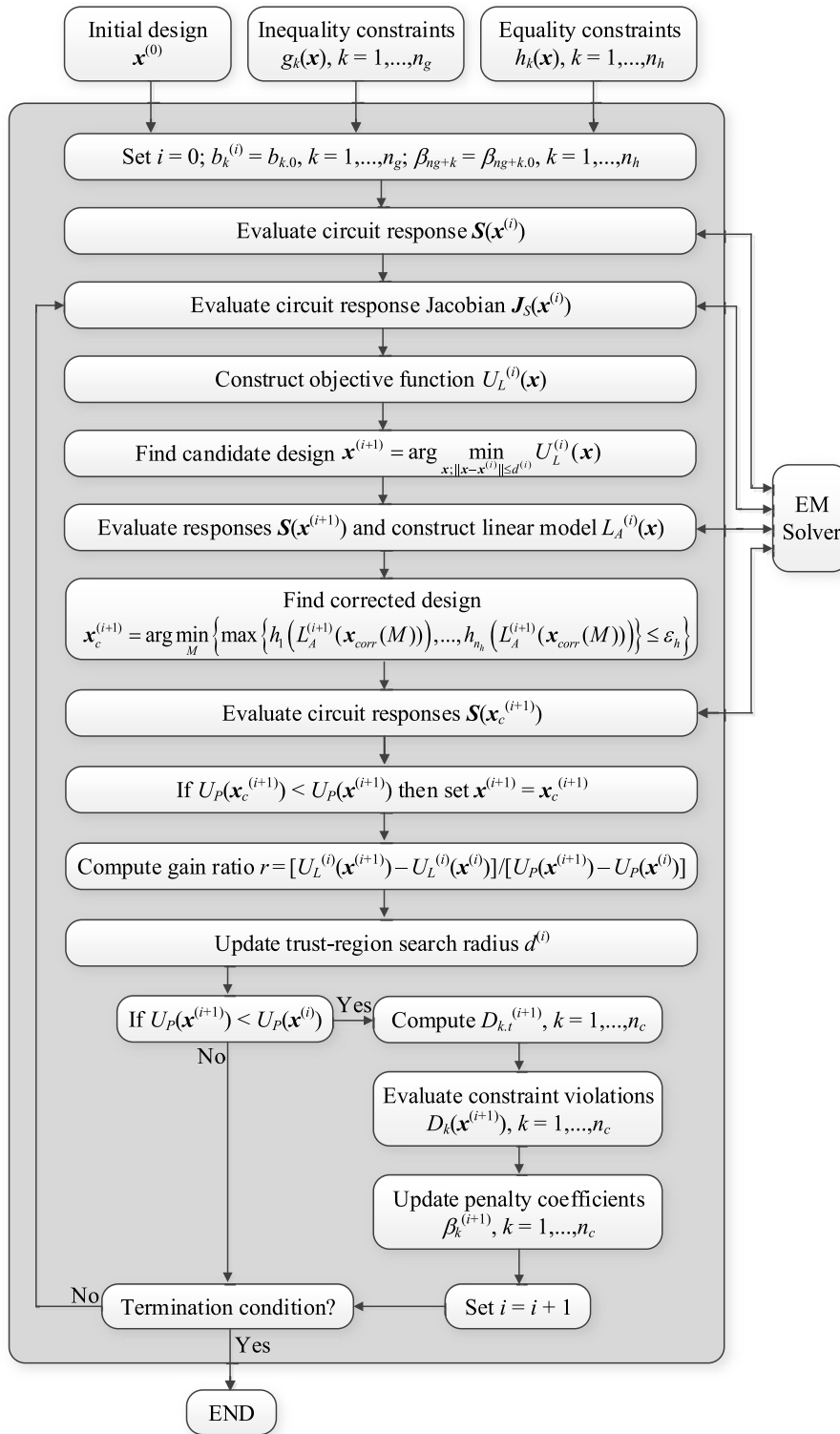
Fig. 4. Flow diagram of the proposed size reduction framework with adaptive adjustment of penalty terms for inequality constraints and optimization-based correction of equality constraints.

- Equality constraint for the power split ratio: $h_1(\boldsymbol{x}) = ||S_{31}(\boldsymbol{x}, f_0) - S_{21}(\boldsymbol{x}, f_0)|| - K_P$, where K_P is the required power split (in dB);

For Circuit I and II, we consider two scenarios corresponding to different operating bandwidths. For Circuit III and IV, we consider one scenario with design specifications setup up over a single frequency, i.e., $F = f_0$ (target operating frequency).

3.2. Numerical results

The optimization procedure introduced in Section 2 has been applied to reduce the size of Circuits I through IV. For the first two circuits, we consider two design scenarios, corresponding to different target operating bands for circuit matching and isolation. For Circuits III and IV, we are only interested in the performance at the centre frequency f_0. For comparison, all circuits are
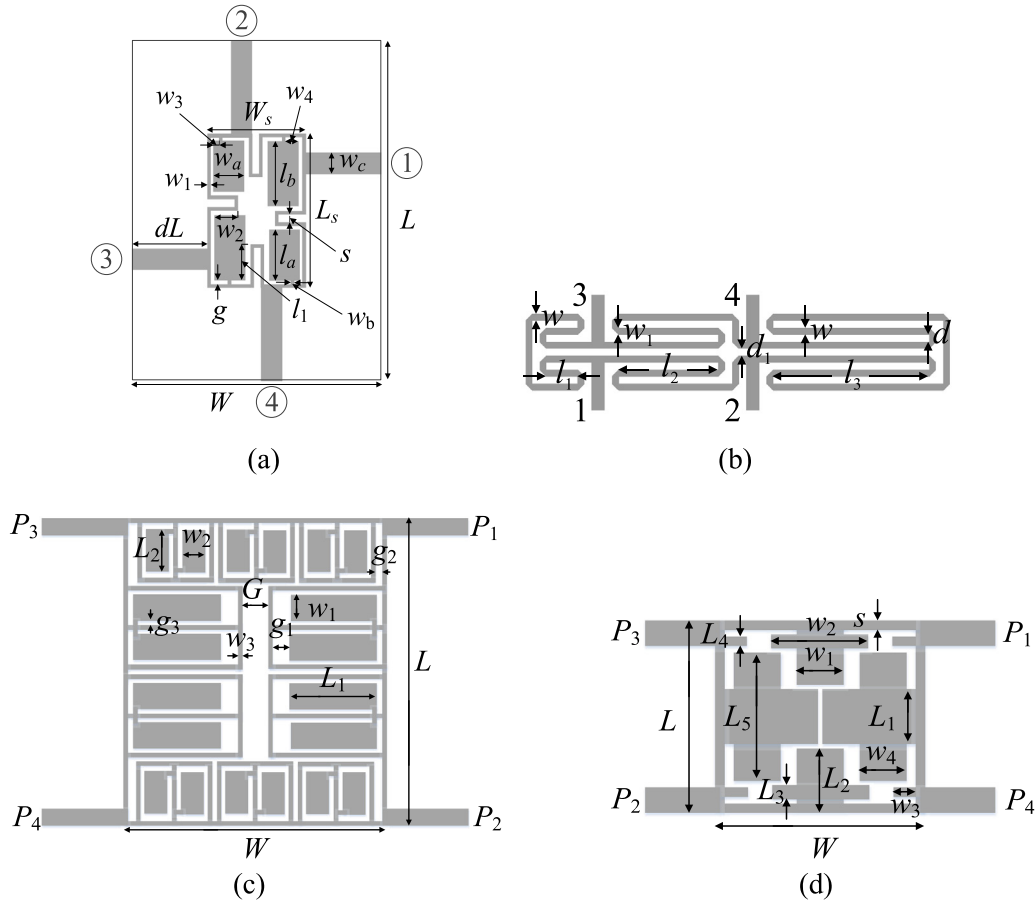
Fig. 5. Microstrip components utilized as test cases for verification of the proposed size-reduction procedure: (a) compact branch-line coupler (Circuit I) [78], (b) rat-race coupler with folder transmission lines (Circuit II) [79], (c) branch-line coupler with microstrip cells (Circuit III) [80], (d) compact branch-line coupler with unequal power division (Circuit IV) [81].

Table 4
Setup of the benchmark size reduction algorithm.

Setup component	Value/description
Core algorithm	Trust-region gradient-based search (cf. Section 2.5)
Objective function	$U(\boldsymbol{x}) = A(\boldsymbol{x}) + \beta_1 c_1(\boldsymbol{x})^2 + \beta_2 c_2(\boldsymbol{x})$
Penalty functions[a]	$c_1(\boldsymbol{x}) = \left\| \frac{h_1(\boldsymbol{x})}{0.1} \right\|$, $c_2(\boldsymbol{x}) = \max \left\{ 0, \frac{g_1(\boldsymbol{x})+20}{20} \right\}$
Penalty coefficients for equality constraint[b]	$\beta_1 \in \{10, 100, 1000, 10000\}$
Penalty coefficients for inequality constraint[b]	$\beta_2 \in \{10, 100, 1000, 10000\}$

[a] Equality constraint is normalized to 0.1 dB.
[b] The benchmark algorithm is run for all combinations of β_1 and β_2.

also optimized using the penalty function approach with fixed penalty coefficients, as well as particle swarm optimizer (PSO), a representative nature-inspired algorithm [77].

The details concerning the benchmark algorithm with adaptive penalty factors setup can be found in Table 4. It has been executed sixteen times, for all combinations of the β_1 and β_2 listed in the Table. This allows us to demonstrate the importance of the penalty coefficient arrangements, along with the fact that identifying the optimum combination is a non-trivial task with profound effects on the optimization process performance. Whereas for PSO algorithm, we have: swarm size $N_s = 10$, maximum number of iterations $k_{\max} = 100$, and weight coefficients for velocity updating: $\chi = 0.73$, $c_1 = c2 = 2.05$ [77]. Because of the high computational cost, six independent runs of PSO procedure

have been executed. For each run, the computational budget has been set to 1,000 objective function evaluations.

Tables 5 through 7 contain the numerical results for the introduced and the benchmark algorithms. Figs. 6, 7, 8, and 9, show the circuit S-parameters at the initial and optimized designs. We also illustrate the evolution of the circuit size and constraint violations, along with the evolution of the penalty factor for the inequality constraint.

3.3. Discussion

The results in Tables 5 through 7 contain comprehensive data allowing us to investigate the properties of the proposed size reduction algorithm, and compare it to the benchmark techniques.

Table 5
Optimization results for Circuit I.

Optimization approach		Performance parameters					
Method	Penalty coefficient values	Design scenario I ($F = [1.45\ 1.55]$ GHz)			Design scenario II ($F = [1.47\ 1.53]$ GHz)		
		Footprint area A [mm^2]	Violation of constraint h_1 [dB]	Violation of constraint g_1 [dB]	Footprint area A [mm^2]	Violation of constraint h_1 [dB]	Violation of constraint g_1 [dB]
Implicit constraint handling (penalty function approach)	$\beta_1 = 10^1,\ \beta_2 = 10^1$	241	0.03	6.8	264	0.07	3.5
	$\beta_1 = 10^1,\ \beta_2 = 10^2$	259	0.06	5.3	264	0.07	3.5
	$\beta_1 = 10^1,\ \beta_2 = 10^3$	301	−0.01	1.9	272	0.02	2.1
	$\beta_1 = 10^1,\ \beta_2 = 10^4$	325	0.01	0.2	293	0.02	0.2
	$\beta_1 = 10^2,\ \beta_2 = 10^1$	247	−0.05	6.6	264	0.07	3.5
	$\beta_1 = 10^2,\ \beta_2 = 10^2$	258	−0.02	5.7	276	0.00	1.7
	$\beta_1 = 10^2,\ \beta_2 = 10^3$	318	0.01	1.0	292	−0.01	0.5
	$\beta_1 = 10^2,\ \beta_2 = 10^4$	319	0.00	0.3	297	−0.08	0.3
	$\beta_1 = 10^3,\ \beta_2 = 10^1$	247	−0.04	7.1	333	−0.00	0.5
	$\beta_1 = 10^3,\ \beta_2 = 10^2$	264	−0.03	53	335	−0.01	1.0
	$\beta_1 = 10^3,\ \beta_2 = 10^3$	318	−0.01	1.3	322	−0.02	−1.1
	$\beta_1 = 10^3,\ \beta_2 = 10^4$	319	0.00	0.2	301	−0.05	0.1
	$\beta_1 = 10^4,\ \beta_2 = 10^1$	242	0.00	6.9	323	−0.00	0.5
	$\beta_1 = 10^4,\ \beta_2 = 10^2$	258	−0.05	5.7	292	−0.06	0.8
	$\beta_1 = 10^4,\ \beta_2 = 10^3$	310	−0.03	1.4	325	−0.00	0.0
	$\beta_1 = 10^4,\ \beta_2 = 10^4$	317	0.00	0.4	302	−0.07	0.1
PSO		534	1.2	0.8	527	0.9	2.9
Size reduction with optimization-based equality constraint correction (this work)		**304**[a]	**0.02**	**0.1**	**286**[a]	**0.02**	**0.2**

[a]The best solution for the respective design scenario considering the combination of the constraint violation (the smaller the absolute value, the better) and the obtained footprint area (the smaller, the better). In other words, the design can be considered optimum if and only if all constraints are satisfied (or close to be satisfied). Otherwise, the design is not optimum regardless of the obtained circuit size.

Table 6
Optimization results for Circuit II.

Optimization approach		Performance parameters					
Method	Setup	Design scenario I ($F = [0.9\ 1.1]$ GHz)			Design scenario II ($F = [0.95\ 1.05]$ GHz)		
		Footprint area A [mm^2]	Violation of constraint h_1 [dB]	Violation of constraint g_1 [dB]	Footprint area A [mm^2]	Violation of constraint h_1 [dB]	Violation of constraint g_1 [dB]
Implicit constraint handling (penalty function approach)	$\beta_1 = 10^1,\ \beta_2 = 10^1$	124	0.01	16.8	114	0.00	16.6
	$\beta_1 = 10^1,\ \beta_2 = 10^2$	104	0.02	17.0	90	0.00	17.6
	$\beta_1 = 10^1,\ \beta_2 = 10^3$	464	0.27	3.2	439	0.21	2.9
	$\beta_1 = 10^1,\ \beta_2 = 10^4$	508	0.17	−0.1	364	−0.09	0.2
	$\beta_1 = 10^2,\ \beta_2 = 10^1$	593	0.04	−3.4	593	0.04	−5.2
	$\beta_1 = 10^2,\ \beta_2 = 10^2$	593	0.04	−3.4	593	0.04	−5.2
	$\beta_1 = 10^2,\ \beta_2 = 10^3$	538	0.07	−1.9	593	0.04	−5.2
	$\beta_1 = 10^2,\ \beta_2 = 10^4$	593	0.04	−3.4	593	0.04	−5.2
	$\beta_1 = 10^3,\ \beta_2 = 10^1$	595	0.04	−3.4	595	0.04	−5.2
	$\beta_1 = 10^3,\ \beta_2 = 10^2$	595	0.04	−3.4	595	0.04	−5.2
	$\beta_1 = 10^3,\ \beta_2 = 10^3$	595	0.04	−3.4	595	0.04	−5.2
	$\beta_1 = 10^3,\ \beta_2 = 10^4$	595	0.04	−3.4	595	0.04	−5.2
	$\beta_1 = 10^4,\ \beta_2 = 10^1$	595	0.04	−3.4	595	0.04	−5.2
	$\beta_1 = 10^4,\ \beta_2 = 10^2$	595	0.04	−3.4	595	0.04	−5.2
	$\beta_1 = 10^4,\ \beta_2 = 10^3$	595	0.04	−3.4	595	0.04	−5.2
	$\beta_1 = 10^4,\ \beta_2 = 10^4$	595	0.04	−3.4	595	0.04	−5.2
PSO		551	−1.1	0.3	547	0.5	0.8
Size reduction with optimization-based equality constraint correction (this work)		**477**[a]	**0.05**	**0.2**	**349**[a]	**0.0**	**0.2**

[a]The best solution for the respective design scenario considering the combination of the constraint violation (the smaller the absolute value, the better) and the obtained footprint area (the smaller, the better). In other words, the design can be considered optimum if and only if all constraints are satisfied (or close to be satisfied). Otherwise, the design is not optimum regardless of the obtained circuit size.

It should be emphasized that, in order to appropriately assess the obtained optimization results, one needs to take into account both the device footprint area and the constraint violations. In other words, for the best design, the combination of the miniaturization rate and the constraint violation should be the most advantageous. Formally speaking, any design that violates either of the constraints cannot be considered optimum.

The following properties and performance indicators of our method can be highlighted:

- The results obtained for all considered circuits and all design scenarios are consistent. In particular, the proposed procedure enables accurate control over both the equality and inequality constraints. For the power split, the average constraint violation is about 0.03 dB, whereas for the matching and isolation constraint, it is about 0.2 dB. Furthermore, it can be observed that the equality constraint violation is kept close to zero for the majority of the optimization run duration, which indicates that the correction procedure

Table 7

Optimization results for Circuit III and IV.

Optimization approach		Performance parameters					
Method	Setup	Circuit III ($f_0 = 1.0$ GHz)			Circuit IV ($f_0 = 2.0$ GHz)		
		Footprint area A [mm^2]	Violation of constraint h_1 [dB]	Violation of constraint g_1 [dB]	Footprint area A [mm^2]	Violation of constraint h_1 [dB]	Violation of constraint g_1 [dB]
Implicit constraint handling (penalty function approach)	$\beta_1 = 10^1$, $\beta_2 = 10^1$	73	0.01	13.8	130	0.03	2.1
	$\beta_1 = 10^1$, $\beta_2 = 10^2$	75	0.01	13.8	135	0.06	1.8
	$\beta_1 = 10^1$, $\beta_2 = 10^3$	305	1.12	2.3	121	1.75	−0.3
	$\beta_1 = 10^1$, $\beta_2 = 10^4$	334	1.22	0.2	146	0.02	0.1
	$\beta_1 = 10^2$, $\beta_2 = 10^1$	73	0.01	13.8	114	0.02	4.7
	$\beta_1 = 10^2$, $\beta_2 = 10^2$	73	0.01	13.8	141	0.04	2.9
	$\beta_1 = 10^2$, $\beta_2 = 10^3$	382	0.20	2.4	135	0.01	0.5
	$\beta_1 = 10^2$, $\beta_2 = 10^4$	428	0.08	−0.1	152	0.01	0.1
	$\beta_1 = 10^3$, $\beta_2 = 10^1$	73	0.01	13.8	141	0.00	1.1
	$\beta_1 = 10^3$, $\beta_2 = 10^2$	268	0.03	11.5	140	0.09	1.2
	$\beta_1 = 10^3$, $\beta_2 = 10^3$	324	0.07	3.7	142	0.02	1.3
	$\beta_1 = 10^3$, $\beta_2 = 10^4$	414	0.04	0.2	148	0.01	0.2
	$\beta_1 = 10^4$, $\beta_2 = 10^1$	262	0.00	11.7	164	0.00	1.3
	$\beta_1 = 10^4$, $\beta_2 = 10^2$	303	0.00	9.6	200	0.00	−6.1
	$\beta_1 = 10^4$, $\beta_2 = 10^3$	448	0.00	0.0	203	0.01	−5.0
	$\beta_1 = 10^4$, $\beta_2 = 10^4$	419	0.01	0.3	207	0.01	−0.3
PSO		417	0.8	1.4	225	3.17	1.2
Size reduction with optimization-based equality constraint correction (this work)		**362**[a]	**0.03**	**0.2**	**129**[a]	**0.03**	**−0.4**

[a]The best solution for the respective circuit considering the combination of the constraint violation (the smaller the absolute value, the better) and the obtained footprint area (the smaller the better). In other words, the design can be considered optimum if and only if all constraints are satisfied (or close to be satisfied). Otherwise, the design is not optimum regardless of the obtained circuit size.

works as expected. At the same time, adaptive adjustment of the penalty coefficient for the inequality constraint seems essential because the optimum value thereof is very much dependent on the circuit but also design specifications, and can vary from 10^0 to almost 10^4.

- Optimization-based correction of equality constraints seems to be important for achieving competitive miniaturization ratios, because it facilitates the exploration of the feasible region. This can be concluded by comparing the results obtained for the proposed algorithm with the procedure with fixed penalty factors, executed with the penalty coefficient values close to the optimum ones. For these, the optimization process is either stuck at the designs corresponding to relatively large footprint areas (especially for Circuits I and II) or ends up at the designs that violate the inequality constraint (Circuit III) or are allocated deeply in the feasible region (Circuit IV).
- It can be observed that, in contrast to the proposed approach, the performance of the algorithm exploiting fixed penalty factors is immensely dependent on the penalty coefficient setup. Only few (typically one to three) combinations of penalty coefficient values lead to reasonable results, which are still inferior to those rendered by our procedure, especially in terms of the final circuit size.
- The results obtained using PSO algorithm feature significantly poorer miniaturization rates than that of the proposed procedure for all the considered microwave structures. The average values of the footprint area of each device are from fifteen percent to over eighty percent larger than that of our approach (around fifty percent on average). Moreover, the constraint violations obtained with PSO are worse than that of our procedure, especially for Circuit I, III and IV, which are more challenging in terms of the number of designable parameters examples.

The scrutiny of the results shown in Figs. 6 through 9 allow us to draw additional conclusions on the performance of the proposed size reduction framework. First, all of the optimal designs for the considered benchmark microwave structures satisfy the design specifications and also feature satisfactory footprint areas. In all cases, the constraint violations have been diminished throughout the algorithm run: from the initial infeasible values exceeding 1 dB to nearly zero (for the equality constraints), and small fractions of dB (for the inequality ones). Moreover, the values of the penalty coefficients for the inequality constraints are shown to be problem-specific to a large extent, which corroborates the relevance of employing the adaptive adjustment of penalty factors scheme. An important advantage of the proposed algorithm with correction-based control of equality constraints is that – for all the considered circuits and for all the design tasks – the ultimate equality constraint violation at the optimal design has been close to zero, which is in contrast with the results rendered by the benchmark algorithm.

