Wireless communications assisted by reconfigurable intelligent surfaces : an electromagnetic model

Fadil Danufane

To cite this version:

Fadil Danufane. Wireless communications assisted by reconfigurable intelligent surfaces : an electromagnetic model. Networking and Internet Architecture [cs.NI]. Université Paris-Saclay, 2021. English. NNT : 2021UPASG038 . tel-03271277

HAL Id: tel-03271277
https://theses.hal.science/tel-03271277

Submitted on 25 Jun 2021

Wireless Communications Assisted by Reconfigurable Intelligent Surfaces: An Electromagnetic Model

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 580, Sciences et technologies de l'information et de la communication (STIC)
Spécialité de doctorat: Réseaux, Information et Communications
Unité de recherche: Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France.
Référent: : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Gif-sur-Yvette, le 20 Mai 2021, par

Fadil Habibi DANUFANE

Composition du jury:

Mathias Fink Professeur, PSL, France	Président
Andrea Massa Professeur, University of Trento, Italy	Rapporteur
Alain Sibille Professeur, IP Paris, France	Rapporteur
Martine Liénard Professeur, University of Lille, France	Examinatrice
Dominique Lesselier DR CNRS Emeritus, UPSaclay, France	Examinateur
Marco Di Renzo DR CNRS, UPSaclay, France	Directeur

Thèse de doctorat

NNT: 2021UPASG038

université
PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies
de l'information
et de la communication (STIC)

Titre: Communications sans fil assistées par des surfaces intelligentes reconfigurables: un modèle électromagnétique

Mots clés: Communications sans fil, surfaces intelligentes reconfigurables, environnements radio intelligents, modélisation électromagnétique, diffusion.

Résumé: L'émergence de l'environnement radio intelligent (SRE) en tant que nouveau paradigme qui remet en question le statu quo dans la communication sans fil a motivé l'utilisation des surface intelligentes reconfigurables (RIS) basées sur des métasurfaces pour améliorer la limite de performance des systèmes de communication sans fil. L'objectif principal de cette thèse est la modélisation d'un système de communication assisté par un RIS à l'aide d'approches basées sur l'électromagnétisme.

Le Chapitre 1 présente le concept d'environnement radio intelligent. Nous donnons également la définition du RIS et comment le RIS peut être utilisé dans le contexte du SRE. Nous donnons aussi quelques perspectives historiques et discutons le développement des activités de recherche dans ce contexte.

Le Chapitre 2 présente les concepts théoriques nécessaires pour comprendre les résultats des chapitres suivants. Ce chapitre est divisé en deux parties. La première partie traite de la modélisation des métasurfaces où nous passons d'une description microscopique basée sur la physique d'une métasurface et introduisons une représentation macroscopique de celle-ci, qui s'avère appropriée pour une application aux communications sans fil. La deuxième partie présente plusieurs approches analytiques qui nous permettent de calculer le champ l'électromagnétique dans un certain volume.

Le Chapitre 3 fournit une comparaison des performances entre les RISs fonctionnant comme des réflecteurs anormaux et un schéma de relais basé sur le décodage et la transmission. La comparaison est qualitative et couvre une multitude de métriques. En outre, une comparaison quantitative en termes de débits de données réalisables est également présentée. En particulier, les résultats numériques montrent que des RISs suffisamment grands peuvent être plus performants que les systèmes assistés par relais en termes de débit de données, tout en ré- duisant la complexité.

Le Chapitre 4 propose une caractérisation analytique du champ electromagnetic en présence d'un RIS qui est modélisé comme une métasurface homogénéisée et qui fonctionnent soit en réflexion ou transmission. Des expressions de forme fermée sont également obtenues dans deux régimes asymptotiques qui sont représentatifs de deux déploiements en champ lointain et en champ proche. Sur la base de l'approche proposée, l'impact de plusieurs paramètres est etudié.

Enfin, le Chapitre 5 résume la thèse et discute des perspectives futures qui méritent d'être étudiées pour mieux comprendre les avantages et les inconvénients des RISs pour application aux communications sans fil.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies
de l'information
et de la communication (STIC)

Title: Wireless Communications Assisted by Reconfigurable Intelligent Surfaces: An Electromagnetic Model

Keywords: Wireless communications, reconfigurable intelligent surfaces, smart radio environments, electromagnetic-based modeling, scattering.

Abstract: The emergence of smart radio environment (SRE) as a new paradigm that challenges the status quo in wireless communication has motivated the use of metasurface-based reconfigurable intelligent surface (RIS) to improve the performance limits in wireless communication systems. The main focus of this thesis is the modeling of reconfigurable intelligent surfaces (RIS)-aided communication systems using electromagnetic based methods.

Chapter 1 introduces the concept of smart radio environment. We also give the definition of RIS and how RIS can be used in context of SRE. To give some historical perspectives, we also discuss several important milestone papers throughout the development of research activities that lead to the current state of the art.

Chapter 2 introduces theoretical concepts that are necessary to understand the results in the subsequent chapters. This chapter is divided into two parts. The first part discusses the metasurfaces modeling where we move from a physics-based microscopic description of a metasurface and introduce a macroscopic representation for it, which is shown to be suitable for application in wireless communications. The second part introduces several analytical approaches that allow us to compute the EM field at any point of a given volume that contains the metasurface.

Chapter 3 provides a performance comparison between RISs operating as anomalous reflectors and a decode-and-forward relaying scheme that is representative of competing candidate technologies to realize SREs. The comparison is qualitative and covers multitude metrics. Furthermore, a quantitative comparison in terms of achievable data rates is presented. In particular, the numerical results show that sufficiently large RISs can outperform relay-aided systems in terms of data rate, while reducing the implementation complexity.

Chapter 4 proposes a electromagnetic based analytical characterization of the free-space path-loss of a wireless link in the presence of a RIS that is modeled as a two-dimensional homogenized metasurfaces made of sub-wavelength scattering elements and that operate either in reflection or transmission mode. The analytical method of physical optics is employed. Closed-form expressions are also obtained in two asymptotic regimes that are representative of far-field and near-field deployments. Based on the proposed approach, the impact of several design parameters and operating regimes is unveiled.

Finally, Chapter 5 summarizes the main findings of the thesis and discusses possible future directions that are worth investigating to unlock the full potential of RIS and bring it into practice.

Acknowledgment

First and foremost, I would like to thank my supervisor, Prof. Marco Di Renzo, for his invaluable advice, continuous support, and patience during my Ph.D. His immense expertise in research and plentiful experience has encouraged me in all the time of my academic research and daily life. I would also like to thank Prof. Mathias Fink, Prof. Andrea Massa, Prof. Alain Sibille, Prof. Martine Liénard, and Prof. Dominique Lesselier for accepting to become members of my thesis committee and for their valuable comments.

Doing a Ph.D. in a new country has always been difficult without support from friends. Therefore, I would like to thank all the currents and former labmates in L2S, CentraleSupelec: Jiang Liu, Dr. Xuewen Qian, Dr. Alexis Aravanis, Mohamed Abdelhamed, Arzhang Shahbazi, Dr. Jian Song, Dr. Xiaojun Xi, Dr. Shanshan Wang, Dr. Tu Lam Thanh, Dr. Kishor Chandra Joshi, Dr. Farshad Shams, and Romain Fara. They have been very helpful to my research by sharing their insightful advice and being friends these past three years. I would also like to thank all my Indonesian friends in Paris for being very helpful with their advice on living and studying in France and for providing me help whenever I need them. I enjoy very much having social interaction with them on weekly basis.

I would like to thank the European Commission for generously supporting my research under the auspices of the H2020 5G STEP-FWD project. Thanks to this project, I got to know a lot of excellent researchers and fellow Ph.D. students from across Europe and had ample opportunities to participate in professional training that would benefit me in the future. I would also like to acknowledge the support from the Indonesian Embassy in Paris, especially during this COVID-19 pandemic situation.

Last but not the least, I would like to thank my family: my father Fauzan Ali, my mother Mulharnetti Syas, my brother Fauzil Halim Mufassa, my father-in-law Arifin Noor Sugiharto, my mother-in-law Ratna Orbaningtyas. Their prayers and moral support from Indonesia are what allows me to finish my Ph.D. journey. Finally, a special token of gratefulness goes to my wife, Hana Akari Arifin, for her unconditional love and support during the highs and lows of my life in Paris. This thesis would not have been there without her.

Gif sur Yvette, 7 March 2021 Fadil Habibi Danufane

Contents

Acknowledgements **v**

Acronyms **v**

Notation **vii**

List of figures **ix**

List of tables **xi**

1 Introduction **1**

1.1 Introduction 3
1.2 Smart Radio Environment 4
1.2.1 From Adaptation to Control and Programmability 5
1.2.2 Illustrative Example of Smart Radio Environment 5
1.3 What is an RIS? 7
1.4 Historical Perspective on the State of Research on Reconfigurable Intelligent Surfaces 8
1.4.1 Metamaterials and Metasurfaces (1967, 2000 – Now) 8
1.4.2 Smart Radio Environment (2003 – Now) 9
1.4.3 Theoretical Modeling of RIS (2014 – Now) 11
1.5 Thesis Overview and Major Contributions 12
1.6 Publications 13

2 Macroscopic Homogenized Modeling of Metasurfaces - An Introduction **17**

2.1 Introduction 19
2.2 Surface Electromagnetics 20
2.2.1 Definition of SEM 20
2.2.2 From uniform to quasi-periodic metasurfaces 22
2.2.3 Theoretical Foundation of Surface Electromagnetics 23
2.3 Modeling Metasurfaces: A Macroscopic Homogenized Approach 27
2.3.1 From a microscopic to a macroscopic representation 27
2.3.2 EM-Based Model of a Metasurface 28
2.3.3 Microscopic Description of a Metasurface 31
2.3.4 Macroscopic Description of a Metasurface 34

2.4 Modeling Radio Wave Propagation in the Presence of Metasurfaces 37
  2.4.1 Reference operating regimes 37
  2.4.2 Theory of Electromagnetic Diffraction 40
  2.4.3 Methods of Physical Optics 46
2.5 Conclusion 50

3 Reconfigurable Intelligent Surfaces vs. Relays: Performance Comparison 51
3.1 Introduction 53
3.2 Reconfigurable Intelligent Surfaces vs. Relaying 54
  3.2.1 Hardware Complexity 54
  3.2.2 Noise 55
  3.2.3 Spectral Efficiency 55
  3.2.4 Power Budget 55
3.3 Performance Comparison: Achievable Data Rate 56
  3.3.1 System Model and Analytical Formulation of Achievable Data Rate . . . 56
  3.3.2 Quantitative Comparison 61
  3.3.3 Average Signal-to-Noise Ratio vs. Transmission Distance 62
  3.3.4 Takeaway Messages from the Comparison 63
3.4 Numerical Results 64
  3.4.1 RISs vs. Relays as a Function of the Transmission Distance 66
  3.4.2 RISs: Anomalous Mirrors vs. Focusing Lenses 67
  3.4.3 RISs vs. Relays as a Function of the Carrier Frequency 67
  3.4.4 RISs vs. Relays as a Function of the Size of the RIS 67
3.5 Conclusions 68
3.6 Appendices 68
  3.6.1 Proof of Theorem 3.1 68
  3.6.2 Proof of Proposition 3.1 69
  3.6.3 Proof of Proposition 3.2 70

4 On the Path-Loss of Reconfigurable Intelligent Surfaces: Physical Optics Analytical Formulation and Scaling Laws 73
4.1 Introduction 76
4.2 System Model 76
  4.2.1 Source Modeling 78
  4.2.2 Metasurface Modeling 79
4.3 Preliminaries 81
  4.3.1 Received Field at Rx 81
  4.3.2 Approximations and Asymptotic Regimes 85
  4.3.3 Application to Communication Systems 88
4.4 Electric Field In the Presence of a Reflecting Surface 89
  4.4.1 $\mathscr{S}$ is Configured for Specular Reflection 90
  4.4.2 $\mathscr{S}$ is Configured for Anomalous Reflection 91
  4.4.3 $\mathscr{S}$ is Configured for Focusing 93

4.4.4 Comparison with other Path-Loss Models 95
4.5 Electric Field In the Presence of a Transmitting Surface 95
4.5.1 $\mathscr{S}$ is Configured for Specular Transmission 96
4.5.2 $\mathscr{S}$ is Configured for Anomalous Transmission 97
4.5.3 $\mathscr{S}$ is Configured for Focusing 98
4.6 Numerical Results 100
4.6.1 Anomalous Reflection and Focusing 100
4.6.2 Transmission Distance 102
4.6.3 Surface Size 104
4.6.4 Operating Frequency 105
4.7 Conclusion 106
4.8 Appendices 106
4.8.1 Proof of Theorem 4.1 106
4.8.2 Proof of Theorem 4.2 107
4.8.3 Proof of Lemma 4.4 108
4.8.4 Proof of Lemma 4.5 108
4.8.5 Proof of Proposition 4.1 109
4.8.6 Proof of Corollary 4.1 110
4.8.7 Proof of Proposition 4.2 110

5 Conclusions and Future Work **111**
5.1 Conclusion 112
5.2 Future Work 113
5.2.1 Physics-based Modeling 113
5.2.2 Experimental Validation 114
5.2.3 Constrained System Design 114
5.2.4 Statistics of the RIS Channel Model 114
5.2.5 Information and Communication Theory 115

Reference **117**

Acronyms

AC	Alternate current
C-SWaP	Cost, size, weight, and power
DC	Direct current
DF	Decode and forward
DL	Downlink
eMBB	Enhanced mobile broadband
EM	Electromagnetic
EMT	Effective medium theory
FD	Full-duplex
GSTC	Generalized sheet transition condition
GO	Geometrical optics
IoT	Internet of Things
IP	Internet protocol
mMTC	Massive machine type communications
MT	Mobile terminal
PO	Physical optics
RIS	Reconfigurable intelligent surface
SNR	Signal-to-noise ratio
SPM	Stationary phase method
SRE	Smart radio environment
THz	Terahertz

UL	Uplink
ULA	Uniform linear array
URLLC	Ultra-reliable and low latency communication
VLOS	Virtual line-of-sight

Notation

The following notation is used throughout this thesis.

$()^*$	Complex conjugate
$\mathrm{Re}\{\cdot\}$	Real part
$()^T$	Matrix transpose
$()^H$	Hermitian conjugate
$()^\dagger$	Pseudo inverse
$\cdot$	Scalar product
$\times$	Vector product
$\widehat{()}$	Unit vector
λ	Wavelength
ω	Angular frequency
ε	Permittivity in vacuum
μ	Permeability in vacuum
$\eta = \sqrt{\mu/\varepsilon}$	Impedance in vacuum
k	Wavenumber
$\mathbf{P}(x, y)$	Electric surface polarization density
$\mathbf{M}(x, y)$	Magnetic surface polarization density
$\mathbf{P}_{\parallel}\left(x, y\right)$	Longitudinal component of $\mathbf{P}(x, y)$
$\mathbf{M}_{\parallel}\left(x, y\right)$	Longitudinal component of $\mathbf{M}(x, y)$
$\mathbf{E}(x, y, z)$	Electric field
$\mathbf{H}(x, y, z)$	Magnetic field
$\mathbf{D}(x, y, z)$	Electric displacement field
$E_{x,y}(x, y)$	Tangential components of the electric field
$H_{x,y}(x, y)$	Tangential components of the magnetic field
i, r, t	Incident, reflected, transmitted field
$\theta_i, \theta_r, \theta_t$	Angle of incidence, reflection, transmission
$\overline{\overline{\chi}}_{\mathrm{ab}}\left(x, y\right)$	Surface susceptibility dyadics
$R(x, y)$	Surface reflection coefficient
$\delta(\cdot, \cdot)$	Dirac delta function
$\mathrm{Hess}(\cdot)$	Hessian matrix
$\mod(\cdot)$	modulo operator
$\lvert C \rvert, \angle C$	Modulus and argument of complex number C

$\nabla^2 G(x,y,z)$	$= \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)G(x,y,z)$, Laplacian of scalar function $G(x,y,z)$
$\nabla G(x,y,z)$	$= \frac{\partial G(x,y,z)}{\partial x}\hat{\mathbf{x}}+\frac{\partial G(x,y,z)}{\partial y}\hat{\mathbf{y}}+\frac{\partial G(x,y,z)}{\partial z}\hat{\mathbf{z}}$, Gradient of scalar function $G(x,y,z)$
$\nabla\times\mathbf{F}$	$= \left(\frac{\partial F_z}{\partial y}-\frac{\partial F_y}{\partial z}\right)\hat{\mathbf{x}}+\left(\frac{\partial F_x}{\partial z}-\frac{\partial F_z}{\partial x}\right)\hat{\mathbf{y}}+\left(\frac{\partial F_y}{\partial x}-\frac{\partial F_x}{\partial y}\right)\hat{\mathbf{z}}$, Curl of $\mathbf{F}$
$\nabla\cdot\mathbf{F}$	$= \frac{\partial F_x}{\partial x}+\frac{\partial F_y}{\partial y}+\frac{\partial F_z}{\partial z}$, Divergence of $\mathbf{F}$
$\vec{\nabla}^2\mathbf{F}$	$= \nabla^2 F_x\hat{\mathbf{x}}+\nabla^2 F_y\hat{\mathbf{y}}+\nabla^2 F_z\hat{\mathbf{z}}$, Vector Laplacian of $\mathbf{F}$
$\nabla^2_{\mathbf{r}} G(x,y,z)$	Laplacian of $G(x,y,z)$ evaluated at $\mathbf{r}$
$\nabla_{\mathbf{r}} G(x,y,z)$	Gradient of $G(x,y,z)$ evaluated at $\mathbf{r}$
$\vec{\nabla}^2_{\mathbf{r}}\mathbf{F}$, $\nabla_{\mathbf{r}}\cdot\mathbf{F}$	Vector Laplacian and divergence of $\mathbf{F}$ evaluated at $\mathbf{r}$
$G(\mathbf{r}_1,\mathbf{r}_2)$	$= \frac{\exp(-jk\lvert\mathbf{r}_1-\mathbf{r}_2\rvert)}{4\pi\lvert\mathbf{r}_1-\mathbf{r}_2\rvert}$, Green's function solution of (4.8)
$f(x)\rvert_{x=x_1}^{x=x_2}$	$= f(x_2)-f(x_1)$, Shorthand notation
$g(x,y)\rvert_{x=x_1}^{x=x_2}\rvert_{y=y_1}^{y=y_2}$	$= g(x_2,y_2)-g(x_2,y_1)-g(x_1,y_2)+g(x_1,y_1)$, Shorthand notation

List of Figures

1.1 Radio environments vs. smart radio environments. 4
1.2 Example of smart radio environment. 6
1.3 Possible uses of reconfigurable intelligent surfaces. (i) Anomalous reflection: a radio wave incident at an angle of 90 degrees is reflected towards an angle of 45 degrees. (ii) Focusing lens: a radio wave incident at an angle of 90 degrees is focused (beamforming) towards a specified location in order to maximize the energy at that point. (iii) An RIS illuminated by a feeder reflects two phase-modulated signals by mimicking a two-antenna transmitter, thus encoding information on the reflections of the meta-surface. These functions can be obtained by appropriately configuring the phase response of the RIS (i.e., by optimizing $\Phi(x)$ in [1, Eq. (4)]). 7

2.1 Conceptual structure of a reconfiguration intelligent surface. 19
2.2 Classification of electromagnetics in the spatial domain (reproduced from [2]). 21
2.3 The road to quasi-periodic metasurfaces (reproduced from [2]). 22
2.4 Illustration of the surface equivalent theorem applied to the scattering of volumetric electromagnetic sources (a) and to the concept of metamaterials (b) (reproduced from [2]). 24
2.5 Illustration of homogenization: A metasurface consisting of arbitrary sub-wavelength unit cells is modeled as a continuous sheet of surface-averaged (over the unit cell) currents (reproduced from [2]). 26
2.6 Example of induced electric dipole moment and electric polarization density (**P**) in volumetric materials. 28
2.7 Equivalent analytical representation of a metasurface: A zero-thickness material sheet whose EM properties are modeled through surface susceptibility functions that introduce EM discontinuities (or “jumps”) quantified by the generalized sheet transition conditions. 30
2.8 Homogenization modeling of a metasurface based on surface-averaged impedances (reproduced from [2]). 36
2.9 Near-field vs. far-field of a metasurface and far-field of the array microstructure. The far-field boundary of the metasurface is obtained from the Fraunhofer distance. 38
2.10 Electrically large vs. electrically small metasurfaces. 39

2.11 The Huygens-Fresnel principle. 40
2.12 Considered system model for finding scalar electric field in a volume V 43
2.13 Scattering from a material body according to PO. The scattered field at the observation point P is obtained by integration over all secondary sources on the illuminated portion of the scatterer (gray) which are described in the GO approximation. On the shadowed portion of the scatterer (black) the secondary sources are assumed to be zero. 49

3.1 System model for RIS-assisted communication. 57
3.2 Data rate of RISs and relays versus the transmission distance. 64
3.3 Data rate of RISs and relays versus the transmission frequency. 65
3.4 Data rate of RISs and relays versus the size of the RIS. 66

4.1 System model. 77
4.2 Volume V and closed boundary ∂V. 82
4.3 Anomalous transmission vs. focusing lens (transmitting surface). Setup: $2L_x = 2L_y = 1$ m. 100
4.4 Anomalous reflection: Impact of discretization. Setup: $2L_x = 2L_y = 0.5$ m; $d_{\text{Tx0}} = d_{\text{Rx0}} = 5$ m. 101
4.5 Absolute error difference corresponding to Fig. 4.4. 102
4.6 Anomalous reflection vs. focusing lens: Impact of transmission distance. 103
4.7 Impact of surface size. 104
4.8 Impact of transmission frequency. Reflecting surface. Setup: $d_{\text{Tx0}} = d_{\text{Rx0}} = 50$ m. 105

List of Tables

3.1 Rate (R) for relays and RISs. 64

4.1 Main operators ($G(x, y, z)$ is a scalar function, $\mathbf{F} = F_x\hat{\mathbf{x}} + F_y\hat{\mathbf{y}} + F_z\hat{\mathbf{z}}$ is a vector field with $\mathbf{F} = \mathbf{F}(x, y, z)$ and $F_a = F_a(x, y, z)$ for $a = x, y, z$). Symbols in bold denote vectors. Unit-norm vectors are denoted by $\hat{(\cdot)}$. 78
4.2 Approximate conditions for which the electrically-large and electrically-small asymptotic regimes hold. 88
4.3 Simulation setup 99

1 Introduction

One of the main assumptions in the classical communication theory is that the wireless environment is seen as something random and we have to adapt with when designing a wireless communication system. In other words, we do not have the control over the radio waves propagation within the environment, especially regarding the reflection and refraction by objects. This is exactly the vision of smart radio environment (SRE), where one can control the behaviour of the radio waves upon reflection by some objects such that most of the reflected power is maximized towards the user and minimized in parasitic directions. One way to realize the concept of smart radio environment is by employing a recently emerging technology called reconfigurable intelligent surfaces (RISs).

In this chapter, we introduce the concept of SRE and RIS as its enabling technology, highlighting its properties, functionalities, and benefits compared to the existing technologies. We also touch upon the challenge and approaches in modeling of RIS for wireless communication applications. The major contributions in this thesis work, the organization of the thesis, and the list of publications produced during Ph.D. are also presented in this chapter.

1.1 Introduction 3

1.2 Smart Radio Environment 4

  1.2.1 From Adaptation to Control and Programmability 5

  1.2.2 Illustrative Example of Smart Radio Environment 5

1.3 What is an RIS? 7

1.4 Historical Perspective on the State of Research on Reconfigurable Intelligent Surfaces 8

  1.4.1 Metamaterials and Metasurfaces (1967, 2000 – Now) 8

  1.4.2 Smart Radio Environment (2003 – Now) 9

  1.4.3 Theoretical Modeling of RIS (2014 – Now) 11

1.5 Thesis Overview and Major Contributions 12

1.6 Publications 13

1.1 Introduction

Future wireless networks are expected to support very high data rates, since between 2020 and 2030, it is forecast that the data traffic of the global Internet protocol (IP) will increase by 55% each year, eventually reaching 5,016 exabytes. Furthermore, it is expected to offer several other heterogeneous services:

- **Enhanced mobile broadband (eMBB)**: This type of service is similar to the ones existing in 4G but with improved performance (in terms of data rate) and and much more seamless user experience (in the terms of coverage). This scenario covers a plethora of use-cases, such as wide-area coverage and hotspots. For the wide area case, seamless coverage and high mobility are desired, with much improved user data rates compared to that offered today. For hotspots, the support of high user density, and very high traffic capacity is needed, but the requirement for mobility is at pedestrian speeds only.

- **Ultra-reliable and low latency communication (URLLC)**: This type of service promises to provide low latency and ultra-high reliability for mission critical applications such as industrial internet, smart grids, remote surgery and intelligent transportation systems. Thanks to this, delay and reliability emerge as important metrics to consider, in addition to traditional metrics such as capacity and spectral efficiency, in communication systems design and their requirements are becoming more stringent than ever. According to the 3GPP standard, the end-to-end delay of 5G communications is required to be less than 1 ms, which is about 1/200 of delay requirement for 4G, and the target reliability must be as high as 10^{-2} to 10^{-7} .

- **Massive machine type communications (mMTC)**: This type of service involves the connectivity among a huge number of devices with a relatively low volume of non-delay-sensitive data. Devices are required to be low cost, and have a very long battery life. An example of this use-case scenario include Internet of Things (IoTs).

Unfortunately, despite the overarching targets that 5G is expected to achieve, various recent test-beds show that the current 5G technology is still not able to achieve all of them. Moreover, most of the tests are still far from using a complete system which means that the impacts of other cell interference or the benefits of the multiple layers are largely ignored. In addition, most of them consider only the downlink (DL) but not the uplink (UL) scenarios.

In order to fulfill all the 5G requirements, which have not been achieved yet by current advances, one sensible approach is to rethink about the current communication system design paradigm and overcome its limitations.