It should be emphasized that the presented verification of the algorithm of Section 2 is comprehensive (four circuits, six design scenarios altogether), which makes the aforementioned observations conclusive. The proposed methodology ensures superior miniaturization rates while effectively addressing the problem of precise control over both inequality and equality constraints. At the same time, the extra cost related to equality constraint correction is only one EM analysis per algorithm iteration. Furthermore, the algorithm is straightforward to set up, and it does not require problem-specific tuning (in contrast to the benchmark procedure). The latter normally involves engineering experience and incurs additional computational expenses.

4. Conclusion

This work introduced a novel algorithm for EM-driven size reduction of microwave components. The distinctive feature of our methodology is the optimization-based procedure for correcting possible violations of the equality constraints, which facilitates exploration of the feasible space. The latter is normally challenging due to the geometry of feasible sets associated with the equality constraints. At the same time, the correction is realized at a reduced computational cost of only one extra EM analysis per iteration. Furthermore, the presented algorithm incorporates

Fig. 6. Initial (grey) and optimized (black) S-parameters of Circuit I. The vertical and horizontal lines mark the target operating bandwidth and the acceptance level for the matching $|S_{11}|$ and isolation $|S_{41}|$ responses. Also shown is the evolution of the circuit size and constraint violations (in case of feasibility, violations shown as zero), as well as evolution of the penalty coefficient for the inequality constraint: (a) design scenario I (bandwidth 1.45 GHz to 1.55 GHz), (b) design scenario II (bandwidth 1.47 GHz to 1.53 GHz).

adaptive adjustment of penalty coefficients for accurate control of the inequality constraints. The developed mechanisms are based on the intrinsic knowledge of numerical challenges pertinent to both types of design constraints, combined into decision-making strategies selecting the best possible arrangements in terms of the specific approaches to constraint handling. The latter include the adjustment of the search region for equality constraint correction, and proportionality coefficient values for inequality constraint enforcement. The presented technique has been comprehensively validated with the use of four microstrip couplers with acceptance thresholds imposed on their matching

and isolation characteristics, as well as the power division ratios. The obtained results demonstrate consistent performance, with respect to constraint handling and the achievable size reduction, both being superior to the benchmark procedures. The algorithm proposed in this paper can be considered a viable alternative to the conventional optimization techniques, especially whenever precise handling of design constraints of both equality and inequality type is required. The future research will be concerned with the application of the procedure for other challenging design cases (e.g., simultaneous control of several equality conditions
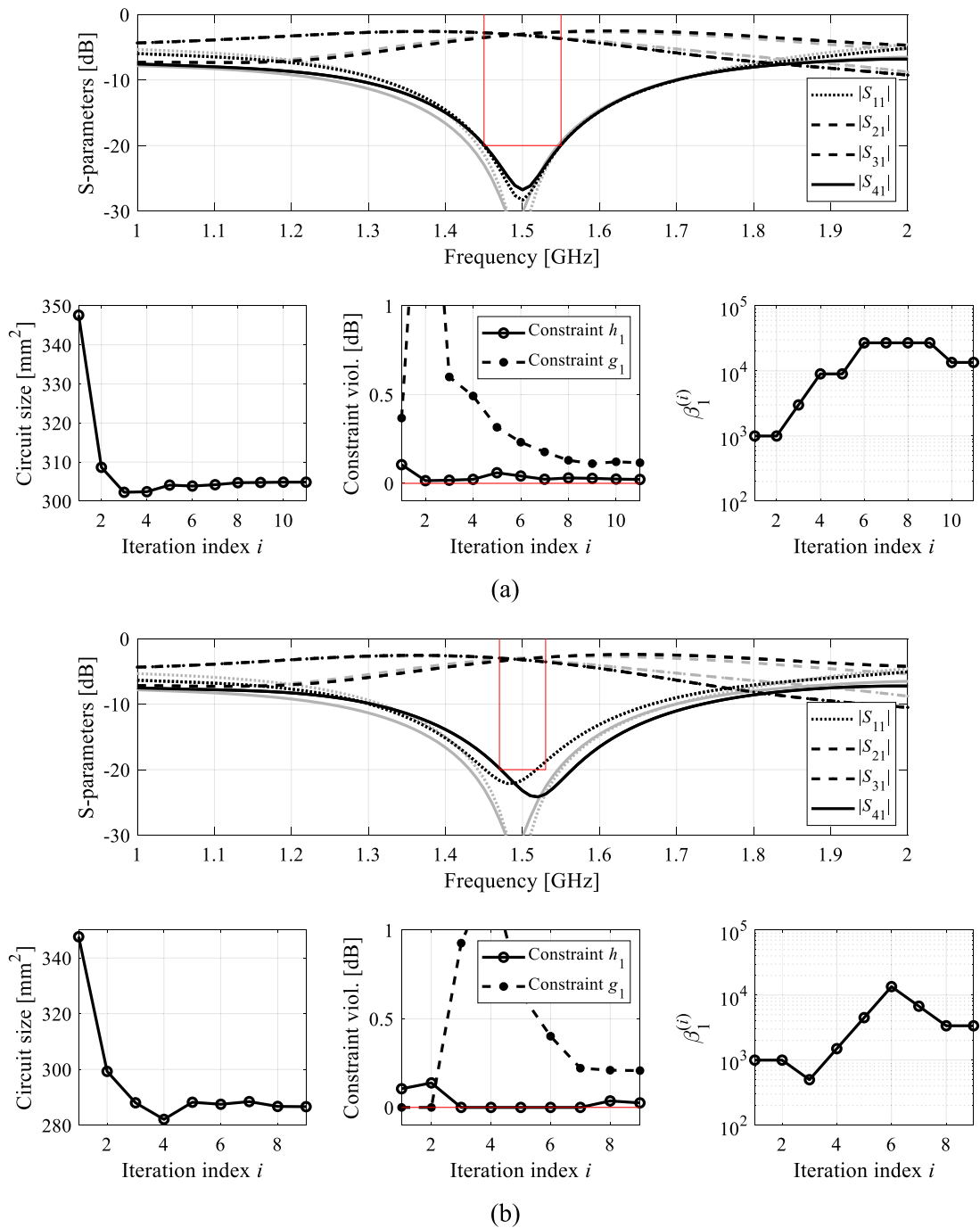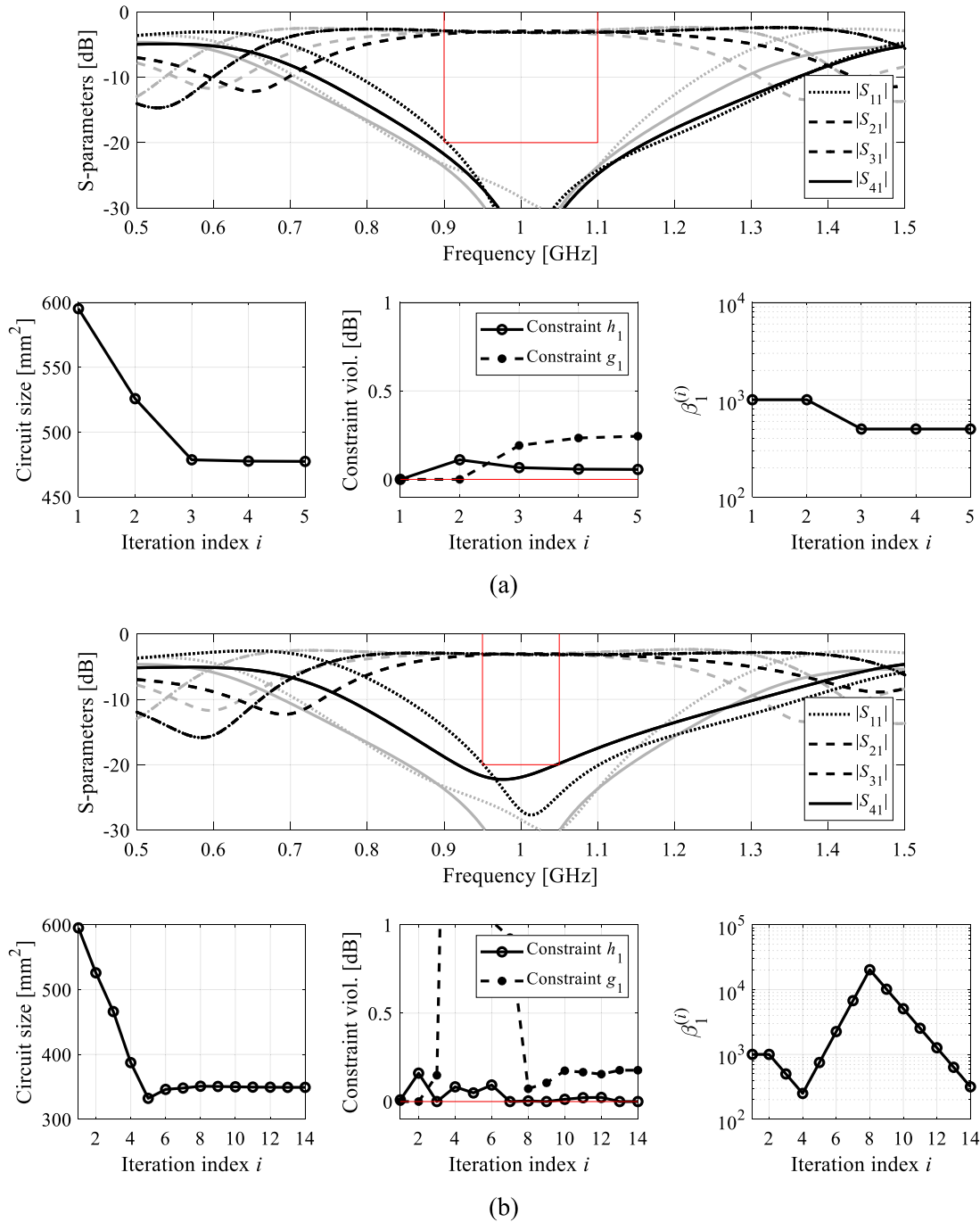
Fig. 7. Initial (grey) and optimized (black) *S*-parameters of Circuit II. The vertical and horizontal lines mark the target operating bandwidth and the acceptance level for the matching |S_{11}| and isolation |S_{41}| responses. Also shown is the evolution of the circuit size and constraint violations (in case of feasibility, violations shown as zero), as well as evolution of the penalty coefficient for the inequality constraint: (a) design scenario I (bandwidth 0.9 GHz to 1.1 GHz), (b) design scenario II (bandwidth 0.95 GHz to 1.05 GHz).

such as power split and phase difference), as well as its further improvements in terms of the computational efficiency.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements
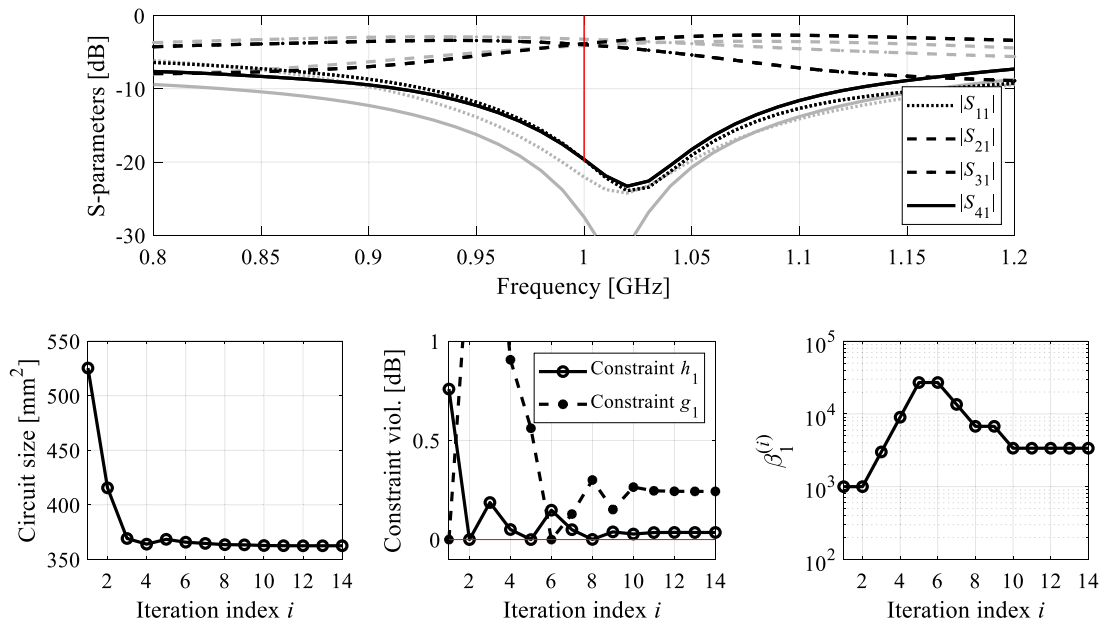
Fig. 8. Initial (grey) and optimized (black) S-parameters of Circuit III. Also shown is the evolution of the circuit size and constraint violations (in case of feasibility, violations shown as zero), as well as evolution of the penalty coefficient for the inequality constraint. The vertical line shows the target operating frequency of 1.0 GHz.

Fig. 9. Initial (grey) and optimized (black) S-parameters of Circuit III. Also shown is the evolution of the circuit size and constraint violations (in case of feasibility, violations shown as zero), as well as evolution of the penalty coefficient for the inequality constraint. The vertical line shows the target operating frequency of 2.0 GHz.

References

[1] J. Zhang, S. Yan, X. Hu, G.A.E. Vandenbosch, Dual-band dual-polarized wearable button array with miniaturized radiator, IEEE Trans. Biomed. Circuits Syst. 13 (6) (2019) 1583–1592.

[2] M.S. Khan, A. Iftikhar, R.M. Shubair, A. Capobianco, B.D. Braaten, D.E. Anagnostou, Eight-element compact UWB-MIMO/diversity antenna with WLAN band rejection for 3G/4G/5G communications, IEEE Open J. Ant. Propag. 1 (2020) 196–206.

[3] M. Kim, S. Kim, Design and fabrication of 77-GHz radar absorbing materials using frequency-selective surfaces for autonomous vehicles application, IEEE Microw. Wirel. Comp. Lett. 29 (12) (2019) 779–782.

[4] Z. He, C. Liu, A compact high-efficiency broadband rectifier with a wide dynamic range of input power for energy harvesting, IEEE Microw. Wirel. Comp. Lett. 30 (4) (2020) 433–436.

[5] Z.H. Jiang, M.D. Gregory, D.H. Werner, Design and experimental investigation of a compact circularly polarized integrated filtering antenna for wearable biotelemetric devices, IEEE Trans. Biomed. Circuits Syst. 10 (2) (2016) 328–338.

[6] J. Kracek, M. Švanda, M. Mazanek, J. Machac, Implantable semi-active UHF RFID tag with inductive wireless power transfer, IEEE Ant. Wirel. Propag. Lett. 15 (2016) 1657–1660.

[7] H. Zhang, M. Li, F. Yang, S. Xu, H. Zhou, Y. Yang, L. Chen, A low-profile compact dual-band L-shape monopole antenna for microwave thorax monitoring, IEEE Ant. Wirel. Propag. Lett. 19 (3) (2020) 448–452.

[8] L. Martinez, A. Belenguer, V.E. Boria, A.L. Borja, Compact folded bandpass filter in empty substrate integrated coaxial line at S-band, IEEE Microw. Wirel. Comp. Lett. 29 (5) (2019) 315–317.

[9] Z.J. Hou, Y. Yang, X. Zhu, Y.C. Li, E. Dutkiewicz, Q. Xue, A compact and low-loss bandpass filter using self-coupled folded-line resonator with capacitive

feeding technique, IEEE Electron Dev. Lett. 39 (10) (2018) 1584–1587.

[10] G. Shen, W. Che, Q. Xue, Compact microwave and millimeter-wave band-pass filters using LTCC-based hybrid lumped and distributed resonators, IEEE Access 7 (2019) 104797–104809.

[11] W. Qin, Q. Xue, Elliptic response bandpass filter based on complementary CMRC, Electr. Lett. 49 (15) (2013) 945–947.

[12] S. Chen, M. Guo, K. Xu, P. Zhao, L. Dong, G. Wang, A frequency synthesizer based microwave permittivity sensor using CMRC structure, IEEE Access 6 (2018) 8556–8563.

[13] J. Luo, J. He, A. Apriyana, G. Feng, Q. Huang, Y.P. Zhang, Tunable surface-plasmon-polariton filter constructed by corrugated metallic line and high permittivity material, IEEE Access 6 (2018) 10358–10364.

[14] F. Wei, Y. Jay Guo, P. Qin, X. Wei Shi, Compact balanced dual- and tri-band bandpass filters based on stub loaded resonators, IEEE Microw. Wirel. Comp. Lett. 25 (2) (2015) 76–78.

[15] W. Zhang, Z. Shen, K. Xu, J. Shi, A compact wideband phase shifter using slotted substrate integrated waveguide, IEEE Microw. Wirel. Comp. Lett. 29 (12) (2019) 767–770.

[16] S. Sen, T. Moyra, Compact microstrip low-pass filtering power divider with wide harmonic suppression, IET Microw. Ant. Propag. 13 (12) (2019) 2026–2031.

[17] S. Liu, F. Xu, Compact multilayer half mode substrate integrated waveguide 3-dB coupler, IEEE Microw. Wirel. Comp. Lett. 28 (7) (2018) 564–566.

[18] D. Yang, H. Zhai, C. Guo, H. Li, A compact single-layer wideband microstrip antenna with filtering performance, IEEE Antennas Wirel. Propag. Lett. 19 (5) (2020) 801–805.

[19] F. Feng, W. Na, W. Liu, S. Yan, L. Zhu, Q.-J. Zhang, Parallel gradient-based EM optimization for microwave components using adjoint-sensitivity-based neuro-transfer function surrogate, IEEE Trans. Microw. Theory Techn. 68 (9) (2020) 3606–3620.

[20] W. Na, K. Liu, H. Cai, W. Zhang, H. Xie, D. Jin, Efficient EM optimization exploiting parallel local sampling strategy and Bayesian optimization for microwave applications, IEEE Microw. Wirel. Comp. Lett. 31 (10) (2021) 1103–1106.

[21] S. Koziel, A. Pietrenko-Dabrowska, M. Al-Hasan, Improved-efficacy optimization of compact microwave passives by means of frequency-related regularization, IEEE Access 8 (2020) 195317-195326.

[22] B. Liu, H. Yang, M.J. Lancaster, Global optimization of microwave filters based on a surrogate model-assisted evolutionary algorithm, IEEE Trans. Microw. Theory Techn. 65 (6) (2017) 1976–1985.

[23] A. Pietrenko-Dabrowska, S. Koziel, Globalized parametric optimization of microwave components by means of response features and inverse metamodels, Sc. Rep. 11 (2021) 23718.

[24] Y. Zhang, D.-.W. Gong, J.-.H. Zhang, Robot path planning in uncertain environment using multi-objective particle swarm optimization, Neurocomputing 103 (2013) 172–185.

[25] W. Guo, J. Li, G. Chen, Y. Niu, C. Chen, A PSO-optimized real-time fault-tolerant task allocation algorithm in wireless sensor networks, IEEE Trans. Parall. Distrib. Syst. 26 (12) (2015) 3236–3249.

[26] G. Liu, W. Guo, Y. Niu, G. Chen, X. Huang, A PSO-based timing-driven octi-linear steiner tree algorithm for VLSI routing considering bend reduction, Soft. Comput. 19 (2015) 1153–1169.

[27] F. Güneş, A. Uluslu, P. Mahouti, Pareto optimal characterization of a microwave transistor, IEEE Access 8 (2020) 47900–47913.

[28] S. Koziel, A. Pietrenko-Dabrowska, Recent advances in accelerated multi-objective design of high-frequency structures using knowledge-based constrained modeling approach, Knowl. Based Syst. 214 (2021) 106726.

[29] S. Koziel, A. Pietrenko-Dabrowska, Constrained multi-objective optimization of compact microwave circuits by design triangulation and Pareto front interpolation, Eur. J. Op. Res. 299 (1) (2022) 302–312.

[30] Z. Lv, L. Wang, Z. Han, J. Zhao, W. Wang, Surrogate-assisted particle swarm optimization algorithm with Pareto active learning for expensive multi-objective optimization, IEEE/CAA J. Autom. Sinica 6 (3) (2019) 838–849.

[31] A. Pietrenko-Dabrowska, S. Koziel, M. Al-Hasan, Expedited yield optimization of narrow- and multi-band antennas using performance-driven surrogates, IEEE Access (2020) 143104–143113.

[32] A.K. Prasad, M. Ahadi, S. Roy, Multidimensional uncertainty quantification of microwave/RF networks using linear regression and optimal design of experiments, IEEE Trans. Microw. Theory Techn. 64 (8) (2016) 2433–2446.

[33] M.A.E. Sabbagh, M.H. Bakr, J.W. Bandler, Adjoint higher order sensitivities for fast full-wave optimization of microwave filters, IEEE Trans. Microw. Theory Techn. 54 (8) (2006) 3339–3351.

[34] S. Koziel, F. Mosler, S. Reitzinger, P. Thoma, Robust microwave design optimization using adjoint sensitivity and trust regions, Int. J. RF Microw. CAE 22 (1) (2012) 10–19.

[35] S. Koziel, A. Pietrenko-Dabrowska, Efficient gradient-based algorithm with numerical derivatives for expedited optimization of multi-parameter miniaturized impedance matching transformers, Radioengineering 28 (3) (2019) 572–578.

[36] A. Pietrenko-Dabrowska, S. Koziel, Expedited antenna optimization with numerical derivatives and gradient change tracking, Eng. Comp. 37 (4) (2019) 1179–1193.

[37] A. Pietrenko-Dabrowska, S. Koziel, Computationally-efficient design optimization of antennas by accelerated gradient search with sensitivity and design change monitoring, IET Microw. Ant. Prop. 14 (2) (2020) 165–170.

[38] F. Feng, J. Zhang, W. Zhang, Z. Zhao, J. Jin, Q. Zhang, Coarse- and fine-mesh space mapping for EM optimization incorporating mesh deformation, IEEE Microw. Wirel. Comp. Lett. 29 (8) (2019) 510–512.

[39] S. Koziel, A. Bekasiewicz, Multi-Objective Design of Antennas using Surrogate Models, World Scientific, Singapore, 2016.

[40] S. Koziel, A. Pietrenko-Dabrowska, Performance-Driven Surrogate Modeling of High-Frequency Structures, Springer, 2020.

[41] J.E. Rayas-Sanchez, S. Koziel, J.W. Bandler, Advanced RF and microwave design optimization: a journey and a vision of future trends, IEEE J. Microw. 1 (1) (2021) 481–493.

[42] Z. Zhang, H. Chen, Y. Yu, F. Jiang, Q.S. Cheng, Yield-constrained optimization design using polynomial chaos for microwave filters, IEEE Access 9 (2021) 22408–22416.

[43] D.K. Lim, D.K. Woo, H.K. Yeo, S.Y. Jung, J.S. Ro, H.K. Jung, A novel surrogate-assisted multi-objective optimization algorithm for an electromagnetic machine design, IEEE Trans. Magn. 51 (3) (2015) 8200804.

[44] B. Xia, Z. Ren, C.S. Koh, Utilizing kriging surrogate models for multi-objective robust optimization of electromagnetic devices, IEEE Trans. Magn. 50 (2) (2014) 7017104.

[45] S. An, S. Yang, O.A. Mohammed, A kriging-assisted light beam search method for multi-objective electromagnetic inverse problems, IEEE Trans. Magn. 54 (3) (2018) 7001104.

[46] N. Taran, D.M. Ionel, D.G. Dorrell, Two-level surrogate-assisted differential evolution multi-objective optimization of electric machines using 3-D FEA, IEEE Trans. Magn. 54 (11) (2018) 8107605.

[47] G. Bramerdorfer, A.C. Zavoianu, Surrogate-based multi-objective optimization of electrical machine designs facilitating tolerance analysis, IEEE Trans. Magn. 53 (8) (2017) 8107611.

[48] J.A. Easum, J. Nagar, P.L. Werner, D.H. Werner, Efficient multi-objective antenna optimization with tolerance analysis through the use of surrogate models, IEEE Trans. Ant. Prop. 66 (12) (2018) 6706–6715.

[49] B. Liu, H. Aliakbarian, Z. Ma, G.A.E. Vandenbosch, G. Gielen, P. Excell, An efficient method for antenna design optimization based on evolutionary computation and machine learning techniques, IEEE Trans. Ant. Propag. 62 (1) (2014) 7–18.

[50] I. Couckuyt, F. Declercq, T. Dhaene, H. Rogier, L. Knockaert, Surrogate-based infill optimization applied to electromagnetic problems, Int. J. RF Microw. Computt.-Aided Eng. 20 (5) (2010) 492–501.

[51] A.M. Alzahed, S.M. Mikki, Y.M.M. Antar, Nonlinear mutual coupling compensation operator design using a novel electromagnetic machine learning paradigm, IEEE Ant. Wirel. Prop. Lett. 18 (5) (2019) 861–865.

[52] J.E. Rayas-Sanchez, Power in simplicity with ASM: tracing the aggressive space mapping algorithm over two decades of development and engineering applications, IEEE Microw. Mag. 17 (4) (2016) 64–76.

[53] S. Koziel, J.W. Bandler, K. Madsen, Space mapping with adaptive response correction for microwave design optimization, IEEE Trans. Microw. Theory Tech. 57 (2) (2009) 478–486.

[54] D.I.L. de Villiers, I. Couckuyt, T. Dhaene, Multi-objective optimization of reflector antennas using kriging and probability of improvement, in: Int. Symp. Ant. Prop, San Diego, USA, 2017, pp. 985–986.

[55] Q. Zhou, Y. Wang, P. Jiang, X. Shao, S.-K. Choi, J. Hu, L. Cao, X. Meng, An active learning radial basis function modeling method based on self-organization maps for simulation-based design problems, Knowl.-Based Syst. 131 (2017) 10–27.

[56] D. Kim, M. Kim, W. Kim, Wafer edge yield prediction using a combined long short-term memory and feed-forward neural network model for semiconductor manufacturing, IEEE Access 8 (2020) 215125–215132.

[57] J.P. Jacobs, Characterization by Gaussian processes of finite substrate size effects on gain patterns of microstrip antennas, IET Microw. Ant. Prop. 10 (11) (2016) 1189–1195.

[58] A. Toktas, D. Ustun, M. Tekbas, Multi-objective design of multi-layer radar absorber using surrogate-based optimization, IEEE Trans. Microw. Theory Techn. 67 (8) (2019) 3318–3329.

[59] H.M. Torun, M. Swaminathan, High-dimensional global optimization method for high-frequency electronic design, IEEE Trans. Microw. Theory Techn. 67 (6) (2019) 2128–2142.

[60] S. Xiao, G.Q. Liu, K.L. Zhang, Y.Z. Jing, J.H. Duan, P. Di Barba, J.K. Sykulski, Multi-objective Pareto optimization of electromagnetic devices exploiting kriging with Lipschitzian optimized expected improvement, IEEE Trans. Magn. 54 (3) (2018) 7001704.

[61] S. Koziel, A. Pietrenko-Dabrowska, Global EM-driven optimization of multi-band antennas using knowledge-based inverse response-feature surrogates, Knowl. Based Syst. 227 (2021) 107189.

[62] S. Li, X. Fan, P.D. Laforge, Q.S. Cheng, Surrogate model-based space mapping in postfabrication bandpass filters' tuning, IEEE Trans. Microw. Theory Tech. 68 (6) (2020) 2172–2182.

[63] S. Koziel, S.D, Unnsteinsson expedited design closure of antennas by means of trust-region-based adaptive response scaling, IEEE Antennas Wirel. Prop. Lett. 17 (6) (2018) 1099–1103.

[64] Y. Su, J. Li, Z. Fan, R. Chen, Shaping optimization of double reflector antenna based on manifold mapping, in: Int. Applied Comp. Electromagnetics Soc. Symp., ACES, Suzhou, China, 2017, pp. 1–2.

[65] S. Koziel, L. Leifsson, Simulation-Driven Design By Knowledge-Based Response Correction Techniques, Springer, New York, 2016.

[66] S. Koziel, A. Pietrenko-Dabrowska, P. Plotka, Reduced-cost microwave design closure by multi-resolution EM simulations and knowledge-based model management, IEEE Access 9 (2021) 116326–116337.

[67] M.A. Haq, S. Koziel, Q.S. Cheng, Miniaturization of wideband antennas by means of feed line topology alterations, IET Microw. Ant. Prop. 12 (13) (2018) 2128–2134.

[68] D.O. Johanesson, S. Koziel, Feasible space boundary search for improved optimization-based miniaturization of antenna structures, IET Microw. Ant. Prop. 12 (8) (2018) 1273–1278.

[69] A.V. Fiacco, G.P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York, 1969.

[70] U. Ullah, S. Koziel, I.B. Mabrouk, Rapid re-design and bandwidth/size trade-offs for compact wideband circular polarization antennas using inverse surrogates and fast EM-based parameter tuning, IEEE Trans. Ant. Prop. 68 (1) (2019) 81–89.

[71] S. Koziel, A. Pietrenko-Dabrowska, Reliable EM-driven size reduction of antenna structures by means of adaptive penalty factors, IEEE Trans. Ant. Propag. Early View (2022).

[72] M. Mahrokh, S. Koziel, Explicit size-reduction of circularly polarized antennas through constrained optimization with penalty factor adaptation, IEEE Access 9 (2021) 132390–132396.

[73] S. Koziel, A. Pietrenko-Dabrowska, On EM-driven size reduction of antenna structures with explicit constraint handling, IEEE Access 9 (2021) 165766–165772.

[74] X. Cai, H. Sun, Q. Zhang, Y. Huang, A grid weighted sum Pareto local search for combinatorial multi and many-objective optimization, IEEE Trans. Cybern. 49 (9) (2019) 3586–3598.

[75] R.T. Marler, J.S. Arora, The weighted sum method for multi-objective optimization: new insights, Struct. Multidistip. Opt. 41 (2010) 853–862.

[76] A.R. Conn, N.I.M. Gould, P.L. Toint, Trust region methods, in: MPS-SIAM Series on Optimization, 2000.

[77] J. Kennedy, R.C. Eberhart, Swarm Intelligence, Morgan Kaufmann, San Francisco, USA, 2001.

[78] C. Tseng, C. Chang, A rigorous design methodology for compact planar branch-line and rat-race couplers with asymmetrical T-structures, IEEE Trans. Microw. Theory Techn. 60 (7) (2012) 2085–2092.

[79] S. Koziel, A. Pietrenko-Dabrowska, Reduced-cost surrogate modeling of compact microwave components by two-level kriging interpolation, Eng. Opt. 52 (6) (2019) 960–972.

[80] D.A. Letavin, S.N. Shabunin, Miniaturization of a branch-line coupler using microstrip cells, in: Int. Scientific-Technical Conf. Actual Problems of Electronics Instrument Engineering, APEIE, 2018, pp. 62–65.

[81] D.A. Letavin, Y.E. Mitelman, V.A. Chechetkin, Compact microstrip branch-line coupler with unequal power division, in: European Conf. Ant. Propag., EuCAP, 2017, pp. 1162–1165.

7.1 Incorporation of Phase Difference in the Correction-Based Equality Constraint Handling Scheme

This section provides a discussion on the extension of the proposed correction-based equality correction scheme. The extended version incorporates coupler phase difference as an additional equality constraint, a formulated description of which is provided in the following.

In most of the presented cases, the main optimization engine in the proposed procedure is the TR algorithm.

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x} \in X} U_p(\boldsymbol{x}) \tag{7.1}$$

where the objective function takes the form of

$$U_P(\boldsymbol{x}) = A(\boldsymbol{x}) + \sum_{k=1}^{n_{ineq}} \beta_{ineq,k} c_{ineq,k}(\boldsymbol{x})^2 + \sum_{k=1}^{n_{eq}} \beta_{eq,k} c_{eq,k}(\boldsymbol{x})^2 \tag{7.2}$$

The penalty term for the inequality constraints $\beta_{ineq,k}$ are determined using the adaptive penalty factor adjustment procedure described in Section 2.4 of Paper # 4. The penalty factors $\beta_{ineq,k}$ are fixed at low values (e.g., $\beta_{ineq,k} = 10$, $k = 1, \ldots, n_{eq}$), because the enforcement of the equality constraints is mainly realized using the correction procedure of Section 2.4 of the paper, breifly described here as well.

Equation (7.2) is a generic form of the objective function which can take any number of either inequality or equality constraints.

Turning the focus on the equality constraints, the verification examples in the paper consider power split ratio as the only equality constraint. The corresponding correction scheme includes a single equality constraint formulated as the following minimization sub-problem

$$\boldsymbol{x}_{corr}(M) = \arg \min_{\boldsymbol{x}, \|\boldsymbol{x} - \boldsymbol{x}^{(i+1)}\| \leq M} \left\{ h_1(L_A^{(i+1)}(\boldsymbol{x})) \right\} \tag{7.3}$$

with constraints

$$A(\boldsymbol{x}) \leq A(\boldsymbol{x}^{(i+1)}) \tag{7.4}$$

$$g_j(\boldsymbol{x}) \leq g_j(\boldsymbol{x}^{(i+1)}), \quad j = 1, \ldots, n_g \tag{7.5}$$

The corrected design is obtained as

$$\boldsymbol{x}_c^{(i+1)} = \arg \min_{M} \left\{ \left\{ h_1 \left(L_A^{(i+1)}(\boldsymbol{x}_{corr}(M)) \right) \right\} \leq \varepsilon_h \right\} \tag{7.6}$$

The extended version includes phase difference between ports 2 and 3 as an additional equality constraint defined as $h_2(\boldsymbol{x}) = \angle S_{31}(\boldsymbol{x},f) - \angle S_{21}(\boldsymbol{x},f) = 90°$.

The corresponding minimization sub-problem is reformulated as

$$\boldsymbol{x}_{corr}(M) = \arg \min_{\boldsymbol{x}, \|\boldsymbol{x} - \boldsymbol{x}^{(i+1)}\| \leq M} \max \left\{ h_1(L_A^{(i+1)}(\boldsymbol{x})), h_2(L_A^{(i+1)}(\boldsymbol{x})) \right\} \tag{7.7}$$

with constraints (7.4) and (7.5).

The corrected design is obtained as

$$\boldsymbol{x}_c^{(i+1)} = \arg \min_{M} \left\{ \max \left\{ h_1 \left(L_A^{(i+1)}(\boldsymbol{x}_{corr}(M)) \right), ..., h_2 \left(L_A^{(i+1)}(\boldsymbol{x}_{corr}(M)) \right) \right\} \le \varepsilon_h \right\} \qquad (7.8)$$

Chapter 8

8   Paper # 5

Slawomir Koziel, Anna Pietrenko-Dabrowska, and Marzieh Mahrokh

Globalized Simulation-Driven Miniaturization of Microwave Circuits by Means of Dimensionality-Reduced Constrained Surrogates

scientific reports

OPEN

Globalized simulation-driven miniaturization of microwave circuits by means of dimensionality-reduced constrained surrogates

Slawomir Koziel[1,2], Anna Pietrenko-Dabrowska[2✉] & Marzieh Mahrokh[1]

Small size has become a crucial prerequisite in the design of modern microwave components. Miniaturized devices are essential for a number of application areas, including wireless communications, 5G/6G technology, wearable devices, or the internet of things. Notwithstanding, size reduction generally degrades the electrical performance of microwave systems. Therefore, trade-off solutions have to be sought that represent acceptable compromises between the ability to meet the design targets and physical compactness. From an optimization perspective, this poses a constrained task, which is computationally expensive because a reliable evaluation of microwave components has to rely on full-wave electromagnetic analysis. Furthermore, due to its constrained nature, size reduction is a multimodal problem, i.e., the results are highly dependent on the initial design. Thus, utilization of global search algorithms is advisable in principle, yet, often undoable in practice because of the associated computational expenses, especially when using nature-inspired procedures. This paper introduces a novel technique for globalized miniaturization of microwave components. Our technique starts by identifying the feasible region boundary, and by constructing a dimensionality-reduced surrogate model therein. Global optimization of the metamodel is followed by EM-driven local tuning. Application of the domain-confined surrogate ensures low cost of the entire procedure, further reduced by the incorporation of variable-fidelity EM simulations. Our framework is validated using two microstrip couplers, and compared to nature-inspired optimization, as well as gradient-based size reduction. The results indicate superior miniaturization rates and low running cost, which make the presented algorithm a potential candidate for efficient simulation-based design of compact structures.

Design of contemporary microwave passive circuits is a non-trivial endeavour. Performance and functionality demands have been continuously growing to satisfy the needs of the emerging application areas such as mobile communications[1], internet of things[2], remote sensing[3], microwave imaging[4], energy harvesting[5], autonomous vehicles[6], or implantable device[7]. Some of the requirements include multi-band operation[8], reconfigurability[9], harmonic suppression[10], or custom phase characteristics[11]. Furthermore, many applications impose constraints on the physical size of the devices, which fosters miniaturization[12–15]. Miniaturization is essentially a two-stage process. Initially, a basic circuit architecture is selected to ensure compact dimensions[16,17], often with the use of techniques such as transmission line (TL) folding/meandering[18], utilization of the slow-wave phenomenon[19] (typically, in the form of compact microwave resonant cells, CMRCs[20]), multi-layer realizations[21], or incorporation of various supplementary components (stubs[22], defected ground structures[10], substrate-integrated waveguides[23], shorting pins[24]). All of these methods result in geometrically complex structures, whose accurate evaluation requires full-wave electromagnetic (EM) analysis due to the presence of cross-coupling effects in densely arranged circuit layouts. At the same time, geometrical modifications lead to the increase of the number of parameters that have to be simultaneously tuned in order to control both the circuit size and electrical

[1]Engineering Optimization & Modeling Center, Reykjavik University, 102 Reykjavik, Iceland. [2]Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland. ✉email: anna.dabrowska@pg.edu.pl

figures of merit. As size reduction is detrimental to electrical performance of the system, any practical design is a trade-off between compactness and functionality. Initial circuit dimensions can usually be obtained using a combination of equivalent networks and parametric studies, yet rigorous numerical optimization is indispensable to significantly enhance the system performance.

Nowadays, parameter tuning is more and more often carried out using rigorous numerical optimization methods, which is recommended due to their ability to handle multiple parameters, objectives and constraints[25–27]. Optimization is not only used for the purpose of design closure (final tuning of geometry parameters, often using local algorithms[28]), but also multi-criterial design[29], uncertainty quantification (tolerance analysis[30], design centering[31]), and global optimization[32]. Whatever the purpose, microwave circuit optimization is a challenging endeavor. Perhaps the most significant bottleneck is its high computational cost when executed at the level of EM simulation models, otherwise necessary to ensure reliability of the process. While the costs are often manageable in the case of local (e.g., gradient-based) tuning, global or multi-objective optimization, as well as statistical design, are considerably more demanding[33,34]. Consequently, there have been numerous techniques developed to improve computational efficiency of EM-driven optimization. Some of these methods include utilization of adjoint sensitivities[35,36], restricted sensitivity updates[37–39], the employment of (fast) dedicated solvers[40], mesh deformation approaches[41], feature-based optimization[42], or cognition-driven design[43]. Yet, one of the most important developments in making simulation-based design more practical in terms of CPU expenses, has been the incorporation of surrogate modeling methodologies[44–47].

Surrogate-assisted optimization (SBO) has attracted a considerable attention in the design of high-frequency circuits, including microwave and antenna components, primarily because of its ability to accelerate simulation-based procedures, including local[48], and global optimization[49], robust design[50], or multi-criterial optimization[51]. Surrogate-assisted procedures utilize both data-driven[52] or physics-based metamodels[53]. Data-driven techniques are versatile and readily transferrable between the problem domains[54], which make them the most popular class of modeling methods. Specific approaches often employed in the context of high-frequency engineering include kriging[55], radial basis functions[56], many variations of artificial neural networks[57–59], support vector regression[60], Gaussian process regression[61], or polynomial chaos expansion (PCE)[62]. Data-driven models are cheap to evaluate, but they are affected by the curse of dimensionality: the number of training data samples necessary to construct reliable models quickly grows with the number of parameters and parameter ranges, and may become unmanageable even for medium-size problems. Physics-based surrogates are constructed using a lower-fidelity representation of the system of interest (e.g., equivalent network[63], or coarse-discretization EM analysis[64]). The problem-specific knowledge embedded in the low-fidelity model enhances generalization capability of the surrogates of this class[65]. At the same time, it limits the applicability range because each problem requires its own low-fidelity model. Some of popular techniques include space mapping[66], and response correction methods[67–69], most of which are typically used for local optimization purposes. Surrogate-assisted frameworks allowing for solving expensive constrained optimization problems have been recently proposed in[70] and[71].

As mentioned earlier, size reduction constitutes a prerequisite in the design of contemporary microwave components. It is normally addressed at the level of selecting the circuit architecture[72–74], yet appropriate parameter tuning plays just as important part. From numerical perspective, size reduction is a constrained task with expensive constraints that require evaluating through EM analysis (e.g., acceptance thresholds imposed on the circuit bandwidth, power split ratio, or phase responses)[75]. As size reduction is detrimental to electrical performance, at least some of the constraints remain active at the optimal solution, which emphasizes the role of feasible region boundary exploration in the search process[76]. These challenges can be addressed by implicit constraint handling using a penalty function approach[77], where the problem is reformulated into a formally unconstrained one. However, performance of the optimization process turns out to be contingent upon the appropriate choice of penalty coefficients[78], which are normally selected by trial and error. This gave rise to adaptive penalty coefficient strategies[79,80]. Recently, explicit constraint handling methods have been proposed[81], along with the techniques for customized treatment of equality constraints, based on correction procedures[82]. Another approach to constraint handling in the context of design of antennas and antenna arrays using evolutionary algorithms that allows for circumventing the issue of an appropriate setting of the penalty coefficients, has been proposed in[83,84].

The optimization techniques outlined in the previous paragraph are local search procedures, which are highly dependent on the available starting points. At the same time, miniaturized microwave components are often developed using transmission line meandering[85], CMRCs[20], or various geometrical modifications[86,87], which leads to parameter redundancy (e.g., a typical number of geometry parameters of CMRC is four to six versus two for a conventional TL). The increased number of degrees of freedom enables the necessary flexibility; yet, its handling calls for global optimization. Conventional global search methods (e.g., nature-inspired population-based algorithms[88,89]) are not suitable for the purpose due to poor computational efficiency. This work proposes a novel procedure for globalized miniaturization of passive microwave circuits, which is designed to lessen the cost of the search process while maintaining reliability. The presented technique is a multi-stage process, which starts by (roughly) approximating the feasible region boundary using a set of randomly generated parameter vectors coupled with initial (local) optimization runs. Subsequently, a reduced-dimensionality domain is established in the feasible boundary region, along with a fast surrogate model, the latter utilized to conduct a globalized size reduction. The search process is concluded by final miniaturization-oriented parameter tuning of the circuit. The abovementioned dimensionality reduction is achieved using the spectral analysis of the pre-optimized parameter vectors. The initial steps of the search process are executed using low-fidelity model to lower the CPU cost even further. Our methodology has been validated using a compact rat-race coupler and a dual-band power divider. The numerical results demonstrate superior performance of the proposed routine, with regard to both the computational efficiency and reliability, as well as constraint control, as compared to the nature-inspired optimization and multiple-start local search.

Constraint	Type	Analytical description$^\$$
Input matching $\lvert S_{11}\rvert$ not exceeding –20 dB over the operating bandwidth $[f_1\,f_2]$	Inequality	$\lvert S_{11}(\boldsymbol{x},f)\rvert \leq -20$ dB for $f \in [f_1\,f_2]$
Port isolation $\lvert S_{41}\rvert$ not exceeding –20 dB over the operating bandwidth $[f_1\,f_2]$	Inequality	$\lvert S_{41}(\boldsymbol{x},f)\rvert \leq -20$ dB for $f \in [f_1\,f_2]$
In-band transmission ripple not exceeding 0.2 dB over the operating bandwidth $[f_1\,f_2]$	Inequality	$\lvert S_{21}(\boldsymbol{x},f)\rvert \geq -0.2$ dB for $f \in [f_1\,f_2]$
Power split ratio between output ports 2 and 3 equal to K_P at the center frequency f_0	Equality	$\lvert S_{31}(\boldsymbol{x},f)\rvert - \lvert S_{21}(\boldsymbol{x},f)\rvert = K_P$ at $f = f_0$;
Phase difference between output ports 2 and 3 equal to 90° at the center frequency f_0	Equality	$\angle S_{31}(\boldsymbol{x},f) - \angle S_{21}(\boldsymbol{x},f) = 90°$ at $f = f_0$;

Table 1. Example constraints in size-reduction of microwave components. $^\$$The symbol $\lvert S_{jk}(\boldsymbol{x},f)\rvert$ stands for the modulus of the S-parameter S_{jk} at the design $\boldsymbol{x}$, and frequency f.