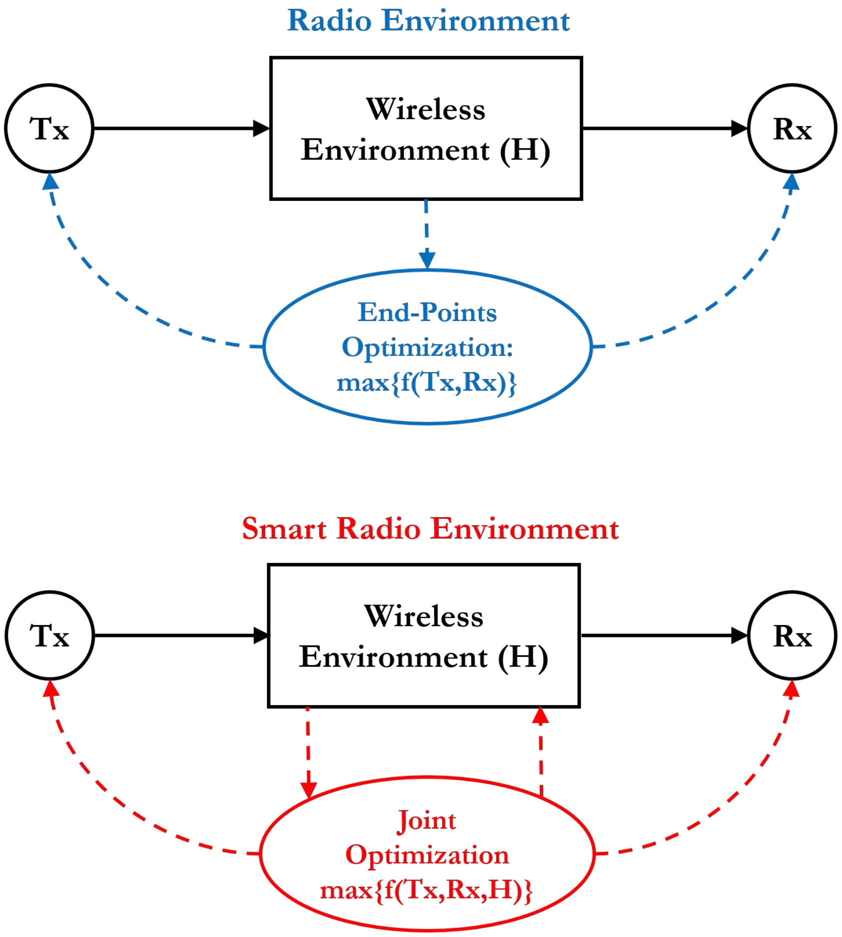


Figure 1.1 – Radio environments vs. smart radio environments.

1.2 Smart Radio Environment

During the first five generations of wireless networks, the design of wireless communication systems have been adhering to the postulates that the wireless environment between communicating devices (i) is controlled by nature, (ii) cannot be modified, (iii) can be only compensated through the design of sophisticated transmission and reception schemes. The sixth generation (6G) of communication networks is, on the other hand, envisioned to break through these postulates by assuming that one can control the wireless environment to achieve major performance gains. For example, a typical base station transmits radio waves of the order of magnitude of Watts while a user equipment detects signals of the order of magnitude of μWatts. The rest of the power is, in general, wasted in different ways through the environment by, e.g., generating interference to other network elements or creating security threats, since the propagation of radio waves through the wireless channel cannot be controlled and customized after they are emitted from the transmitters and before they are received by the receivers. This new paradigm which sees environment as a controllable element in wireless system design is termed as the **Smart Radio Environment** (SRE). Conceptually, the vision of SREs is depicted in Fig. 1.1.

1.2.1 From Adaptation to Control and Programmability

From the viewpoint of the communication engineer, the wireless environment is conventionally modeled as an exogenous entity that cannot be controlled, but only adapted to. To this end, communication engineers can only design the transmitters, the receivers, and the transmission protocols in order to achieve the desired performance. Common approaches to capitalize on the properties of the wireless environment and to mitigate its impairments include using multiple antennas, employing complex encoding and decoding algorithms at the end-points of the communication link, and adding additional network infrastructure, e.g., relays, in an attempt to make the transmission of signals more reliable. These solutions, however, may increase the network complexity, the network power consumption, and the network deployment cost [3].

RISs provide wireless researchers and engineers with a different **view of the wireless environment**. Since RISs are capable of shaping the wavefront of the radio waves throughout the network, the **wireless environment can be in principle customized** to suit the system requirements. The wireless environment is not to be treated as a random uncontrollable entity, but rather as part of the network design parameters that are subject to optimization in order to support diverse performance metrics, such as rate, latency, reliability, energy efficiency, privacy, and massive connectivity. The overarching vision consists of coating environmental objects and devices with digitally-controlled RISs, and programming them, through environmental sensing and SDN-based protocols, for shaping the radio propagation environment and meeting the desired system requirements [4], [5].

1.2.2 Illustrative Example of Smart Radio Environment

An example of smart radio environment is sketched in Fig. 1.2, where four application scenarios are identified.

- **Signal engineering**: Assume that small cell 1 wishes to communicate with mobile terminal (MT) 1, but the LOS link is blocked by an object. In this case, small cell 1 redirects the transmitted beam towards RIS 1 that coats object 1, and assists the communication by shaping the incident wave towards MT 1 so that the received signal strength is maximized.
- **Interference engineering**: While small cell 1 communicates with MT 1, small cell 2 communicates with MT 2. Thus, an interfering signal reaches MT 1 from small cell 2. To suppress it at MT 1, RIS 2 is programmed to shape the impinging radio wave from small

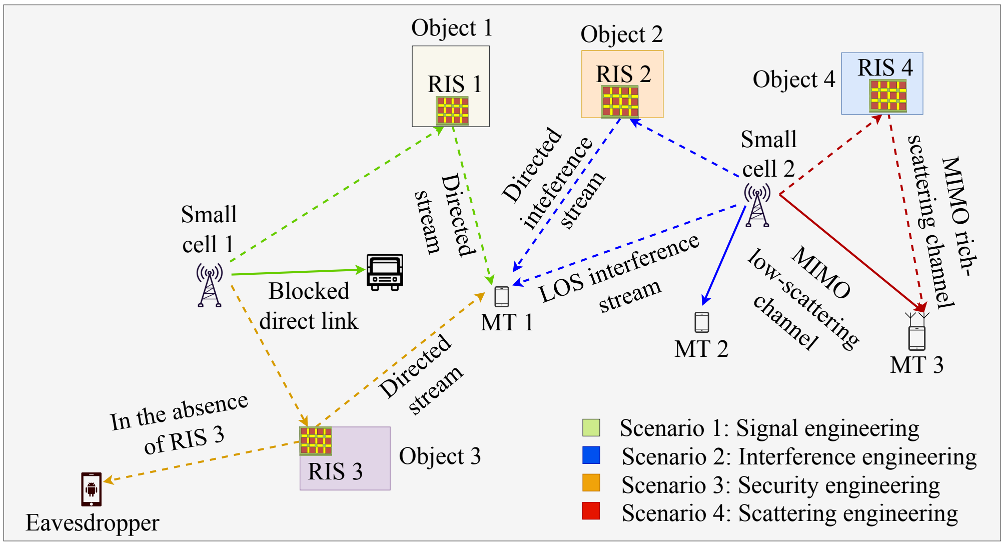


Figure 1.2 – Example of smart radio environment.

cell 2 towards MT 1 in a way that the two signals are destructively combined at MT 1.

- **Security engineering**: In the absence of RIS 3, the signal emitted by small cell 1 and intended to MT 1 is reflected from object 3 towards a malicious user that overhears it. To avoid this, RIS 3 is programmed to shape the reflection towards MT 1 so that it is steered away from the malicious user while being decoded more reliably, via diversity combining, at MT 1.

- **Scattering engineering**: The multiple-antenna small cell 2 wishes to convey information to the multiple-antenna MT 3 with the aid of multiple-input multiple-output transmission. The channel between small cell 2 and MT 3 has, however, a low rank (low scattering environment), which negatively affects the attainable data rate. To avoid this issue, small cell 2 directs the signal intended to MT 3 towards RIS 4, which appropriately shapes it so as to create a rich-scattering environment (high rank channel) for high data rate transmission.

From the analysis of these four scenarios, it is apparent that, with the aid of RISs, the propagation of radio waves in wireless networks may be engineered and optimized, at a low complexity, in a way that benefits the network. In contemporary wireless networks, transmitters and receivers can be programmed and controlled for optimizing the system performance. The environmental objects (buildings, walls, ceilings, etc.) that constitute the wireless environment cannot, on the other hand, be customized based on the network conditions. This status quo has recently been challenged by the emerging technology of reconfigurable intelligent surfaces (RISs) – Thin sheets of electromagnetic materials that are capable of shaping the radio waves in arbitrary ways [5], [6]. The overarching vision consists of coating the environmental objects with RISs and optimizing their properties, in order to, e.g., reflect an impinging radio

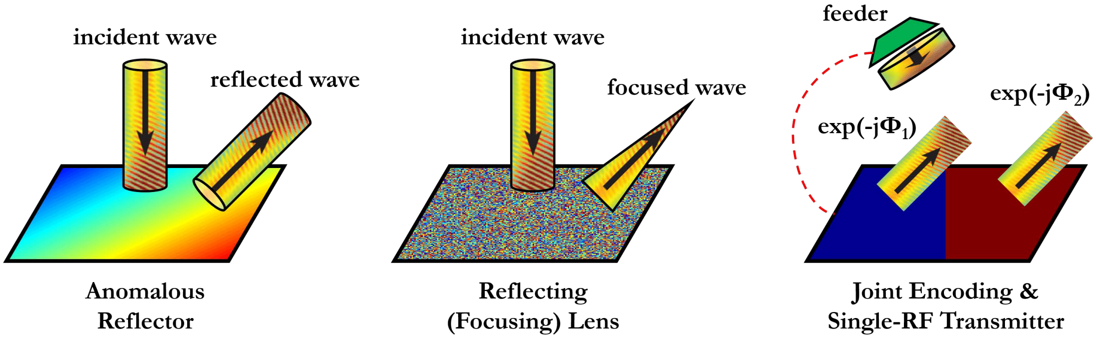


Figure 1.3 – Possible uses of reconfigurable intelligent surfaces. (i) Anomalous reflection: a radio wave incident at an angle of 90 degrees is reflected towards an angle of 45 degrees. (ii) Focusing lens: a radio wave incident at an angle of 90 degrees is focused (beamforming) towards a specified location in order to maximize the energy at that point. (iii) An RIS illuminated by a feeder reflects two phase-modulated signals by mimicking a two-antenna transmitter, thus encoding information on the reflections of the meta-surface. These functions can be obtained by appropriately configuring the phase response of the RIS (i.e., by optimizing $\Phi(x)$ in [1, Eq. (4)]).

wave towards a desired direction with the objective of capitalizing from multipath propagation rather than being negatively affected by it [4, 7–10].

1.3 What is an RIS?

RISs are an emerging transmission technology for application to wireless communications [11] that is considered as a key enabler towards SRE. In simple terms, an RIS can be viewed as the two-dimensional equivalent of a reconfigurable metamaterial, and is made of elementary elements called scattering particles or meta-atoms. By configuring the phase-shift of each unit cell, an RIS is theoretically capable of controlling the behavior of the radio wave that impinges upon it. In the recent literature, the following four major uses have been considered (see Fig. 1.3 for illustrations).

- **Anomalous reflection/transmission** [12]: The RIS is configured in order to reflect or refract the impinging radios waves towards specified *directions* that do not necessarily adhere to the laws of reflection and refraction. The advantage of this application is that the operation of the RIS is independent of the fading channels and the locations of the receivers. The limitation is that, in general, the signal-to-noise-ratio is not maximized and the system capacity is not achieved.
- **Beamforming/focusing** [3]: The RIS is configured in order to focus the impinging radio waves towards specified *locations*. The advantage of this application is that the signal-to-noise-ratio is maximized at the locations of interest. The challenge is that, in general, the optimization of the RIS depends on the fading channels and the locations of the

receivers. Also, the system capacity is usually not achieved.

- **Joint transmitter/RIS encoding** [10]: The RIS is configured in order to optimize the system capacity. The advantage of this application is that the specific status of the meta-atoms is exploited to modulate additional data. The challenge is that, in general, the transmitter and the RIS need to be jointly optimized. In addition, the setup of the RIS depends on the fading channels and the locations of the receivers.
- **Single-RF multi-stream transmitter design** [13]: This operation is similar to the previous one, with the difference that the transmitter is a simple RF feeder located in close vicinity of the RIS. The feeder emits an unmodulated carrier towards the RIS, which reflects multiple data-modulated signals. This approach is suitable to realize multi-stream transmitters by employing a limited number of (even a single) RF chains.

Other than its ability to manipulate the wavefront of the impinging electric fields, RIS can be practically implemented since it is relatively inexpensive, energy-efficient, and very easy to deploy, especially in the wall of buildings due to its 2D-shape. Compared with, e.g., phased arrays, multi-antenna transmitters, and relays, RISs require the largest number of scattering elements, but each of them needs to be backed by the fewest and least costly components. Also, since RISs are nearly passive in nature, no signal processing capability is expected and thus they do not require power amplifiers or highly energy-consuming components such as RF chain in MIMO. Thanks to these properties, RIS is receiving major attention from the wireless community, which is indicated by the exploding number of research papers in the literature during the last few years, and is considered to be the key enabler of the emerging concept of smart radio environments.

1.4 Historical Perspective on the State of Research on Reconfigurable Intelligent Surfaces

Despite the recent spike of interest in RIS among researchers in wireless communication community, the research on metasurface as well as their application for wireless communication has its decades-long history, ranging from theoretical investigation to experiments and field measurements. In this section, we briefly mention several works that become the milestones which lead to the current state of the art of research on RIS. For ease of reading, we categorize them into several aspects that represent some of the most important facets of RIS research.

1.4.1 Metamaterials and Metasurfaces (1967, 2000 – Now)

Electromagnetic metamaterials are generally defined as *artificial electromagnetic structures consisting of a subwavelength lattice of scattering particles – or metaparticles – that are engineered to provide electromagnetic properties that are not found by other materials in nature* [14]. The history of metamaterials is started by a seminal paper [15] in 1967 in which the authors envision the existence of material which has negative values of permittivity ϵ and permeability μ. These materials are called the "left-handed" (LH) substances in [15], owing to the fact that that they would allow the propagation of electromagnetic waves with the electric field, the magnetic field, and the phase constant vectors building a left-handed triad, compared with conventional materials where this triad is known to be right-handed [16]. This kind of material, as predicted by the author of [15], will trigger unprecedented phenomenas such as reversed Doppler effect and reversed Snell's law.

It was not until more 30 years after [15] that the first LH material was proposed and demonstrated experimentally by the authors of [17] at University of California, San Diego (UCSD), in 2000, which is inspired by a seminal work in the same year [18] whose ambitious goal is to create a so-called perfect lens. These works later set into a motion a large body of research going and, in the span of 20 years, new more materials were proposed with additional effects such as cloaking [19, 20], magnet-less magnetism [21], and metasurfaces [22, 23], which, in its early introduction, was introduced as a two-dimensional metamaterial with new phase, magnitude and polarization transformation capabilities. As of late, thanks to recent advances in fields such as nanotechnology and bio/chemico-technology, the notion of metasurface has been extended to dynamic/reconfigurable metasurfaces, which consist of reconfigurable materials that can experience spatio-temporal change of electromagnetic properties through the electrical excitation by and external source of energy.

1.4.2 Smart Radio Environment (2003 – Now)

The reconfigurability property of dynamic metasurfaces made them found their use in wireless communications to make a communication process – which is normally heavily impaired by the spatio-temporal randomness of communication channels (e.g., in forms of fading, shadowing, and noise) – more reliable by transforming the wireless environment (which was classically as randomness) into something that can be controlled. Furthermore, reconfigurable metasurfaces can also be designed to modulate, transmit, and receive their own information (instead of acting as passively reflectors), making them suitable for antenna applications. These advances in wireless technologies accumulate and take form into a grand vision of Smart Radio Environment (SRE) which, as we have discussed previously, is a technological concept which envisions a wireless environment that can be controlled to enhance the performance gains of the ongoing communication process.

Despite being a recently hot research field, the term of *smart environment* has, in fact, been envisioned as early as in 2003. In particular, in 2003 and 2005 respectively, the authors of [24] and [25] introduced the term "smart controllable surface" to refer to a surface of electrically small scatterers, which are controllable by controlling the polarization densities of the scatterers or through the inclusion of bias field, for instance, in order to to switch electronically between reflecting and transmitting states. Between 2010 and 2012, the authors of [26–28] propose methods of controlling radio propagation environments by using *intelligent walls*, which are defined as conventional walls situated inside a building, but equipped with an active frequency selective surfaces and sensors. These surfaces are made of special materials such as ferro-electric films, liquid crystals, and even new materials such as graphene, and their transmission and reflection characteristics can be adjusted depending on the traffic demand or to avoid interference between transmitters.

In 2017, the authors of [29] propose the use of Dynamic Metasurface Aperture (DMA) to realize the concept of *Smart Homes* in which two probe antennas in an electromagnetic cavity are used for motion detection and extraction of useful information from the motions. DMA is defined as *"a device that is capable of producing arbitrary radiation patterns that can be varied with simple electronic controls"*. In the paper, the authors use the disordered three-dimensional DMA design which enables wavefront shaping. By using a DMA as one of the probe antennas in an electromagnetic cavity, the authors show that one is able to measure the transmitted field between the two antennas to reveal any motion of objects inside the cavity – even outside the line-of-sight (LOS) of the two probe antennas – and obtain useful information about the motion, without having to add more antenna pairs at locations where the local patterns of the cavity modes are different or having to implement complex hardware dseign that might arise from working with an antenna pair that operate in multiple frequencies.

It is also worth mentioning that the vision of "smart" or "reconfigurable" environment has also been mentioned in other works which make use of technologies that do not necessarily utilize metamaterials/metasurface. For example, the authors of [30] propose the use of passive relaying arrays (PRAs) to mitigate the channel impairements caused by small-scale and large-scale fading and overcome the latency issues and instability of the conventional relaying schemes such as decode-and-forward (DAF) and amplify-and-forward (AAF). Similar to metasurfaces/RISs, PRAs work by scattering energy on a directional and configurable basis while staying in a nearly passive domain by keeping the dissipative losses low. However, different from RISs, PRAs use electronically tunable reactive loads instead of metamaterials.

We end this subsection by mentioning that, as far as the term SRE is concerned, it was first used in 2019 by the authors in [5] to refer to *"a wireless environment that is turned into a smart reconfigurable space and that plays an active role in transferring and processing information"*. The acronym RIS was first introduced later in the same year in [8].

1.4.3 Theoretical Modeling of RIS (2014 – Now)

In pursuit of understanding the full potential of the RIS, one of the biggest questions to ask is: "how to theoretically model the RIS?". Theoretical modeling is important in order to predict the performance of an RIS, optimize it depending on the specific goals to achieve (e.g., broader coverage, higher data rate, and lower latency), and assess the advantages and limitations RIS-empowered wireless networks. In terms of theoretical modeling, in particular, a major open research issue is the development of simple but sufficiently accurate models for the power received at a given location in space when a transmitter emits radio waves that illuminate an RIS. To answer this question, in general, there exist several approaches that one can employ:

- **Measurement campaign**: This is the most straightforward and a traditional way to asses the performance of a given technology prior to real-world applications. The advantage of this approach is that one can obtain a realistic performance assessment, depending on the measurement settings. As an example, in 2014, the authors of seminal work [31] conducted experiments by using a RIS with size of 0.4 m^2 and thickness of 1.5mm that consists of 102 controllable EM reflectors and operates at a working frequency of 2.47 GHz. The 102 reflectors are controlled by using two Arduino 54-channel digital controllers. Through the experiments, the authors demonstrate that spatial microwave modulators are capable of shaping, in a passive way, complex microwave fields in complex wireless environments, by using only binary phase state tunable metasurfaces. In particular, the author showed that the RIS can improve spatial focusing of the radiated EM wave or minimize the EM field on the receiving antenna (i.e., nulling). A similar experiment was also conducted in 2019 by the authors of [32]. In particular, a similar RIS design as in [31] was deployed in a 1.45m × 1m × 0.75m chaotic aluminium cavity – 4% of which is covered by the RIS – to improve the communication of a MIMO system where two antenna arrays with 8 element each are located inside the cavity. Through the experiments, the authors showed that passive RISs can enrich multipath scattering, thus, enabling throughput boosted wireless connectivity. A more recent example is given in [33] in 2020, where the authors perform a measurement campaign in an anechoic chamber to model the power reflected from an RIS and investigate its trends with respect to many parameters, including the size of the RIS, the mutual distances between the transmitter/receiver and the RIS (i.e., near-field vs. far-field), and whether the RIS is used for beamforming or broadcasting.

- **Simulation**: The advantage of this approach, especially compared to measurement, is that it has lower cost to operate, and can be used to predict the performance of the technology of interest in many different scenarios. The drawback of this approach is, on the other hand, is that it does not provide one with a closed-form formulation and therefore, one cannot predict the trend or scaling law of the performance depending on typical wireless system parameters, such as the size and the inter-distances between the devices. A recent example of this approach is [34], in which the authors combine the theory of antenna and simulation technique to compute the electric field in the near-field and far-field of a finite-size RIS, and prove that an RIS is capable of acting as an anomalous mirror in the near-field of the array.

- **Physics-based modeling**: In this approach, the performance evaluation of a RIS starts from the root of the electromagnetic theory, namely the Maxwell equations, and, with aid of necessary tools such as geometrical optics and numerical approximations, develops into an accurate mathematical formulation. The advantage of this approach is that the mathematical formulation obtained by this approach can unveil the impact of certain system parameters to the performance, thus allowing one to optimize the RIS and the overall system performance. The formulation can also sometimes be obtained in a closed-form, in which case the performance of a RIS can be efficiently computed. In [35], the power measured from passive reflectors in the millimeter-wave frequency band is compared against ray tracing simulations. By optimizing the area of the surface that is illuminated, it is shown that a finite-size passive reflector can act as an anomalous mirror. The authors of [36] and [37] study the path-loss modeling of an RIS under the assumption of plane waves, which is essentially applicable in the far-field of the RIS. The author of [38] proposes a model that is applicable to continuous RISs, and holds in the near-field and far-field of the RIS. It is important to note, however, that the author in [38] specifically focuses on characterizing the available spatial degrees of freedom of two RISs communicating with each other, instead of RISs that are utilized for reflection or transmission. In [1], we propose a path-loss model that is applicable only to one-dimensional RISs that are deployed in a two-dimensional space. Also, the approach in [1] does not account for the vectorial nature of the electromagnetic waves.

1.5 Thesis Overview and Major Contributions

A huge portion of this thesis work revolves around the performance evaluation of RISs by using physic-based approach. In particular, we tackle the problem of path-loss modeling of a RIS and compare its performance against existing competing technology (e.g., relay). We also attempt to find the answer the problem of phase-shift optimization of RIS in order to achieve the most energy-efficient performance. The main contributions of this thesis are as follows:

1. This thesis provides the detailed introduction of physic-based modeling of RIS, including the theory of scattering of electromagnetic field, method of physical optics, and homogenized representation of a RIS.
2. This thesis introduce simple closed-form expressions to compute the power reflected from a RIS, by leveraging the general scalar theory of diffraction and the Huygens-Fresnel principle, as a function of the distance between the transmitter/receiver and the RIS, the size of the RIS, and the phase transformation applied by the RIS. Sufficient conditions under which an RIS acts as an anomalous mirror are also identified. In addition, using the formulation of the reflected power by a RIS, we compare the performance of RIS against full-duplex (FD) and half-duplex relay in terms of data rate.

3. This thesis proposes an approach, which adheres to the principles of physical optics, for calculating the free-space path-loss of an RIS-aided transmission link. The proposed path-loss model leverages the vector generalization of Green's theorem [39], and it is formulated in terms of a computable integral that depends on the transmission distances, the polarization of the radio waves, the size of the RIS, and the desired surface transformations. Based on the proposed model, the impact of several design parameters is unveiled, and the differences and similarities between the far-field and near-field asymptotic regimes are discussed.

1.6 Publications

Journal Papers The following is a list of publications in refereed journals produced from the research outcomes of this thesis. These journal papers are used as the basis for this thesis.

(*J*1) M. Di Renzo, K. Ntontin, J. Song, **F. H. Danufane**, X. Qian, F. Lazarakis, J. De Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, M. Debbah, G. Lerosey, M. Fink, S. Tretyakov, S. Shamai, "Reconfigurable Intelligent Surfaces vs. Relaying: Differences, Similarities, and Performance Comparison," in *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798-807, 2020.

Abstract: *Reconfigurable intelligent surfaces (RISs) have the potential of realizing the emerging concept of smart radio environments by leveraging the unique properties of metamaterials and large arrays of inexpensive antennas. In this article, we discuss the potential applications of RISs in wireless networks that operate at high-frequency bands, e.g., millimeter wave (30-100 GHz) and sub-millimeter wave (greater than 100 GHz) frequencies. When used in wireless networks, RISs may operate in a manner similar to relays. The present paper, therefore, elaborates on the key differences and similarities between RISs that are configured to operate as anomalous reflectors and relays. In particular, we illustrate numerical results that highlight the spectral efficiency gains of RISs when their size is sufficiently large as compared with the wavelength of the radio waves. In addition, we discuss key open issues that need to be addressed for unlocking the potential benefits of RISs for application to wireless communications and networks.*

(*J*2) **F. H. Danufane**, M. Di Renzo, J. de Rosny, and S. Tretyakov, "On the Path-Loss of Reconfigurable Intelligent Surfaces: An Approach Based on Green's Theorem Applied to Vector Fields", submitted to IEEE Transactions on Communications.

Abstract: In this paper, we introduce a physics-based analytical characterization of the free-space path-loss of a wireless link in the presence of a reconfigurable intelligent surface. The proposed approach is based on the vector generalization of Green's theorem. The obtained path-loss model can be applied to two-dimensional homogenized metasurfaces, which are made of sub-wavelength scattering elements and that operate either in reflection or transmission mode. The path-loss is formulated in terms of a computable integral that depends on the transmission distances, the polarization of the radio waves, the size of the surface, and the desired surface transformation. Closed-form expressions are obtained in two asymptotic regimes that are representative of far-field and near-field deployments. Based on the proposed approach, the impact of several design parameters and operating regimes is unveiled.

Conference Papers The following is the refereed conference that originated from the main findings of this thesis.