The primary technical contributions of the paper can be summarized as follows: (i) the development of a novel framework for globalized EM-driven miniaturization of passive microwave circuits, which incorporates several mechanisms (variable-fidelity EM analysis, surrogate modelling, and dimensionality reduction), (ii) a demonstration of the competitive performance of the presented method as compared to the state-of-the-art methods (both local and global), also in terms of achievable miniaturization rates, (iii) a demonstration of the search process reliability, especially low variance of the optimization results (equivalent to consistent repeatability). According to the authors' knowledge, the literature does not offer any size-reduction framework featuring comparable properties and performance. Consequently, the proposed approach may become an interesting alternative to existing methods, particularly in terms of combining computational efficiency and achievable miniaturization rates.

Globalized EM-driven miniaturization using variable-fidelity models and spectral analysis

This section provides the details of the globalized optimization procedure for passive microwave components introduced in the paper. The EM-driven size reduction problem is formulated in "EM-driven size reduction: problem statement" Section. The concept of the optimization algorithm is described in Globalized size reduction: explanation of the concept Section. "Feasible Region boundary approximation" Section elaborates on feasible region boundary approximation, one of the keystones of the presented methodology. The surrogate modeling stage is outlined in "Surrogate model construction", "Surrogate model optimization for size reduction" Sections. delineates global optimization of surrogate model, whereas "Final parameter adjustment" Section discusses the final (gradient-based) design closure. The entire optimization framework is summarized in "Globalized EM-driven size reduction: complete procedure" Section using a pseudocode and a flow diagram.

EM-driven size reduction: problem statement. Design of compact microwave components consists of the two major stages: (i) a selection of the circuit geometry, and (ii) parameter tuning. The first stage is essential to ensure structural miniaturization (e.g., by replacing TLs with their abbreviated counterparts such as CMRCs[90]), whereas the second allows for exploring further the size reduction potential of the chosen architecture, in particular, to push the design as much as possible towards feasible region boundary, where the electrical performance parameters are barely satisfied in exchange for additional reduction of the circuit physical dimensions.

In the following, we will use $\boldsymbol{x} = [x_1 \ldots x_n]^T$ for a vector of design variables, and by $A(\boldsymbol{x})$ the circuit size (e.g., footprint area). The miniaturization problem is simply stated as

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in X_f} A(\mathbf{x}) \tag{1}$$

where X_f is a feasible space, i.e., the region in which all design constraints are satisfied. The constraints can be of inequality type, $g_k(\boldsymbol{x}) \leq 0$, $k = 1, \ldots, n_g$ (e.g., acceptance threshold for the circuit operating bandwidth), and equality constraints $h_k(\boldsymbol{x}) = 0$, $k = 1, \ldots, n_h$ (e.g., target power split ratio).

The constraints imposed on electrical characteristics of the circuit are expensive to evaluate (require EM simulation). Consequently, their explicit handling might be problematic, although some recent strategies demonstrated promising results (e.g.[81],). A convenient alternative is implicit handling using a penalty functions[77]. According to this approach, the original objective function is supplemented by scaled constraint violations. We have

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} U(\mathbf{x}) \tag{2}$$

where the merit function U is given by

$$U(\mathbf{x}) = A(\mathbf{x}) + \sum_{k=1}^{n_g+n_h} \beta_k c_k(\mathbf{x}) \tag{3}$$

The second term in (3) consist of penalty functions $c_k(\boldsymbol{x})$ and proportionality coefficients β_k; $n_c = n_g + n_h$ is the overall number of constraints. Table 1 provides a few examples of constraints that may be encountered in size reduction tasks. Table 2 shows example definitions of the penalty functions, often expressed through relative violations. It should be noted that the formulation (2), (3) corresponds to soft constraint handling, i.e., it does not guarantee their exact satisfaction. As a matter of fact, the constraint control is reliant on coefficients β_k,

Constraint	Penalty function
Input matching $|S_{11}|$ not exceeding -20 dB over the operating bandwidth $[f_1 f_2]$	$c(\mathbf{x}) = \left[\frac{\max\{\max\{f_1 \leq f \leq f_2 : |S_{11}(\mathbf{x},f)|\} + 20, 0\}}{20} \right]^2$
Port isolation $|S_{41}|$ not exceeding -20 dB over the operating bandwidth $[f_1 f_2]$	$c(\mathbf{x}) = \left[\frac{\max\{\max\{f_1 \leq f \leq f_2 : |S_{41}(\mathbf{x},f)|\} + 20, 0\}}{20} \right]^2$
In-band transmission ripple not exceeding 0.2 dB over the operating bandwidth $[f_1 f_2]$	$c(\mathbf{x}) = \left[\frac{\max\{-\min\{f_1 \leq f \leq f_2 : |S_{21}(\mathbf{x},f)|\} - 0.2, 0\}}{0.2} \right]^2$
Power split ratio between output ports 2 and 3 equal to K_P at the center frequency f_0	$c(\mathbf{x}) = \left[|S_{31}(\mathbf{x},f_0)| - |S_{21}(\mathbf{x},f_0)| - K_P \right]^2$
Phase difference between output ports 2 and 3 equal to 90° at the center frequency f_0	$c(\mathbf{x}) = \left[\angle S_{31}(\mathbf{x},f_0) - \angle S_{21}(\mathbf{x},f_0) - 90° \right]^2$

Table 2. Possible formulation of penalty functions for constraints of Table 1.

which should be adjusted appropriately. Too low values result in an insufficient control over constraint violations, whereas the values that are too high lead to numerical problems as the objective function becomes very steep at the feasible region boundary. This issue has been addressed by adaptive coefficient adjustment schemes[78,80], where the values of β_k are changed based on currently-detected violations, as well as the algorithm convergence status[78].

Globalized size reduction: explanation of the concept. Miniaturization of microwave components is typically obtained by appropriate selection of the circuit architecture (line folding[18], slow-wave phenomenon[19], defected ground[10]). Any deviation from conventional structures results in increasing the number of geometry parameters, and creating complex relations between those parameter and electrical characteristics, which are often counter-intuitive. From the perspective of optimization as considered in "EM-driven size reduction: problem statement" section, this leads to multimodal tasks potentially exhibiting a number of local optima. Appropriate treatment of such problems requires global search methods. However, as mentioned in "Introduction" section, conventional algorithms (e.g., population-based metaheuristics[91]) are just too expensive. On the other hand, surrogate-assisted methods[92] are hindered by dimensionality issues and high-nonlinearity of microwave component responses. This paper proposes an alternative methodology, designed to improve the efficacy of the optimization-based size reduction process, which includes making the search less dependent on the initial design quality as compared to local methods.

The central concept of the proposed technique is the boundary X_b of the feasible region X_f. We have the following definitions (here, X is the space of design parameters, usually, delimited by lower and upper bounds):

$$X_f = \left\{ \mathbf{x} \in X : g_k(\mathbf{x}) \leq 0 \text{ for } k = 1, \dots, n_g \text{ AND } h_k(\mathbf{x}) = 0 \text{ for } k = 1, \dots, n_h \right\} \quad (4)$$

and

$$X_b = \left\{ \begin{array}{l} \mathbf{x} \in X : g_k(\mathbf{x}) = 0 \text{ for at least one } k = 1, \dots, n_g \\ \text{OR } h_k(\mathbf{x}) = 0 \text{ for at least one } k = 1, \dots, n_h \end{array} \right\} \quad (5)$$

As miniaturization generally degrades the circuit performance (e.g., reduces the operating bandwidth), minimum-size designs normally reside in X_b as at least one of the constraints is active. Therefore, (approximate) identification of the spatial allocation of X_b allows for narrowing down the part of the parameter space that needs to be explored. The exploration involves surrogate modeling techniques, as well as final EM-driven parameter tuning. Figure 1 shows the overall concepts of the proposed optimization methodology. Figure 2 briefly explains the search stages. Detailed description will be provided in the remaining parts of this section.

To improve computational efficiency of the process, Stages 1 through 3 are carried out using the low-fidelity model R_c, which is based on coarse-discretization EM analysis. At these stages, the accuracy is not of a major concern. Stages 5 and 7 are executed using the high-fidelity model R_f, which is to ensure reliability of the search process. "Feasible region boundary approximation" Section through "Final parameter adjustment" provide the details of how all the stages are implemented. "Globalized EM-driven size reduction: complete procedure" section summarizes the complete framework.

Feasible region boundary approximation. The parameter space for the microwave circuit of interest is conventionally assumed to be an interval $X = [l \, u]$. Therein, the vectors l and u represent the lower and upper parameter bounds. At the component level it may be written as $l_k \leq x_k \leq u_k$, $k = 1, \dots, n$. Stages 1 through 3 of the search process (cf. Fig. 2) are arranged as follows. We start by generating N_1 random observables $\mathbf{x}_r^{(j)}$ that satisfy the following conditions:

- $\mathbf{x}_r^{(j)} \in X = [l \, u]$;
- $A(\mathbf{x}_r^{(j)}) \leq A_1$;
- $A(\mathbf{x}_r^{(j)}) \geq A_2$;
- $\mathbf{x}_r^{(j)}$ satisfy other possible constraints (problem dependent).

Therein, A_1 and A_2 are optional maximum and minimum circuit size values. These might be available from the previous design work with the same circuit, and give the idea of what level of physical sized are achievable for the circuit.
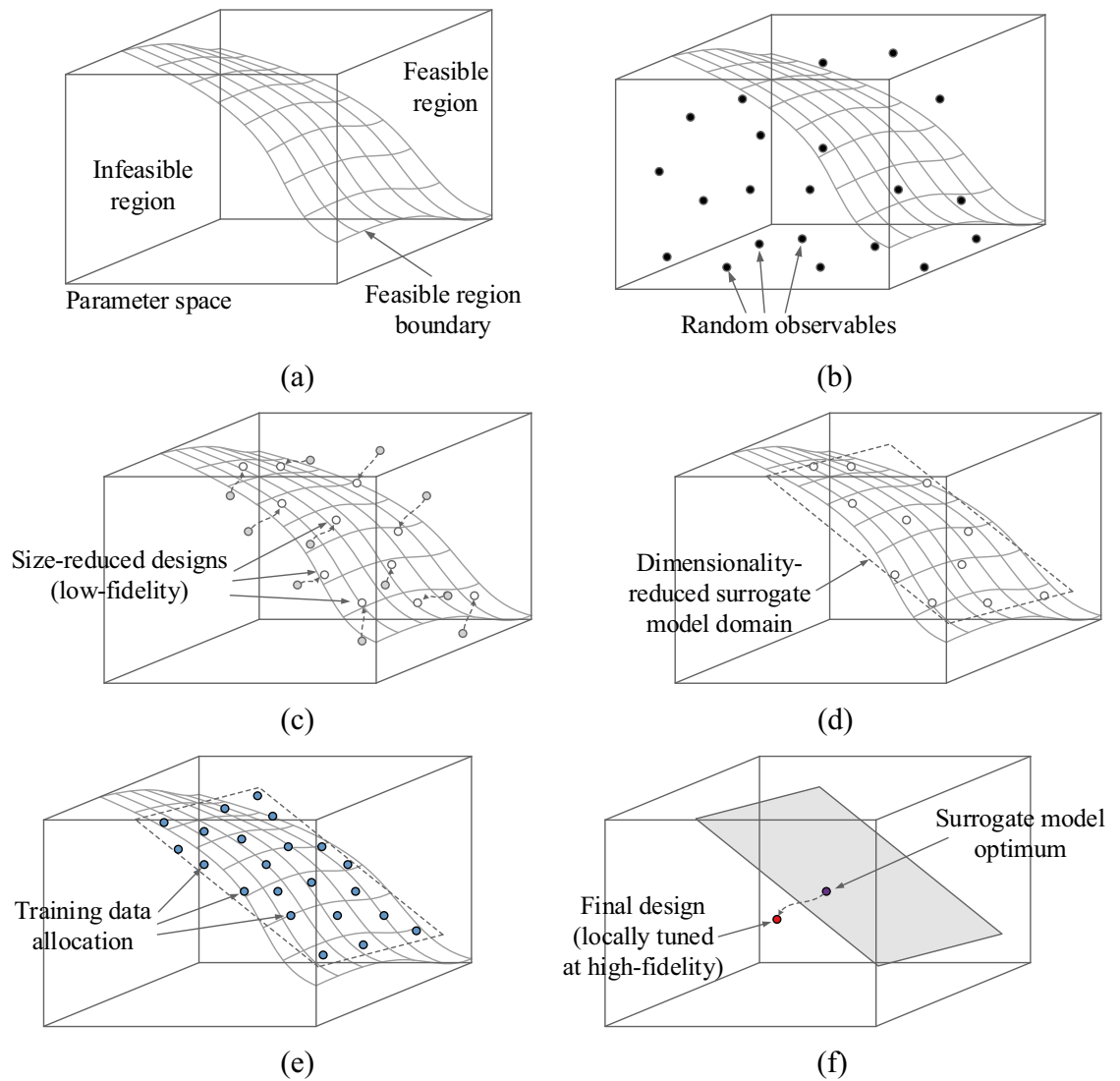
Figure 1. Conceptual illustration of the proposed globalized size reduction procedure involving variable-resolution EM models and dimensionality reduction: (**a**) Exemplary parameter space with feasible and infeasible region indicated along with the boundary region marked in grey, (**b**) Stage 1: allocation of random observables; the acquired EM data will be used to approximate the feasible region boundary, (**c**) Stages 2 and 3: selected observables are optimized for size reduction at low-fidelity EM level, (**d**) Stage 4: spectral analysis of the pre-optimized observables is used to define the domain of the surrogate model in the boundary area, (**e**) Stage 5: training data is allocated in the domain, and kriging interpolation model is constructed, (**f**) Stages 6 and 7: the design obtained through global optimization of surrogate model is finally tuned at high-fidelity level using gradient-based routine.

In other words, having such data, we may initially filter out samples that correspond to circuit sizes that are clearly too small or too large. One may also impose additional constraints for the sake of restricting the parameter space regions to be sampled even further. Such constraints should be based on the designer's knowledge and/or available data. The number of samples N_1 is a user-defined control parameter, typically set to 500.

Having the set of samples, the low-fidelity model is evaluated to obtain the circuit characteristics $\boldsymbol{R}_c(\boldsymbol{x}_r^{(j)})$, $j = 1$, ..., N_1. The best subset of N_2 samples, $\{\boldsymbol{x}_i^{(j)}\}_{j=1,...,N2} \subset \{\boldsymbol{x}_r^{(j)}\}_{j=1,...,N1}$ is selected based on the corresponding values of penalty-based objective function (3). Here, we set $N_2 = 20$. This number is a reasonable trade-off between the computational cost of subsequent stages and the data on the feasible region boundary X_b that can be obtained therefrom.

The parameter vectors $\boldsymbol{x}_i^{(j)}$ are used as initial designs for EM-driven size reduction at the low-fidelity model level. Thus, for $j = 1$, ..., N_2, we solve

$$x_c^{(j)} = \operatorname{argmin}\{x : U(x)\} \tag{6}$$

Stage	Name	Action undertaken
1	Allocation of random observables	Allocation of random vectors in the parameter space. Circuit responses at these vectors are evaluated through EM analysis, cf. Fig. 1(b)
2	Sample selection	A small subset of observables generated in Stage 1 are selected based on their distance to the feasible region boundary, e.g., by evaluating the values of design constraints therein
3	Pre-optimization	The designs selected in Stage 2 are optimized for minimum size according to (2), (3) at the level of low-fidelity EM model; cf. Fig. 1(c)
4	Spectral analysis and surrogate model definition	The pre-optimized designs obtained in Stage 3 undergo spectral analysis (here, using principal component analysis [90]), and the reduced-dimensionality subset is defined as the domain of the surrogate model to be constructed, cf. Fig. 1(d). The domain is spanned by the most dominant eigenvectors of the pre-optimized design set
5	Data acquisition and surrogate model construction	The training data is allocated in the domain, and kriging interpolation model representing circuit responses is identified therein, cf. Fig. 1(e)
6	Global optimization of the surrogate	The surrogate model is optimized for minimum circuit size within its domain
7	Final tuning	Local (gradient-based) tuning of the circuit parameters is performed to yield the final design

Figure 2. Conceptual stages of globalized size reduction of microwave components.

Again, U is the objective function (3) incorporating the penalty terms. Because the accuracy is not of the fundamental importance at this stage, the problem (6) uses relaxed termination criteria to reduce the CPU cost. In this work, the underlying optimization method is a trust-region (TR) gradient-based algorithm[94]; the circuit response sensitivity is estimated using finite differentiation[95] (cf. "Final parameter adjustment" section).

Upon solving (6) for $j = 1, \ldots, N_2$, an N_3- element subset of $\{x_c^{(j)}\}_{j=1,\ldots,N2}$ is selected that consists of designs being of sufficient quality in terms of constraint violation. This is to filter out designs for which (6) was unsuccessful. Later on, the selected subset will be referred to as $\{x_c^{(j)}\}_{j=1,\ldots,N3}$.

Surrogate model construction. In the proposed global optimization framework, the surrogate model is constructed to represent the circuit responses. As the objective function (3) is a function of these responses, the surrogate-predicted response is employed for its evaluation. Next, global optimization of the surrogate model is carried out, and the approximate design is rendered, which further undergoes a local refinement as shown in Fig. 1f.

The vectors $x_c^{(j)}$, $j = 1, \ldots, N_3$, have been obtained by optimizing the circuit for minimum size. Also, due to the definition of the objective function, they exhibit low constraint violations. Consequently, these designs reside in the vicinity of the boundary X_b of the feasible region. Based on $\{x_c^{(j)}\}$, we will set up the domain of the surrogate model to be employed for global search purposes. Further, using the spectral analysis of the set $\{x_c^{(j)}\}$, the dimensionality of the domain will be reduced as compared to the dimensionality of the original parameter space X, which is to limit the computational cost of the surrogate model rendition.

Figure 3 summarizes the process of defining the surrogate model domain. It follows the procedure proposed in[96] for domain-confined modelling of high-frequency devices. The main idea is to define the domain of the surrogate model as the smallest set spanned by the most dominant eigenvectors that contains all vectors in $\{x_c^{(j)}\}$. In practice, the designs $x_c^{(j)}$ are strongly correlated (in the spatial sense), therefore, the dimensionality p of the domain can be kept small without losing too much of information. In this work, we keep $p = 3$ for the verification circuits considered in "Demonstration examples" section. Dimensionality reduction is essential for reducing the number of training data samples (here, denoted as N_4) necessary to build the surrogate model. In this work, we set $N_4 = 200$, which results in good predictive power of the model (at the level of a few percent of relative RMS error). The training data is obtained from the high-fidelity EM model R_f. Figure 4 provides a graphical illustration of the surrogate model domain definition.

The surrogate model is constructed using kriging interpolation[54], although a particular selection of the modeling method is not critical. The training samples, denoted as $x_B^{(j)} \in X_p$, $j = 1, \ldots, N_4$, are distributed using Latin Hypercube Sampling (LHS)[97]. The design of experiments procedure (cf. Fig. 5) has to account for the fact that the domain is not aligned with the coordinate system axes. The surrogate model will be used to perform global optimization of the circuit within its domain X_p.