$(C1)$ M. Di Renzo, **F. H. Danufane**, X. Xi, J. de Rosny and S. Tretyakov, "Analytical Modeling of the Path-Loss for Reconfigurable Intelligent Surfaces – Anomalous Mirror or Scatterer ?," *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Atlanta, GA, USA, 2020

Abstract: Reconfigurable intelligent surfaces (RISs) are an emerging field of research in wireless communications. A fundamental component for analyzing and optimizing RIS-empowered wireless networks is the development of simple but sufficiently accurate models for the power scattered by an RIS. By leveraging the general scalar theory of diffraction and the Huygens-Fresnel principle, we introduce simple formulas for the electric field scattered by an RIS that is modeled as a sheet of electromagnetic material of negligible thickness. The proposed approach allows us to identify the conditions under which an RIS of finite size can be approximated as an anomalous mirror or as a scatterer. Numerical results are shown to confirm the proposed approach.

Journal Papers The following is a list of publications the author has contributed to and is not related to the topic of this thesis.

$(J1)$ **F. H. Danufane** and M. Di Renzo, “Latency Distribution Analysis in Wireless Networks: Stochastic Geometry Approach", to be submitted to IEEE Transactions on Wireless Communications.

Abstract: In this paper, we introduce analytical frameworks to compute the delay performances in homogeneous cellular network with the aid of stochastic geometry under various different coverage criterion. In particular, we derive the formula of the local delay, defined as the average number of time slots required for a successful transmission, and show its limitation, namely the fact that it is infinite under some condition and its inability to capture the distribution of delay among users in the network. Motivated by this limitation, we propose and develop the frameworks of several delay distribution metrics, each of which give a better insight on the delay performance of the network and is related to relevant performance metric in URLLC applications. To mitigate the inherent computation difficulties of the formulations, we employ efficient numerical approximations based on numerical inversion method, Riemann sum, and Beta distribution. Finally, we provide analysis of packet loss probability which represent the probability of a user experiencing an infinite waiting time.

2 Macroscopic Homogenized Modeling of Metasurfaces - An Introduction

This chapter introduces theoretical concepts that are necessary to understand the results in the subsequent chapters. This chapter is divided into two parts. The first part discusses the metasurfaces modeling where we move from a physics-based *microscopic* description of a metasurface and introduce a *macroscopic* representation for it, which is shown to be suitable for application in wireless communications. The second part introduces analytical approaches, based on the *theory of diffraction* and the *Huygens-Fresnel principle*, which allows us to compute the EM field at any point of a given volume that contains the metasurface.

2.1 Introduction 19

2.2 Surface Electromagnetics 20

  2.2.1 Definition of SEM 20

  2.2.2 From uniform to quasi-periodic metasurfaces 22

  2.2.3 Theoretical Foundation of Surface Electromagnetics 23

2.3 Modeling Metasurfaces: A Macroscopic Homogenized Approach 27

  2.3.1 From a microscopic to a macroscopic representation 27

  2.3.2 EM-Based Model of a Metasurface 28

  2.3.3 Microscopic Description of a Metasurface 31

  2.3.4 Macroscopic Description of a Metasurface 34

2.4 Modeling Radio Wave Propagation in the Presence of Metasurfaces 37

  2.4.1 Reference operating regimes 37

  2.4.2 Theory of Electromagnetic Diffraction 40

  2.4.3 Methods of Physical Optics 46

2.5 Conclusion 50

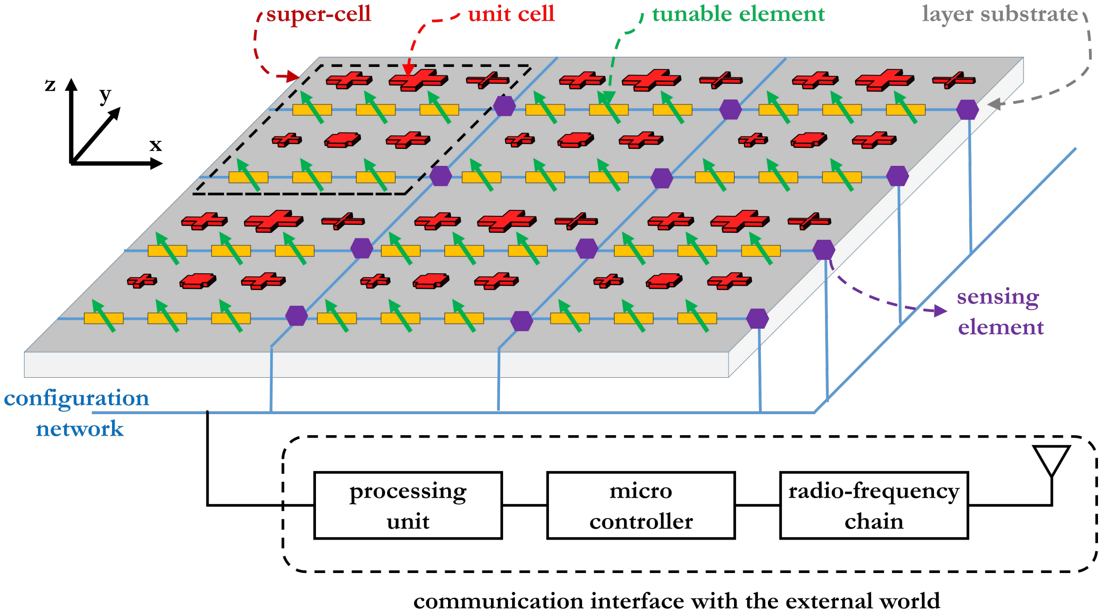


Figure 2.1 – Conceptual structure of a reconfiguration intelligent surface.

2.1 Introduction

As discussed in Chapter 1, unleashing the full potential of RISs requires a thorough understanding of their behavior in the context of communication systems. To this end, an accurate model of RISs is necessary to predict and optimize the performance of an RIS for given performance target. To achieve this objective, one of the most rigorous methods is to use physics-based modeling by using the theory of electromagnetic (EM) waves. As we will discuss later in this chapter, a large body of EM waves theory such as surface electromagnetics, scattering theories, and physical optics are very powerful analysis tools that can help us in obtaining an model of the RISs that is accurate yet still provides us insight on its performance limit in terms of relevant system parameters. For ease of understanding, the rest of this chapter is structured as follows:

- ***Surface Electromagnetics***. In this section, we briefly introduce the research discipline of surface electromagnetics (SEM), which is the enabling tool for modeling, analyzing, and synthesizing metasurfaces. The content of this sub-section is, in particular, based on [2]. We first give a definition of SEM and the three milestones that characterize the evolution of the spatial variations of a metasurface. We also provide the theoretical foundations of SEM which ultimately allows us to treat the metasurfaces as *homogenizable surface*, as explained later in detail.
- ***Modeling metasurfaces***. In this section, we move from a physics-based *microscopic* description of a metasurface and introduce a *macroscopic* representation for it, which is shown to be suitable for application in wireless communications. In particular, a

metasurface is represented by using continuous inhomogeneous functions that allow one to describe the signal transformations applied by a metasurface directly on the EM fields. The use of continuous functions is allowed, even though the (conceptual) physical structure of the metasurface in Fig. 2.1 is made of discrete unit cells (scatterers), because a metasurface can be homogenized. Therefore, its EM properties can be completely described through macroscopic parameters. The homogenized macroscopic representation of a metasurface based on continuous inhomogeneous functions is shown to be useful for the *synthesis* (i.e., to design a metasurface based on specified/desired signal transformations) and for the *analysis* (i.e., to compute the reflected and refracted EM fields in close proximity of a metasurface, for a given incident EM field and the physical structure of the metasurface) of RISs.

- ***Modeling radio waves propagation.*** In this section, we consider the interaction between the radio waves emitted by a source and a metasurface. We introduce analytical approaches, based on the *theory of diffraction* and the *Huygens-Fresnel principle*, which allows us to compute the EM field at any point of a given volume that contains the metasurface. In particular, we show that the EM field at any point of the volume can be formulated in terms of the EM field in close proximity of the metasurface (i.e., on its surface), as specified in the first sub-section.

2.2 Surface Electromagnetics

2.2.1 Definition of SEM

Electromagnetics is a fundamental discipline of sciences that describes the temporal and spatial behavior of the electric and magnetic fields. Broadly speaking, electromagnetics can be defined as the theory of EM fields and waves. SEM is a sub-discipline of electromagnetics. From the temporal point of view, electromagnetics is usually classified into different categories according to the oscillation frequency of the EM fields, such as direct current (DC), RF, microwaves, terahertz (THz), optics, X-rays, and beyond. The definition of SEM can be traced back to the classification of electromagnetics from the spatial point of view. Four regimes can be identified in the space domain, and each of them is usually modeled through *effective parameters*, which are viewed as an adequate simplification of Maxwell's equations under the corresponding spatial regime [2, Chapter 1].

- ***Zero-dimensional EM phenomena.*** This regime occurs when the spatial variations of a device or an EM phenomenon are much smaller than the wavelength of the radio waves in all three spatial dimensions. *Circuit theory* is considered to be an accurate and efficient approach for modeling zero-dimensional EM phenomena. The effective parameters that are usually employed in circuit theory are the resistor, the inductor, and the capacitor.

- ***One-dimensional EM phenomena.*** This regime occurs when the longitudinal dimen-

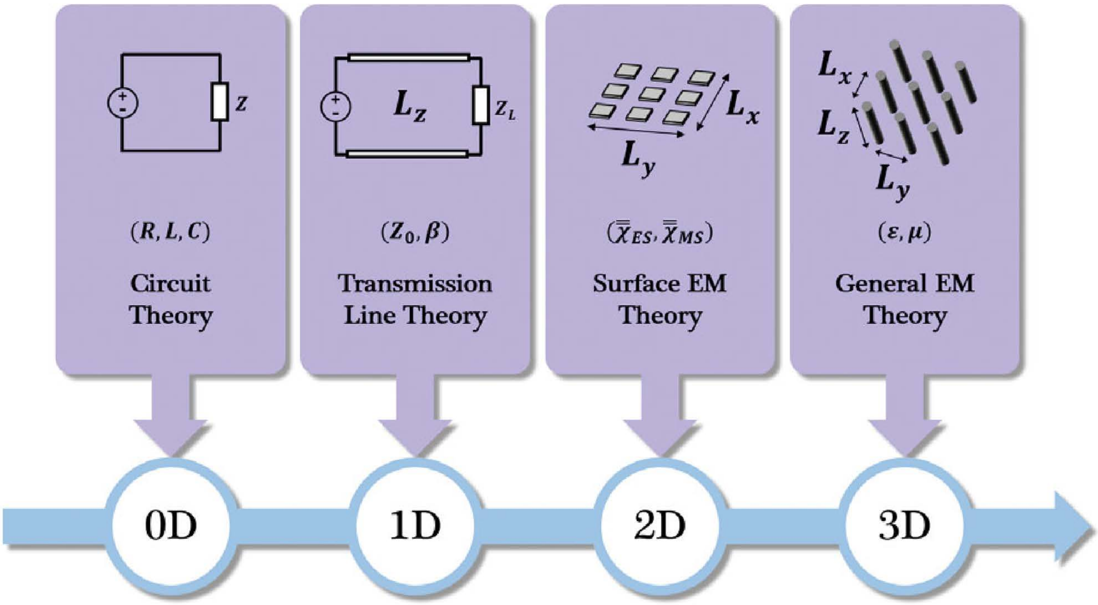


Figure 2.2 – Classification of electromagnetics in the spatial domain (reproduced from [2]).

sion and the transverse dimensions of a device or an EM phenomenon are comparable to and much smaller than, respectively, the wavelength of the radio waves. In this regime, circuit theory is no longer valid and it is replaced by *transmission line theory*. The effective parameters that are usually employed in transmission line theory are the characteristic impedance and the propagation constant.

- ***Two-dimensional EM phenomena***. This regime occurs when the longitudinal dimension and the transverse dimensions of a device or an EM phenomenon are much smaller than and comparable to, respectively, the wavelength of the radio waves. This is the regime of interest of this thesis and it will be comprehensively elaborated in further text through its corresponding effective (surface-averaged homogenized) parameters.

- ***Three-dimensional EM phenomena***. This is the most general regime in which the variations of the EM fields are comparable to the wavelength of the radio waves in all three spatial dimensions. Maxwell's equations are usually employed to analyze this regime. The effective parameters that are employed in this regime are the electric permittivity and magnetic permeability of the volumetric material.

A graphical comparison among these four operating spatial regimes is sketched in Fig. 2.2. Broadly speaking, effective parameters allow one to (approximately) model natural and artificial structures and EM phenomena as a whole, instead of modeling their many constituent elements individually. Therefore, they are a convenient tool for studying complex EM phenomena and structures from the macroscopic point of view.

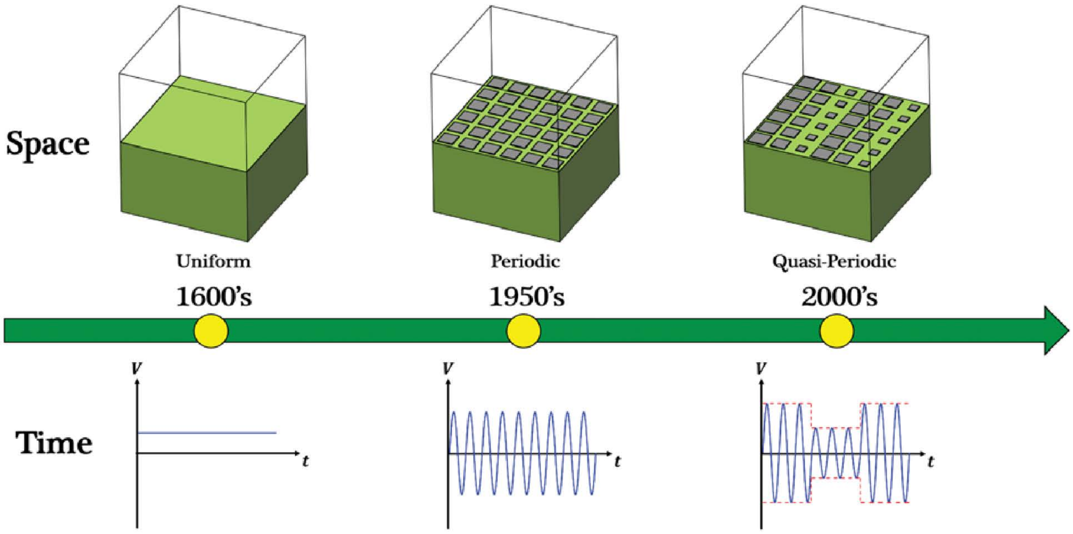


Figure 2.3 – The road to quasi-periodic metasurfaces (reproduced from [2]).

2.2.2 From uniform to quasi-periodic metasurfaces

In the context of SEM research and development, it is important to distinguish three milestones that characterize the evolution of the spatial variations along the transverse dimensions of a metasurface.

- ***Uniform metasurfaces***. By definition, (natural) surfaces are uniform surfaces. They are characterized by variations of the properties of the medium, which surrounds the surface, along the longitudinal direction. The properties of the surface do not change, on the other hand, along with the tangential directions, i.e., on the surface itself ($z = 0$ in Fig. 2.1).
- ***Periodic metasurfaces***. In contrast to uniform surfaces, periodic surfaces exhibit spatial variations along the tangential directions. Concretely, this implies that the unit cells in Fig. 2.1 are arranged in a periodic lattice and each unit cell has the same geometry and size.
- ***Quasi-periodic metasurfaces***. Similar to periodic surfaces, quasi-periodic surfaces exhibit spatial variations along with the tangential directions. In contrast to periodic surfaces, the unit cells are still arranged in a periodic lattice, but each unit cell differs from the others in terms of, e.g., geometrical variations, shape, size, orientation angle, etc. This is the configuration reported in Fig. 2.1.

The evolution from the uniform, periodic, and quasi-periodic feature of the transverse dimensions of a metasurface can be considered to be equivalent to the evolution of the analysis of signals in circuits, i.e., from DC signals, to alternate current (AC) signals, and eventually to modulated signals. A graphical comparison among the uniform, periodic, and quasi-periodic design structures is reported in Fig. 2.3. The quasi-periodic structure of the smart surface in Fig. 2.1 is, therefore, apparent and completely justified.

The core concept behind the development and widespread utilization of quasi-periodic metasurfaces lies in their more versatile, unprecedented, and exotic capabilities of manipulating EM waves. Naively, the operating principle of quasi-periodic metasurfaces can be thought of as a two-step manipulation of the radio waves: (i) first, each unit cell manipulates the incident radio waves locally and individually; and (ii) next, the spatial distribution of the obtained wave manipulations collectively alters the wavefront of the incident radio waves, leading to extraordinary EM responses and transformations of EM fields.

2.2.3 Theoretical Foundation of Surface Electromagnetics

Volumetric (bulk) vs. surface effective parameters

The behavior and the properties of metamaterials, which are volumetric engineered structures, are determined by the electric and magnetic properties of its constituent scatterers. The traditional and most convenient approach for modeling metamaterials is the *effective medium theory* (EMT) that pertains to the theoretical and analytical models and methods for describing the macroscopic properties of composite materials.

At the constituent level, composite materials can be viewed as a micro-inhomogeneous medium. The precise calculation and analysis of the many elements that constitute a composite material is a nearly impossible task. EMT provides one with a set of effective parameters that describe, approximately, a composite material as a whole. These parameters are, e.g., the effective electric permittivity and the effective magnetic permeability of the composite material, which are obtained by averaging (over a small volume) the response of the multiple constituents that compose the material. More precisely, the notions of electric permittivity and magnetic permeability result from the volumetric averaging of microscopic electric and magnetic currents over volumes that are small compared to the wavelength of the radio waves.

A similar bulk-parameters representation is, in general, not appropriate for metasurfaces [2, Chapter 3]. In this case, surface-averaged effective parameters are a more physically sound and accurate choice, as compared with volumetric-averaged effective parameters that account for an arbitrary non-zero thickness parameter in order to model the sub-wavelength thickness of metasurfaces. If a metasurface is, in particular, modeled by using an effective permittivity and effective permittivity, then these two parameters have to be infinite. This is because the reflection coefficient from a material slab tends to zero when its thickness tends to zero. Therefore, a model based on effective permittivities and permeabilities implies the need of having a surface of finite (non-zero) thickness.

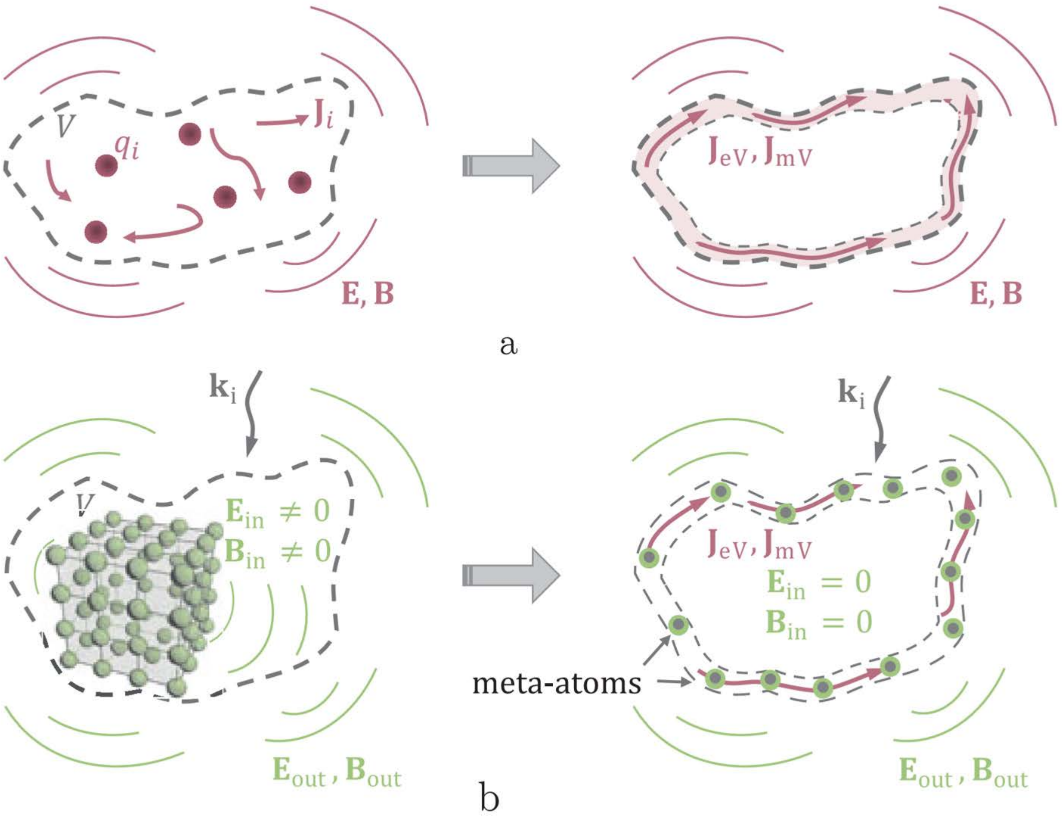


Figure 2.4 – Illustration of the surface equivalent theorem applied to the scattering of volumetric electromagnetic sources (a) and to the concept of metamaterials (b) (reproduced from [2]).

In detail, surface-averaged effective parameters result from the surface averaging of microscopic currents over a surface area of the order of the wavelength. This implies that the conventional notions of electric permittivity and magnetic permeability lose their meaning because there is no volume over which the EM fields and their corresponding induced currents can be averaged out. Surface averaging methods lead to the introduction of effective surface parameters, which include surface susceptibility functions and sheet impedances. These parameters are formally introduced later along with the concept of homogenized models for metasurfaces. We anticipate that surface-averaged effective parameters constitute an essential notion for developing macroscopic models and representations of metasurfaces.

The surface equivalence theorem

The surface equivalence theorem can be regarded as the physics-based foundation of SEM [2]. The basic idea of the surface equivalence theorem originates from (i) the *Huygens's principle*, which states that "each point on a primary wavefront can be considered to be a new source of a secondary spherical wave, and that a secondary wavefront can be constructed as the envelope of these secondary spherical waves", and (ii) the *uniqueness theorem*, which, in the context of this thesis, can be stated as "if the tangential electric and magnetic fields are completely known over a closed surface, the fields in the source-free region can be determined" [2, Chapter 1]. The surface equivalence theorem is illustrated in Fig. 2.4.

Let us consider a volume V that is filled with arbitrary sources of EM radiation, e.g., some charges q_i and some currents $\mathbf{J}_i$. These sources create an electric field, $\mathbf{E}$, and a magnetic induction field, $\mathbf{B}$, outside the volume V. According to the Huygens principle, the system of scatterers can be replaced by an arbitrarily thin layer of specific electric currents, $\mathbf{J}_{\mathrm{eV}}$, and magnetic currents, $\mathbf{J}_{\mathrm{mV}}$, that encloses the volume V. The thickness of the layer can be electrically small but non-zero since magnetic currents can be generated only via loops of electric currents with finite thickness. The equivalent (surface) currents $\mathbf{J}_{\mathrm{eV}}$ and $\mathbf{J}_{\mathrm{mV}}$ scatter EM fields only outward of the volume V, and these EM fields are the same as those created by the original system of sources. Such currents, which scatter EM fields only in one side, are referred to as Huygens's surfaces or Huygens's sources. This concept can be applied to metamaterials.

Let us consider an arbitrary volumetric metamaterial sample that is excited by an arbitrary external EM wave whose wavevector is $\mathbf{k}_i$. The external wave induces some charges q_i and currents $\mathbf{J}_i$ in the inclusions (or unit cells) of the sample, which irradiate secondary EM fields $\mathbf{E}_{\mathrm{out}}$ and $\mathbf{B}_{\mathrm{out}}$ into the space outside the volume V that encloses the metamaterial sample. According to the Huygens principle, one can replace the bulky metamaterial sample with induced polarization charges and currents by equivalent surface currents $\mathbf{J}_{\mathrm{eV}}$ and $\mathbf{J}_{\mathrm{mV}}$ that scatter the same EM fields $\mathbf{E}_{\mathrm{out}}$ and $\mathbf{B}_{\mathrm{out}}$ outside the volume V. By knowing these equivalent currents, one can determine appropriate topologies of unit cells (meta-atoms in the figure), which are placed along the surface of the volume V and generate, if illuminated by an external EM wave with wavevector $\mathbf{k}_i$, the same currents $\mathbf{J}_{\mathrm{eV}}$ and $\mathbf{J}_{\mathrm{mV}}$ as those of the original volumetric setup. The resulting arrangement of unit cells is the metasurface structure that yields the same EM response as that of the original metamaterial sample.

Broadly speaking, therefore, the surface equivalence theorem states that volumetric metamaterials can be replaced by electrically thin and, in general, curved metasurfaces, i.e., structures whose surface is engineered to provide field transformation capabilities. The surface equivalence theorem is, therefore, the core tenet behind the development and evolution from metamaterials to metasurfaces.

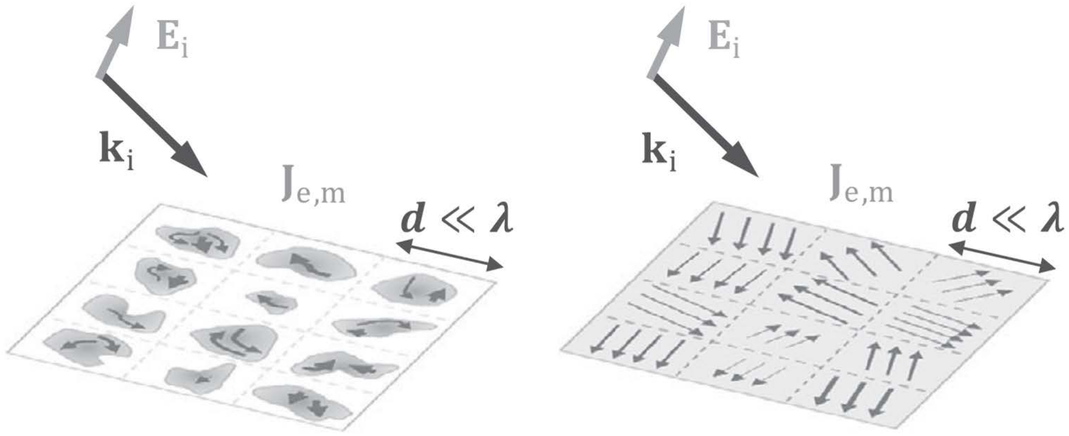


Figure 2.5 – Illustration of homogenization: A metasurface consisting of arbitrary sub-wavelength unit cells is modeled as a continuous sheet of surface-averaged (over the unit cell) currents (reproduced from [2]).