Surrogate model optimization for size reduction. The domain of the surrogate model covers the vicinity of the feasible region boundary along the most important directions, as determined using the spectral

Step	Action	Comment
1	Define the center of gravity $\boldsymbol{x}_m = N_3^{-1}\sum_{j=1,\ldots,N3}\boldsymbol{x}_c^{(j)}$ of the set $\{\boldsymbol{x}_c^{(j)}\}$	-
2	Define covariance matrix $$\boldsymbol{S}_c = \frac{1}{N_3-1}\sum_{j=1}^{N_3}(\boldsymbol{x}_c^{(j)}-\boldsymbol{x}_m)(\boldsymbol{x}_c^{(j)}-\boldsymbol{x}_m)^T$$	It is assumed that $N_3 > n$ (parameter space dimensionality)
3	Perform spectral analysis of $\boldsymbol{S}_c$ to yield eigenvectors (principal components) $\boldsymbol{a}_k$, $k = 1, \ldots, n$, of $\{\boldsymbol{x}_c^{(j)}\}$, and the corresponding eigenvalues λ_k	Eigenvalues represent the variance of $\{\boldsymbol{x}_c^{(j)}\}$ in the eigenspace; the eigenvalues are arranged in a descending order, i.e., we have $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$
4	Define matrices $\boldsymbol{A}_k = [\boldsymbol{a}_1 \ldots \boldsymbol{a}_k]$	
5	Compute expansions $\boldsymbol{x}_c^{(j)} = \sum_{k=1}^{n} b_{jk}\boldsymbol{a}_k$	Expansion is unique as $\{\boldsymbol{a}_k\}$ forms a basis in the parameter space X
6	Define: $b_{j.\max} = \max\{k : b_{kj}\}$, $b_{j.\min} = \min\{k : b_{kj}\}$, $b_{j.0} = \dfrac{b_{j.\min}+b_{j.\max}}{2}$, $j = 1, \ldots, n$,	-
7	Define: $\boldsymbol{b}_0 = [b_{1.0} \ldots b_{n.0}]^T$ and $\boldsymbol{\lambda}_b = [\lambda_{b1} \ldots \lambda_{bn}]^T$ with $\lambda_{bj} = (b_{j.\max} - b_{j.\min})/2$	-
8	Define the center point $\boldsymbol{x}_c = \boldsymbol{x}_m + \boldsymbol{A}_n\boldsymbol{b}_0$	-
9	Define p-dimensional model domain: $$X_p = \left\{ \begin{array}{l} \boldsymbol{x} = \boldsymbol{x}_c + \sum_{k=1}^{p}(2\lambda_k-1)\lambda_{b_k}\boldsymbol{a}_k \\ 0 \leq \lambda_k \leq 1, \ k = 1,\ldots,p \end{array} \right\}$$	The set X_p contains all vectors $\boldsymbol{x}_c^{(j)}$, $j = 1, \ldots, N_3$, in the directions $\boldsymbol{a}_1$ through $\boldsymbol{a}_p$

Figure 3. Definition of reduced-dimensionality surrogate model domain.
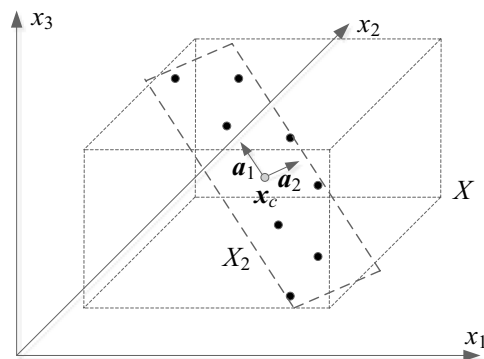
Figure 4. Conceptual illustration of reduced-dimensionality surrogate model domain. Here, a two-dimensional domain X_2 spanned by the two most dominant eigenvectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$; the gray circle represents the center point $\boldsymbol{x}_c$ (cf. Figure 3).

analysis described in "Surrogate model construction" section (cf. Fig. 3). Having the surrogate, the next stage is to optimize it in a global sense within X_p. Due to low dimensionality of the domain, the search process is conducted in two phases:

Exhaustive search on the grid M_p given in the form of a complete set of vectors

Design of experiments (domain X_p):

1. Allocate normalized samples $\mathbf{z}^{(j)} = [z_1^{(j)} \ldots z_p^{(j)}]^T$ in the p-dimensional unity interval $X_U = [0\ 1] \times \ldots \times [0\ 1] = [0\ 1]^p$ (i.e., $0 \le z_j \le 1$, $j = 1, \ldots, p$) using LHS [94];

2. Define mapping $h : X_U \rightarrow X_p$ as $h(\mathbf{z}) = \mathbf{x}_c + \sum_{k=1,\ldots,p}(2z_k - 1)\lambda_k \mathbf{a}_k$, where the coefficients λ_k are as in the definition of X_p (cf. Fig. 3);

3. Obtain $\mathbf{x}_B^{(j)} = h(\mathbf{z}^{(j)})$ for $j = 1, \ldots, N_4$.

(a)

(b)

Figure 5. Design of experiments (data sampling) in reduced-dimensionality domain (here, two dimensional): (**a**) sampling proced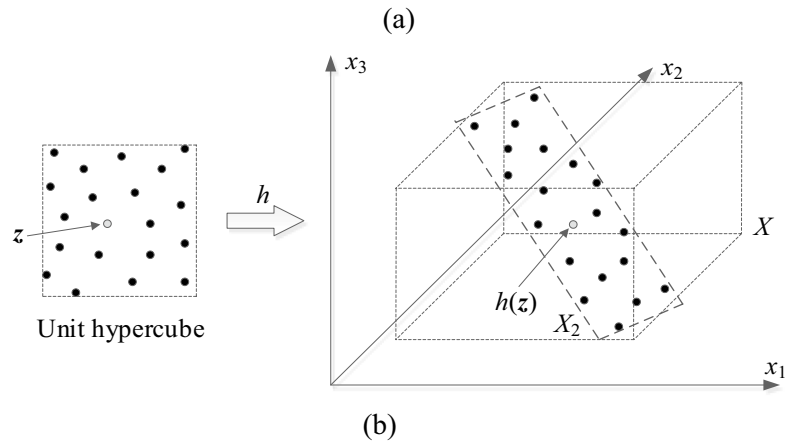ure, (**b**) graphical illustration: normalized samples are uniformly distributed in the unity interval using LHS, and mapped into X_2 using the transformation h.

$$M_p = \left\{ \begin{array}{l} \mathbf{x} = \mathbf{x}_c + \sum_{k=1}^{p}(2\lambda_k - 1)\lambda_{b_k}\mathbf{a}_k \\ \lambda_k \in \{0, 1/K, 2/K, \ldots, 1\}, \quad k = 1, \ldots, p \end{array} \right\} \tag{7}$$

where K is the grid resolution (we use $K = 20$). The initial design $\mathbf{x}_g^{(0)}$ is found by solving

$$\mathbf{x}_g^{(0)} = \arg\min \left\{ \mathbf{x} \in M_p \cap X : U(\mathbf{x}) \right\} \tag{8}$$

Note that $\mathbf{x}_g^{(0)}$ is the design that minimizes (surrogate-evaluated) U over the intersection of the search grid and parameter space X (in general, X_p may extend beyond the original domain X);

Local size-reduction-oriented optimization of the surrogate within $X_p \cap X$, according to (2). The optimization algorithm is a trust-region gradient search described in "Final parameter adjustment" section. For notational simplicity, the design found at this stage will be also denoted as $\mathbf{x}_g^{(0)}$.

Final parameter adjustment. The final stage of the global optimization procedure proposed in this paper is a local tuning of the circuit parameter. For accuracy reasons, it is performed at the level of the high-fidelity model R_f. This step is again executed using the trust-region (TR) gradient-based routine[94], which was also used for initial tuning ("Feasible region boundary approximation" section), and surrogate optimization (" Surrogate model optimization for size reduction" section). The formulation of the TR algorithm has been recalled in Fig. 6.

Globalized EM-driven size reduction: complete procedure. This section puts together the building blocks of the globalized size reduction algorithm discussed in " Globalized size reduction: explanation of the concept" section through "Final parameter adjustment", and summarizes the operating flow of the entire framework. The algorithm control parameters are gathered in Table 3, their meaning has been already elaborated on earlier. Here we provide general guidelines for their setup. Four parameters of Table 3, i.e., N_1 through N_4, pertain to the computational budget of the entire optimization framework. The number N_1 of samples used for initial approximation of the feasible region boundary is typically set to 500, because, in most practical cases, this value is sufficient and allows for a satisfactory estimation of the said boundary. The next parameter, N_2, i.e., the number of samples for which optimization-based size reduction is carried out, is typically set to 20. This value constitutes a reasonable trade-off between the computational cost of subsequent tuning these deigns and the precision of assessing the surrogate domain. The number N_3 of the refined designs of sufficient quality should somewhat exceed a half *of* N_2, as this allows for discarding the designs for which the tuning procedure has failed. The fourth parameter controlling the computational budget, i.e., the number N_4 of data samples used for setting
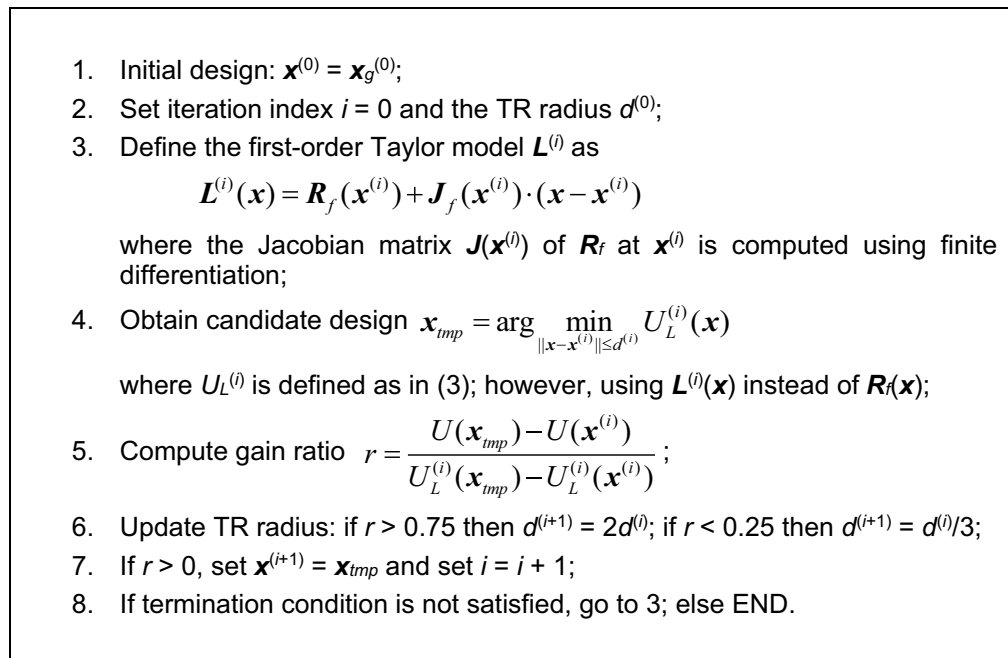
1. Initial design: $\boldsymbol{x}^{(0)} = \boldsymbol{x}_g^{(0)}$;

2. Set iteration index $i = 0$ and the TR radius $d^{(0)}$;

3. Define the first-order Taylor model $\boldsymbol{L}^{(i)}$ as

$$\boldsymbol{L}^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_f(\boldsymbol{x}^{(i)}) + \boldsymbol{J}_f(\boldsymbol{x}^{(i)}) \cdot (\boldsymbol{x} - \boldsymbol{x}^{(i)})$$

where the Jacobian matrix $\boldsymbol{J}(\boldsymbol{x}^{(i)})$ of $\boldsymbol{R}_f$ at $\boldsymbol{x}^{(i)}$ is computed using finite differentiation;

4. Obtain candidate design $\boldsymbol{x}_{tmp} = \arg \min\limits_{\|\boldsymbol{x}-\boldsymbol{x}^{(i)}\| \leq d^{(i)}} U_L^{(i)}(\boldsymbol{x})$

where $U_L^{(i)}$ is defined as in (3); however, using $\boldsymbol{L}^{(i)}(\boldsymbol{x})$ instead of $\boldsymbol{R}_f(\boldsymbol{x})$;

5. Compute gain ratio $r = \dfrac{U(\boldsymbol{x}_{tmp}) - U(\boldsymbol{x}^{(i)})}{U_L^{(i)}(\boldsymbol{x}_{tmp}) - U_L^{(i)}(\boldsymbol{x}^{(i)})}$;

6. Update TR radius: if $r > 0.75$ then $d^{(i+1)} = 2d^{(i)}$; if $r < 0.25$ then $d^{(i+1)} = d^{(i)}/3$;

7. If $r > 0$, set $\boldsymbol{x}^{(i+1)} = \boldsymbol{x}_{tmp}$ and set $i = i + 1$;

8. If termination condition is not satisfied, go to 3; else END.

Figure 6. Formulation of the trust-region gradient-based algorithm. The termination condition is based on convergence in argument, $\|\boldsymbol{x}^{(i+1)}-\boldsymbol{x}^{(i)}\| < \varepsilon$, and reduction of the TR radius, $d^{(i)} < \varepsilon$ (whichever occurs first). The termination threshold ε is set to 10^{-3} for final tuning of the high-fidelity model, but it is relaxed to 10^{-2} for low-fidelity optimization runs.

Parameter	Meaning	Recommended value
N_1	The number of random observables generated to obtain initial approximation of the feasible region boundary ("Feasible region boundary approximation" section)	500
N_2	The number of observables selected to conduct size reduction optimization runs at low-fidelity level ("Feasible region boundary approximation" section)	20
N_3	The number of designs selected from the outcome of low-fidelity model optimization runs, and used to define the surrogate model domain ("Feasible region boundary approximation" section)	$> \lceil N_2/2 \rceil$
N_4	The number of training data samples for surrogate model construction ("Surrogate model construction" section)	200
p	Dimensionality of the surrogate model domain ("Surrogate Model Optimization for Size Reduction" section)	3

Table 3. Control parameters of the proposed globalized size reduction algorithm.

up the surrogate model, should be adjusted to ensure the required accuracy of this model (e.g., at the level of a few percent of relative RMS error).

As for the last parameter p, which refers to the surrogate domain dimensionality, it should be kept small (of around one third or half of the number of design variables) to maintain the training data acquisition cost at a reasonable level. The values provided in the Table 3 will be used in the verification experiments of "Demonstration examples" section. The pseudocode of the algorithm can be found in Fig. 7, whereas Fig. 8 shows the flow diagram of the method.

It should also be emphasized that while utilization of the low-fidelity EM model at the early stages of the search process leads to certain inaccuracies (including identification of the feasible region boundary, where the constrained optimum is normally allocated), these are corrected at the final stages, where the high-fidelity EM model is employed to fine-tune the geometry parameters of the circuit.

Demonstration examples

The proposed globalized size reduction framework is validated with the use of two examples of microstrip circuits, a rat-race coupler (RRC) and a branch-line coupler (BLC). The structures are designed for minimum size, under the constraints imposed on their operating frequency, operating bandwidth, and power split ratio. The performance of the algorithm is compared to nature-inspired optimization using particle swarm optimizer (PSO), as a representative technique of this category, as well as multiple-start gradient search. This remainder of this Section is arranged in a following manner. "Test cases and experimental setup" Section delineates the test

1.   Define the parameter space X and design constraints (cf. Section 2.1);
2.   Generate N_1 random observables $\boldsymbol{x}_r^{(j)}$ within the parameter space X (cf. Section 2.3);
3.   Evaluate low-fidelity responses $\boldsymbol{R}_c(\boldsymbol{x}_r^{(j)})$, $j = 1, ..., N_1$;
4.   Select N_2-element subset $\{\boldsymbol{x}_i^{(j)}\}_{j=1,...,N2} \subset \{\boldsymbol{x}_r^{(j)}\}_{j=1,...,N1}$ to be used as initial designs for size reduction; selection based on constraint violation levels (the smaller, the better);
5.   For each $j = 1, ..., N_2$, find $\boldsymbol{x}_c^{(j)} = \arg\min\{\boldsymbol{x} : U(\boldsymbol{x})\}$ starting from $\boldsymbol{x}_i^{(j)}$;
6.   Select an N_3-element subset of $\{\boldsymbol{x}_c^{(j)}\}_{j = 1,...,N2}$ based on the lowest constraint violation (cf. Section 2.5);
7.   Use designs $\{\boldsymbol{x}_c^{(j)}\}$, $j = 1, ..., N_3$ to construct the surrogate model domain X_p (cf. Section 2.4):
     - Define covariance matrix $\boldsymbol{S}_c$ (cf. Fig. 3);
     - Perform spectral analysis of $\boldsymbol{S}_c$;
     - Use p most significant eigenvectors $\boldsymbol{a}_k$ of $\boldsymbol{S}_c$ to define domain X_p (cf. Table 3);
8.   Perform design of experiments in X_p (cf. Section 2.4);
9.   Acquire training data and identify the kriging surrogate model using high-fidelity EM model (cf. Section 2.4);
10. Perform global size-reduction-oriented surrogate model optimization (Section 2.5):
     - Define search grid M_p as in (7);
     - Obtain global surrogate model optimum found as $\boldsymbol{x}_g^{(0)} = \arg\min\left\{\boldsymbol{x} \in M_p \cap X : U(\boldsymbol{x})\right\}$;
     - Improve $\boldsymbol{x}_g^{(0)}$ through local gradient-based optimization;
11. Find the final design $\boldsymbol{x}^*$ through final parameter tuning using the TR algorithm; optimization is performed using high-fidelity EM model $\boldsymbol{R}_f$ (cf. Section 2.6);
12. END.

Figure 7. Operating flow of the proposed globalized size reduction algorithm.

cases and the most important experimental settings. "Numerical results" section gathers the numerical results. "Discussion" section contains a discussion that includes qualitative comparisons between the introduced and the benchmark techniques concerning reliability and computational efficiency.

Test cases and experimental setup.   Verification of the proposed algorithm involves two microstrip circuits, both shown in Fig. 9, and referred to as Circuit I and II, respectively. The evaluation models are rendered in CST Microwave Studio, and simulated with the use of its time-domain solver. The design task is posed as follows:

- Minimize the footprint area $A(\boldsymbol{x})$ of the circuit under design;
- Satisfy inequality constraint for matching and port isolation, $g_1(\boldsymbol{x}) = \max\{f \in F : \max\{|S_{11}(\boldsymbol{x},f)|, |S_{41}(\boldsymbol{x},f)|\}\} + 20$ dB;
- Satisfy equality constraint for the power split ratio: $h_1(\boldsymbol{x}) = |\,|S_{31}(\boldsymbol{x},f_0)| - |S_{21}(\boldsymbol{x},f_0)|\,| = 0$ (both transmission responses are in dB);

The first constraint corresponds to a condition that both $|S_{11}(\boldsymbol{x},f)|$ and $|S_{41}(\boldsymbol{x},f)|$ should not be greater than -20 dB over the operating band F. The second constraint requires the circuit to maintain an even power split ratio at its operating frequency f_0. The objective function is formulated as in (3) with the penalty functions defined as in Tables 1 and 2. Table 4 provides essential parameters for both circuits, including design variables, parameter spaces, operating frequencies, etc.
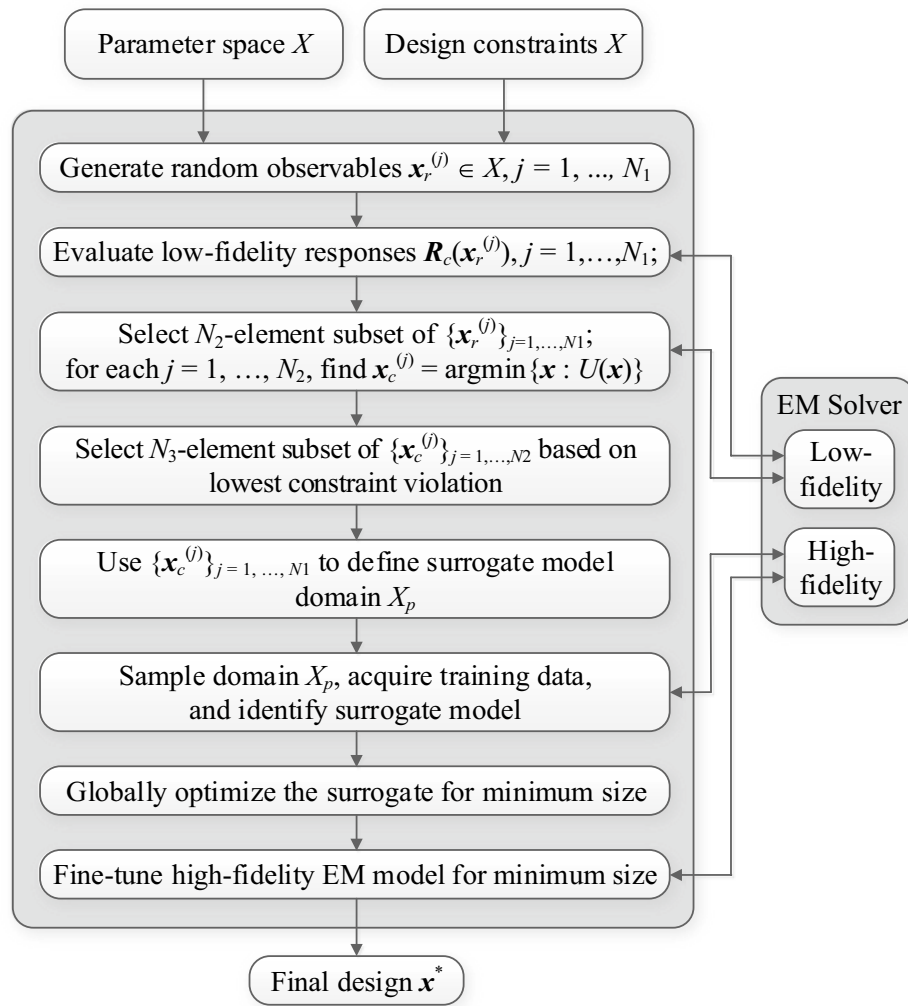
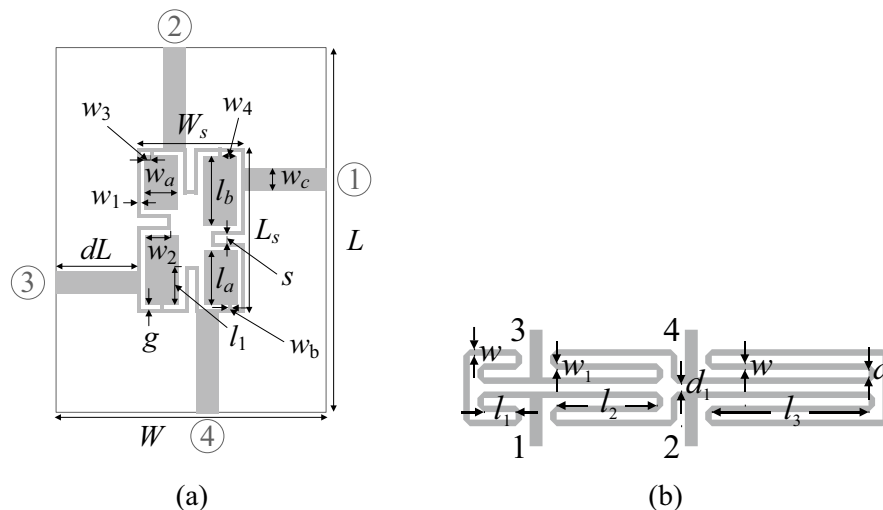**Figure 8.** Flow diagram of the proposed globalized size reduction framework.

Figure 9. Microstrip structures employed as test cases for verification of the proposed size reduction framework: (**a**) compact branch-line coupler (Circuit I)[98], (**b**) rat-race coupler with folded transmission lines (Circuit II)[99].

Circuit	I[98]	II[99]
Substrate	AD300 ($\varepsilon_r = 2.97$, $h = 0.76$ mm)	RO4003 ($\varepsilon_r = 3.38$, $h = 0.762$ mm)
Designable Parameters [mm]	$x = [g\ l_{1r}\ l_a\ l_b\ w_1\ w_{2r}\ w_{3r}\ w_{4r}\ w_a\ w_b]^T$	$x = [l_1\ l_2\ l_3\ d\ w\ w_1]^T$
Other Parameters [mm]	$L = 2dL + L_s$, $L_s = 4w_1 + 4\ g + s + l_a + l_b$, $W = 2dL + W_s$, $W_s = 4w_1 + 4\ g + s + 2w_a$, $l_1 = l_b l_{1r}$, $w_2 = w_a w_{2r}$, $w_3 = w_{3r} w_a$, and $w_4 = w_{4r} w_a$, $w_c = 1.9$ mm	$d_1 = d + |w - w_1|$, $d = 1.0$, $w_0 = 1.7$, and $l_0 = 15$ mm
Parameter space X	$l = [0.4\ 0.1\ 3.0\ 3.0\ 0.4\ 0.1\ 0.1\ 0.1\ 2.0\ 0.2]^T$ $u = [1.0\ 0.99\ 15.0\ 25.0\ 1.5\ 0.99\ 0.9\ 0.9\ 12.0\ 1.0]^T$	$l = [0.1\ 5.0\ 5.0\ 0.2\ 0.2\ 0.5]^T$ $u = [15.0\ 30.0\ 50.0\ 2.0\ 2.0\ 2.0]^T$
Operating parameters	$f_0 = 1.5$ GHz $F = [1.45\ 1.55]$ GHz	$f_0 = 1.0$ GHz $F = [0.95\ 1.05]$ GHz
Low-fidelity EM model	~24,000 mesh cells Simulation time 110 s	~50,000 mesh cells Simulation time 55 s
High-fidelity EM model	~160,000 mesh cells Simulation time 240 s	~200,000 mesh cells Simulation time 160 s

Table 4. Essential parameters of Circuits I and II of Fig. 9.

Algorithm	Description
I	Local gradient-based size reduction using the trust region algorithm (cf. "Final parameter adjustment" section). The optimization problem is formulated as in (2), (3)
II	Particle swarm optimizer (PSO)[100], employed as a representative nature-inspired technique. The algorithm setup is as follows: swarm size of 10, maximum number of iterations 100, standard setup of control parameters ($\chi = 0.73$, $c_1 = c_2 = 2.05$), cf.[100]. The problem formulated as in (2), (3)

Table 5. Benchmark algorithms.

	Performance figure						
			Inequality constraint		Equality constraint		
Optimization algorithm	Circuit size A [mm²] [1]	Std(A) [2]	Violation D_1 [dB] [3]	Std(D_1) [dB] [4]	Violation D_2 [dB] [5]	Std(D_2) [dB] [6]	CPU cost[7]
Algorithm I	295.1	24.7	3.6	1.9	0.2	0.1	$77 \times R_f$ [5.2 h]
Algorithm II	541.5	240.4	5.5	6.8	0.7	0.1	$1,000 \times R_f$ [66.7 h]
Globalized search with dimensionality reduction (this work)	301.8	3.9	0.4	0.2	0.1	0.03	$852 \times R_f$ [56.8 h]

Table 6. Optimization results for Circuit I. [1] Optimized footprint area of the circuit averaged over ten algorithm runs. [2] Standard deviation of the optimized footprint area averaged over ten algorithm runs. [3] Violation of inequality constraint, defined as $D_1 = \max\{f \in F : \max\{|S_{11}(x,f)|, |S_{41}(x,f)|\}\} + 20$ dB, averaged over ten algorithm runs. [4] Standard deviation of the constraint violation D_1, averaged over ten algorithm runs. [5] Violation of equality constraint, defined as $D_2 = |\ |S_{31}(x,f_0)| - |S_{21}(x,f_0)|\ |$ dB, averaged over ten algorithm runs. [6] Standard deviation of the constraint violation D_2, averaged over ten algorithm runs. [7] Cost expressed in terms of equivalent number of high-fidelity EM analyzes. Numbers in brackets correspond to the running time in hours.

The low-fidelity models of both verification circuits are obtained by reducing discretization density of the structure. The proportion of simulation times between the high- and low-fidelity model is 2.2 and 2.9 for Circuit I and II, respectively, which will carry over to computational savings of the entire optimization procedure.

It should be emphasized that the search spaces are large in terms of the ranges of geometry parameters (average upper-to-lower bound ratio is almost seven in the case of Circuit I and over thirty for Circuit II). Furthermore, both circuits feature parameter redundancy, i.e., additional variables related to the specific circuit geometries (utilization of CMRCs for Circuit I, and transmission line meandering for Circuit II). Both factors make the design tasks multimodal, in particular, size reduction outcome will very much depend on the initial design. At the same time, global search methods are likely to exhibit limited repeatability of solutions due to the parameter space dimensionality and overall size. In order to take this into account, verification experiments are carried out in a statistical sense, by running multiple instances of the proposed and benchmark algorithms, and comparing statistical moments of the outcomes. More specifically, each algorithm is run ten times. The figures of interest to be compared are average circuit size along with the standard deviation of the size, as well as average violation of design constraints (and the corresponding standard deviations). Another factor to be compared is the computational cost of the optimization process. Table 5 briefly outlines the two benchmark methods utilized in this work, multiple-start gradient search, and the particle swarm optimizer (PSO).

Optimization algorithm	Performance figure						
	Circuit size A [mm²]¹	Std(A) ²	Inequality constraint		Equality constraint		CPU cost⁷
			Violation D_1 [dB] ³	Std(D_1) [dB] ⁴	Violation D_2 [dB] ⁵	Std(D_2) [dB] ⁶	
Algorithm I	378.0	59.3	4.5	4.3	0.2	0.2	$63 \times R_f$ [2.8 h]
Algorithm II	543.1	86.8	− 1.0	1.6	0.1	0.1	$1000 \times R_f$ [44.4 h]
Globalized search with dimensionality reduction (this work)	370.7	20.8	0.0	0.8	0.1	0.05	$584 \times R_f$ [25.9 h]

Table 7. Optimization results for Circuit II. ¹ Optimized footprint area of the circuit averaged over ten algorithm runs. ² Standard deviation of the optimized footprint area averaged over ten algorithm runs. ³ Violation of inequality constraint, defined as $D_1 = \max\{f \in F : \max\{|S_{11}(\mathbf{x},f)|, |S_{41}(\mathbf{x},f)|\}\} + 20$ dB, averaged over ten algorithm runs. ⁴ Standard deviation of the constraint violation D_1, averaged over ten algorithm runs. ⁵ Violation of equality constraint, defined as $D_2 = | |S_{31}(\mathbf{x},f_0)| - |S_{21}(\mathbf{x},f_0)| |$ dB, averaged over ten algorithm runs. ⁶ Standard deviation of the constraint violation D_2, averaged over ten algorithm runs. ⁷ Cost expressed in terms of equivalent number of high-fidelity EM analyzes. Numbers in brackets correspond to the running time in hours.
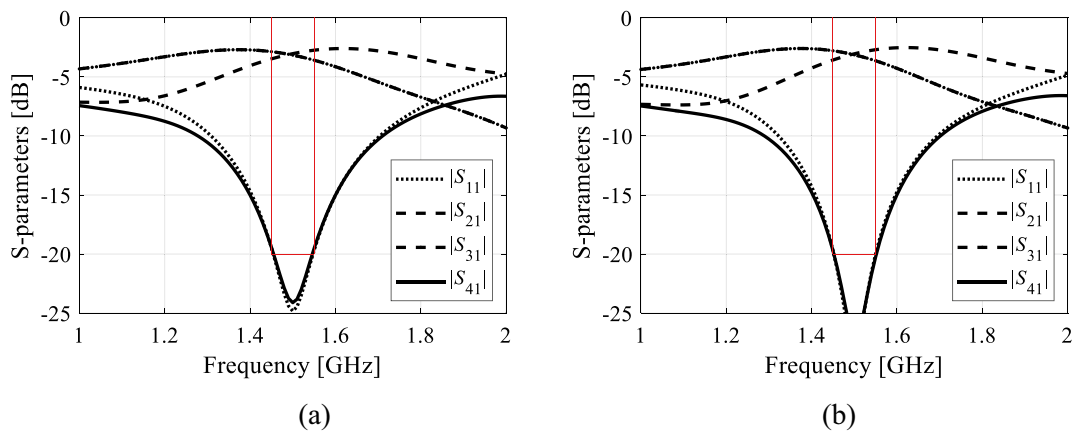

Figure 10. Circuit I: EM-simulated scattering parameters for two selected designs obtained using the proposed size reduction algorithm: (**a**) design 1 (footprint area 305.1 mm²), (**b**) design 2 (footprint area 302.4 mm²). Target operating frequency and bandwidth indicated using the vertical and horizontal lines, respectively.

Figure 11. Circuit II: EM-simulated scattering parameters for two selected designs obtained using the proposed size reduction algorithm: (**a**) design 1 (footprint area 370 mm²), (**b**) design 2 (footprint area 364 mm²). Target operating frequency and bandwidth indicated using the vertical and horizontal lines, respectively.
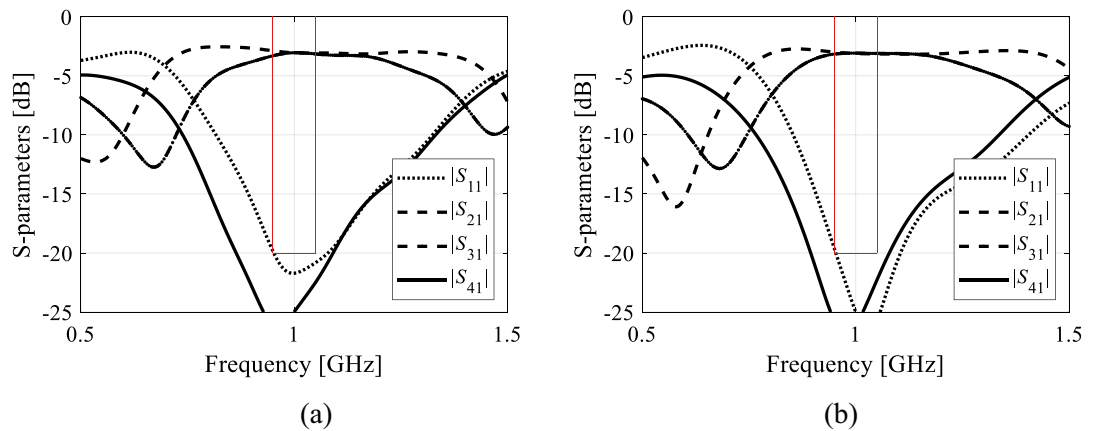
The reason for incorporating gradient search is to demonstrate multi-modality of the considered design tasks. On the other hand, PSO is employed to verify whether the proposed algorithm is capable to bring any advantages over nature-inspired procedures, both in terms of computational efficiency and design quality. Note that the computational budget of PSO has been limited to 1000 EM simulations, which is clearly insufficient from numerical perspective, yet this number can be considered borderline from the perspective of practicality: even for relatively low-cost computational models of Circuit I and II, the PSO runs take a few days each.

Numerical results. The results obtained for the proposed framework and the benchmark algorithms have been gathered in Tables 6 and 7 for Circuit I and II, respectively. Figures 10 and 11 show the circuit S-parameters at the final designs found during the selected runs of the proposed procedure. As mentioned earlier, the data contains the mean values of the circuit size, violations of the inequality and equality constraints, as well as standard deviations thereof, all computed over the ten runs of each algorithm. The mean figures can be viewed as performance metrics, whereas standard deviations quantify the repeatability of solutions.

Discussion. The performance analysis of the proposed algorithm, and the comparison with the benchmark methods will be carried out using the results contained in Tables 6 and 7. One can formulate the following observations:

- The results obtained using Algorithm I (multiple-start gradient-based optimizer) demonstrate that the considered design problems are indeed multimodal. The standard deviation of the footprint area is close to ten percent of the average area (Circuit I), and it exceeds fifteen percent (Circuit II). This means that the optimization results are highly dependent on the initial design, which—in turn—indicates the need for global search. It should also be noted that although Algorithm I produces designs that exhibit small size on the average, the constraint control is poor. In particular, a typical violation of the first constraint is around four decibels.
- The performance of nature-inspired optimization (here, using PSO) is poor. The circuit sizes achieved with Algorithm II are significantly larger than for the remaining methods with high standard deviation. Also, constraint control is inferior and inconsistent between the algorithm runs. These results are partially associated with a limited computational budged assigned for Algorithm II (1000 objective function evaluations). It appears that achieving usable results would require significantly larger budgets, probably at the level of 5000 to 10,000 EM simulations, which is not practical.
- The proposed algorithm exhibits the best consistency out of the entire benchmark set. The average circuit size is small (and comparable with Algorithm I); however, the average constraint violations are much smaller (only 0.4 dB and 0.0 dB for the first constraint, and 0.1 dB for the second constraint, on the average). At the same time, the standard deviation of the circuit area is considerable lower than for the benchmark methods: it is only about 1.3 percent (in relation to the average size) in the case of Circuit I, and only about five percent in the case of Circuit II. This corroborates truly global search capabilities of the presented method.
- Computational overhead of the presented algorithm is clearly much higher than that of local optimization, yet it is lower than for Algorithm II. As mentioned earlier, achieving reasonable results with the PSO algorithm would require increasing its computational budget by a factor five to ten, which means that the cost of the proposed algorithm can be estimated as one order of magnitude lower than for the nature-inspired methods.

The overall efficacy of the proposed size reduction procedure is superior over the benchmark. Within reasonable computational budget, the algorithm produces consistent results in terms of the circuit footprint areas with remarkably low standard deviation over the set of repetitive runs. At the same time, it exhibits excellent control of the design constraints: the average violations are around a small fraction of a decibel. Competitive computational cost is a result of employing variable-resolution EM models but also due to dimensionality reduction at the stage of constructing the surrogate model for globalized search stage of the optimization process.

Conclusion

In this work, we introduced a technique for EM-driven miniaturization of passive microwave components. The foundation of the presented methodology is parameter pre-screening and initial optimization runs (both carried out using low-fidelity simulation model), oriented towards identification of the special location of the feasible region boundary. The reduced-dimensionality surrogate model established in this region is employed to perform global size reduction, followed by gradient-based parameter tuning. The last two stages are executed using high-fidelity EM model for reliability reasons. The combination of the developed algorithmic approaches results in an optimization framework that enables globalized size reduction at low computational expenses. Comprehensive validation involving two microstrip couplers corroborates the efficacy of the proposed technique, and its superiority over local (gradient-based) parameter tuning as well as nature-inspired optimization, here, represented by the particle swarm optimization algorithm. The numerical results demonstrate global search capability, as well as consistent results, both in terms of the achieved circuit footprint, constraint control, and the computational cost. The latter is a consequence of the implemented mechanisms, i.e., dimensionality reduction and variable-fidelity EM simulations. One of the objectives of the future work will be to improve the feasible region boundary identification stage of the algorithm, as well as extending the range of applicability to include a larger variety of microwave components and antenna structures.

Data availability

References

1. Khan, M. S. *et al.* Eight-element compact UWB-MIMO/diversity antenna with WLAN band rejection for 3G/4G/5G communications. *IEEE Open J. Antenna Propag.* **1**, 196–206 (2020).
2. Zhang, J., Yan, S., Hu, X. & Vandenbosch, G. A. E. Dual-band dual-polarized wearable button array with miniaturized radiator. *IEEE Trans. Biomed. Circuits Syst.* **13**(6), 1583–1592 (2019).
3. Ma, C. *et al.* Quantifying uncertainties in passive microwave remote sensing of soil moisture via a Bayesian probabilistic inversion method. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–18 (2022).
4. Zhang, H. *et al.* A low-profile compact dual-band L-shape monopole antenna for microwave thorax monitoring. *IEEE Antenna Wirel. Propag. Lett.* **19**(3), 448–452 (2020).
5. He, Z. & Liu, C. A compact high-efficiency broadband rectifier with a wide dynamic range of input power for energy harvesting. *IEEE Microw. Wirel. Compon. Lett.* **30**(4), 433–436 (2020).
6. Kim, M. & Kim, S. Design and fabrication of 77-GHz radar absorbing materials using frequency-selective surfaces for autonomous vehicles application. *IEEE Microw. Wirel. Compon. Lett.* **29**(12), 779–782 (2019).
7. Kracek, J., Švanda, M., Mazanek, M. & Machac, J. Implantable semi-active UHF RFID tag with inductive wireless power transfer. *IEEE Antenna Wirel. Propag. Lett.* **15**, 1657–1660 (2016).
8. Ameen, M., Thummaluru, S. R. & Chaudhary, R. K. A compact multilayer triple-band circularly polarized antenna using anisotropic polarization converter. *IEEE Antenna Wirel. Propag. Lett.* **20**(2), 145–149 (2021).
9. Shirazi, M., Li, T., Huang, J. & Gong, X. A reconfigurable dual-polarization slot-ring antenna element with wide bandwidth for array applications. *IEEE Trans. Antenna Propag.* **66**(11), 5943–5954 (2018).
10. Sen, S. & Moyra, T. Compact microstrip low-pass filtering power divider with wide harmonic suppression. *IET Microw. Antenna Propag.* **13**(12), 2026–2031 (2019).
11. Chi, P.-L., Lin, H.-M. & Chien, C.-P. A tunable balanced coupler with improved phase balance and extended bandwidth. *IEEE Access* **7**, 37927–37935 (2019).
12. Zhu, F., Luo, G. Q., Liao, Z., Dai, X. W. & Wu, K. Compact dual-mode bandpass filters based on half-mode substrate-integrated waveguide cavities. *IEEE Microw. Wirel. Compon. Lett.* **31**(5), 441–444 (2021).
13. Lin, G. & Dong, Y. A compact, hybrid SIW filter with controllable transmission zeros and high selectivity. *IEEE Trans Circuit Syst. II Express Briefs* **69**(4), 2051–2055 (2022).
14. Cano, J. L., Ceccato, G., Fernandez, T., Mediavilla, A. & Perregrini, L. An ultra-compact full-band waveguide quadrature hybrid coupler. *IEEE Microw. Wirel. Compon. Lett.* **32**(1), 9–12 (2022).
15. Chen, Z., Wu, Y., Yang, Y. & Wang, W. A novel unequal lumped-element coupler with arbitrary phase differences and arbitrary impedance matching. *IEEE Trans Circuit Syst. II Express Briefs* **69**(2), 369–373 (2022).
16. Zhu, Y. & Dong, Y. A compact dual-band quasi-elliptic filter based on hybrid SIW and microstrip technologies. *IEEE Trans. Circuit Syst. II Express Briefs* **69**(3), 719–723 (2022).
17. Tang, S.-C., Chu, P.-C., Kuo, J.-T., Wu, L.-K. & Lin, C.-H. Compact microstrip wideband cross-coupled inline bandpass filters with miniaturized stepped-impedance resonators (SIRs). *IEEE Access* **10**, 21328–21335 (2022).
18. Martinez, L., Belenguer, A., Boria, V. E. & Borja, A. L. Compact folded bandpass filter in empty substrate integrated coaxial line at S-Band. *IEEE Microw. Wirel. Compon. Lett.* **29**(5), 315–317 (2019).
19. Qin, W. & Xue, Q. Elliptic response bandpass filter based on complementary CMRC. *Electr. Lett.* **49**(15), 945–947 (2013).
20. Chen, S. *et al.* A frequency synthesizer based microwave permittivity sensor using CMRC structure. *IEEE Access* **6**, 8556–8563 (2018).
21. Shen, G., Che, W. & Xue, Q. Compact microwave and millimeter-wave bandpass filters using LTCC-based hybrid lumped and distributed resonators. *IEEE Access* **7**, 104797–104809 (2019).
22. Wei, F., Guo, Y. J., Qin, P. & Shi, X. W. Compact balanced dual- and tri-band bandpass filters based on stub loaded resonators. *IEEE Microw. Wirel. Compon. Lett.* **25**(2), 76–78 (2015).
23. Liu, S. & Xu, F. Compact multilayer half mode substrate integrated waveguide 3-dB coupler. *IEEE Microw. Wirel. Compon. Lett.* **28**(7), 564–566 (2018).
24. Yang, D., Zhai, H., Guo, C. & Li, H. A compact single-layer wideband microstrip antenna with filtering performance. *IEEE Antennas Wirel. Propag. Lett.* **19**(5), 801–805 (2020).
25. Rayas-Sanchez, J. E., Koziel, S. & Bandler, J. W. Advanced RF and microwave design optimization: A journey and a vision of future trends. *IEEE J. Microw.* **1**(1), 481–493 (2021).
26. Zhang, F., Li, J., Lu, J. & Xu, C. Optimization of circular waveguide microwave sensor for gas-solid two-phase flow parameters measurement. *IEEE Sens. J.* **21**(6), 7604–7612 (2021).
27. Koziel, S., Pietrenko-Dabrowska, A. & Plotka, P. Reduced-cost microwave design closure by multi-resolution EM simulations and knowledge-based model management. *IEEE Access* **9**, 116276–116337 (2021).
28. Feng, F. *et al.* Parallel gradient-based EM optimization for microwave components using adjoint- sensitivity-based neuro-transfer function surrogate. *IEEE Trans. Microw. Theory Technol.* **68**(9), 3606–3620 (2020).
29. Han, H., Chen, C., Sun, H., Du, S. & Qiao, J. Multi-objective model predictive control with gradient eigenvector algorithm. *Inf. Sci.* **601**, 114–128 (2022).
30. Ochoa, J. S. & Cangellaris, A. C. Random-space dimensionality reduction for expedient yield estimation of passive microwave structures. *IEEE Trans. Microw. Theory Technol.* **61**(12), 4313–4321 (2013).
31. Spina, D., Ferranti, F., Antonini, G., Dhaene, T. & Knockaert, L. Efficient variability analysis of electromagnetic systems via polynomial chaos and model order reduction. *IEEE Trans. Compon. Packag. Manuf. Technol.* **4**(6), 1038–1051 (2014).
32. Koziel, S., Pietrenko-Dabrowska, A. & Al-Hasan, M. Improved-efficacy optimization of compact microwave passives by means of frequency-related regularization. *IEEE Access* **8**, 195317–195326 (2020).
33. Liu, B., Yang, H. & Lancaster, M. J. Global optimization of microwave filters based on a surrogate model-assisted evolutionary algorithm. *IEEE Trans. Microw. Theory Technol.* **65**(6), 1976–1985 (2017).
34. Pietrenko-Dabrowska, A. & Koziel, S. Globalized parametric optimization of microwave components by means of response features and inverse metamodels. *Sci. Rep.* **11**(1), 1–18 (2021).
35. Sabbagh, M. A. E., Bakr, M. H. & Bandler, J. W. Adjoint higher order sensitivities for fast full-wave optimization of microwave filters. *IEEE Trans. Microw. Theory Technol.* **54**(8), 3339–3351 (2006).
36. Koziel, S., Mosler, F., Reitzinger, S. & Thoma, P. Robust microwave design optimization using adjoint sensitivity and trust regions. *Int. J. RF Microw. CAE* **22**(1), 10–19 (2012).