Homogenized modeling of metasurfaces

A metasurface can be homogenized owing to the sub-wavelength inter-distance among their constituent sub-wavelength unit cells. Concretely, this implies that the EM properties of a metasurface can be completely described by using macroscopic effective (surface) parameters, similar to the macroscopic effective (volumetric) parameters that describe three-dimensional materials, e.g., the effective permittivity and permeability. Knowledge of the macroscopic effective (surface) parameters of a metasurface allows one to formulate its EM response to arbitrary (in terms of wavefront, incidence angle, and polarization) impinging EM waves. This implies that macroscopic models for metasurfaces, which are described in the next subsections, do not assume that the impinging EM fields are plane wave. It is necessary, however, that the incident EM fields do not change significantly over the scale of one unit cell. This condition is usually fulfilled in metasurface structures whose unit cells have sub-wavelength inter-distances and sub-wavelength sizes. It is worth mentioning that, on the other hand, it is not possible to introduce effective macroscopic parameters for non-homogenizable structures, even if such structures are electrically thin.

Constructing a homogenization model corresponds to (i) determining the effective parameters that can appropriately describe the response of a metasurface to incident EM fields and (ii) identifying such effective parameters from the physical and EM properties of the unit cells. The last step can be performed by using experimental data or simulations of the metasurface response as well. Homogenization is a fundamental prerequisite for understanding, modeling, and designing metasurfaces. By using effective parameters, the complexity of the problem is remarkably reduced, since modeling the collective response of many individually small unit cells is a complex brute-force numerical optimization problem. In (volumetric) metamaterials, as mentioned, homogenized models are obtained through effective parameters that are averaged over small volumes that contain many unit cells. Due to the sub-wavelength thickness of metasurfaces, homogenized models are obtained through effective parameters that are averaged over small surface areas (whose dimensions are of the order of one wavelength) that contain several unit cells. If the impinging EM fields are plane waves, surface averaging over the area of one unit cell is usually sufficient.

The concept of homogenized equivalent of a metasurface is sketched in Fig. 2.5. The figure depicts a metasurface that consists of an array of sub-wavelength unit cells. Since the inter-distance between adjective unit cells is sufficiently small compared to the wavelength of the impinging EM fields, the incident radio waves can be assumed to be uniform or homogeneous along the size of a single unit cell. The impinging EM fields induce electric and magnetic currents in each unit cell. In homogenized models, the position-dependent induced electric and magnetic currents in each unit cell are replaced by surfaced-averaged (hence macroscopic) effective parameters. By using homogenization, therefore, a metasurface is modeled as a planar thin sheet of electric and magnetic surface-averaged current densities or, as described in further text, polarization surface densities. This implies that the modeling and analysis of the scattering from a metasurface that is made of finite-size (or discrete) unit cells is turned into the modeling and analysis of the scattering from an equivalent continuous sheet of electric and magnetic currents.

2.3 Modeling Metasurfaces: A Macroscopic Homogenized Approach

Starting from the introductory material on SEM, this section is focused on reporting tractable analytical models for metasurfaces. Special emphasis is, in particular, put on the theory that leads from a microscopic to a macroscopic representation of metasurfaces through surface-averaged homogenized effective parameters.

2.3.1 From a microscopic to a macroscopic representation

We consider an RIS made of a metasurface that is electrically thin, electrically large, homogenizable, and has a sub-wavelength structure. Under these assumptions, we begin our development by introducing a microscopic description of a metasurface that is based on and accounts for the spatial coupling between adjacent unit cells, which is due to their sub-wavelength inter-distance. The obtained physics-based model may not, however, be sufficiently flexible for designing and analyzing complex metasurfaces. Subsequently, for this reason, we introduce a macroscopic description of a metasurface, which is based on continuous tensor functions, that yields a more suitable representation of a metasurface, especially for application to wireless networks. The macroscopic approach, however, still allows us to obtain a direct connection with the material parameters, i.e., the surface susceptibility of the smart surface.

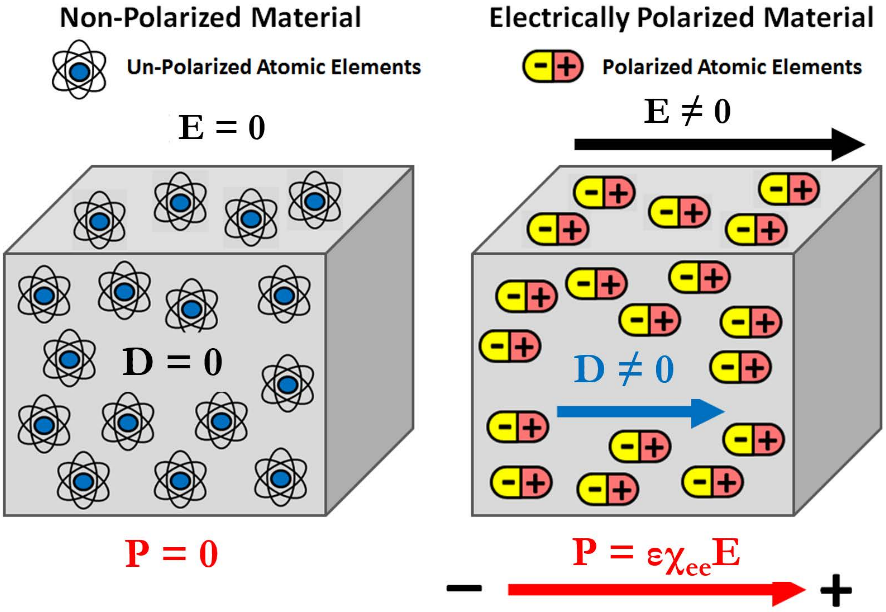


Figure 2.6 – Example of induced electric dipole moment and electric polarization density (**P**) in volumetric materials.

It is worth mentioning that the approach reported in this sub-section is widely employed in the fields of electromagnetism and metamaterials, and our treatise is mainly based on the theories available in [40], [41], [42], [43], [44]. These methods are, however, not necessarily known to wireless researchers. The main objective of this section is to fill this gap of knowledge in a simple but rigorous manner.

2.3.2 EM-Based Model of a Metasurface

The departing point is the EM-based model of a metasurface reported in [40], [41].

A metasurface as a surface distribution of electrically small resonant scatterers

In the most general sense, a metasurface (also called a metafilm in [40], [41]) is viewed as a surface distribution of electrically small resonant scatterers, and it is characterized by electric and magnetic surface polarization densities. In simple terms, the electrically small scatterers are the unit cells in Fig. 2.1. In volumetric, three-dimensional materials, the electric polarization density is a vector field that expresses the density of permanent or induced electric dipole moments in a dielectric material. When a dielectric is placed in an external electric field, its molecules gain electric dipole moment and the dielectric is said to be polarized. The electric dipole moment induced per unit volume of the dielectric material is referred to as the electric polarization of the dielectric. The electric polarization density describes how a material responds to an applied electric field and how the material changes the electric field. More precisely, an external electric field that is applied to a dielectric material causes a displacement of bound charged elements, which are elements bound to molecules and, hence, are not free to move around the material. Positive charged elements are displaced in the direction of the electric field and negative charged elements are displaced in the opposite direction of the field. The molecules may remain neutral in charge, but an electric dipole moment is formed. The electric polarization density corresponds to the induced dipole moment per unit volume of the material. For ease of understanding, this concept is sketched in Fig. 2.6.

Similarly, the magnetic polarization density describes how a material responds to an applied magnetic field. To model metasurfaces as sheets of electrically negligible thickness, the concept of electric and magnetic surface polarization densities is used. The electric (magnetic) surface polarization density is the electric (magnetic) dipole moment per unit area of the surface, while the usual volumetric electric (magnetic) polarization density is the electric (magnetic) dipole moment per unit volume.

A metasurface as an array of polarizable unit cells

A metasurface can be broadly defined as an array of polarizable unit cells that induce discontinuities of the electric and magnetic fields at the two sides of the surface, i.e., at $z = 0^+$ and $z = 0^-$ in Fig. 2.1. The objective of a macroscopic description of a metasurface consists of formulating and expressing the discontinuities of the electric and magnetic fields as a function of surface-averaged electric and magnetic surface polarization densities. Since a metasurface is homogenizable, i.e., the distance between adjacent unit cells is much smaller than the wavelength of the radio waves, the approach consists of replacing the discrete distribution of unit cells with a continuous distribution, which, in turn, results in a continuous function of electric and magnetic surface polarization densities. It is worth nothing that the unit cells may be of arbitrary shape and are not infinitely thin. Their thickness is only required to be small in comparison with the wavelength of the radio waves.

Towards a macroscopic description: Generalized sheet transition conditions

A macroscopic description of a metasurface consists of replacing the actual spatial distribution of the unit cells on the smart surface with analytical relations between the electric and magnetic fields at the two sides of the metasurface (at $z = 0^+$ and $z = 0^-$ in Fig. 2.1). Conceptually, this equivalence is illustrated in Fig. 2.7. The size, the shape, and the physical characteristics of the unit cells are directly incorporated into these analytical relations through the homogenized (continuous) electric and magnetic surface polarization densities. The analytical relations that express the electric and magnetic fields at the two sides of a metasurface as functions of the homogenized (continuous) electric and magnetic surface polarization densities are referred to as ***generalized sheet transition conditions (GSTCs)***. As far as the GSTCs are concerned, three remarks are worth mentioning.

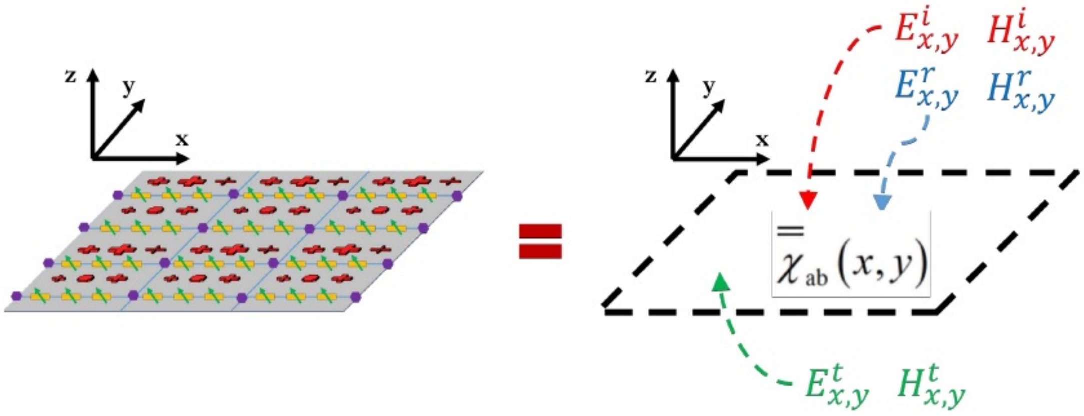


Figure 2.7 – Equivalent analytical representation of a metasurface: A zero-thickness material sheet whose EM properties are modeled through surface susceptibility functions that introduce EM discontinuities (or "jumps") quantified by the generalized sheet transition conditions.

- ***The tangential components of the EM fields are sufficient.*** The electric and magnetic fields that appear in the GSTCs involve only their tangential (or transverse) components across the surface. In other words, the Cartesian or longitudinal component of the EM fields along the z-axis in Fig. 2.1 does not explicitly appear. This is because they can be uniquely determined by the transverse components by virtue of the uniqueness theorem [42].

- ***The EM fields have variations on scales larger than the wavelength.*** The electric and magnetic fields that appear in the GSTCs do not exhibit variations on a length scale that is comparable with the dimension of the unit cells and with their inter-distance. They only exhibit variations on a length scale that is larger than the wavelength of the radio waves. In this thesis, however, the corresponding model is still referred to as microscopic because it requires an appropriate definition of the spatial coupling between adjacent unit cells [40]. We use the term macroscopic modeling to identify similar analytical relations in which this coupling is implicitly taken into account by using surface-averaged susceptibility tensor functions [42]. This concept is better elaborated in further text.

- ***Generalization of conventional interface conditions.*** The GSTCs, which best characterize a metasurface as an EM discontinuity, constitute a generalization of conventional interface conditions for the EM fields at the interface of two media. For example, it is known that the tangential components of the electric field are continuous at the interface of two media. This is not true if the interface is constituted by a zero-thickness metasurface that acts as an EM discontinuity. The relations between the tangential components of the electric and magnetic fields are, in fact, specified by the GSTCs. In general, the tangential components of the electric and magnetic fields are different at the two sides of a metasurface, i.e., they are discontinuous. It is usual jargon to say that the GSTCs formulate the "jumps" (i.e., discontinuities) of the electric and magnetic fields at the two sides of a metasurface.

2.3.3 Microscopic Description of a Metasurface

After introducing the general definition of GSTCs and the general meaning of macroscopic homogenized description of a metasurface, we are ready to formulate these two concepts in analytical terms. To this end, we depart from the physical structure of a metasurface and its microscopic representation.

Metasurface structure

We consider a two-dimensional metasurface that lies on the xy-plane at $z = 0$. The metasurface is centered at the origin and has a finite size equal to $2L_x$ and $2L_y$ along the x-axis and y-axis, i.e., $-L_x \leq x \leq L_x$ and $-L_y \leq y \leq L_y$, respectively. Without loss of generality, we assume that the incidence-reflection side of the metasurface is $z = 0^+$ and the transmission side of the metasurface is $z = 0^-$. Also, we assume that the medium surrounding the metasurface is vacuum. Our analytical formulation is mostly based on [42], [43], [44]. For simplicity, we consider the sufficient general case of metasurfaces whose longitudinal components of the electric ($\mathbf{P}$) and magnetic ($\mathbf{M}$) surfaces polarization densities are equal to zero, i.e., $P_z = 0$ and $M_z = 0$, respectively. We show in further text that this assumption allows one to design metasurfaces that realize EM-based functions that are sufficiently general for wireless applications, e.g., reflection and refraction towards arbitrary directions. The advantage of this assumption is that closed-form analytical expressions for the EM fields are obtained.

Analytical formulation of GSTCs

Let us consider an EM field that illuminates a metasurface. Under the assumption $P_z = 0$ and $M_z = 0$, the GSTCs, i.e., the "jumps" of the electric and magnetic fields at the two sides of a metasurface, can be formulated as follows:

$$\begin{aligned} \hat{z} \times \Delta\mathbf{H}\left(x, y\right) &= j\omega \mathbf{P}_{\parallel}\left(x, y\right) \\ \Delta\mathbf{E}\left(x, y\right) \times \hat{z} &= j\omega\mu \mathbf{M}_{\parallel}\left(x, y\right) \end{aligned} \tag{2.1}$$

where $\mathbf{P}_{\parallel}\left(x, y\right)$ and $\mathbf{M}_{\parallel}\left(x, y\right)$ are the longitudinal components of the surface-averaged electric and magnetic surface polarization densities, respectively, and $\Delta\mathbf{E}\left(x, y\right)$ and $\Delta\mathbf{H}\left(x, y\right)$ are the differences between the surface-averaged electric and magnetic fields at the two sides of the metasurface, whose Cartesian components can be written, respectively, as follows:

$$\begin{aligned} \Delta\Psi_u\left(x, y\right) &= \hat{\mathbf{u}} \cdot \Delta\Psi\left(x, y\right)\Big|_{z=0^-}^{z=0^+} \\ &= \Psi_u^t\left(x, y\right) - \left(\Psi_u^i\left(x, y\right) + \Psi_u^r\left(x, y\right)\right) \end{aligned} \tag{2.2}$$

where $\Psi = \{\mathbf{E}, \mathbf{H}\}$, $u = \{x, y, z\}$, and the superscripts i, r and t denote the incident, reflected, and transmitted components of the fields, respectively.

Computing P and M

The GSTCs depend on the longitudinal components of the electric and magnetic surface polarization densities. In order to formulate and compute them, we need to investigate the working operation of a metasurface at the microscopic level and we need to introduce the concept of acting (or local) fields.

Acting (or local) field: Excitation field at the location of an individual unit cell

At the microscopic level, a metasurface is viewed as an array of electrically small polarizable unit cells. When an EM field illuminates a metasurface, the unit cells get polarized and, since they are small (sub-wavelength) in size, the induced polarizations of each individual unit cell can be modeled through induced electric and magnetic dipole moments (see Fig. 2.6). Since the unit cells are arranged in a dense array, i.e., the inter-distance between adjacent unit cells is sub-wavelength, knowing the polarization induced on a single unit cell is not sufficient to determine the response of the entire metasurface. This is because the EM field that excites a unit cell that occupies a specific location, which is referred to as the acting or local EM field, is given by the summation of the incident EM field and the, so-called, interaction EM field, which is the field created by the induced electric and magnetic dipoles of all the unit cells of the metasurface with the exception of the unit cell under analysis.

Acting field: Modeling the mutual coupling among unit cells

By definition, therefore, the acting electric and magnetic fields account for the contribution of all the unit cells of a metasurface with the exception of the given unit cell under consideration. This implies that they account for the mutual (spatial) coupling among the unit cells of a metasurface and that they yield a microscopic description of a metasurface, since the acting fields are referred to individual unit cells. If the acting fields are known, the electric (**P**) and magnetic (**M**) surfaces polarization densities in (2.1) can be formulated as follows [40]:

$$\mathbf{P}\left(x,y\right) = \varepsilon N \langle \overline{\overline{\alpha}}_{\text{ee}}\left(x,y\right)\rangle \mathbf{E}_{\text{act}}\left(x,y\right) + \sqrt{\mu\varepsilon} N \langle \overline{\overline{\alpha}}_{\text{em}}\left(x,y\right)\rangle \mathbf{H}_{\text{act}}\left(x,y\right) \tag{2.3}$$

$$\mathbf{M}\left(x,y\right) = \sqrt{\varepsilon/\mu} N \langle \overline{\overline{\alpha}}_{\text{me}}\left(x,y\right)\rangle \mathbf{E}_{\text{act}}\left(x,y\right) + N \langle \overline{\overline{\alpha}}_{\text{mm}}\left(x,y\right)\rangle \mathbf{H}_{\text{act}}\left(x,y\right) \tag{2.4}$$

where N is the number of unit cells per unit area, and $\langle \overline{\overline{\alpha}}_{\text{ab}}\left(x,y\right)\rangle$, for $a = \{\text{e},\text{m}\}$ and $b = \{\text{e},\text{m}\}$, are the average (electric-electric, electric-magnetic, magnetic-electric, and magnetic-magnetic) polarizability dyadics of the unit cell in which the electric and magnetic surface polarization densities are computed. In particular, the average $\langle\cdot\rangle$ is calculated over the unit cells in the vicinity of the unit cell being considered and where the electric and magnetic surfaces polarization densities are evaluated. The polarizability dyadics depend on the physics and EM properties of the unit cells that can be obtained from analysis, simulations, or measurements.

Relating local (microscopic) fields to average (macroscopic) fields

Based on (2.3) and (2.4), the electric and magnetic surface polarization densities can be formulated in terms of the acting electric ($\mathbf{E}_{\mathrm{act}}\left(x,y\right)$) and magnetic ($\mathbf{H}_{\mathrm{act}}\left(x,y\right)$) fields at the position of a given unit cell. By definition, the acting fields can be formulated as follows [40]:

$$\begin{aligned}\mathbf{E}_{\mathrm{act}}\left(x,y\right)&=\mathbf{E}_{\mathrm{av}}\left(x,y\right)-\mathbf{E}^{\text{unit cell}}_{\text{scattering}}\left(x,y\right)\\ \mathbf{H}_{\mathrm{act}}\left(x,y\right)&=\mathbf{H}_{\mathrm{av}}\left(x,y\right)-\mathbf{H}^{\text{unit cell}}_{\text{scattering}}\left(x,y\right)\end{aligned}\tag{2.5}$$

where $\mathbf{E}_{\mathrm{av}}\left(x,y\right)$ and $\mathbf{H}_{\mathrm{av}}\left(x,y\right)$ are the average electric and magnetic fields, respectively, at the two sides of the metasurface, which are defined as follows:

$$\begin{aligned}E_{\mathrm{av},u}\left(x,y\right)&=\frac{E^{t}_{u}\left(x,y\right)+\left(E^{i}_{u}\left(x,y\right)+E^{r}_{u}\left(x,y\right)\right)}{2}\\ H_{\mathrm{av},u}\left(x,y\right)&=\frac{H^{t}_{u}\left(x,y\right)+\left(H^{i}_{u}\left(x,y\right)+H^{r}_{u}\left(x,y\right)\right)}{2}\end{aligned}\tag{2.6}$$

where $u=\left\{x,y,z\right\}$, and $\mathbf{E}^{\text{unit cell}}_{\text{scattering}}\left(x,y\right)$ and $\mathbf{H}^{\text{unit cell}}_{\text{scattering}}\left(x,y\right)$ are the electric and magnetic fields scattered by the single unit cell under consideration, respectively.

Contribution of a single unit cell

In practice, the contribution of a single unit cell can be calculated by expressing the unit cell as a combination of electric and magnetic dipoles that are contained within a small disk of a given radius. The electric and magnetic fields scattered by the resulting small disk are, by definition, $\mathbf{E}^{\text{unit cell}}_{\text{scattering}}\left(x,y\right)$ and $\mathbf{H}^{\text{unit cell}}_{\text{scattering}}\left(x,y\right)$.

Surface-averaged fields

It is worth emphasizing that (2.5) relates local, hence microscopic, fields to average, hence macroscopic, fields. More specifically, $\mathbf{E}_{\mathrm{av}}\left(x,y\right)$ and $\mathbf{H}_{\mathrm{av}}\left(x,y\right)$ in (2.6) are surface-averaged fields, where the average is computed over a small surface area of the order of the wavelength (including one or more unit cells depending on the setup). In other words, the rapid variations of the EM fields over distances of the order of the typical separation between adjacent unit cells along the surface are eliminated (i.e., averaged out) in macroscopic EM fields. This is not the case for microscopic EM fields.

The need for a macroscopic representation

By inserting (2.3) and (2.4) into (2.1), one can obtain the analytical relations between the transverse components of the electric and magnetic fields at the two sides of a metasurface. The issue with this approach is that the acting electric and magnetic fields need to be estimated, which depend on a specific unit cell, i.e., $\mathbf{E}_{\text{scattering}}^{\text{unit cell}}(x, y)$ and $\mathbf{H}_{\text{scattering}}^{\text{unit cell}}(x, y)$, and may not be easy to deal with. This issue is overcome next, by introducing an explicit macroscopic description of a metasurface, which depends only on the average (macroscopic) fields.

2.3.4 Macroscopic Description of a Metasurface

Explicit macroscopic formulation

In [40], the authors succeed in calculating a closed-form expression for $\mathbf{E}_{\text{scattering}}^{\text{unit cell}}(x, y)$ and $\mathbf{H}_{\text{scattering}}^{\text{unit cell}}(x, y)$. Based on their analytical formulation, the electric and magnetic surface polarization densities in (2.3) and (2.4) can be formulated in terms of the average electric and magnetic fields in (2.6), which, in contrast with the acting fields, yield a macroscopic description of a metasurface that is viewed as an electromagnetic discontinuity in space. This is because the average fields do not exclude the fields generated by individual unit cells.

Surface-averaged susceptibility functions: The effective parameters of metasurfaces

More specifically, a macroscopic description of a metasurface, which depends only on average (macroscopic) electric and magnetic fields, can be obtained by formulating the electric and magnetic surface polarization densities in terms of *susceptibility tensors* rather than in terms of polarizabilities. Similar to the average fields, the susceptibility tensors are macroscopic surface-averaged parameters. They can can be viewed as the effective parameters of metasurfaces, in analogy with the electric permittivity and magnetic permeability of volumetric metamaterials, and they depend on the topology and geometry of the unit cells, the inter-distance among the unit cells, the properties of the material, and the wavelength. In particular, the following relations hold true [42]:

$$\mathbf{P}(x, y) = \varepsilon \overline{\overline{\chi}}_{\text{ee}}(x, y) \mathbf{E}_{\text{av}}(x, y) + \sqrt{\mu\varepsilon} \overline{\overline{\chi}}_{\text{em}}(x, y) \mathbf{H}_{\text{av}}(x, y) \tag{2.7}$$

$$\mathbf{M}(x, y) = \sqrt{\varepsilon/\mu} \overline{\overline{\chi}}_{\text{me}}(x, y) \mathbf{E}_{\text{av}}(x, y) + \overline{\overline{\chi}}_{\text{mm}}(x, y) \mathbf{H}_{\text{av}}(x, y) \tag{2.8}$$

where $\overline{\overline{\chi}}_{\text{ab}}(x, y)$, for $a = \{\text{e}, \text{m}\}$ and $b = \{\text{e}, \text{m}\}$, are the (electric-electric, electric-magnetic, magnetic-electric, and magnetic-magnetic) surface-averaged (or simply surface) susceptibility dyadics, which are the macroscopic quantities of interest for a simple but accurate synthesis and analysis of metasurfaces.