37. Koziel, S. & Pietrenko-Dabrowska, A. Efficient gradient-based algorithm with numerical derivatives for expedited optimization of multi-parameter miniaturized impedance matching transformers. *Radioengineering* **28**(3), 572–578 (2019).
38. Pietrenko-Dabrowska, A. & Koziel, S. Expedited antenna optimization with numerical derivatives and gradient change tracking. *Eng. Comput.* **37**(4), 1179–1193 (2019).
39. Pietrenko-Dabrowska, A. & Koziel, S. Computationally-efficient design optimization of antennas by accelerated gradient search with sensitivity and design change monitoring. *IET Microw. Antenna Propag.* **14**(2), 165–170 (2020).
40. F. Arndt, "WASP-NET®: Recent advances in fast full 3D EM CAD of waveguide feeds and aperture antennas," *IEEE Int. Symp. Ant. Propag., APS-URSI*, Spokane, WA, pp. 2724-2727, 2011
41. Feng, F. *et al.* Coarse- and fine-mesh space mapping for EM optimization incorporating mesh deformation. *IEEE Microw. Wirel. Comput. Lett.* **29**(8), 510–512 (2019).
42. Pietrenko-Dabrowska, A., & Koziel, S. Generalized formulation of response features for reliable optimization of antenna structures. *IEEE Trans. Antenna Propag. Early View* (2021).
43. Zhang, C., Feng, F., Gongal-Reddy, V., Zhang, Q. J. & Bandler, J. W. Cognition-driven formulation of space mapping for equal-ripple optimization of microwave filters. *IEEE Trans. Microw. Theory Technol.* **63**(7), 2154–2165 (2015).
44. Koziel, S. & Ogurtsov, S. *Simulation-Based Optimization of Antenna Arrays* (World Scientific, Singapore, 2019).
45. Koziel, S. & Pietrenko-Dabrowska, A. *Performance-Driven Surrogate Modeling of High-Frequency Structures* (Springer, Cham, 2020).
46. Easum, J. A., Nagar, J., Werner, P. L. & Werner, D. H. Efficient multi-objective antenna optimization with tolerance analysis through the use of surrogate models. *IEEE Trans. Antenna Propag.* **66**(12), 6706–6715 (2018).
47. Zhang, Z., Chen, H., Yu, Y., Jiang, F. & Cheng, Q. S. Yield-constrained optimization design using polynomial chaos for microwave filters. *IEEE Access* **9**, 22408–22416 (2021).
48. Li, S., Fan, X., Laforge, P. D. & Cheng, Q. S. Surrogate model-based space mapping in postfabrication bandpass filters' tuning. *IEEE Trans. Microw. Theory Techol.* **68**(6), 2172–2182 (2020).
49. Cheng, Q. S., Rautio, J. C., Bandler, J. W. & Koziel, S. Progress in simulator-based tuning—the art of tuning space mapping. *IEEE Microw. Mag.* **11**(4), 96–110 (2010).
50. Du, J. & Roblin, C. Stochastic surrogate models of deformable antennas based on vector spherical harmonics and polynomial chaos expansions: Application to textile antennas. *IEEE Trans. Antenna Propag* **66**(7), 3610–3622 (2018).
51. Li, Q., Chu, Q., Chang, Y. & Dong, J. Tri-objective compact log-periodic dipole array antenna design using MOEA/D-GPSO. *IEEE Trans. Antenna Propag.* **68**(4), 2714–2723 (2020).
52. Tak, J., Kantemur, A., Sharma, Y. & Xin, H. A 3-D-printed W-band slotted waveguide array antenna optimized using machine learning. *IEEE Antenna Wirel. Propag. Lett.* **17**(11), 2008–2012 (2018).
53. Koziel, S., Cheng, Q. S. & Bandler, J. W. Fast EM modeling exploiting shape-preserving response prediction and space mapping. *IEEE Trans. Microw. Theory Techol.* **62**(3), 399–407 (2014).
54. Queipo, N. V. *et al.* Surrogate-based analysis and optimization. *Progag. Aerosp. Sci.* **41**(1), 1–28 (2005).
55. Leifsson, L., Du, X. & Koziel, S. Efficient yield estimation of multi-band patch antennas by polynomial chaos-based kriging. *Int. J. Numer. Model.* **33**(6), e2722 (2020).
56. Goh, P. Y., Tan, S. C., Cheah, W. P. & Lim, C. P. Adaptive rough radial basis function neural network with prototype outlier removal. *Inf. Sci.* **505**, 127–143 (2019).
57. Zhang, W., Feng, F., Jin, J. & Zhang, Q. J. Parallel multiphysics optimization for microwave devices exploiting neural network surrogate. *IEEE Microw. Wirel. Compon. Lett.* **31**(4), 341–344 (2021).
58. Feng, F. *et al.* Multifeature-assisted neuro-transfer function surrogate-based EM optimization exploiting trust-region algorithms for microwave filter design. *IEEE Trans. Microw. Theory Technol.* **68**(2), 531–542 (2020).
59. Egrioglu, E., Baş, E. & Chen, M.-Y. Recurrent dendritic neuron model artificial neural network for time series forecasting. *Inf. Sci.* **607**, 572–584 (2022).
60. Cai, J., King, J., Yu, C., Liu, J. & Sun, L. Support vector regression-based behavioral modeling technique for RF power transistors. *IEEE Microw. Wirel. Compon. Lett.* **28**(5), 428–430 (2018).
61. Jiang, P., Cheng, Y., Yi, J. & Liu, J. An efficient constrained global optimization algorithm with a clustering-assisted multiobjective infill criterion using Gaussian process regression for expensive problems. *Inf. Sci.* **569**, 728–745 (2021).
62. Petrocchi, A. *et al.* Measurement uncertainty propagation in transistor model parameters via polynomial chaos expansion. *IEEE Microw. Wirel. Compon. Lett.* **27**(6), 572–574 (2017).
63. Gustrau, F. *RF and Microwave Engineering. Fundamentals of Wireless Communications* (John Wiley & Sons, Hoboken, NJ, USA, 2012).
64. Pietrenko-Dabrowska, A. & Koziel, S. Nested kriging with variable domain thickness for rapid surrogate modeling and design optimization of antennas. *Electronics* **9**(10), 1621 (2020).
65. Bandler, J. W., Koziel, S. & Madsen, K. Space mapping for engineering optimization. *SIAG/Optim. Views News Spec. Issue Surrog. Deriv. Free Optim.* **17**(1), 19–26 (2006).
66. Koziel, S., Bandler, J. W. & Madsen, K. Space mapping with adaptive response correction for microwave design optimization. *IEEE Trans. Microw. Theory Techol.* **57**(2), 478–486 (2009).
67. Koziel, S. & Unnsteinsson, S. D. Expedited design closure of antennas by means of trust-region-based adaptive response scaling. *IEEE Antennas Wirel. Propog. Lett.* **17**(6), 1099–1103 (2018).
68. Su, Y., Li, J., Fan, Z. & Chen, R. Shaping optimization of double reflector antenna based on manifold mapping. In *Int. Appl. Compon. Electromagn. Soc. Symp. (ACES)* 1–2 (Suzhou, China, 2017).
69. Koziel, S. & Leifsson, L. *Simulation-Driven Design by Knowledge-Based Response Correction Techniques* (Springer, Cham, 2016).
70. Jiao, R., Zeng, S., Li, C., Jiang, Y. & Jin, Y. A complete expected improvement criterion for Gaussian process assisted highly constrained expensive optimization. *Inf. Sci.* **471**, 80–96 (2019).
71. Jiao, R., Xue, B. & Zhang, M. Investigating the correlation amongst the objective and constraints in Gaussian process-assisted highly-constrained expensive optimization. *IEEE Trans. Evol. Comput.* https://doi.org/10.1109/TEVC.2021.3120980 (2021).
72. Tang, D. & Luo, X. Compact filtering balun with wide stopband and low radiation loss using hybrid microstrip and substrate-integrated defected ground structure. *IEEE Microw. Wirel. Compon. Lett.* **31**(6), 549–552 (2021).
73. Pan, B. C., Yu, P., Liao, Z., Zhu, F. & Luo, G. Q. A compact filtering power divider based on spoof surface plasmon polaritons and substrate integrated waveguide. *IEEE Microw. Wirel. Compon. Lett.* **32**(2), 101–104 (2022).
74. Chen, C. A compact wideband endfire filtering antenna inspired by a uniplanar microstrip antenna. *IEEE Antenna Wirel. Propag. Lett.* **21**(4), 853–857 (2022).
75. Haq, M. A., Koziel, S. & Cheng, Q. S. Miniaturization of wideband antennas by means of feed line topology alterations. *IET Microw. Antenna Propag.* **12**(13), 2128–2134 (2018).
76. Johanesson, D. O. & Koziel, S. "Feasible space boundary search for improved optimization-based miniaturization of antenna structures. *IET Microw. Antenna Propag.* **12**(8), 1273–1278 (2018).
77. Ullah, U., Koziel, S. & Mabrouk, I. B. Rapid re-design and bandwidth/size trade-offs for compact wideband circular polarization antennas using inverse surrogates and fast EM-based parameter tuning. *IEEE Trans. Antenna Propag.* **68**(1), 81–89 (2019).
78. Koziel, S. & Pietrenko-Dabrowska, A. Reliable EM-driven size reduction of antenna structures by means of adaptive penalty factors. *IEEE Trans. Antenna Propag.* **70**(2), 1380–1401 (2021).

79. Mahrokh, M. & Koziel, S. Improved-efficacy EM-based antenna miniaturization by multi-fidelity simulations and objective function adaptation. *Energies* **15**(2), 403 (2021).
80. Mahrokh, M. & Koziel, S. Explicit size-reduction of circularly polarized antennas through constrained optimization with penalty factor adaptation. *IEEE Access* **9**, 132390–132396 (2021).
81. Koziel, S. & Pietrenko-Dabrowska, A. On EM-driven size reduction of antenna structures with explicit constraint handling. *IEEE Access* **9**, 165766–165772 (2021).
82. Koziel, S., Pietrenko-Dabrowska, A. & Mahrokh, M. On decision-making strategies for improved-reliability size reduction of microwave passives: Intermittent correction of equality constraints and adaptive handling of inequality constraints. *Knowl. Based Syst.* (2022).
83. Jiao, R., Sun, Y., Sun, J., Jiang, Y. & Zeng, S. Antenna design using dynamic multi-objective evolutionary algorithm. *IET Microw. Antennas Propag.* **12**, 2065–2072 (2018).
84. Xu, Q., Zeng, S., Zhao, F., Jiao, R. & Li, C. On formulating and designing antenna arrays by evolutionary algorithms. *IEEE Trans. Antennas Propag.* **69**(2), 1118–1129 (2021).
85. Wang, Y., Ma, K. & Mou, S. A compact branch-line coupler using substrate integrated suspended line technology. *IEEE Microw. Wirel. Compon. Lett.* **26**(2), 95–97 (2016).
86. Hassona, A., Vassilev, V., Zaman, A. U., Belitsky, V. & Zirath, H. Compact low-loss chip-to-waveguide and chip-to-chip packaging concept using EBG structures. *IEEE Microw. Wirel. Compon. Lett.* **31**(1), 9–12 (2021).
87. Zhang, W., Shen, Z., Xu, K. & Shi, J. A compact wideband phase shifter using slotted substrate integrated waveguide. *IEEE Microw. Wirel. Compon. Lett.* **29**(12), 767–770 (2019).
88. Sharma, A. Nature inspired algorithms with randomized hypercomputational perspective. *Inf. Sci.* **608**, 670–695 (2022).
89. Li, H. *et al.* Newly emerging nature-inspired optimization - algorithm review, unified framework, evaluation, and behavioural parameter optimization. *IEEE Access* **8**, 72620–72649 (2020).
90. Kurgan, P. & Koziel, S. Selection of circuit geometry for miniaturized microwave components based on concurrent optimization of performance and layout area. *AEU Int. J. Electr. Commun.* **108**, 287–294 (2019).
91. Yang, S. H. & Kiang, J. F. Optimization of sparse linear arrays using harmony search algorithms. *IEEE Trans. Antenna Propag.* **63**(11), 4732–4738 (2015).
92. Lv, Z., Wang, L., Han, Z., Zhao, J. & Wang, W. Surrogate-assisted particle swarm optimization algorithm with Pareto active learning for expensive multi-objective optimization. *IEEE/CAA J. Autom. Sin.* **6**(3), 838–849 (2019).
93. Jolliffe, I. T. *Principal Component Analysis* 2nd edn. (Springer, Cham, 2002).
94. Conn, A. R., Gould, N.I . M. & Toint, P.L. *Trust Region Methods*. (MPS-SIAM Series on Optimization, 2000).
95. Levy, H. & Lessman, F. *Finite Difference Equations* (Dover Publications Inc., Mineola, 1992).
96. Koziel, S. & Pietrenko-Dabrowska, A. Fast multi-objective optimization of antenna structures by means of data-driven surrogates and dimensionality reduction. *IEEE Access* **8**, 183300–183311 (2020).
97. Beachkofski, B. & Grandhi, R. Improved distributed hypercube sampling. In *American Institute of Aeronautics and Astronautics*. 2002–1274 (AIAA ,2002).
98. Koziel, S. & Pietrenko-Dabrowska, A. Reduced-cost surrogate modeling of compact microwave components by two-level kriging interpolation. *Eng. Opt.* **52**(6), 960–972 (2019).
99. Tseng, C. & Chang, C. A rigorous design methodology for compact planar branch-line and rat-race couplers with asymmetrical T-structures. *IEEE Trans. Microw. Theory Technol.* **60**(7), 2085–2092 (2012).
100. Kennedy, J. & Eberhart, R. C. *Swarm Intelligence* (Morgan Kaufmann, Burlington, 2001).

Acknowledgements

Author contributions

Conceptualization, S.K. and A.P.; methodology, S.K. and A.P.; software, S.K. and A.P.; validation, S.K., A.P. and M.M.; formal analysis, S.K.; investigation, S.K. and A.P.; resources, S.K.; data curation, S.K., A.P. and M.M.; writing—original draft preparation S.K. and A.P.; writing—review and editing, S.K. and A.P.; visualization, S.K. and A.P.; supervision, S.K.; project administration, S.K.; funding acquisition, S.K All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.P.-D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

8.1 Dimensionality Reduction Using Principal Component Analysis

This section provides a detailed description of the surrogate model dimensionality reduction using principal component analysis (PCA) as well as how the dominant dimensions are determined.

Consistent with the notation used in Paper # 5, the vectors $x_c^{(j)}$, $j = 1, ..., N_3$ includes the designs resulted from the circuit optimization for minimum size. They all feature low constraint violations based on the objective function definition used in the optimization process. Therefore, they are located in the vicinity of the feasible region boundary X_b. Using this fact, we determine the domain of the surrogate model based on a spectral analysis of the set $\{x_c^{(j)}\}$.

Let us denote

$$x_m = N_3^{-1} \sum j = 1,...,N_3 x_c^{(j)} \tag{8.1}$$

as the center of gravity of the set $\{x_c^{(j)}\}$. The covariance matrix of the set is defined as

$$S_c = \frac{1}{N_3 - 1} \sum_{j=1}^{N_3} \left(x_c^{(j)} - x_m \right) \left(x_c^{(j)} - x_m \right)^T, N_3 > n \tag{8.2}$$

Let us also denote the eigenvectors and eigenvalues of the set $\{x_c^{(j)}\}$ as v_k and λ_k, $k = 1, ..., n$. The matrix $V_p = [v_1, ..., v_p]$ contains the first p eigenvectors as columns, where p is the selected number of the principal components. The surrogate model domain X_p will be constructed using the information contained in V_p. Given a fast decrease as the typical behavior of the eigenvalues, it is usually sufficient to use $p = 3$ or 4 without the risk of missing important information. Defining X_p requires an expansion of the design vectors $\{x_c^{(j)}\}$ using the principal components

$$x_c^{(j)} = \sum_{k=1}^{p} b_{jk} v_k \tag{8.3}$$

The coefficients b_{jk} are calculated as follows

$$b_{jk} = v_k \left[v_k^T (x_c^j - x_m) \right] \tag{8.4}$$

Using the following notation

$$b_{k.max} = \max \left\{ j = b_{jk} \right\}, b_{k.min} = \min \left\{ j = b_{jk} \right\} \tag{8.5}$$

$$b_{k.0} = \frac{b_{k.min} + b_{k.max}}{2}, j = 1,...,p \tag{8.6}$$

$$b_0 = \left[b_{1.0},...,b_{p.0} \right] \tag{8.7}$$

$$\lambda_b = \left[\lambda_{b.1},...,\lambda_{b.p} \right]^T \tag{8.8}$$

$$\lambda_{bk} = (b_{k.max} - b_{k.min}) / 2 \tag{8.9}$$

The center point is defined as

$$x_c = x_m + Vb_0 \tag{8.10}$$

The surrogate model domain, X_p, can be therefore defined as

$$X_p = \begin{cases} x = x_c + \sum_{k=1}^{p} b_{jk} v_k (2\lambda_k - 1)\lambda_{bk} v_k \\ 0 \le \lambda_k \le 1, \, j = 1,...,p \end{cases} \tag{8.11}$$

Chapter 9

9   Summary of Findings

This chapter represents a brief discussion on the findings of the thesis and outlines potential future directions that might originate from the work carried out so far. The latter includes a number of open problems related to the automated design optimization and miniaturization of high-frequency circuit structures.

9.1   Conclusion

The main focus of this work was to develop cost-efficient and reliable algorithms for optimization-based size reduction of high-frequency circuits which are intended to contribute to the state of the art of high-frequency CAD. The obtained numerical results demonstrate that the goals of this thesis have been accomplished, and the proposed procedures have been positively verified.

As demonstrated, the developed solutions in the form of an appropriate adjustment of the penalty functions and proper treatment of the equality constraints, as well as the employment of fast-to-evaluate surrogate models and variable-fidelity simulations, allow for improving the reliability and efficiency of simulated-driven size reduction of high-frequency structures. Furthermore, a variety of high-frequency benchmark structures that have been employed as verification case studies validate the performance efficacy of the proposed algorithmic methodologies as compared to the existing size reduction algorithms.

It should be emphasized that the simulation-driven size reduction algorithms developed in this study offer several advantages over the conventional procedures. These include superior performance in terms of the obtainable size reduction rates, precise control over the design constraints (both equality or inequality type), time and computational efficiency, and low-cost global search capability. Furthermore, the developed size reduction algorithms permitted to obtain significant miniaturization rates of high-frequency circuit components of various types, namely branch-line and rat-race couplers, narrowband CP antennas, and UWB antennas.

9.2   Future Directions

The research conducted under this project can be expanded in many different directions. Some of potential extensions include:

1. Widening the range of applicability of the methods developed so far. A representative example would be origami-based antennas, which can be optimized

for minimum volume with degrees of freedom provided by the configuration of the origami folds and fold angles in addition to those of the geometry/material parameters. Another example is optimal design of wideband electromagnetic field (EMF) probes with a travelling-wave antenna as the EM sensor. Given the flatness of the reflection coefficient response as one of the critical EMF probe performance figures, the constraints could include maintaining the target level of the reflection coefficient as well as minimum variability thereof over the entire desired operating band.

2. An enhancement to the proposed correction-based treatment of the equality constraints. Here, one of potential options would be to expand the utility of the algorithm to more complex high-frequency circuit elements with multiple equality constraints. These include but are not limited to power coupler/divider structures with equality constraints on their phase shift, and power split ratio.

3. Incorporation of acceleration mechanisms into the optimization procedures. One of such mechanisms, already implemented, was the employment of variable-resolution EM simulations. Others include restricted sensitivity updating schemes, as well as utilization of response features.

Bibliography

[1]   T. T. Le and T. -Y. Yun, "Miniaturization of a dual band wearable antenna for dual-band WBAN applications," *IEEE Trans. Antennas Propag,* vol. 19, no. 8, pp. 1452-1456, 2020.

[2]   S. Agneessens and H. Rogier, "Compact half diamond dual-band textile HMSIW on-body antenna," *IEEE Trans. Antennas Propag,* vol. 62, no. 5, pp. 2374-2381, 2014.

[3]   M. Kim and S. Kim, "Design and fabrication of 77-GHz radar absorbing materials using frequency-selective surfaces for autonomous vehicles application," *IEEE Microwave Wireless Comp. Lett,* vol. 29, no. 12, pp. 779-782, 2019.

[4]   M. S. Khan, A. Iftikhar, R. M. Shubair, A. Capobianco, B. D. Braaten and D. E. Anagnostou, "Eight-element compact UWB-MIMO/diversity antenna with WLAN band rejection for 3G/4G/5G communications," *IEEE Open J. Ant. Propag,* vol. 1, pp. 196-206, 2020.

[5]   F. Amin, Y. Liu, Y. Zhao and S. Hu, "Compact and low-loss phase shifters and multibit phase shifters based on inverted-E topology," *IEEE Trans. Microw. Theory Techn.,* vol. 69, no. 4, pp. 2120-2129, 2021.

[6]   I. Piekarz, J. Sorocki, R. Smolars, S. Gruszczynski and K. Wincza, "Four-node antenna feeding network for interfacing with differential front-end electronics," *IEEE Access,* vol. 9, pp. 103728-103736, 2021.

[7]   J. -G. Chi and Y. -J. Kom, "A compact wideband millimeter-wave quadrature hybrid coupler using artificial transmission lines on a glass substrate," *IEEE Microw. Wirel. Compon. Lett,* vol. 30, no. 11, pp. 1037-1040, 2020.

[8]   C. W. Byeon and C. S. Park, "Low-loss compact millimeter-wave power divider/combiner for phased array systems," *IEEE Microw. Wirel. Compon. Lett,* vol. 29, pp. 312-314, 2019.

[9]   J. Oh and K. Sarabandi, "A topology-based miniaturization of circularly polarized patch antennas," *IEEE Trans. Antennas Propag,* vol. 61, pp. 1422-1426, 2013.

[10]   W. Zhang, Z. Shen, K. Xu and J. Shi, "A compact wideband phase shifter using slotted substrate integrated waveguide," *IEEE Microw. Wirel. Comp. Lett.,* vol. 29, no. 12, pp. 767-770, 2019.

[11]   Y. Peng and L. Sun, "A Compact Broadband Phase Shifter Based on HMSIW Evanescent Mode," *IEEE Microwave and Wireless Components Letters,* vol. 31, no. 7, pp. 857-860, 2021.

[12]   W. He, Y. He, Y. Li, S. -W. Wong and L. Zhu, "A compact ultrawideband circularly polarized antenna array with shared partial patches," *IEEE Antennas Wireless Propag. Lett.,* vol. 20, no. 12, pp. 2280-2284, 2021.

[13]   S. Koziel and A. Pietrenki-Dabrowska, "Performance-based nested surrogate modeling of antenna input characteristics," *IEEE Trans. Antennas Propag.,* vol. 67, no. 5, pp. 2904-2912, 2019.

[14]   Y. Song, Q. S. Cheng and S. Koziel, "Multi-fidelity local surrogate model for computationally efficient microwave component design optimization," *Sensors,* vol. 19, p. 3023, 2019.

[15]   A. A. Al-Azza and A. Al-Jodah, "A Spider monkey optimization: a novel technique for antenna optimization," *IEEE Antennas Wireless Propag. Lett.,* vol. 15, pp. 1016-1019, 2016.

[16]   A. Lalbakhsh, M. U. Afzal and K. P. Esselle, "Multiobjective particle swarm optimization to design a time-delay equalizer metasurface for an electromagnetic band-gap resonator antenna," *IEEE Antennas Wireless Propag. Lett.,* vol. 16, pp. 912-915, 2017.

[17]   S. K. Goudos, K. Siakavara, T. Samaras, E. E. Vafiadis and J. N. Sahalos, "Self-adaptive differential evolution applied to real-valued antenna and microwave design problems," *IEEE Trans. Antennas Propag,* vol. 59, pp. 1286-1298, 2011.

[18]   S. Koziel and A. Bekasiewicz, "A structure and simulation-driven design of compact CPW-fed UWB antenna," *IEEE Antennas Wireless Propag. Lett.,* vol. 15, pp. 750-753, 2016.

[19]   D. O. Johanesson and S. Koziel, "EM-driven constrained miniaturization of antennas using adaptive in-band reflection acceptance threshold," *Int. J. Numerical Modeling,* vol. 32, no. 2, p. e2513, 2019.

[20]   S. Koziel, "Objective Relaxation Algorithm for Reliable Simulation-Driven Size Reduction of Antenna Structures," *IEEE Antennas Wireless Propag. Lett.,* vol. 16, pp. 1949-1952, 2017.

[21] A. Bekasiewicz and S. Koziel, "Structure and computationally efficient simulation-driven design of compact UWB monopole antenna," *IEEE Antennas Wireless Propag. Lett.,* vol. 14, pp. 1282-1285, 2015.

[22] A. Bekasiewicz and S. Koziel, "Structure and EM-driven design of novelcompact UWB slot antenna," *IET Microw. Antennas Propag.,* vol. 11, no. 2, pp. 219-223, 2017.

[23] Y. Song, Q. S. Cheng and S. Koziel, "Multi-fidelity local surrogate model for computationally efficient microwave component design optimization," *Sensors,* vol. 19, no. 13, p. 3023, 2019.

[24] S. Director and R. Rohrer, "The generalized adjoint network and network sensitivities," *IEEE Trans. Circuit Theory,* vol. 16, pp. 318-323, 1969.

[25] B. Tessema and G. G. Yen, "A self adaptive penalty function based algorithm for constrained optimization," in *IEEE International Conference on Evolutionary Computation*, Vancouver, BC, Canada, 2006.

[26] F. Feng, W. Na, W. Liu, S. Yan, L. Zhu and Q. -J. Zhang, "Parallel gradient-based EM optimization for microwave components using adjoint- sensitivity-based neuro-transfer function surrogate," *IEEE Trans. Microwave Theory Techn.,* vol. 68, no. 9, pp. 3606-3620, 2020.

[27] M. A. E. Sabbagh, M. H. Bakr and J. W. Bandler, "Adjoint higher order sensitivities for fast full-wave optimization of microwave filters," *IEEE Trans. Microwave Theory Techn,* vol. 54, no. 8, pp. 3339-3351, 2006.

[28] B. Liu, H. Yang and M. J. Lancaster, "Global optimization of microwave filters based on a surrogate model-assisted evolutionary algorithm," *IEEE Trans. Microwave Theory Techn. ,* vol. 65, no. 6, pp. 1976-1985, 2017.

[29] H. M. Torun and M. Swaminathan, "High-dimensional global optimization method for high-frequency electronic design," *IEEE Trans. Microwave Theory Techn.,* vol. 67, no. 6, pp. 2128-2142, 2019.

[30] J. Huang, W. Li, Y. He, L. Zhang and S. -W. Wong, "Optimization of antenna design using the artificial neural network and the simulated annealing algorithm," in *2021 Computing, Communications and IoT Applicaitons (ComComAp)*, Shenzhen, China, 2021.

[31] C. Hu, S. Zeng, Y. Jiang, J. Sun, Y. Sun and S. Gao, "A robust technique without additional computational cost in evolutionary antenna optimization," *IEEE Trans. Antennas Propag.,* vol. 67, no. 4, pp. 2252-2259, 2019.

[32] J. Bi, H. Yuan, S. Duanmu, M. Zgou and A. Abusorrah, "Energy-optimized partial computation offloading in mobile-edge computing with genetic simulated-annealing-based particle swarm optimization," *IEEE Internet of Things Journal,* vol. 8, no. 5, pp. 3774-3785, 2021.

[33] Z. Lv, L. Wang, Z. Han, J. Zhao and W. Wang, "Surrogate-assisted particle swarm optimization algorithm with Pareto active learning for expensive multi-objective optimization," *IEEE/CAA J. Automatica Sinica,* vol. 6, no. 3, pp. 838-849, 2019.

[34] A. Pietrenko-Dabrowska and S. Koziel, "Accelerated multiobjective design of miniaturized microwave components by means of nested kriging surrogates," *Int. J. RF Microwave CAE,* vol. 30, no. 4, p. e22124, 2020.

[35] J. A. Easum, J. Nagar, P. L. Werner and D. H. Werner, "Efficient multiobjective antenna optimization with tolerance analysis through the use of surrogate models," *IEEE Trans. Antennas Propag.,* vol. 66, no. 12, pp. 6706-6715, 2018.

[36] J. Tak, A. Kantemur, Y. Sharma and H. Xin, "A 3-D-printed W-band slotted waveguide array antenna optimized using machine learning," *IEEE Ant. Wireless Prop. Lett.,* vol. 17, no. 11, pp. 2008-2012, 2018.

[37] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidynathan and P. K. Tucker, "Surrogate based analysis and optimization," *Progress in Aerospace Sciences,* vol. 41, no. 1, pp. 1-28, 2005.

[38] L. Leifsson, X. Du and S. Koziel, "Efficient yield estimation of multi-band patch antennas by polynomial chaos-based kriging," *Int. J. Numerical Modeling,* vol. 33, no. 6, p. e2722, 2020.

[39] Q. Zhou, Y. Wang, P. Jiang, X. Shao, S. -K. Choi, J. Hu, L. Cao and X. Meng, "An active learning radial basis function modeling method based on self-organization maps for simulation-based design problems," *Knowledge-Based Systems,* vol. 131, pp. 10-27, 2017.

[40] W. Zhang, F. Feng, J. Jin and Q. J. Zhang, "Parallel multiphysics optimization for microwave devices exploiting neural network surrogate," *IEEE Microwave Wireless Comp. Lett.,* vol. 31, no. 4, pp. 341-344, 2021.

[41] F. Feng, W. Na, W. Liu, S. Yan, L. Zhu, J. Ma and Q. J. Zhang, "Multifeature-assisted neuro-transfer function surrogate-based EM optimization exploiting trust-region algorithms for microwave filter design," *IEEE Trans. Microwave Theory Techn.,* vol. 68, no. 2, pp. 531-542, 2020.

[42] D. Ki, M. Kim and W. Kim, "Wafer edge yield prediction using a combined long short-term memory and feed- forward neural network model for semiconductor manufacturing," *IEEE Access,* vol. 8, pp. 215125-215132, 2020.

[43] Cai, J. King, C. Yu, J. Liu and L. Sun, "Support vector regression-based behavioral modeling technique for RF power transistors," *IEEE Microwave Wireless Comp. Lett.,* vol. 28, no. 5, pp. 428-430, 2018.

[44] J. P. Jacobs, "Characterization by Gaussian processes of finite substrate size effects on gain patterns of microstrip antennas," *IET Microwaves Ant. Prop,* vol. 10, no. 11, pp. 1189-1195, 2016.

[45] A. Petrocchi, A. Kaintura, G. Avolio, D. Spina, T. Dhaene, A. Raffo and D. M. P. -P. Schreurs, "Measurement uncertainty propagation in transistor model parameters via polynomial chaos expansion," *IEEE Microwave Wireless Comp. Lett.,* vol. 27, no. 6, pp. 572-574, 2017.

[46] L. S. Kalantari and M. H. Bakr, "Wideband cloaking of objects with arbitrary shapes exploiting adjoint sensitivities," *IEEE Trans. Antennas Propag.,* vol. 64, no. 5, pp. 1963-1968, 2016.

[47] O. Paronneau, Optimal Shape Design for Elliptic Systems, Berlin/Heidelberg, Germany: Springer, 1982.

[48] A. Jameson, "Aerodynamic design via control theory," *J. Sci. Comput,* vol. 3, no. 3, pp. 233-260, 1988.

[49] M. A. E. Sabbagh, M. H. Bakr and N. K. Nilolova, "Sensitivity analysis of the scattering parameters of microwave filters using the adjoint network method," *Int. J. RF Microw. CAE,* vol. 16, no. 6, pp. 569-606, 2006.

[50] D. Papadimitrou and K. Giannakoglou, "Aerodynamic shape optimization using first and second order adjoint and direct approaches," *Arch. Comput. Methods Eng,* vol. 15, pp. 447-488, 2008.

[51] J. I. Toivann, R. A. E. Mäkinen, S. Järvenpää, P. Ylä-Oijala and J. Rahola, "Electromagnetic sensitivity analysis and shape optimization using method of moments and automatic differentiation," *IEEE Trans. Antennas Propag.,* vol. 57, pp. 168-175, 2009.

[52] A. Pietrenko-Dabrowska and S. Koziel, "Fast EM-driven parameter tuning of microwave circuits with sparxed sensitivity updates via principal directions, Knowledge-Based Systems," *Knowledge-Based Systems,* vol. 252, 2022.

[53] M. Diehl and A. Walther, "An adjoint-based SQP algorithm with quasi-Newton Jacobin updates for iequality constrained optimization," *Opt. Methods Softw.,* vol. 25, pp. 531-552, 2009.

[54] J. A. Tomasson, S. Koziel and A. P. Dabrowska, "Quasi-global optimization of antenna structures using principal components and affine subspace-spanned surrogates," *IEEE Access,* pp. 50078-50084, 2020.

[55] S. Koziel and A. Pietrenko-Dabrowska, "Expedited feature-based quasi-global optimization of multi-band antenna input characteristics with jacobian variability tracking," *IEEE Access,* vol. 8, pp. 83907-83915, 2020.

[56] S. Koziel and A. Pietrenko-Dabrowska, "Reduced-cost electromagnetic-driven optimization of antenna structures by means of trust-region gradient-search with sparse Jacobian updates," *IET Microw. Ant. Propag.,* vol. 13, pp. 1646-1652, 2019.

[57] S. Koziel and S. Ogurtsov, "Robust multi-fidelity simulation-driven design optimization of microwave structures," in *IEEE MTT-S International Microwave Symposium*, 2010.

[58] S. Koziel, "Multi-Fidelity Multi-Grid Design Optimization of Planar Microwave Structures with Sonnet," in *26th Annual Review of Progress in Applied Computational Electromagnetics*, Tampere, Finland, 2010.

[59] S. Koziel and S. Ogurtsov, "Model management for cost-efficient surrogate-based optimisation of antennas using variable-fidelity electromagnetic simulations," *IET Microw. Antennas Propag.,* vol. 6, no. 15, pp. 1643-1650, 2012.

[60] A. Bekasiewicz and S. Koziel, "Efficient multi-fidelity design optimization of microwave filters using adjoint sensitivity," *Int. J.RF Microwave CAE,* vol. 25, no. 2, pp. 178-183, 2015.

[61] S. Koziel and A. Bekasiewicz, "EM-simulation-driven design optimization of compact microwave structures using multi-fidelity simulation models and adjoint sensitivities," *Int. J. RF Microwave CAE,* vol. 26, no. 5, pp. 442-448, 2016.

[62] Y. Song, Q. S. Cheng and S. Koziel, "Multi-fidelity local surrogate model for computationally efficient microwave component design optimization," *Sensors,* vol. 19, p. 3023, 2019.

[63] S. Koziel and S. Ogurtsov, "Computational-budget-driven automated microwave design optimization using variable-fidelity electromagnetic simulations," *Int. J. RF Microwave CAE,* vol. 23, no. 3, pp. 349-356, 2013.

[64] S. Koziel and A. Bekasiewicz, "Low-fidelity model considerations for EM-driven design of antenna structures," *J. Electromag. Waves Applic.,* vol. 30, no. 18, pp. 2444-2458, 2016.

[65] S. Koziel, S. D. Unnsteinsson and A. Bekasiewicz, "Low-fidelity model considerations for simulation-based optimisation of miniaturised wideband antennas," *IET Microw. Antennas Propag.,* vol. 12, no. 10, pp. 1613-1619, 2018.

[66] L. Leifsson, S. Koziel and P. Kurgan, "Automated Low-Fidelity Model Setup for Surrogate-Based Aerodynamic Optimization," in *Solving Computationally Expensive Electromagnetic Problems*, vol. 97, Springer, Cham, 2014.

[67] S. Koziel and S. Ogurtsov, " Multi-level design optimization of microwave structures with automated model fidelity adjustment," in *IEEE MTT-S International Microwave Symposium Digest (MTT)*, Seattle, WA, USA, 2013.

[68] "Talaria Two Modules," InnoPhase Inc, [Online]. Available: https://innophaseinc.com/talaria-two-modules/. [Accessed 29 August 2022].

[69] "Electronics lab," [Online]. Available: https://www.electronics-lab.com/innophase-releases-inp1010-inp1011-talaria-two-modules/. [Accessed 29 August 2022].

[70] D. D. Battista, S. G. Fabri, M. K. Bugeja and M. A. Azzopardi, "PocketQube pico-satellite attitude control: implementation and testing," *IEEE J. Miniat. Air Space Systems,* vol. 1, no. 2, pp. 90-102, 2020.

[71] R. Li, B. Li, G. Du, X. Sun and H. Sun, "A compact broadband antenna with dual-resonance for implantable devices," *Micromachines,* vol. 10, no. 1, p. 59, 2019.

[72] C. Song, P. Lu and S. Shen, "Highly efficient omnidirectional integrated multiband wireless energy harvesters for compact sensor nodes of internet-of-things," *IEEE Trans. industrial Electr.,* vol. 68, no. 9, pp. 8128-8140, 2021.

[73] M. Wagih, Y. Wei, A. Komolafe, R. Torah and S. Beeby, "Reliable UHF long-range textile-integrated RFID tag based on a compact flexible antenna filament," *MDPI: Sensors,* vol. 20, no. 12, p. 3435, 2020.

[74] A. Desai, T. Upadhyaya and R. Patel, "Compact wideband transparent antenna for 5G communicaiton systems," *Microwave Optical Techn. Lett.,* vol. 61, no. 3, pp. 781-786, 2019.

[75] G. -L. Huang, C. -Z. Han, W. Xu, T. Yuan and X. Zhang, "A compact 16-way high-power combiner implemented via 3-D metal printing technique for advanced radio-frequency electronics system applications," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 6, pp. 4767-4776, 2018.

[76] M. Kumar, S. N. Islam, G. Sen and S. K. Parui, "Design of compact Wilkonson power divider and branch line coupler using hairpin based line," *AEU - Int. J. Electronics Comm.,* vol. 110, p. 152825, 2019.

[77] C. Li, X. -W. Zhu, P. Liu, C. Yu and W. Hong, "A metasurface-based multilayer wideband circularly polarized patch antenna array with a parallel feeding network for Q-Band," *IEEE Antennas and Wireless Propagation Letters,* vol. 18, no. 6, pp. 1208-1212, 2019.

[78] J. Shi, Y. Nie, P. Han, W. Zhang and Q. Cao, "Compact filtering phase shifter with simple structure," *IEEE Microwave Wireless Comp. Lett.,* vol. 31, no. 12, pp. 1263-1266, 2021.

[79] A. S. Dixit, S. Kumar, S. Urooj and A. Malibari, "A highly compact antipodal vivaldi antenna array for 5G millimeter wave applications," *Sensors,* vol. 21, no. 7, p. 2360, 2021.

[80] C. Chen, "A compact wideband endfire filtering antenna inspired by a uniplanar microstrip antenna," *IEEE Antennas and Wireless Propagation Letters,* vol. 21, no. 4, pp. 853-857, 2022.

[81] R. Lu, C. Yu, Y. Zhu and W. Hong, "Compact millimeter-wave endfire dual-polarized antenna array for low-cost multibeam applications," *IEEE Antennas and Wireless Propagation Letters,* vol. 19, no. 12, pp. 2526-2530, 2020.

[82] K. Xue, D. Yang, C. Guo, H. Zhai, H. Li and Y. Zeng, "A dual-polarized filtering base-station antenna with compact size for 5G applications," *IEEE Antennas and Wireless Propagation Letters,* vol. 19, no. 8, pp. 1316-1320, 2020.

[83] A. Sieganschin, B. Tegowski and A. F. Jacob, "A Compact Filter With Dual-Mode Folded Circular SIW Cavities," in *51st European Microwave Conference (EuMC)*, London, United Kingdom, 2021.

[84] Z. Liu and R. M. Weikle, "A compact quadrature coupler based on coupled artificial transmission lines," *IEEE Microwave Wireless Comp. Lett.,* vol. 15, no. 12, pp. 889-891, 2005.

[85] J. Coromina, P. Vélez, J. Bonache and F. Martín, "Branch line couplers with small size and harmonic suppression based on non-periodic step impedance shunt stub (SISS) loaded lines," *IEEE Access,* vol. 8, pp. 67310-67320, 2020.

[86] A. G. N. T. P. Conn, "A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds," *SIAM Journal,* vol. 28, no. 2, p. 545–572, 1991.

[87] J. Nocedal and S. Wright, Numerical Optimization, Berlin: Springer, 2006.

[88] N. I. M. G. P. L. T. A. R. Conn, Heidelber: Computational Mathematics. Springer, 1992.

[89] J. W. Stephen and R. Benjamin, Optimization for Data Analysis, Cambridge University Press, 2022.

[90] A. R. Conn, N. I. M. Gould and P. L. Toint, Trust Region Methods, Philadelphia, PA, USA: SIAM: MPS-SIAM Series on Optimization, 2000.

[91] A. Antoniou and W.-S. Lu, Practical Optimization: Algorithms and Engineering Applications, New York: Springer, 2021.

[92] D. P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, 1982.

[93] H. Kuhn and A. Tucker, Nonlinear Programming. Proceedings of the 2nd Berkeley Symposium on Mathematics, Statistics and Probability, Berkeley: University of California Press, 1951.

[94] P. J. M. v. Laarhoven and E. H. L. Aarts, Simulated Annealing: Theory and Applications, Dordrecht: Springer, 1987.

[95] K. Marti, Random Search Procedures for Global Optimization, Cham: Springer, 2020.

[96] E. -G. Talbi, Metaheuristics: From Design to Implementation, Hoboken: John Wiley & Sons, 2009.

[97] T. Beck, D. B. Fogel and Z. Michalewicz, Evolutionary Computation 1: Basic Algorithms and Operators, Boca Raton, FL: CRC Press, 2000.

[98]   D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Boston, MA: Addison Wesley, 1989.

[99]   R. C. Eberhart, J. Kennedy and Y. Shi, Swarm Intelligence, New York: Elsevier, 2001.

[100] M. Dorigo and T. Stützle, The Ant Colony Optimization Metaheuristic: Algorithms, Applications, and Advances, Boston, MA: Springer, 2003.

[101] S. Koziel and X. S. Yang, Computational Optimization, Methods and Algorithms. Studies in Computational Intelligence, Berlin: Springer, 2011.

[102] F. J. Ares-Pena, J. A. Rodriguez-Gonzalez, E. Villanueva-Lopez and S. R. Rengarajan, "Genetic algorithms in the design and optimization of antenna array patterns," *IEEE Trans. Antennas & Propag.,* vol. 47, no. 3, p. 506–510, 1999.

[103] P. J. Bevelacqua and C. A. Balanis, "Optimizing antenna array geometry for interference suppression," *IEEE Trans. Antennas & Propag.,* vol. 55, no. 3, p. 637–641, 2007.

[104] S. Koziel and S. Ogurtsov, "Model management for cost-efficient surrogate-based optimization of antennas using variable-fidelity electromagnetic simulations," *IET Microw. Ant. Propag.,* vol. 6, pp. 1643-1650, 2012.

[105] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences,* vol. 45, pp. 50-79, 2009.

[106] T. W. Simpson, J. D. Poplinski, P. N. Koch and J. K. Allen, "Metamodels for computer-based engineering design: survey and recommendations," *Engineering Computations,* vol. 17, no. 2, p. 129–150, 2001.

[107] S. Haykin, Neural Network: A Comprehensive Foundation, New York: Prentice Hall, 1998.

[108] S. Koziel and A. Pietrenko-Dabrowska, Performance-driven surrogate modeling of high-frequency structures, Cham, Switzerland: Springer, 2020.

[109] Y. H. Y. E. W. M. N. S. Prasad, "Going beyong Chu harrington limit: ULF radiation with a spinning magnet array," in *General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, Montreal, 2017.