GSTCs: A macroscopic algebraic formulation

With the aid of the analytical formulation of the electric and magnetic surface polarization densities as a function of only macroscopic fields and surface susceptibility dyadics, one can obtain an explicit analytical (algebraic) formulation of the GSTCs. More specifically, by inserting (2.7) and (2.8) into (2.1), the following explicit relation can be obtained (for ease of notation, the dependency on (x, y) for all fields and susceptibility dyadics is explicitly omitted but implied):

$$\begin{bmatrix} -H_y^t + H_y^i + H_y^r \\ H_x^t - H_x^i - H_x^r \end{bmatrix} = \frac{j\omega\varepsilon}{2} \begin{bmatrix} \chi_{\text{ee}}^{xx} & \chi_{\text{ee}}^{xy} \\ \chi_{\text{ee}}^{yx} & \chi_{\text{ee}}^{yy} \end{bmatrix} \begin{bmatrix} E_x^t + E_x^i + E_x^r \\ E_y^t + E_y^i + E_y^r \end{bmatrix} \tag{2.9}$$

$$+ \frac{j\omega\sqrt{\mu\varepsilon}}{2} \begin{bmatrix} \chi_{\text{em}}^{xx} & \chi_{\text{em}}^{xy} \\ \chi_{\text{em}}^{yx} & \chi_{\text{em}}^{yy} \end{bmatrix} \begin{bmatrix} H_x^t + H_x^i + H_x^r \\ H_y^t + H_y^i + H_y^r \end{bmatrix} \tag{2.10}$$

$$\begin{bmatrix} E_y^t - E_y^i - E_y^r \\ -E_x^t + E_x^i + E_x^r \end{bmatrix} = \frac{j\omega\mu}{2} \begin{bmatrix} \chi_{\text{mm}}^{xx} & \chi_{\text{mm}}^{xy} \\ \chi_{\text{mm}}^{yx} & \chi_{\text{mm}}^{yy} \end{bmatrix} \begin{bmatrix} H_x^t + H_x^i + H_x^r \\ H_y^t + H_y^i + H_y^r \end{bmatrix} \tag{2.11}$$

$$+ \frac{j\omega\sqrt{\mu\varepsilon}}{2} \begin{bmatrix} \chi_{\text{me}}^{xx} & \chi_{\text{me}}^{xy} \\ \chi_{\text{me}}^{yx} & \chi_{\text{me}}^{yy} \end{bmatrix} \begin{bmatrix} E_x^t + E_x^i + E_x^r \\ E_y^t + E_y^i + E_y^r \end{bmatrix} \tag{2.12}$$

where, for $a = \{\text{e}, \text{m}\}$ and $b = \{\text{e}, \text{m}\}$, we have introduced the notation:

$$\overline{\overline{\chi}}_{\text{ab}}(x, y) = \begin{bmatrix} \chi_{\text{ab}}^{xx}(x, y) & \chi_{\text{ab}}^{xy}(x, y) \\ \chi_{\text{ab}}^{yx}(x, y) & \chi_{\text{ab}}^{yy}(x, y) \end{bmatrix} \tag{2.13}$$

GSTCs: A mathematical tool for analysis and synthesis

The analytical expressions in (2.9) and (2.11) provide one with the necessary equations for designing metasurfaces, which are capable of applying specified transformations to the impinging radio waves, and for analyzing the performance of wireless networks in the presence of metasurfaces. We close this sub-section by emphasizing four important properties of the constitutive relations in (2.9) and (2.11):

1. the constitutive relations are formulated only in terms of macroscopic surface-averaged EM fields and macroscopic surface-averaged effective parameters;
2. the analysis of the surface-averaged susceptibility functions allows one to understand whether the metasurface is locally lossy, i.e., their imaginary part is always negative, or whether some local gains are needed, i.e., their imaginary part takes positive values;
3. even though the constitutive relations offer a macroscopic description of a metasurface, they inherently account for the mutual coupling among the unit cells, since they originate from a microscopic description of a metasurface. In particular, the mutual coupling between adjacent unit cells is implicitly taken into account in the surface-averaged

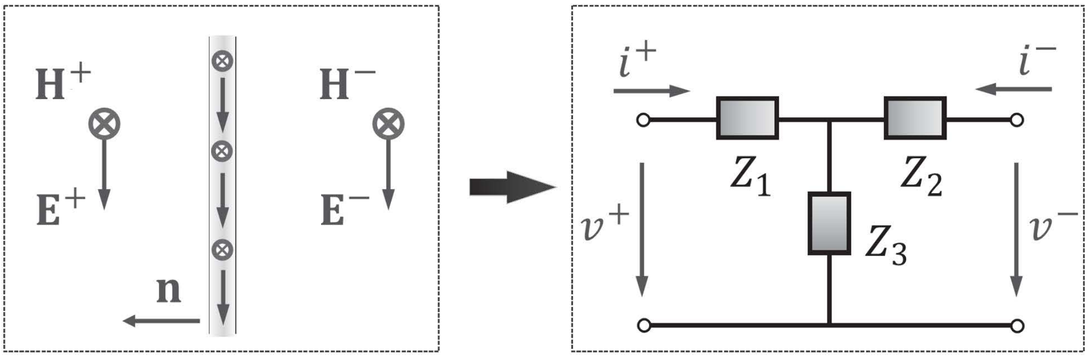


Figure 2.8 – Homogenization modeling of a metasurface based on surface-averaged impedances (reproduced from [2]).

susceptibility matrices in (2.13).

4. the constitutive relations in (2.9) and (2.11) can be applied to formulate the EM response of a metasurface to arbitrary (in terms of wavefront, incidence angle, and polarization) impinging EM waves. This implies that (2.9) and (2.11) do not assume that the impinging EM fields are plane waves. It is necessary, however, that the incident EM fields do not change significantly over the scale of one unit cell. This condition is usually fulfilled in metasurface structures whose unit cells have sub-wavelength inter-distances and sub-wavelength sizes.

In summary, from an analytical point of view, a metasurface can be approximately replaced by its corresponding constitutive relations in (2.9) and (2.11), which are sufficient for the analysis and design of wireless networks (see Fig. 2.7).

Equivalent representation in terms of surface-averaged impedances

An alternative homogenization model for a metasurface can be obtained by establishing an analogy between the propagation of radio waves in vacuum and the propagation of signals in transmission lines [2, Chapter 2]. In this analogy, more precisely, the electric and magnetic fields of a radio wave that propagates in vacuum are matched to the voltages and currents of a signal that propagates in an equivalent transmission line, respectively. The impedances that determine the equivalent transmission line constitute a macroscopic surface-averaged representation of a metasurface, which is equivalent to the representation of a metasurface in terms of surface-averaged susceptibility functions.

An example of representation of a metasurface in terms of transmission line equivalent is reported in Fig. 2.8. The T-circuit transmission line in Fig. 2.8 ensures, in particular, that there exist “jumps” (discontinuities) of both the electric and magnetic fields, i.e., the voltages and currents represented in the figure, in agreement with the GSTCs in (2.1). By applying Kirchhoff's circuit laws to the equivalent transmission line representation of the metasurface, and by inserting the resulting equations in (2.9) and (2.11), one can compute analytical relations between the surface-averaged susceptibility functions and the surface-averaged impedances of the transmission line equivalent in Fig. 2.8. An example of this computation and an explicit relation between the two representations for a metasurface can be found in [45, Appendix].

The homogenization model formulated in terms of surface-averaged impedances provides one with useful engineering insights on the properties of a metasurface structure. For example, a metasurface is capacitive and inductive if the imaginary part of the impedances is negative and positive, respectively, and a metasurface is lossy and active if the real part of the impedances is positive and negative, respectively. As remarked in [45], more in general, a representation of a metasurface in terms of circuital parameters, i.e., surface-averaged impedances, may be potentially more useful to engineering-oriented communities, while a representation of a metasurface in terms of surface-averaged susceptibility functions may be more widely used in physics-oriented communities. Both representations are, however, homogenized, macroscopic, and equivalent with each other.

2.4 Modeling Radio Wave Propagation in the Presence of Metasurfaces

In the previous section, we have shown that a metasurface can be described in terms of surface susceptibility functions and that the electric and magnetic (surface) fields scattered by the metasurface can be formulated in an algebraic form under sufficiently general modeling assumptions. In this sub-section, we show that the knowledge of the EM field at the two sides of a metasurface, i.e., only at $z = 0^+$ and $z = 0^-$, is sufficient for computing the EM field at any point of a given volume.

2.4.1 Reference operating regimes

We first introduce some relevant operating regimes and concepts. To help us clarify them, the illustrations in Figs. 2.9 and 2.10 are utilized.

- ***Near-field vs. far-field.*** As depicted in Fig. 2.1, a metasurface is made of unit cells whose size is usually smaller than the wavelength of the impinging radio waves. At distances from the metasurface that are larger than two-three times the size of a unit cell, which is still small compared with the wavelength, the EM fields of the evanescent surface modes of the metasurface structure, which may be present, can be considered to be negligible already. From the macroscopic point of view, in particular, the radio waves impinging upon a metasurface can be assumed to be locally plane waves regardless of the actual

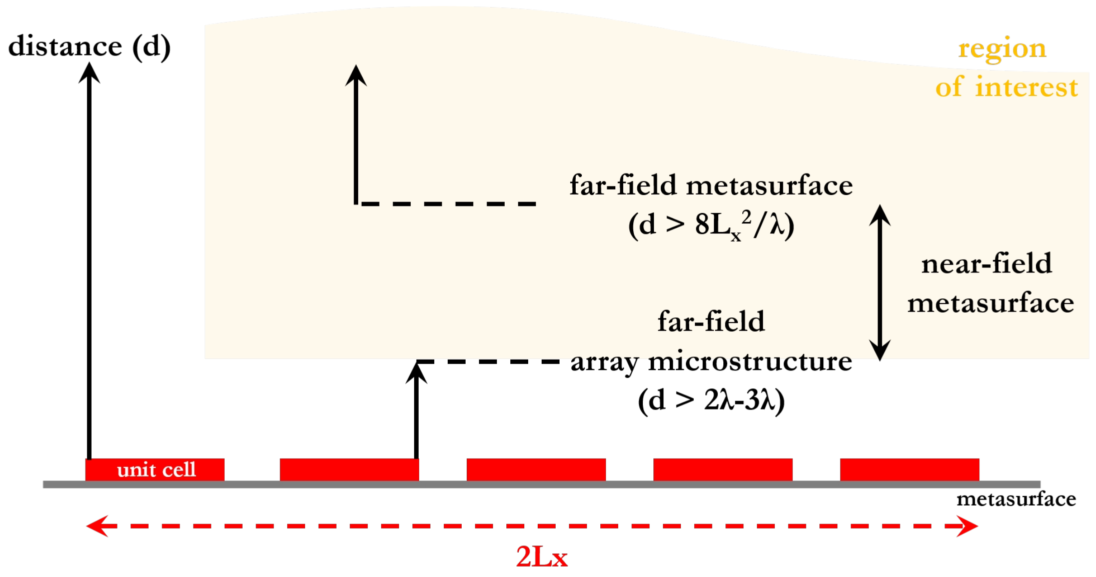


Figure 2.9 – Near-field vs. far-field of a metasurface and far-field of the array microstructure. The far-field boundary of the metasurface is obtained from the Fraunhofer distance.

characteristics of the source (e.g., the source may not necessarily emit plane waves). In simple terms, this implies that, e.g., a spherical wavefront impinging upon a unit cell can be approximated by its tangent provided that the size of the unit cell is sufficiently small that the phase of the impinging radio wave does not change along the unit cell. Since we are interested in EM fields that are evaluated at transmission distances from the metasurface that are larger than two-three times the size of a unit cell, we can ignore the evanescent fields that may be present along the metasurface structure, and we can utilize the surface-averaged reflection coefficient for locally plane waves. In this thesis, this region is referred to as “the far-field of the array microstructure”. Physically, as mentioned, this region corresponds to distances from the metasurface structure at which the evanescent EM fields are negligible. If we consider the example of an anomalous perfect reflector, this implies that the EM field scattered by the metasurface is approximately equal to the sum of two plane waves (as modeled in the previous example). At these distances, one can define and use the surface-averaged reflection coefficient $R_{\mathrm{EM}}(x)$ as defined in the previous example. As far as the whole metasurface structure is concerned, on the other hand, the operating regime is not unique. The transverse size of a metasurface can, in fact, be tens or hundreds (or more) times larger than the wavelength of the impinging radio waves. This implies that it is not possible to ignore, in general, the phase change of the impinging radio waves along the entire metasurface. Depending on the actual network geometry, the size of the metasurface, the operating wavelength, therefore, a metasurface can operate either in the near-field regime or in the far-field regime. For clarity, these operating regimes are conceptually illustrated in Fig. 2.9.

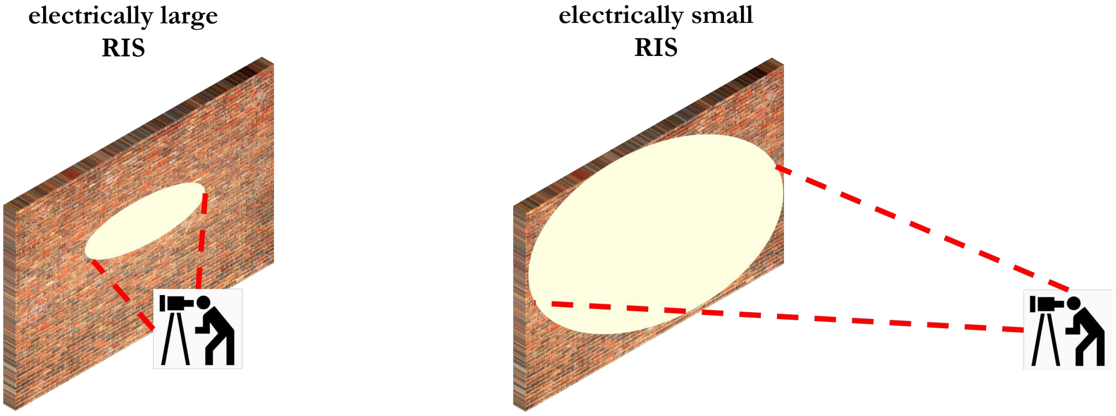


Figure 2.10 – Electrically large vs. electrically small metasurfaces.

- ***Electrically large and electrically small metasurfaces***. As mentioned, we cannot assume, a priori, that a metasurface operates in the far-field regime. This implies that the signal models and the corresponding analytical frameworks need to be sufficiently general to take this into account. This can be considered to be a not so usual situation in mainstream wireless communications, where the far-field operating regime is usually *de facto* implied, even though often not explicitly stated.

  In this chapter, we do not attempt to give a formal definition of near-field and far-field for a metasurface as a whole; they will be given in Chapter III and Chapter IV. We provide, on the other hand, arguments to distinguish these two operating regimes from a practical standpoint. Instead of using the terms near-field and far-field, we introduce the concepts of electrically large and electrically small metasurfaces. For clarity, these two concepts are illustrated in Fig. 2.10.

  A metasurface is referred to as electrically large if its geometric size is large enough as compared with the wavelength of the radio waves, and as compared with the transmission distance from the source to the metasurface and from the metasurface to the observation point. If this is not the case, then the metasurface is referred to as electrically small. The term “large enough” is deliberately left a bit vague, since a formal definition would depend on a large number of system parameters. In simple terms, based on Fig. 2.10, a metasurface can be considered to be electrically large if the size of a metasurface and the transmission distances are such that one sees the metasurface as being infinitely large, and, therefore, one cannot see the edges of the metasurface structure. If the opposite holds true, a metasurface can be considered to be electrically small. By considering a metasurface of a given and fixed size, with a similar line of thought, a metasurface can likely be considered to be electrically large if the transmission distances are sufficiently short. If the transmission distances are, on the other hand, sufficiently long, then a metasurface can likely be considered to be electrically small.

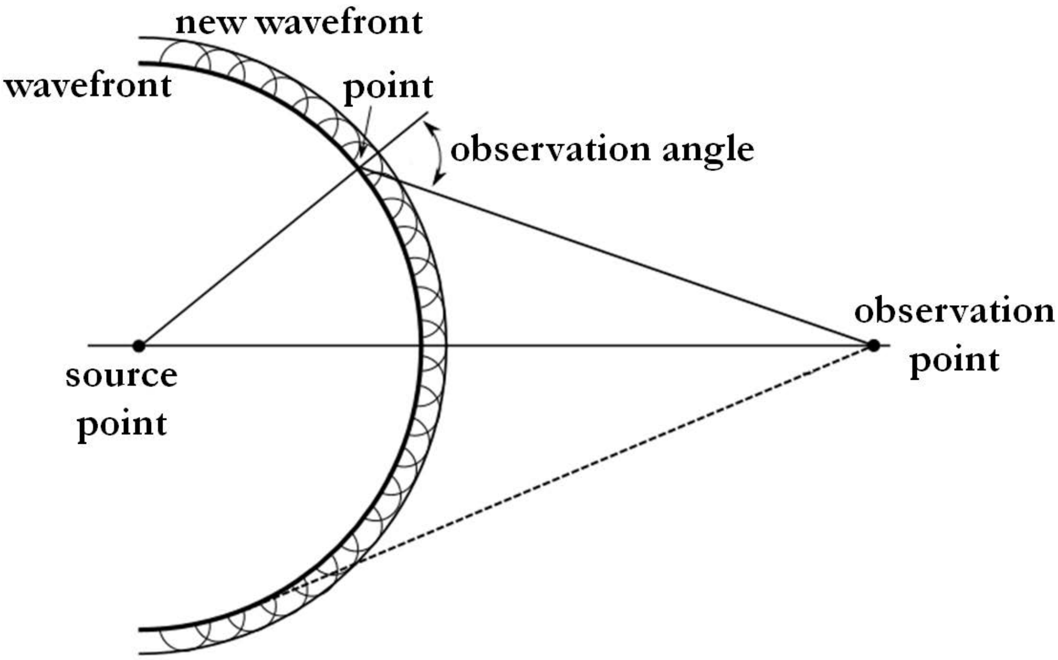


Figure 2.11 – The Huygens-Fresnel principle.

2.4.2 Theory of Electromagnetic Diffraction

With these preliminary definitions at hand, we can now discuss how to appropriately model the propagation of radio waves in the presence of a metasurface. As recently proved in [1] and [46], the EM field at any point of a volume and in the presence of a metasurface can be obtained by invoking the theory of diffraction and the Huygens-Fresnel principle. In general terms, the theory of diffraction provides one with the mathematical tools for modeling the bending of the radio waves when they encounter an object or a discontinuity, which in our case is a metasurface characterized by its specific effective parameters. In classical physics, more precisely, diffraction-based phenomena, including the reflection and transmission from objects, are described by the Huygens-Fresnel principle. According to this principle, every point of a propagating wavefront is viewed as a collection of individual spherical wavelets. In particular, every point on a wavefront is itself the source of spherical wavelets, and the secondary wavelets emanated from different points interfere with each other. The sum of the emanated spherical wavelets forms a new wavefront. This concept is sketched in Fig. 2.11.

When the radio waves emitted by a source impinge upon a metasurface, based on the Huygens-Fresnel principle, the wavefront of the impinging radio wave in correspondence of a metasurface, i.e., at each point of a metasurface, becomes the source of secondary wavelets that determine the EM field at any point of the volume of interest. Based on the divergence theorem (or, its more general formulation, the Green theorem [39]), which states that the surface integral of a vector field over a closed surface is equal to the volume integral of the divergence of the vector field over the region inside the surface, the EM field of interest is uniquely determined by the EM fields on the two sides of a metasurface, i.e., the surface EM fields. The readers are referred to [1], [46] for the analytical details. As far as the chapter is concerned, it is sufficient to understand that this implies that the EM field at any point of a volume can be obtained from the GSTCs given in the previous section, which, in fact, formulate the tangential components of the EM (surface) fields at the two sides of a metasurface.

Having considered this fact, the most suitable tool to model the EM field propagation in presence of metasurface is the theory of electromagnetic diffraction. Electromagnetic diffraction refers to phenomena where when an EM wave encounters an obstacle or opening. In this section, we derive several formulations that gives us the quantity of the electric field observed at a given observation point due to the presence of a scattering body, which, in our case is the metasurface/RIS. For the sake of rigorous analysis, we start our discussion from the set of Maxwell equations before proceeding to the derivation of important scalar and vector diffraction formulations.

Inhomogeneous Helmholtz equations

Consider the Maxwell equation for electromagnetic field when the nonzero sources in forms of charge density and current density are taken into account (e.g., [47, Eq. (7.40)])

$$\nabla_{\mathbf{r}} \cdot \mathbf{E}(\mathbf{r}, t) = \frac{\rho(\mathbf{r}, t)}{\epsilon_0} \tag{2.14}$$

$$\nabla_{\mathbf{r}} \cdot \mathbf{H}(\mathbf{r}, t) = 0 \tag{2.15}$$

$$\nabla_{\mathbf{r}} \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t} \tag{2.16}$$

$$\nabla_{\mathbf{r}} \times \mathbf{H}(\mathbf{r}, t) = \epsilon_0 \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} + \mathbf{J}(\mathbf{r}, t) \tag{2.17}$$

where we use the following notation:

$$\nabla_{\mathbf{r}} f(\mathbf{r}) = \left(\frac{\partial}{\partial x}\hat{\mathbf{x}} + \frac{\partial}{\partial y}\hat{\mathbf{y}} + \frac{\partial}{\partial z}\hat{\mathbf{z}}\right) f(x, y, z)\Big|_{x\hat{\mathbf{x}}+y\hat{\mathbf{y}}+z\hat{\mathbf{z}}=\mathbf{r}} \tag{2.18}$$

i.e., $\nabla_{\mathbf{r}}$ denotes the Laplace operator evaluated at $\mathbf{r}$, $\rho(\mathbf{r}, t)$ and $J(\mathbf{r}, t)$ are the charge and current densities at location $\mathbf{r}$ and time t, respectively, ϵ_0 is the vacuum permittivity, μ_0 is the vacuum permeability, and $c = 1/\sqrt{\epsilon_0 \mu_0}$ is the speed of EM waves in vacuum. We assume the universal time-dependency of $e^{j\omega t}$, i.e.:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}) e^{j\omega t}, \qquad \mathbf{H}(\mathbf{r}, t) = \mathbf{H}(\mathbf{r}) e^{j\omega t}. \tag{2.19}$$

Substituting (2.19) into (2.14) – (2.17) and canceling out the time-dependent factor, we have

$$\nabla_{\mathbf{r}} \cdot \mathbf{E}(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon_0} \tag{2.20}$$

$$\nabla_{\mathbf{r}} \cdot \mathbf{H}(\mathbf{r}) = 0 \tag{2.21}$$

$$\nabla_{\mathbf{r}} \times \mathbf{E}(\mathbf{r}) = -j\omega\mu_0 \mathbf{H}(\mathbf{r}) \tag{2.22}$$

$$\nabla_{\mathbf{r}} \times \mathbf{H}(\mathbf{r}) = j\omega\epsilon_0 \mathbf{E}(\mathbf{r}) + \mathbf{J}(\mathbf{r}) \tag{2.23}$$

Taking the curl of (2.22) and (2.23) and using the vector identity $\nabla_{\mathbf{r}} \times \nabla_{\mathbf{r}} \times \mathbf{E}(\mathbf{r}) = \nabla_{\mathbf{r}}(\nabla_{\mathbf{r}} \cdot \mathbf{E}(\mathbf{r})) - \nabla_{\mathbf{r}}^2 \mathbf{E}(\mathbf{r})$ yields us with

$$\nabla_{\mathbf{r}}(\nabla_{\mathbf{r}} \cdot \mathbf{E}(\mathbf{r})) - \nabla_{\mathbf{r}}^2 \mathbf{E}(\mathbf{r}) = -j\omega\mu_0 \nabla_{\mathbf{r}} \times \mathbf{H}(\mathbf{r}) \tag{2.24}$$

$$\nabla_{\mathbf{r}}(\nabla_{\mathbf{r}} \cdot \mathbf{H}(\mathbf{r})) - \nabla_{\mathbf{r}}^2 \mathbf{H}(\mathbf{r}) = j\omega\epsilon_0 \nabla_{\mathbf{r}} \times \mathbf{E}(\mathbf{r}, t) + \nabla_{\mathbf{r}} \times \mathbf{J}(\mathbf{r}) \tag{2.25}$$

Finally, by susbtituting (2.20) and (2.21) into (2.24) and (2.25), as well as using the fact that $\omega^2\epsilon_0\mu_0 = k^2$, we obtain the following inhomogeneous Helmholtz equations:

$$\nabla_{\mathbf{r}}^2 \mathbf{E}(\mathbf{r}) + k^2 \mathbf{E}(\mathbf{r}) = j\omega\mu_0 \mathbf{J}(\mathbf{r}) + \frac{\rho(\mathbf{r})}{\epsilon_0} \tag{2.26}$$

$$\nabla_{\mathbf{r}}^2 \mathbf{H}(\mathbf{r}) + k^2 \mathbf{H}(\mathbf{r}) = -\nabla_{\mathbf{r}} \times \mathbf{J}(\mathbf{r}) \tag{2.27}$$

Kirchhoff diffraction formula

The Kirchhoff diffraction formula is the foundation of the scalar diffraction theory. The formula is based on the principle that the scalar-valued components of the **E** and **H** fields also obey the scalar version of the inhomogeneous Helmholtz equations in (2.26). For example, assume that the current source $\mathbf{J}(\mathbf{r})$ and the charge source $\rho(\mathbf{r})$ can be modeled as an infinitesimal volume located at $\mathbf{r}_{\text{Tx}}$. If $E_x(\mathbf{r}')$ is the x-component of the electric field $\mathbf{E}(\mathbf{r}')$ at a point $\mathbf{r}'$, then we have from (2.26) that

$$\nabla_{\mathbf{r}}^2 E_x(\mathbf{r}) + k^2 E_x(\mathbf{r}) = -\mathbb{1}_{(\mathbf{r}=\mathbf{r}_{\text{Tx}})} A_x(\mathbf{r}_{\text{Tx}}) \tag{2.28}$$

where $A_x(\mathbf{r}_{\text{Tx}})$ is the scalar electric field at the source. The identity function $\mathbb{1}_{\mathbf{r}=\mathbf{r}_{\text{Tx}}}$ is present due to the fact that the sources $\mathbf{J}(\mathbf{r})$ and $\rho(\mathbf{r})$ are non-zero only at $\mathbf{r} = \mathbf{r}_{\text{Tx}}$ and zero otherwise.

We consider the problem of finding the scalar electric field in a volume $V \in \mathbb{R}^3$ bounded by a closed surface S and an infinite spherical surface S_∞, as shown in Fig. 2.12. Note that S can be regarded as a solid body that scatters the EM wave from the source. Let $G(\mathbf{r}_{\text{Rx}}, \mathbf{r}') = \frac{e^{-jk|\mathbf{r}_{\text{Rx}}-\mathbf{r}'|}}{4\pi|\mathbf{r}_{\text{Rx}}-\mathbf{r}'|}$ be the Green function in a 3-dimensional space where $\mathbf{r}_{\text{Rx}}$ is an observation point (e.g. a receiver antenna), away from the sources, i.e., $\mathbf{J}(\mathbf{r}_{\text{Rx}}) = 0$ and $\rho(\mathbf{r}_{\text{Rx}}) = 0$, and $\mathbf{r} \in S$ is a point in

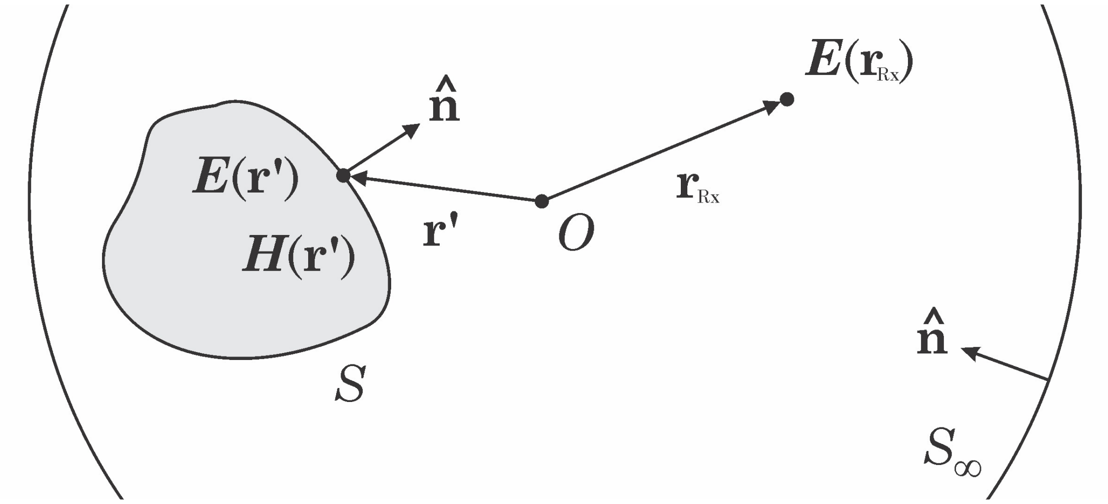


Figure 2.12 – Considered system model for finding scalar electric field in a volume V

the boundary. The Green function satisfies the following differential equation

$$\nabla^2_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}') + k^2 G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}') = -\delta(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}') \tag{2.29}$$

Applying Green's second identity [48, Eq. (C.27)] to $G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')$ and $E_x(\mathbf{r}_{\mathrm{Rx}})$, we have

$$\begin{aligned}&\int_V \big(G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})\nabla^2_{\mathbf{r}} E_x(\mathbf{r}) - E_x(\mathbf{r})\nabla^2_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})\big)\, d\mathbf{r}\\ &\quad = \int_{S+S_\infty} \big(G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')(\hat{\mathbf{n}}_{\mathrm{out}}\cdot\nabla_{\mathbf{r}'})E_x(\mathbf{r}') - E_x(\mathbf{r}')(\hat{\mathbf{n}}_{\mathrm{out}}\cdot\nabla_{\mathbf{r}'})G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')\big)\, d\mathbf{r}'\end{aligned} \tag{2.30}$$

where $\hat{\mathbf{n}}_{\mathrm{out}}$ is the normal vector at any point in S such that it points towards the outside of V (see Fig. 2.12). The integral over the infinite surface S_∞ is taken to be zero, which can justified rigorously [48]. Thus, by dropping the S_∞ term and adding as well as subtracting $k^2 G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})E_x(\mathbf{r})$ in the left-hand side of (2.30), we obtain:

$$\begin{aligned}&\int_V \big(G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})(\nabla^2_{\mathbf{r}}+k^2)E_x(\mathbf{r}) - E_x(\mathbf{r})(\nabla^2_{\mathbf{r}}+k^2)G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})\big)\, d\mathbf{r}\\ &\quad = \int_{S} \big(G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')(\hat{\mathbf{n}}_{\mathrm{out}}\cdot\nabla_{\mathbf{r}'})E_x(\mathbf{r}') - E_x(\mathbf{r}')(\hat{\mathbf{n}}_{\mathrm{out}}\cdot\nabla_{\mathbf{r}'})G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')\big)\, d\mathbf{r}'\end{aligned} \tag{2.31}$$

From (2.29), we have that

$$-\int_V E_x(\mathbf{r})(\nabla^2_{\mathbf{r}}+k^2)G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})d\mathbf{r} = \int_V E_x(\mathbf{r})\delta(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})d\mathbf{r} = E_x(\mathbf{r}_{\mathrm{Rx}}) \tag{2.32}$$

Also, from (2.28), we have that

$$\begin{aligned}\int_V G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})(\nabla^2_{\mathbf{r}}+k^2)E_x(\mathbf{r})d\mathbf{r} &= -\int_V G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r})\mathbb{1}_{\mathbf{r}=\mathbf{r}_{\mathrm{Tx}}}A_x(\mathbf{r}_{\mathrm{Tx}})d\mathbf{r}\\ &= -A_x(\mathbf{r}_{\mathrm{Rx}})G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}_{\mathrm{Tx}}).\end{aligned} \tag{2.33}$$

Substituting (2.32) and (2.33) into (2.31) yields:

$$\begin{aligned}E_x(\mathbf{r}_{\text{Rx}}) &= A_x(\mathbf{r}_{\text{Rx}})G(\mathbf{r}_{\text{Rx}},\mathbf{r}_{\text{Tx}}) \\ &\quad + \int_S \left(G(\mathbf{r}_{\text{Rx}},\mathbf{r}')(\hat{\mathbf{n}}_{\text{out}}\cdot\nabla_{\mathbf{r}'})E_x(\mathbf{r}') - E_x(\mathbf{r}')(\hat{\mathbf{n}}_{\text{out}}\cdot\nabla_{\mathbf{r}'})G(\mathbf{r}_{\text{Rx}},\mathbf{r}')\right) d\mathbf{r}'.\end{aligned} \tag{2.34}$$

Note that, in absence of S (the scattering body), the field at the observation point $\mathbf{r}_{\text{Rx}}$ is reduced to the *direct radiated field* from source $\mathbf{r}_{\text{Tx}}$ towards $\mathbf{r}_{\text{Rx}}$, which is given as $E_x(\mathbf{r}_{\text{Rx}}) = A_x(\mathbf{r}_{\text{Rx}})G(\mathbf{r}_{\text{Rx}},\mathbf{r}_{\text{Tx}})$.

Vector diffraction theory

In this sub-section, we extend the problem of finding the electric field at an observation point into the vectorial domain. Historically, the expression of vectorial electric field due to a scattering body can be given in various forms, with one of the most prominent representations being ***the Franz formula***. The advantage of this representation is due to the fact that it uses only the tangential components of the **E** and **H** fields on the integration surface and automatically satisfy Maxwell's equations even with arbitrary boundary values of the fields. In this thesis, especially in Chapter 4, we use mainly another representation called the ***Stratton-Chu formula*** which, despite having the disadvantages of involving the normal components of the fields under the integration sign and of changing their analytical form for surfaces with edges, is easier to manipulate to obtain a physically meaningful interpretation for RIS/metasurface applications since one can distinguish the direct link component from the reflected field by a surface, as well as see the impact of the system's physical parameters.

Let $V \subseteq \mathbb{R}^3$ be an arbitrary volume and ∂V be its boundary. Similar to the previous sub-section, we let $\mathbf{r}_{\text{Rx}} \in V$ be an observation point inside V and $\mathbf{r}_{\text{Tx}}$ be the point at which the source is located, which can be inside or outside V. From Franz formula in [48, Eq.(18.10.11)], the electric field evaluated at $\mathbf{r}_{\text{Rx}} \in V$ is given as follows:

$$\mathbf{E}(\mathbf{r}_{\text{Rx}}) = \frac{1}{j\omega\mu_0\epsilon_0}\left[\nabla_{\mathbf{r}_{\text{Rx}}} \times \left(\nabla_{\mathbf{r}_{\text{Rx}}} \times (\mathbf{A}(\mathbf{r}_{\text{Rx}}) + \mathbf{A}_s(\mathbf{r}_{\text{Rx}}))\right)\right] - \frac{1}{\epsilon_0}\nabla_{\mathbf{r}_{\text{Rx}}} \times (\mathbf{A}_m(\mathbf{r}_{\text{Rx}}) + \mathbf{A}_{ms}(\mathbf{r}_{\text{Rx}})) \tag{2.35}$$

where we define

$$\mathbf{A}(\mathbf{r}_{\text{Rx}}) = \int_V \mu_0 \mathbf{J}(\mathbf{r}) G(\mathbf{r}_{\text{Rx}},\mathbf{r}) d\mathbf{r} \tag{2.36}$$

$$\mathbf{A}_m(\mathbf{r}_{\text{Rx}}) = \int_V \epsilon_0 \mathbf{J}_m(\mathbf{r}) G(\mathbf{r}_{\text{Rx}},\mathbf{r}) d\mathbf{r} \tag{2.37}$$

$$\mathbf{A}_s(\mathbf{r}_{\text{Rx}}) = -\int_{\partial V} \mu_0 [\hat{\mathbf{n}}_{\text{out}} \times \mathbf{H}(\mathbf{r})] G(\mathbf{r}_{\text{Rx}},\mathbf{r}) d\mathbf{r}' \tag{2.38}$$

$$\mathbf{A}_{ms}(\mathbf{r}_{\text{Rx}}) = \int_{\partial V} \epsilon_0 [\hat{\mathbf{n}}_{\text{out}} \times \mathbf{E}(\mathbf{r})] G(\mathbf{r}_{\text{Rx}},\mathbf{r}) d\mathbf{r}' \tag{2.39}$$

Note that

$$\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \mathbf{A}(\mathbf{r}_{\mathrm{Rx}})\right) = \nabla_{\mathbf{r}_{\mathrm{Rx}}}(\nabla_{\mathbf{r}_{\mathrm{Rx}}} \cdot \mathbf{A}(\mathbf{r}_{\mathrm{Rx}})) - \nabla^2_{\mathbf{r}_{\mathrm{Rx}}}\mathbf{A}(\mathbf{r}_{\mathrm{Rx}}) \overset{(b)}{=} \nabla_{\mathbf{r}_{\mathrm{Rx}}}(\nabla_{\mathbf{r}_{\mathrm{Rx}}} \cdot \mathbf{A}(\mathbf{r}_{\mathrm{Rx}})) + k^2\mathbf{A}(\mathbf{r}_{\mathrm{Rx}}) \tag{2.40}$$

where (b) follows from [48, Eq.(18.2.5)]. Also, from (2.36) we have

$$\nabla_{\mathbf{r}_{\mathrm{Rx}}}\left[\nabla_{\mathbf{r}_{\mathrm{Rx}}} \cdot \mathbf{A}(\mathbf{r}_{\mathrm{Rx}})\right] = \int_V \mu_0 \nabla_{\mathbf{r}_{\mathrm{Rx}}}\left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \cdot \mathbf{J}(\mathbf{r})\right) d\mathbf{r} \overset{(c)}{=} \int_V \mu_0 (\mathbf{J}(\mathbf{r}) \cdot \nabla_{\mathbf{r}_{\mathrm{Rx}}})\left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})\right) d\mathbf{r} \tag{2.41}$$

where (c) follows from identity

$$\begin{aligned}\nabla_{\mathbf{r}_{\mathrm{Rx}}}\left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \cdot \mathbf{J}(\mathbf{r})\right) &= (\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \cdot \nabla_{\mathbf{r}_{\mathrm{Rx}}})\mathbf{J}(\mathbf{r}) + (\mathbf{J}(\mathbf{r}) \cdot \nabla_{\mathbf{r}_{\mathrm{Rx}}})\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \\ &\quad + \nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \times (\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \mathbf{J}(\mathbf{r})) + \mathbf{J}(\mathbf{r}) \times (\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})) \\ &\overset{(d)}{=} (\mathbf{J}(\mathbf{r}) \cdot \nabla_{\mathbf{r}_{\mathrm{Rx}}})\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \end{aligned} \tag{2.42}$$

and (d) follows form the fact that $\nabla_{\mathbf{r}_{\mathrm{Rx}}}$ only acts on $\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})$ and that $\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})) = 0$. Substituting (2.36) and (2.41) into (2.40), we have

$$\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \mathbf{A}(\mathbf{r}_{\mathrm{Rx}})\right) = \int_V \left[\mu_0 k^2 \mathbf{J}(\mathbf{r}) G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) + \mu_0 (\mathbf{J}(\mathbf{r}) \cdot \nabla_{\mathbf{r}_{\mathrm{Rx}}})\left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})\right)\right] d\mathbf{r} \tag{2.43}$$

Using similar steps, we can show that

$$\begin{aligned}&\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \mathbf{A}_s(\mathbf{r}_{\mathrm{Rx}})\right) \\ &\quad = -\int_{\partial V} \left[\mu_0 k^2 [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') + \mu_0 ([\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] \cdot \nabla_{\mathbf{r}_{\mathrm{Rx}}})\left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}')\right)\right] d\mathbf{r}' \end{aligned} \tag{2.44}$$

Also, from (2.37) and (2.39), we have

$$\frac{1}{\epsilon_0}\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \mathbf{A}_m(\mathbf{r}_{\mathrm{Rx}}) = \frac{1}{\epsilon_0}\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \int_V \epsilon_0 \mathbf{J}_m(\mathbf{r}) G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) d\mathbf{r} = \int_V \left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \times \mathbf{J}_m(\mathbf{r})\right) d\mathbf{r} \tag{2.45}$$

$$\begin{aligned}\frac{1}{\epsilon_0}\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \mathbf{A}_{ms}(\mathbf{r}_{\mathrm{Rx}}) &= \frac{1}{\epsilon_0}\nabla_{\mathbf{r}_{\mathrm{Rx}}} \times \left(\int_{\partial V} \epsilon_0 [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{E}(\mathbf{r})] G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) d\mathbf{r}\right) \\ &= \int_{\partial V} \left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \times [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{E}(\mathbf{r})]\right) d\mathbf{r} \end{aligned} \tag{2.46}$$

Substituting (2.43), (2.44), (2.45), and (2.46) to (2.35), we obtain

$$\begin{aligned}\mathbf{E}(\mathbf{r}_{\mathrm{Rx}}) &= \frac{1}{j\omega\epsilon_0}\int_V \left[k^2 \mathbf{J}(\mathbf{r}) G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) + (\mathbf{J}(\mathbf{r}) \cdot \nabla_{\mathbf{r}_{\mathrm{Rx}}})\left(\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})\right) - j\omega\epsilon_0 \nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \times \mathbf{J}_m(\mathbf{r})\right] d\mathbf{r} \\ &\quad - \frac{1}{j\omega\epsilon_0}\int_{\partial V} \left[k^2 [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') + ([\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] \cdot \nabla_{\mathbf{r}_{\mathrm{Rx}}})\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \right. \\ &\quad \left. + j\omega\epsilon_0 \nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \times [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{E}(\mathbf{r}')]\right] d\mathbf{r}' \end{aligned} \tag{2.47}$$

Note that $\nabla_{\mathbf{r}_{\mathrm{Rx}}}$ only acts on $G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})$ and $G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}')$. Also, note that $\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) = -\nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})$

and $\nabla_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') = -\nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}')$. Therefore, (2.47) can be transformed into:

$$
\begin{aligned}
\mathbf{E}(\mathbf{r}_{\mathrm{Rx}}) = & \frac{1}{j\omega\epsilon_0} \int_V \left[k^2 \mathbf{J}(\mathbf{r}) G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) + (\mathbf{J}(\mathbf{r}) \cdot \nabla_{\mathbf{r}}) \left(\nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \right) - j\omega\epsilon_0 \nabla_{\mathbf{r}} \mathbf{J}_m(\mathbf{r}) \times G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \right] d\mathbf{r} \\
& - \frac{1}{j\omega\epsilon_0} \int_{\partial V} \left[k^2 [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') + ([\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] \cdot \nabla_{\mathbf{r}'}) \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \right. \\
& \left. + j\omega\epsilon_0 \nabla_{\mathbf{r}'} [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{E}(\mathbf{r}')] \right] \times G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') d\mathbf{r}'
\end{aligned} \tag{2.48}
$$

This expression in (2.48) is called ***Kottler formula*** [48, Eq. (16.10.10)]. We use the relation [48]

$$
\int_V (\mathbf{J}(\mathbf{r}) \cdot \nabla_{\mathbf{r}}) \left(\nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \right) d\mathbf{r} - \int_{\partial V} ([\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] \cdot \nabla_{\mathbf{r}'}) \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') d\mathbf{r}' \tag{2.49}
$$

$$
= \int_V j\omega\rho(\mathbf{r}) \nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) d\mathbf{r} - \int_{\partial_V} j\omega\epsilon_0 [\mathbf{n}_{\mathrm{out}} \cdot \mathbf{E}(\mathbf{r}')] \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') d\mathbf{r}' \tag{2.50}
$$

Applying (2.49) into (2.48) yields:

$$
\begin{aligned}
\mathbf{E}(\mathbf{r}_{\mathrm{Rx}}) = & \int_V \left[-j\omega\mu_0 \mathbf{J}(\mathbf{r}) G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) + \frac{\rho(\mathbf{r})}{\epsilon_0} \nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) - \mathbf{J}_m(\mathbf{r}) \times \nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \right] d\mathbf{r} \\
& - \int_{\partial V} \left[-j\omega\mu_0 [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') + [\mathbf{n}_{\mathrm{out}} \cdot \mathbf{E}(\mathbf{r}')] \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \right. \\
& \left. + [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{E}(\mathbf{r}')] \times \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \right] d\mathbf{r}'
\end{aligned} \tag{2.51}
$$

This expression in (2.51) is called the ***Stratton-Chu formula*** [48, Eq. (16.10.7)]. Note that, if the source location $\mathbf{r}_{\mathrm{Tx}}$ is contained in V, we have

$$
\int_V \left[-j\omega\mu_0 \mathbf{J}(\mathbf{r}) G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) + \frac{\rho(\mathbf{r})}{\epsilon_0} \nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \right] d\mathbf{r} = \mathbf{E}_{\mathrm{inc}}(\mathbf{r})
$$

where $\mathbf{E}_{\mathrm{inc}}(\mathbf{r})$ is the radiated field (i.e., field that propagates directly) from the source at $\mathbf{r}_{\mathrm{Tx}}$ to the observation point at $\mathbf{r}_{\mathrm{Rx}}$. Therefore, assuming $\mathbf{J}_m = 0$, we have

$$
\begin{aligned}
\mathbf{E}(\mathbf{r}_{\mathrm{Rx}}) = & \mathbb{1}_{(\mathbf{r}_{\mathrm{Tx}} \in V)} \mathbf{E}_{\mathrm{inc}}(\mathbf{r}) - \int_{\partial V} \left[-j\omega\mu_0 [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}(\mathbf{r}')] G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') + [\mathbf{n}_{\mathrm{out}} \cdot \mathbf{E}(\mathbf{r}')] \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \right. \\
& \left. + [\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{E}(\mathbf{r}')] \times \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \right] d\mathbf{r}'
\end{aligned} \tag{2.52}
$$

2.4.3 Methods of Physical Optics

The existing methods for solving scattering problems can be classified into three broad categories: (i) **analytical methods**, (ii) **numerical methods**, , and (iii) **approximate analytical models**. The main differences between these methods are that analytical methods provide exact explicit solutions, numerical methods lead to approximate analytical solutions for usually simplified structures, and approximate analytical models transform differential or integral equations into matrix equations by projective approximations and solve them iteratively or by matrix inversion.

Physical optics (PO) can be categorized as an approximate analytical technique, i.e., it provides the exact solution of the electromagnetic fields at a given observer location while also approximate and transform it into a closed-form solutions. This technique allows us to solve scattering problems for electrically large objects. The main advantage of the PO method compared to the discretization-based numerical approaches is that it eliminates the need to solve systems of algebraic equations since the solution of the scattering problem is immediately available in the form of an integral over the scattering surface, whose computation can be efficiently carried out by applying various numerical/asymptotic integration techniques.

The PO method is a high-frequency technique in the sense that the accuracy of the PO solutions improves with the increase in frequency so that for scatterers, apertures, lenses, and reflectors greater than severalwavelength in diameter PO provides solutions with an accuracy that is sufficient for most practical applications. In contrast to geometrical optics (GO), PO solutions correctly (at least qualitatively) describe a variety of diffraction effects, including the smooth transition between the lit and shadowed portions of space at shadow boundaries, the edge-diffracted waves, and the fields at caustics. All these features make the PO method an indispensable tool in optical and microwave engineering.

Reflection and Transmission on Planar Interfaces

Before going into the principles of physical optics, we first look at the theory on reflection and transmission (or refraction) of waves at planar interfaces (e.g., 2-dimensional RIS).

When a plane incident wave arrives at a material interface, a plane reflected wave and an in general inhomogeneous transmitted wave are produced. In each of these waves the field vectors are perpendicular to the respective wave vectors and can be completely represented by two components, implying that relations between the incident, reflected, and transmitted fields at a material interface can be described by 2 × 2 reflection and transmission matrices **R** and **T**. In isotropic media these matrices can be made diagonal by splitting up the fields into the components parallel and perpendicular to the so-called plane of incidence, which is the plane defined by the direction of incidence of the incoming wave and by the normal to the material interface (Figure 4.1). A fundamental feature of reflection and transmission of electromagnetic waves at isotropic and non-chiral material interfaces is that these components are completely decoupled in the sense that the parallel (perpendicular) component of an incident wave is related only to the parallel (perpendicular) components in the reflected and transmitted fields. If we denote the electric fields in the incident, reflected, and transmitted waves by E_{inc}, E_{ref}, and E_{tran}, and their parallel and perpendicular components by the symbols $\parallel$ and $\perp$, respectively, then the relations between the components can be expressed as

$$E_{\mathrm{ref}}^{\parallel} = R^{\parallel} E_{\mathrm{inc}}^{\perp}, \quad E_{\mathrm{ref}}^{\perp} = R^{\perp} E_{\mathrm{inc}}^{\perp} \tag{2.53}$$

$$E^{\parallel}_{\text{tran}} = T^{\parallel} E^{\perp}_{\text{inc}}, \quad E^{\perp}_{\text{tran}} = T^{\perp} E^{\perp}_{\text{inc}} \tag{2.54}$$

which implies that

$$\mathbf{R} = \begin{pmatrix} R^{\perp} & 0 \\ 0 & R^{\parallel} \end{pmatrix}, \qquad \mathbf{T} = \begin{pmatrix} T^{\perp} & 0 \\ 0 & T^{\parallel} \end{pmatrix}. \tag{2.55}$$

The four factors $R^{\perp}$, $R^{\parallel}$, $T^{\perp}$, and $T^{\parallel}$, called the reflection and transmission coefficients, are sufficient to completely determine the relations between the complex amplitudes in the incident, reflected, and transmitted waves at a flat interface of isotropic materials.

Principles of physical optics

We first discuss the fundamental physical assumption behind the PO method: the GO approximation of the fields or currents on the scatterer surface. When a plane wave illuminates a flat object (with an infinite planar surface), the reflected field is also a plane wave. The propagation direction of the reflected wave is given by the GO rule: the reflection angle is equal to the incidence angle. For this case of an infinite flat object, the reflected field values can be easily found. The PO approximation treats the neighborhood of every point on the surface of electrically large arbitrary-shaped bodies as locally flat and tangential to the surface at this point, and uses the solution for the corresponding infinite and flat surface to define the fields or currents on the surface of the scatterer. This leads to the assumption of the local character of the interaction between the incident field and the scatterer: the current at a given point is determined by the incident field at the same point. This is approximately correct if the body is large compared to the wavelength and smooth so that the curvature radii are also large compared to the wavelength.

The PO solution for a scattering problem starts from choosing a suitable integration surface over which the electromagnetic field can be reasonably approximated in the framework of GO, for example by a combination of incident, reflected, and transmitted plane waves. Then, one writes the GO-based approximations for the tangential components of the electric and magnetic fields on the integration surface or, which is the same, for the equivalent or physical surface currents. Finally, these approximated currents are inserted into the original integral representation of the scattered field and integrated over the surface, which gives the scattered field in the PO approximation.

For example, consider a scattering body that is impenetrable so that the direct transmission through it can be neglected. We choose the integration surface Ω such that it coincides with the physical scattering surface (see Figure 2.13). Furthermore, we split the Ω into the lit (grey) and shadowed (black) region in Figure 2.13, and denote them as Ω_{lit} and Ω_{shadow}, respectively. In this case, we have $\Omega = \Omega_{\text{lit}} + \Omega_{\text{shadow}}$. If we know the reflection matrix $\mathbf{R}$ for all incidence

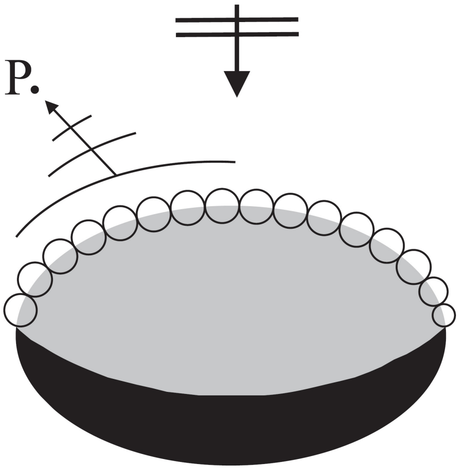


Figure 2.13 – Scattering from a material body according to PO. The scattered field at the observation point P is obtained by integration over all secondary sources on the illuminated portion of the scatterer (gray) which are described in the GO approximation. On the shadowed portion of the scatterer (black) the secondary sources are assumed to be zero.

angles, then the electric field at any point $\mathbf{s}$ in the material body, according to geometrical optics, is approximated as follows:

$$\mathbf{E}_{\mathrm{go}}(\mathbf{s}) \approx \begin{cases} \mathbf{E}_{\mathrm{inc}}(\mathbf{s}) + \mathbf{E}_{\mathrm{ref}}(\mathbf{s}) = (\mathbf{R}+\mathbf{I})\mathbf{E}_{\mathrm{inc}}(\mathbf{s}), & \mathbf{s} \in \Omega_{\mathrm{lit}} \\ 0, & \mathbf{s} \in \Omega_{\mathrm{shadow}} \end{cases} \tag{2.56}$$

Here, $\mathbf{s}$ denotes a point in the material body, $\mathbf{I}$ is a 2×2 identity matrix, $\mathbf{E}_{\mathrm{inc}}(\mathbf{s})$ and $\mathbf{E}_{\mathrm{ref}}(\mathbf{s})$ are the incident and reflected electric fields at $\mathbf{s}$, which are needed to specify the reflected field, are taken for the incidence angle defined at that particular point. The local values of the magnetic field $\mathbf{H}_{\mathrm{go}}(\mathbf{s})$ can be obtained from $\mathbf{E}_{\mathrm{go}}(\mathbf{s})$ by using the relations between the electric and magnetic fields in plane waves:

$$\mathbf{H}_{\mathrm{go}}(\mathbf{s}) = -\nabla_{\mathbf{r}} \times \mathbf{E}_{\mathrm{go}}(\mathbf{s}) / (j\omega\mu_0) \tag{2.57}$$

Derivation of Physical Optics Solutions

Derivation of a PO solution consists of three steps:

1. **choice of integration domain**; In this step, a suitable integration surface Ω in an integral representation of the field is chosen. In problems of electromagnetic scattering by compact scatterers, the most convenient choice is typically the surface of the scatterer; in problems of transmission through or radiation from an opening in a impenetrable

screen a reasonable choice is a surface placed directly behind the screen which coincides either with the aperture or with the front of the transmitted wave.

2. **integral simplification**; In this step, the exact integral representation obtained in the previous step is simplified through approximate evaluation of the differential operators in the integral representation.

3. **approximation of the fields on the integration surface using the GO laws**; In this step, the simplified integral obtained in the last step is approximated by using numerical methods. Typically the resulting integral can be conveniently approximated by stationary phase method (SPM) or the Taylor approximation of distances.

2.5 Conclusion

We conclude this chapter by briefly summarizing the material provided in this chapter and discussing their relations with the subsequent chapters. In this chapter, we have introduced necessary tools for modeling RISs by using physics-based approach i.e., through the theory of EM waves. From the theory of surface electromagnetic, we understand that an RIS can be homogenized thanks to its subwavelength-size of its scattering elements. This allows us to see an RIS as a continuous function even though the RIS is conceptually discrete. Furthermore, using the macroscopic representation of RIS, the electric field at the receiver can be fully described by the knowledge of the electric field at the reflection or transmission side of the RIS. In Chapter 3 and, more specifically, Chapter 4, we will see how this allows us to greatly simplify the analysis of the path-loss modeling of the RIS and, by leveraging the theory of electromagnetic diffraction, to obtain the analytical formulation of the received electric field. Finally, using the methods of physical optics, which overcome the limitation of geometrical optics, the obtained analytical formulation can be obtained in a tractable manner.

3 Reconfigurable Intelligent Surfaces vs. Relays: Performance Comparison

Reconfigurable intelligent surfaces (RISs) have the potential of realizing the emerging concept of smart radio environments by leveraging the unique properties of meta-surfaces. In this chapter, we elaborate on the key differences and similarities between RISs that are configured to operate as anomalous reflectors and relays. We also compare the performance of both technologies, first qualitatively, and then quantitatively through numerical results that highlight their respective spectral efficiency gains.

3.1 Introduction 53

3.2 Reconfigurable Intelligent Surfaces vs. Relaying 54

- 3.2.1 Hardware Complexity 54
- 3.2.2 Noise 55
- 3.2.3 Spectral Efficiency 55
- 3.2.4 Power Budget 55

3.3 Performance Comparison: Achievable Data Rate 56

- 3.3.1 System Model and Analytical Formulation of Achievable Data Rate . . . 56
- 3.3.2 Quantitative Comparison 61
- 3.3.3 Average Signal-to-Noise Ratio vs. Transmission Distance 62
- 3.3.4 Takeaway Messages from the Comparison 63

3.4 Numerical Results 64

- 3.4.1 RISs vs. Relays as a Function of the Transmission Distance 66
- 3.4.2 RISs: Anomalous Mirrors vs. Focusing Lenses 67
- 3.4.3 RISs vs. Relays as a Function of the Carrier Frequency 67
- 3.4.4 RISs vs. Relays as a Function of the Size of the RIS 67

3.5 Conclusions 68

3.6 Appendices 68

- 3.6.1 Proof of Theorem 3.1 68
- 3.6.2 Proof of Proposition 3.1 69
- 3.6.3 Proof of Proposition 3.2 70

3.1 Introduction

By 2022, it is expected that the global mobile data traffic will reach a monthly run of 77 exabytes, which corresponds to a 7-fold growth compared with the monthly run of 2017. Such demands may not be accommodated by current cellular standards that utilize only sub-6 GHz frequency bands. A key feature of future wireless networks is hence the potential migration to higher frequencies, e.g., the millimeter (30-100 GHz) and sub-millimeter (above 100 GHz) wave bands [49]. Extensive measurements have been conducted at the millimeter wave band and, more recently, the sub-millimeter wave band. These measurements have demonstrated that the use of highly directional steerable antennas enables mobile communication at such high frequencies [49]. However, millimeter wave and sub-millimeter wave frequency bands are highly susceptible to blockages from large-size structures, e.g., buildings, on the radio path [49, Tables 4, 5]. In addition, their signals may be severely attenuated by the presence of small-size objects, e.g., human bodies and foliage.

A possible approach for circumventing the unreliability of high-frequency channels is to sense the environment and to identify, on a real-time basis, alternative propagation routes through which the same information-bearing signal can be received. To this end, an established method is the deployment of relays that capitalize on the concept of (distributed) cooperative diversity [50]. The use of relays can effectively turn a single non-line-of-sight (NLOS) link into multiple line-of-sight (LOS) links. This approach requires each relay to be equipped with a dedicated power source and with the necessary front-end circuitry for reception, processing, and re-transmission. For these reasons, the use of relays may result in an increase in the network power consumption and may require a larger capital expenditure for deployment.

In addition, the network spectral efficiency offered by relay-aided systems depends on the duplexing protocol employed for transmission. If a half-duplex (HD) relaying protocol is employed, transmitters and relays are not allowed to transmit concurrently on the same physical resource. This issue can be overcome by employing a full-duplex (FD) relaying protocol, but at the cost of: (i) introducing high *loop-back self-interference* at the relay because of the concurrent transmission and reception of signals; (ii) generating *co-channel interference* at the destination, since relays and transmitters emit different information on the same physical resource; and (iii) increasing the *signal processing complexity* and the *power consumption* of the relays. Relays, therefore, are utilized in an adaptive fashion, depending on channel and interference conditions, for improving the network performance [50].

When the LOS path is of insufficient quality, another approach to establish alternative routes is through *passive non-reconfigurable specular reflectors*, e.g., dielectric or metallic mirrors [35]. This method for coverage enhancement has the potential benefit of being more cost-efficient as compared with relaying, especially in high-frequency bands. However, a main limitation of non-reconfigurable reflectors is that they cannot enable the dynamic shaping of the impinging waves, since their operation cannot be modified after fabrication, i.e., at the time of deployment and operation. Due to the highly dynamic nature of the wireless environment and the nomadic nature of mobile communications, it would be beneficial that such reflectors be capable of adaptively shaping the radio waves based on actual blockage and environmental conditions.

Propitiously, as described in the previous chapter, electromagnetic-based reconfigurable structures that are capable of applying specified transformations to the impinging radio waves do exist and can operate at different frequency bands [4], [8]. In the literature, in addition to **reconfigurable intelligent surfaces (RISs)**, these structures are often referred to as large intelligent surfaces, intelligent reflecting surfaces, digitally controllable scatterers, and software-controllable surfaces. When deployed in wireless networks, RISs have the potential of turning the wireless environment, which is highly probabilistic in nature, into a programmable and partially deterministic space, which is referred to as **smart (or intelligent) radio environment** [5].

3.2 Reconfigurable Intelligent Surfaces vs. Relaying

In this section, we elaborate on differences and similarities between RISs that are employed as anomalous reflectors and relays. The comparison is made here on a qualitative basis, and is complemented, in the next section, with results that compare RISs and relays on a more quantitative basis.

3.2.1 Hardware Complexity

Relays are usually viewed as active devices that need a dedicated power source for operation. They are equipped with active electronic components, such as digital-to-analog converters (DACs) and analog-to-digital converters (ADCs), mixers, power amplifiers for transmission, and low-noise amplifiers for reception. Several electronic components are typically needed for implementing decode-and-forward (DF) and amplify-and-forward (AF) relaying. The deployment of relays may, thus, be costly and power-consuming, especially for realizing multiple-antenna designs at millimeter and sub-millimeter wave frequency bands [49]. If, in addition, FD relays are used, the complexity is further increased due to the need of eliminating the loop-back self-interference by using tailored antennas and analog/digital signal processing methods.

In contrast, RISs are composite material layers that are made of metallic or dielectric patches printed on a grounded dielectric substrate. Their configurability is ensured through low-power and low-complexity electronic circuits (switches or varactors) [12]. RISs are envisioned to be of lower complexity than relays, especially at mass production and if realized by using inexpensive large-area electronics, since no dedicated power amplifiers, mixers, and DACs/ADCs are usually required. A prototype of large-size RIS made of 3,720 inexpensive antennas has recently been realized [3].

3.2.2 Noise

The active electronic components used in relays are responsible for the presence of additive noise that negatively affects the performance of conventional relaying protocols. In AF relaying, for example, the noise is amplified at the relays. The impact of additive noise can be mitigated by employing DF relaying, at the expense of decoding and re-encoding (regeneration) the signal at the relays and increasing the signal processing complexity and power consumption. In FD relaying, the impact of residual loop-back self-interference further deteriorates the system performance.

On the other hand, RISs that behave as anomalous reflectors are not affected by additive noise. However, they may be impaired by phase noises. If they are nearly-passive, in addition, they cannot amplify or regenerate the signals [8].

3.2.3 Spectral Efficiency

The spectral efficiency of relay-aided systems depends on the adopted duplexing protocol. Under HD relaying, the achievable rate is generally scaled down by a factor of two, since different physical resources are used for the data emitted by the transmitter and by the relay. The end-to-end signal-to-noise ratio, on the other hand, can be increased by capitalizing on more favorable propagation conditions for the relayed signal, and by optimally combining the direct and relayed signals. Under FD relaying, the achievable rate is not scaled down by a factor of two, but the relay is affected by the residual loop-back self-interference, and the receiver is impaired by the interference generated by the concurrent transmission of the transmitter and the relay.

RISs that are configured to operate as anomalous reflectors are not subject to the half-duplex constraint and the loop-back self-interference. In addition, the local reflection coefficient of the meta-surface can be designed for optimally combining the signals received from the transmitter and the RIS.

3.2.4 Power Budget

Relays require an independent power source for operation, which is used for transmitting the signals (RF power) and for supplying with power their electronic components.

In contrast, RISs are suitable for nearly passive implementations, since non-reconfigurable meta-surfaces can be realized with fully passive components, and low-power active components (switches or varactors) are needed only for ensuring their reconfigurability. Also, the low-power nature of switches and varactors makes the use of energy harvesting a suitable candidate for realizing close-to-passive implementations.

In relay-aided systems, it is usually assumed that the total RF power is allocated between the transmitter and the relay, so as to ensure a total power constraint. In RISs, the transmitter uses the total RF power. Also, the power reflected and scattered by the RIS depends on its transmittance, which can be optimized through an appropriate design of the meta-surface [12]. In the ideal case, the total power reflected by an RIS is the same as the total power of the impinging radio wave.

3.3 Performance Comparison: Achievable Data Rate

In this section, we compare the performance of RIS-assisted and relay-assisted systems in terms of average signal-to-noise ratio (SNR) at the receiver side. To better understand the comparison, we briefly introduce the system model of both systems and analytical formulation of the achievable data rate. In general, the communication system (both in RIS and relay scenario) is designed to transmit information from a transmitter (Tx) to the receiver (Rx). Tx is modeled as a point source that emits cylindrical electromagnetic (EM) waves through the vacuum. The EM waves emitted by Tx travel at the speed of light c. The frequency of the EM waves is denoted by f, and the wavelength and wavenumber are $\lambda = c/f$ and $k = 2\pi/\lambda$, respectively.

3.3.1 System Model and Analytical Formulation of Achievable Data Rate

Reconfigurable Intelligent Surface

In a two-dimensional space, we consider a system that consists of a transmitter (Tx), a receiver (Rx), and a flat surface ($\mathscr{S}$) of zero-thickness (see Fig. 3.1). Without loss of generality, we assume that $\mathscr{S}$ is located such that its center coincides with the origin. Furthermore, $\mathscr{S}$ lies in the x-axis and spans along $[-L, L]$, i.e.,

$$\mathscr{S} = \{(x,0) : -L \leq x \leq L\}.$$

The locations of Tx and Rx are denoted by (x_T, y_T) and (x_R, y_R), respectively. We consider only the scenario where Tx and Rx are on the same side of the surface $\mathscr{S}$, i.e., we focus our attention on modeling reflections from the surface $\mathscr{S}$.

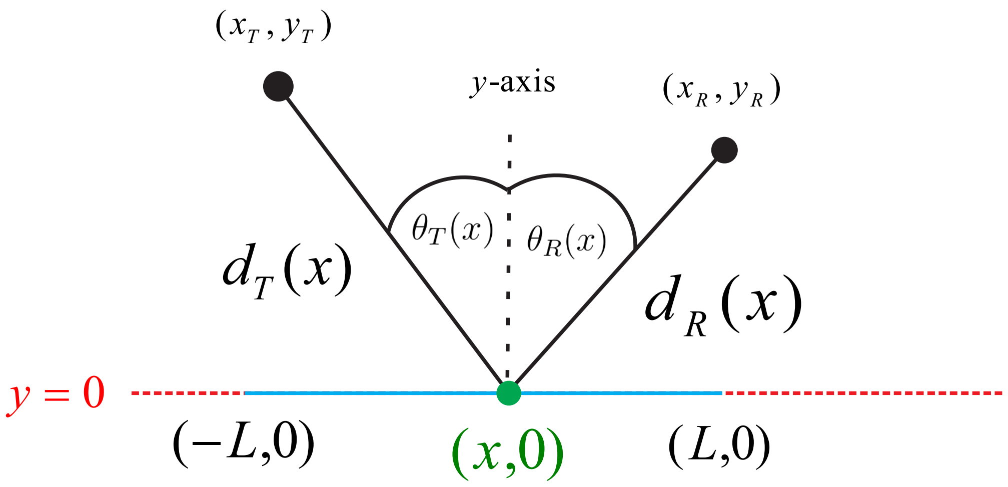


Figure 3.1 – System model for RIS-assisted communication.

For every point $(x, 0) \in \mathscr{S}$, the Tx-to-$\mathscr{S}$ and $\mathscr{S}$-to-Tx distances are denoted by $d_T(x) = \sqrt{(x - x_T)^2 + y_T^2}$ and $d_R(x) = \sqrt{(x_R - x)^2 + y_R^2}$, respectively. In particular, $d_T(x)$ is the radius of the wavefront of the EM wave that is emitted by Tx and intersects $\mathscr{S}$ at $(x, 0)$, and $d_R(x)$ is the radius of the wavefront of the EM wave that originates from $\mathscr{S}$ at $(x, 0)$ and is observed at Rx. With a similar terminology, the angle of incidence of the EM wave at $(x, 0) \in \mathscr{S}$ is denoted by $\theta_T(x)$. It represents the angle formed by the y-axis and the wavefront of the EM wave that originates from Tx and intersects $\mathscr{S}$ at $(x, 0)$. The angle of reflection of the EM wave at $(x, 0) \in \mathscr{S}$ is denoted by $\theta_R(x)$, and it represents the angle formed by the y-axis and the wavefront of the EM wave that is emitted by $\mathscr{S}$ at $(x, 0)$ and is observed at Rx.

For simplicity, we assume $d_T(x) \gg \lambda$ and $d_R(x) \gg \lambda$, which usually hold true in practical setups. Under these assumptions, the electric field emitted by the point source (Tx) and observed at Rx in the absence of $\mathscr{S}$ corresponds to the Green function in the plane, which is well approximated as follows:

$$E_x\left(x_R, y_R\right) \approx E_0 \frac{\exp\left(-jkd_{TR}\left(x_R, y_R\right)\right)}{\sqrt{kd_{TR}\left(x_R, y_R\right)}} \tag{3.1}$$

where $E_0 = -j\sqrt{1/(8\pi)}\exp\left(-j\pi/4\right)$, j is the imaginary unit, and $d_{TR}\left(x_R, y_R\right)$ is the distance between Tx and Rx.

The surface $\mathscr{S}$ is modeled as a spatially-inhomogeneous reflector that is capable of modifying the phase of the incident field. We assume that the electromagnetic properties of the surface $\mathscr{S}$ vary slowly, as compared with the wavelength, along the surface itself. Under this approximation, the surface $\mathscr{S}$ can be well modeled as a local structure: the reflected field at $(x,0) \in \mathscr{S}$ depends, approximately, only on the incident field at $(x,0) \in \mathscr{S}$ [51]. More precisely, the reflection coefficient at $(x,0) \in \mathscr{S}$ can be written as follows:

$$\Gamma_r(x) = C(x)\exp\left(j\Phi(x)\right) \tag{3.2}$$

where $C(x) \in \mathbb{R}^+$ and $\Phi(x) \in [0,2\pi)$ denote the amplitude and phase of the reflection coefficient, respectively. In this chapter, we assume $C(x) = 1$.

In vacuum, the x and y components of the electric field are not coupled, and we assume that $\mathscr{S}$ does not change the polarization of the EM waves. Under these assumptions, we can analyze any components of the electric field. In this chapter, we consider the tangential (to the surface $\mathscr{S}$) component of the electric field, which is denoted by $E_x\left(x_R, y_R\right)$. To this end, the achievable data rate in RIS-assisted communication is given as follows:

$$R_{\text{RIS}} = \log_2\left(1 + \frac{P}{N_0}|E_x(x_R, y_R)|^2\right). \tag{3.3}$$

where P is the transmit power by Tx and N_0 is the noise power at Rx.

In the following, we present analytical results on the intensity of the electric field at Rx due to the presence of RIS, as reported in [1]. In its most general form, based on our system model of RIS, the intensity of the electric field emitted by Tx, reflected by the surface $\mathscr{S}$, and observed at $\left(x_R, y_R\right)$ is given as follows.

Theorem 3.1 *Let us assume $d_T(x) \gg \lambda$ and $d_R(x) \gg \lambda$. The electric field $E_x\left(x_R, y_R\right)$ can be formulated as follows:*

$$E_x\left(x_R, y_R\right) = \mathscr{I}_0 \int_{-L}^{+L} \mathscr{I}(x)\exp\left(-jk\mathscr{P}(x)\right)dx \tag{3.4}$$

where $\mathscr{I}_0 = 1/(8\pi)$, $\mathscr{P}(x) = d_T(x) + d_R(x) - \Phi(x)$, and $\mathscr{I}(x) = \frac{1}{\sqrt{d_T(x)d_R(x)}}\left(\frac{y_T}{d_T(x)} + \frac{y_R}{d_R(x)}\right)$.

Proof 3.1 *See Appendix 3.6.1*

The electric field in (3.4) is formulated in a simple integral form, which, however, does not explicitly unveil the dependency of the electric field as a function of the transmission distances. Also, the electric field depends on the specific phase shift $\Phi(x)$ applied by the surface $\mathscr{S}$. In the following, we consider two case studies for the choice of $\Phi(x)$:

1. $\mathscr{S}$ as ***anomalous reflector***: In this case study, the surface $\mathscr{S}$ operates as an anomalously

reflecting RIS, i.e., it is configured for reflecting the EM waves emitted by Tx towards a given direction. To this end, $\Phi(x)$ in (3.4) is chosen as follows:

$$\Phi(x) = \left(\bar{\phi}_T - \bar{\phi}_R\right) x + \phi_0 / k \tag{3.5}$$

where $\phi_0 \in [0, 2\pi)$ is a fixed phase shift, $\bar{\phi}_T = -\bar{x}_T \Big/ \sqrt{\bar{x}_T^2 + \bar{y}_T^2}$, $\bar{\phi}_R = \bar{x}_R \Big/ \sqrt{\bar{x}_R^2 + \bar{y}_R^2}$, and $\left(\bar{x}_T, \bar{y}_T\right)$ and $\left(\bar{x}_R, \bar{y}_R\right)$ are parameters that are optimized for obtaining desired reflection capability [1]. The results are given in two distinct approximation regimes, namely the electrically-large regime and the electrically-small regime, which are the representative of near-field and far-field, respectively. The regimes are defined as follows:

Definition 3.1 *Let us define* $d_Q(x) = \sqrt{\left(x_Q - x\right)^2 + y_Q^2}$ *and* $\sin\left(\theta_{Q0}\right) = \sin\left(\theta_Q(0)\right) = -q x_Q / d_{Q0}$ *for* $Q = \{T, R\}$, *where* $q = 1$ *if* $Q = T$ *and* $q = -1$ *if* $Q = R$, *as well as* $d_{Q0} = \sqrt{x_Q^2 + y_Q^2}$ *for* $Q = \{T, R\}$. *The system is said to operate in the electrically-small regime if the approximation* $d_Q(x) \approx d_{Q0} + q x \sin\left(\theta_{Q0}\right)$ *holds true for* $Q = \{T, R\}$. *Otherwise, it is said to operate in the electrically-large regime.*

Proposition 3.1 *Assume* $\bar{\phi}_T = \sin(\theta_T(0))$ *and* $\bar{\phi}_R = \sin(\theta_R(0))$. *In the electrically-large regime, the intensity of the electric field in* (3.4) *can be approximated as follows:*

$$\left|E_x\left(x_R, y_R\right)\right| \approx \frac{1}{4\sqrt{2\pi k}} \frac{\sqrt{1-\bar{\phi}_T^2} + \sqrt{1-\bar{\phi}_R^2}}{\sqrt{\left(1-\bar{\phi}_R^2\right) d_{T0} + \left(1-\bar{\phi}_T^2\right) d_{R0}}} \tag{3.6}$$

Proof 3.2 *See Appendix 3.6.2.*

Proposition 3.2 *Assume* $\bar{\phi}_T = \sin(\theta_T(0))$ *and* $\bar{\phi}_R = \sin(\theta_R(0))$. *In the electrically-small regime, the intensity of the electric field in* (3.4) *can be approximated as follows:*

$$\left|E_x\left(x_R, y_R\right)\right| \approx \frac{L}{4\pi} \frac{\sqrt{1-\bar{\phi}_T^2} + \sqrt{1-\bar{\phi}_R^2}}{\sqrt{d_{T0} d_{R0}}} \tag{3.7}$$

Proof 3.3 *See Appendix 3.6.3.*

2. $\mathcal{S}$ as ***focusing lens***: We analyze the case study in which the surface $\mathcal{S}$ operates as an RIS whose phase $\Phi(x)$ is appropriately optimized in order for $\mathcal{S}$ to act as a focusing lens. In particular, we consider that $\mathcal{S}$ acts as a beamformer (or a reflecting lens) that focuses the signal towards a single location $\left(\bar{x}_R, \bar{y}_R\right)$. To this end, $\Phi(x)$ in (3.4) is chosen as follows:

$$\Phi(x) = \sqrt{(x - x_T)^2 + y_T^2} + \sqrt{(x - \bar{x}_R)^2 + \bar{y}_R^2} \tag{3.8}$$

The exact formulation and the electrically-small approximation of the electric field intensity at the receiver is given as the following.

Proposition 3.3 *In case of $\mathscr{S}$ configured as a focusing lens, the intensity of the electric field in* (3.4) *can be approximated as follows:*

$$E_x\left(x_R, y_R\right) = \frac{1}{8\pi}\int_{-L}^{+L}\frac{1}{\sqrt{d_T\left(x\right)d_R\left(x\right)}}\left(\frac{y_T}{d_T\left(x\right)}+\frac{y_R}{d_R\left(x\right)}\right)dx \tag{3.9}$$

Denote $\bar{\phi}_{T0} = \sin\left(\theta_T\left(0\right)\right)$ *and* $\bar{\phi}_{R0} = \sin\left(\theta_R\left(0\right)\right)$. *In the electrically-small regime, the intensity of the electric field in* (3.4) *can be approximated as follows:*

$$\left|E_x\left(x_R, y_R\right)\right| \approx \frac{L}{4\pi}\frac{\sqrt{1-\bar{\phi}_{T0}^2}+\sqrt{1-\bar{\phi}_{R0}^2}}{\sqrt{d_{T0}d_{R0}}} \tag{3.10}$$

Proof 3.4 *The proof follows from substituting* (3.8) *into* (3.4) *and using similar steps as in 3.2.*

We note that, as explained in [1], the electrically-large approximation of the electric field in case of RIS acting as a focusing lens is not available.

Relay

We consider communication from a single transmitter (Tx) antenna to a single receiver (Rx) antenna. We also assume that the communication is assisted by a half-duplex (HD) relay with repetition-coded decode and forward (DF) relaying protocol, in which the transmission procedure is divided into two equal-sized phases:

- In the first phase, Tx transmits the signal containing information which is received at Rx as follows:

  $$y_{\mathrm{1d}} = h_{\mathrm{sd}}\sqrt{P_1}s + n_{\mathrm{1d}}, \tag{3.11}$$

  where $h_{\mathrm{sd}} \in \mathbb{C}$ is the channel gain between Tx (source) and Rx (destination), P_1 is the transmit power in the first phase, s is the unit-power information signal, and $n_{\mathrm{1d}} \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ is the noise power at Rx in the first phase. In the same phase, the received signal at the relay is

  $$y_{\mathrm{1r}} = h_{\mathrm{sr}}\sqrt{P_1}s + n_{\mathrm{1r}}, \tag{3.12}$$

  where $h_{\mathrm{sr}} \in \mathbb{C}$ denotes the channel gain between Tx and relay, while $n_{\mathrm{1r}} \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ is the noise at relay. The DF relay uses y_{1r} to decode the information and then encodes it again for transmission in the second phase.

- In the second phase, the relay transmits the following signal: $\sqrt{P_2}s$ where P_2 is the transmit power by relay in the second phase. In this case, the received signal at Rx (from the relay) is given as follows:

$$y_{2d} = h_{\text{rd}}\sqrt{P_2}s + n_{2\text{d}}, \tag{3.13}$$

where $h_{\text{rd}} \in \mathbb{C}$ denotes the channel between the relay and Rx and $n_{2\text{d}} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ is the noise at Rx in the second phase.

Assuming that the relay-aided system operates in vacuum, the channel gain between Tx and Rx, Tx and relay, as well as relay and Rx are equivalent to the free-space propagation path-loss, which is given by the Green's function in a 2D-space, i.e.:

$$h_{\text{sd}} \approx -\frac{j}{\sqrt{8\pi}}\frac{\exp\left(-jkd_{SD} - j\pi/4\right)}{\sqrt{kd_{SD}}} \tag{3.14}$$

$$h_{\text{sd}} \approx -\frac{j}{\sqrt{8\pi}}\frac{\exp\left(-jkd_{SR} - j\pi/4\right)}{\sqrt{kd_{SR}}} \tag{3.15}$$

$$h_{\text{sd}} \approx -\frac{j}{\sqrt{8\pi}}\frac{\exp\left(-jkd_{RD} - j\pi/4\right)}{\sqrt{kd_{RD}}} \tag{3.16}$$

where d_{SD}, d_{SR}, and d_{RD} are the distance between Tx and Rx, between Tx and relay, and between relay and Rx, respectively. We note, in particular, that the channel gains h_{sd}, h_{sr} and h_{rd} include the impact of transmission distance and the wavelength λ, i.e., we have

$$|h_{\text{sd}}| \propto \left(\frac{\lambda}{d_{\text{SD}}}\right)^{1/2}, \qquad |h_{\text{sr}}| \propto \left(\frac{\lambda}{d_{\text{SR}}}\right)^{1/2}, \qquad |h_{\text{rd}}| \propto \left(\frac{\lambda}{d_{\text{RD}}}\right)^{1/2} \tag{3.17}$$

3.3.2 Quantitative Comparison

Having presented the corresponding analytical formulation of the achievable data rate for RIS-assisted and relay-assisted systems in Sec. 3.3.1, we compare the performance of both systems. We distinguish our comparison into two categories: (i) Average Signal-to-Noise Ratio vs. Number of Elements, and (ii) Average Signal-to-Noise Ratio vs. Transmission Distance.

Average Signal-to-Noise Ratio vs. Number of Elements

Let us consider a multiple-antenna relay that employs maximum ratio weighting for reception and transmission. If N antennas are used at the relay, the average end-to-end signal-to-noise ratio increases *linearly* with N. On the other hand, the average end-to-end signal-to-noise ratio of a RIS made of N individually tunable antennas (or N reconfigurable meta-surfaces, each of them made of an appropriate number of meta-atoms to realize the desired wave transformations) increases *quadratically* with N, while still being subject to the energy conservation principle [3], [52]. Based on existing prototypes for wireless applications, N may be of the order of a few thousands if the RIS is realized by using individually tunable inexpensive antennas [3], and of the order of ten thousands if it is based on meta-surfaces [33].

The different scaling laws as a function of N can be understood as follows. In relays, the available power is allocated among the N antennas so that the total power is kept constant. In RISs, in contrast, each constituent antenna or meta-surface reflects, after scaling the received signal by the transmittance and with no noise addition, the same amount of power received from the transmitter.

It is worth mentioning that, however, the more favorable scaling law as a function on N does not necessarily imply that RISs outperform relays. For a fixed total power constraint, in fact, the path loss as a function of the transmission distance cannot be overlooked. This is discussed next by considering, for ease of exposition and without loss of generality, a free-space propagation model and $N = 1$ for both relays and RISs.

3.3.3 Average Signal-to-Noise Ratio vs. Transmission Distance

For the sake of numerical comparison, we refer to the system model and the analysis in Sec. 3.3.1. To this end, we use the same notation to denote the distance between Tx and Rx (d_{SD}), between Tx and RIS/relay (d_{SR}), and between RIS/relay and Rx (d_{RD}). Under these assumptions, the end-to-end power received from an DF relay scales with the reciprocal of the product of the transmitter-to-relay distance and the relay-to-receiver distance [50], i.e., as $\left(k^2 d_{\mathrm{SR}} d_{\mathrm{RD}}\right)^{-1}$. When considering the effect of noise, the end-to-end signal-to-noise ratio of DF relaying scales with the reciprocal of the distance of the weakest of the two paths, i.e., as $\min\left\{(k d_{\mathrm{SR}})^{-1}, (k d_{\mathrm{RD}})^{-1}\right\}$. On the other hand, the total power reflected by an RIS, and hence the scaling law of the received power as a function of the distance, depend on the relation between the geometric size of the RIS, the wavelength of the radio wave, and the relative transmitter-to-RIS and RIS-to-receiver distances. Based on Sec. 3.3.1, two notable regimes are worth of analysis.

- **Electrically large RISs**: If the geometric size of the RIS is large enough as compared with the wavelength and the transmission distances (d_{SR} and d_{RD}), the RIS behaves, asymptotically, as an anomalous mirror. In this regime, the power received from the RIS and the end-to-end average signal-to-noise ratio at the receiver scale, as function of the distance, as $\left(\alpha k d_{\mathrm{SR}} + \beta k d_{\mathrm{RD}}\right)^{-1}$, where α and β depend on the specified angles of incidence and reflection of the radio waves [1, Eq. (10)];
- **Electrically small RISs**: If the size of the RIS is not large enough as compared with the wavelength and the transmission distances (d_{SR} and d_{RD}), the RIS behaves, asymptotically, as a diffuser. In this regime, the received power and the end-to-end average signal-to-noise ratio at the receiver scale, as a function of the distance, as $4L^2 (d_{\mathrm{SR}} d_{\mathrm{RD}})^{-1}$ [1, Eq. (11)]. This is the same scaling law as for the received power of AF relaying. Notably, the end-to-end average signal-to-noise ratio depends on the length, $2L$, of the RIS.

The analysis of electrically large RISs is a relevant case study because of the large geometric size that some implementations of RISs may have. A recent prototype of RIS reported in [33], whose size is 1 m^2 and whose frequency of operation is 10.5 GHz, is shown to operate in the far-field at distances greater than 70 m based on analytical formulas and at distances of the order of 28 m based on experimental measurements. In typical indoor environments, therefore, an RIS of this kind may be viewed as electrically large by transmitters and receivers.

3.3.4 Takeaway Messages from the Comparison

Based on the considerations and case studies analyzed in the previous sub-section, it is interesting to compare the scaling laws of RISs and relays as a function of the transmission distance. Let us assume, for simplicity, $d_{\mathrm{SR}} = d_{\mathrm{RD}} = d_0$, i.e., the RIS/relay is located equidistantly from the transmitter and receiver. Also, let M_{ma} denote the number of meta-atoms of the RIS and let λ/D with $D > 1$ be their inter-distance. Thus, $2L = M_{\mathrm{ma}}\lambda/D$, and the average end-to-end signal-to-noise ratio scales, as a function of the distance, as follows.

- Relay-aided transmission: $\propto 1/d_0$;
- Electrically large RIS: $\propto 1/\left(\alpha d_0 + \beta d_0\right)$;
- Electrically small RIS: $\propto 4L^2/d_0^2 \propto M_{\mathrm{ma}}^2/d_0^2$.

Accordingly, the following conclusions can be drawn:

- Relay-aided transmission and electrically large RISs (i.e., with a slight abuse of terminology, for short distances d_0) offer a similar scaling law as a function of the distance. Since RISs are not subject to the half-duplex constraint and the loop-back self-interference, they have the potential of providing a better rate than relays if, for a fixed size of the RIS, the distances are not too long;
- Compared with relays, electrically small RISs (i.e., with a slight abuse of terminology, for long distances d_0) offer a less favorable scaling law as a function of the distance. However, the average end-to-end signal-to-noise ratio of electrically small RISs scales quadratically with their size, i.e., quadratically with M_{ma} if D is kept fixed. Thus, a sufficiently large RIS (but still electrically small) has the potential of outperforming relay-aided transmission.

Based on these findings, it can be concluded that RIS-aided transmission may outperform relay-aided transmission provided that the size of the RIS is sufficiently large.

Table 3.1 – Rate (R) for relays and RISs.

Transmission frequency	f_c
Wavelength	λ
Wave number	$k = 2\pi / \lambda$
Electric field (distance d)	$\lvert E(d) \rvert = E_0 \big/ \sqrt{kd}$ [1, Eq. (1)]
Transmit power (RIS)	P
Transmit power (relay)	$P_R = P/2$
Noise power (receiver)	N_0
Self-interference	$I_S = 10 N_0 P_R$
HD DF relay	$R = (1/2)\log_2\left(1 + \left(P_R / N_0\right) \lvert E(d) \rvert^2\right)$
FD DF relay	$R = \log_2\left(1 + \left(P_R / \left(N_0 + I_S\right)\right) \lvert E(d) \rvert^2\right)$
Ideal FD DF relay	$R = \log_2\left(1 + \left(P_R / N_0\right) \lvert E(d) \rvert^2\right)$
RIS - General formula	$R = \log_2\left(1 + \left(P / N_0\right) \left\lvert E_{\mathrm{ris}}(d) \right\rvert^2\right)$
RIS - Mirror (exact)	$E_{\mathrm{ris}}(d)$ in [1, Eq. (3)]
RIS - Mirror (short d)	$E_{\mathrm{ris}}(d)$ in [1, Eq. (10)]
RIS - Mirror (long d)	$E_{\mathrm{ris}}(d)$ in [1, Eq. (11)]
RIS - Lens (exact)	$E_{\mathrm{ris}}(d)$ in [1, Eq. (3)] with $\mathscr{P}(x) = 0$

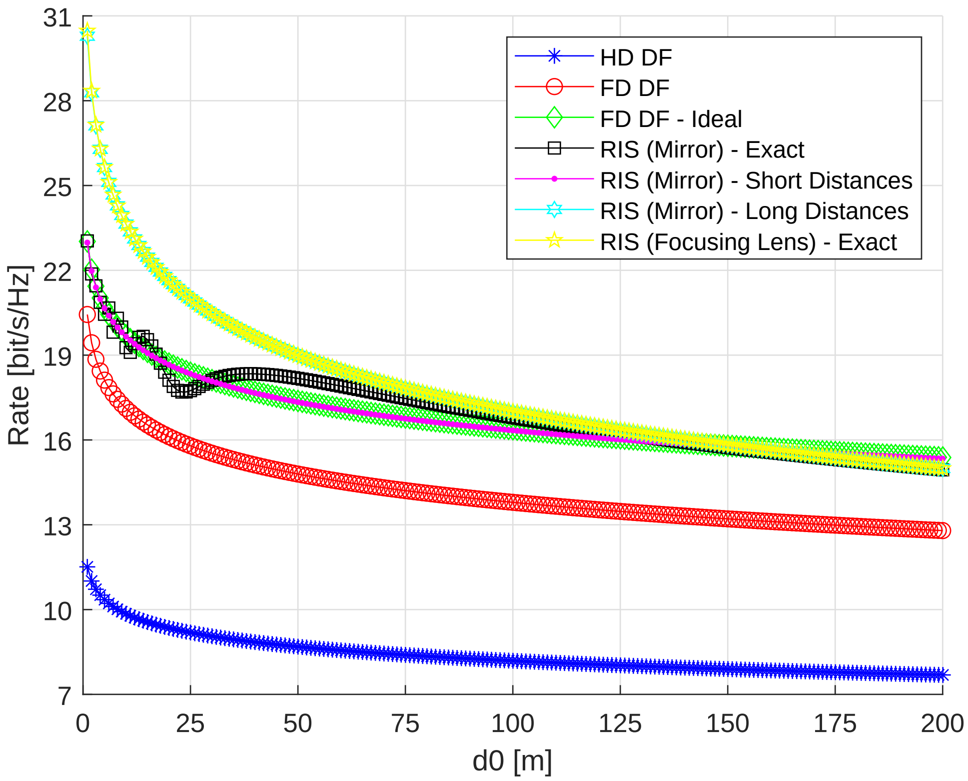


Figure 3.2 – Data rate of RISs and relays versus the transmission distance.

3.4 Numerical Results

In this section, we report some numerical illustrations in order to quantitatively compare RISs and relays. For simplicity, we consider a single relay and a single RIS (i.e., $N = 1$), and

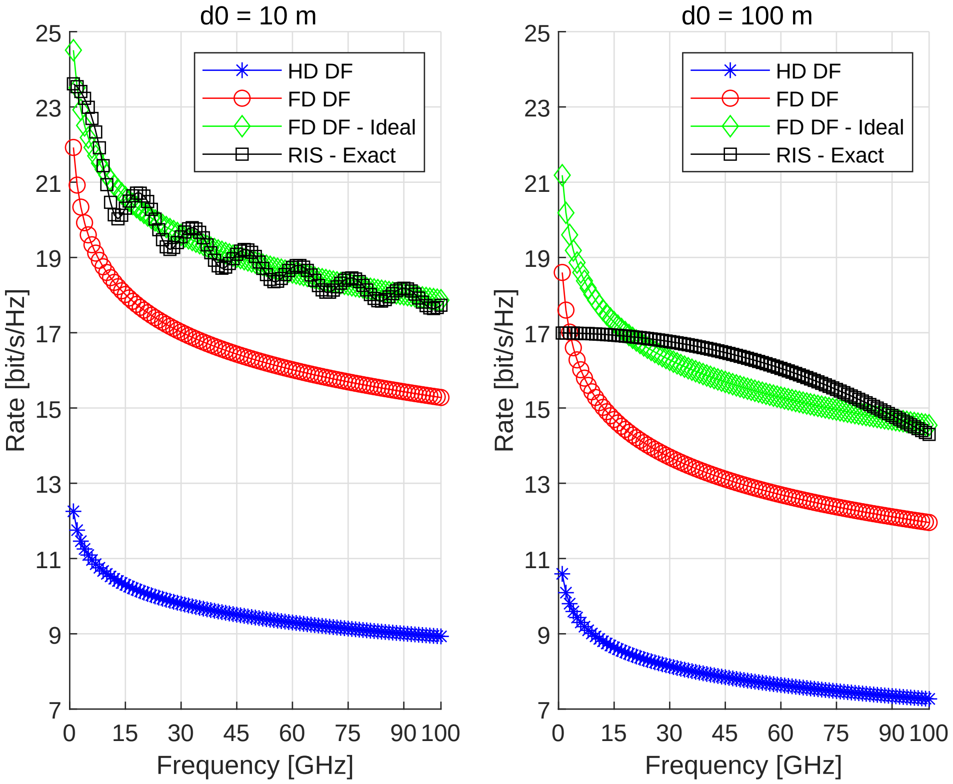


Figure 3.3 – Data rate of RISs and relays versus the transmission frequency.

assume that they are located equidistantly from the transmitter and receiver. For the relay, the results are obtained by using the formulas in Table 3.1. Table 3.1 reports also the rate of an ideal FD relay, in which the residual loop-back self-interference is assumed to be zero. A total power constraint is assumed and, therefore, the total power is equally split between the transmitter and the relay. On the other hand, for the RIS, the intensity of the electric field is given in Sec. 3.3.1 as reported again here in Table 3.1. Also, as in Sec. 3.3.1, a two-dimensional system model is assumed. Therefore, the intensity of the electric field decays with the square root of the distance.

The distance between the transmitter and the relay/RIS, and the relay/RIS and the receiver is denoted by d_0. The RIS is modeled as a straight line centered at the origin, which views the transmitter and receiver under an angle of 45 and 60 degrees with respect to the normal at the origin, respectively. The total length of the RIS is $2L$. The reflection coefficient of the RIS is chosen as elaborated in [1, Eqs. (2), (9)]. Further information about the RIS can be found in [1]. The signal-to-noise ratio at a distance of 1 m is $P/N_0 = 114$ dB.

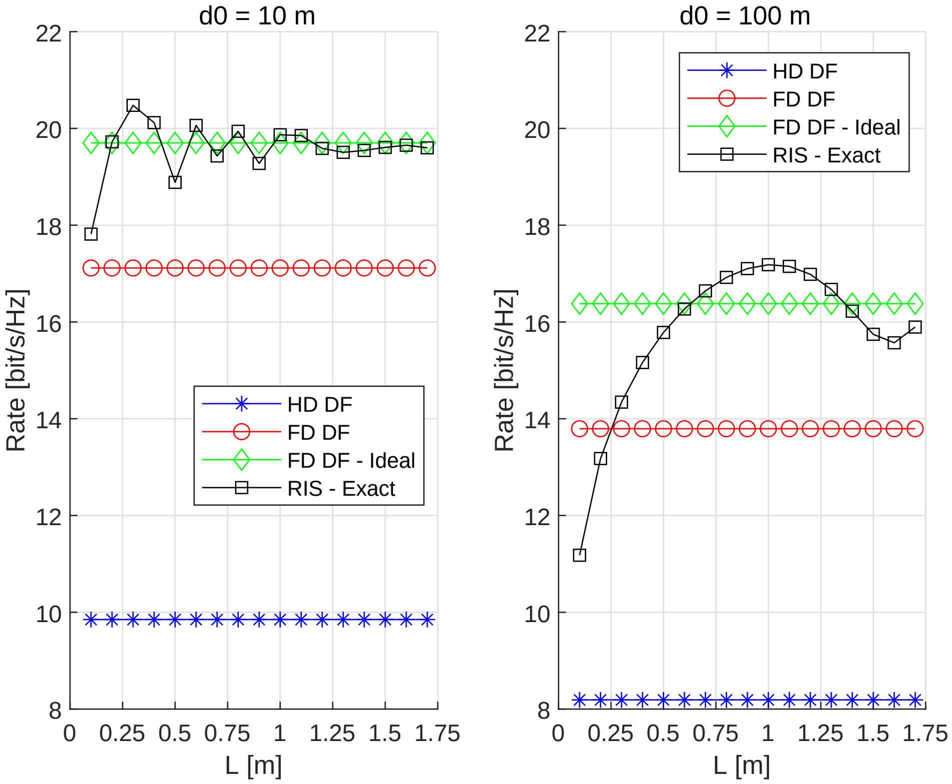


Figure 3.4 – Data rate of RISs and relays versus the size of the RIS.

3.4.1 RISs vs. Relays as a Function of the Transmission Distance

In Fig. 3.2, we compare the data rate of an RIS and a relay as a function of the distance d_0, by assuming a transmission frequency equal to $f_c = 28$ GHz. For comparison, the RIS is configured to operate as an anomalous reflector and as a focusing lens. This latter case study is discussed next. The RIS is of length $2L = 1.5$ m, which corresponds to 140λ. The exact analytical framework in [1, Eq. (3)], and the approximations for short and long transmission distances in [1, Eq. (10)] and [1, Eq. (11)], respectively, are reported. The figure shows that an RIS provides a rate similar to an ideal FD relay without the need of using a power amplifier. This is obtained thanks to the size (effective length) of the RIS. By assuming, for example, that the inter-distance between the meta-atoms of the RIS is in the range $\lambda/5$ and $\lambda/2$, the results in Fig. 3.2 can be obtained if the number of meta-atoms of the RIS is in the range $M_{\text{ma}} = 700$ and $M_{\text{ma}} = 280$, respectively. The specific implementation depends on the technology employed and the range of directions for which specified anomalous reflection capabilities are needed. It is worth noting that, based on Fig. 3.2, the RIS under analysis behaves as an anomalous mirror (i.e., it is viewed as electrically large) for distances d_0 up to 25-50 m and as a diffuse scatterer (i.e., it is viewed as electrically small) for distances d_0 greater than 75-100 m. Figure 3.2 shows, in addition, that an ideal FD relay outperforms an RIS for large transmission distances (greater than 150 m in the considered setup). For long transmission distances, therefore, a larger RIS may be needed for outperforming an ideal FD relay.

3.4.2 RISs: Anomalous Mirrors vs. Focusing Lenses

For completeness, Fig. 3.2 reports the rate of an RIS that is configured to operate as a focusing lens (i.e., a beamformer) In this latter case, the intensity of the received power scales as a function of the product of the distance between the transmitter and the RIS, and the distance between the RIS and the receiver [33]. As expected, Fig. 3.2 shows that an RIS configured to operate as a focusing lens outperforms, in general, an RIS configured to operate as an anomalous reflector. It is interesting to note that, in the setup of Fig. 3.2, an RIS that operates as a focusing lens yields similar rates as the long distance approximation of an RIS that operates as an anomalous reflector (a phase gradient meta-surface [1, Eq. (9)]). The price to pay for this performance gain lies in the need of estimating the exact locations of the transmitter and receiver, and in the need of adapting the phases of the RIS to the wireless channels. An anomalous reflector based on a phase gradient meta-surface requires, on the other hand, the knowledge of only the desired directions of incidence and reflection of the radio waves. It is interesting to observe, however, that a sufficiently long RIS that is designed to operate as a simple phase gradient meta-surface is capable of outperforming an ideal FD relay.

3.4.3 RISs vs. Relays as a Function of the Carrier Frequency

In Fig. 3.3, we compare the data rate of the RIS and relay as a function of the transmission frequency f_c. Two transmission distances are considered, which may be representative of indoor ($d_0 = 10$ m) and outdoor ($d_0 = 100$ m) scenarios. The total length of the RIS is $2L = 1.5$ m. If $d_0 = 10$ m, we obtain findings similar to Fig. 3.2. If $d_0 = 100$ m, in contrast, the performance trend is different: If f_c is not large enough (approximately greater than 20 GHz in the considered example), the length of the RIS is insufficient for outperforming an ideal FD relay. In this case, therefore, an ideal FD relay outperforms an RIS at the price of a higher complexity and power consumption. At higher frequencies, on the other hand, an RIS provides similar rates as an ideal FD relay. This is similar to the findings obtained in Fig. 3.2.

3.4.4 RISs vs. Relays as a Function of the Size of the RIS

In Fig. 3.4, we compare the data rate of the RIS and relay as a function of the size of the RIS L, by assuming $f_c = 28$ GHz. Similar to Fig. 3.3, two transmission distances are analyzed. Once again, we observe that an RIS provides similar rates as an ideal FD relay provided that its is sufficiently (electrically) large as compared with the wavelength λ. If $d_0 = 100$ m, for example, this holds true if the length of the RIS is of the order of $L = 0.5$-0.75 m.

It is worth noting that Figs. 3.2-3.4 show, for short transmission distances, the typical and expected oscillating behavior that is caused by the coherent sum of the many secondary waves, with a different phase, reflected by the RIS [1, Eq. (3)].

3.5 Conclusions

RISs are an emerging and little understood technology with several applications in wireless networks. In this chapter, we have discussed the differences and similarities between relays and RISs that are configured to operate as anomalous reflectors. With the aid of simple scaling laws and numerical simulations, we have provided arguments showing that sufficiently large RISs can outperform relay-aided systems in terms of data rate, while reducing the implementation complexity. The obtained results unveil the advantages and limitations, as compared with relays, of employing RISs that operate as anomalous reflectors in wireless networks.

3.6 Appendices

3.6.1 Proof of Theorem 3.1

A more rigorous proof is given in Chapter 4 where we use the Green's theorem for vector field and by using the general scalar theory of diffraction and by applying appropriate boundary conditions at the surface $\mathscr{S}$. For the moment, it suffices to give the proof starting from the Kirchhoff scalar diffraction theory given in (2.34). According to (2.34), the field reflected by $\mathscr{S}$ and received at Rx is given as follows:

$$E_x(\mathbf{r}_{\mathrm{Rx}}) = \int_S \big(G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')(\hat{\mathbf{n}}_{\mathrm{out}}\cdot\nabla_{\mathbf{r}'})E_x(\mathbf{r}') - E_x(\mathbf{r}')(\hat{\mathbf{n}}_{\mathrm{out}}\cdot\nabla_{\mathbf{r}'})G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')\big)\, d\mathbf{r}'.$$

where $G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')$ and $E_x(\mathbf{r}')$ are the 2D Green's function and the reflected field at $\mathscr{S}$, respectively, given as

$$G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}') = E_0\frac{\exp\big(-jkd_R(x,y)\big)}{\sqrt{kd_R(x,y)}}\Big|_{y=0}, \qquad E_x(\mathbf{r}') = \Gamma_r(x)E_0\frac{\exp\big(-jkd_T(x,y)\big)}{\sqrt{kd_T(x,y)}}\Big|_{y=0} \tag{3.18}$$

Here, we use notations $d_T(x,y) = \sqrt{(x_T - x)^2 + (y_T - y)^2}$ and $d_R(x,y) = \sqrt{(x_R - x)^2 + (y_R - y)^2}$. Note that $\hat{\mathbf{n}}_{\mathrm{out}} = -\hat{\mathbf{y}}$. Also, since the y-component of the incident wave (represented by $E_x(\mathbf{r}')$) is directed towards $-\hat{\mathbf{y}}$ and the y-component of the reflected wave (represented by $G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}')$) is directed towards $+\hat{\mathbf{y}}$, we have

$$(\hat{\mathbf{n}}_{\mathrm{out}}\cdot\nabla_{\mathbf{r}'})E_x(\mathbf{r}') = +\frac{\partial}{\partial y}E_x(\mathbf{r}') = +\Gamma_r(x)E_0\frac{\partial}{\partial y}\frac{\exp\big(-jkd_T(x,y)\big)}{\sqrt{kd_T(x,y)}}\Big|_{y=0} \tag{3.19}$$

$$(\hat{\mathbf{n}}_{\mathrm{out}}\cdot\nabla_{\mathbf{r}'})G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}') = -\frac{\partial}{\partial y}G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}') = -E_0\frac{\partial}{\partial y}\frac{\exp\big(-jkd_R(x,y)\big)}{\sqrt{kd_R(x,y)}}\Big|_{y=0} \tag{3.20}$$

Therefore, we can write (3.18) as follows:

$$\begin{aligned} E_x(x_R, y_R) = \int_S \Bigg(& E_0 \frac{\exp\left(-jkd_R(x,0)\right)}{\sqrt{kd_R(x,0)}} \left(+\Gamma_r(x) E_0 \frac{\partial}{\partial y} \frac{\exp\left(-jkd_T(x,y)\right)}{\sqrt{kd_T(x,y)}} \Big|_{y=0} \right) \\ & -\Gamma_r(x) E_0 \frac{\exp\left(-jkd_T(x,0)\right)}{\sqrt{kd_T(x,0)}} \left(-E_0 \frac{\partial}{\partial y} \frac{\exp\left(-jkd_R(x,y)\right)}{\sqrt{kd_R(x,y)}} \Big|_{y=0} \right) \Bigg) d\mathbf{r}'. \end{aligned} \tag{3.21}$$

The proof follows from computing the derivatives in (3.21), substituting $\Gamma_r(x) = C(x)\exp\left(j\Phi(x)\right)$ and $E_0 = -j\sqrt{1/(8\pi)}\exp\left(-j\pi/4\right)$, as well as using the fact that $k \gg 1/d_T(x,0)$ and $k \gg 1/d_T(x,0)$ and which implies that following approximations hold for $Q \in \{T, R\}$:

$$\frac{\partial}{\partial y} \frac{\exp\left(-jkd_Q(x,y)\right)}{\sqrt{kd_Q(x,y)}} \Big|_{y=0} \approx jk \frac{\exp\left(-jkd_Q(x,0)\right)}{\sqrt{kd_Q(x,0)}} \frac{y_Q}{d_Q(x,0)}. \tag{3.22}$$

3.6.2 Proof of Proposition 3.1

In case of an RIS acting as anomalous reflector, $\Phi(x) = \left(\bar{\phi}_T - \bar{\phi}_R\right) x + \phi_0 / k$ where $\bar{\phi}_T = \sin\theta_{T0}$ and $\bar{\phi}_R = \sin\theta_{R0}$. Thus $\mathscr{P}'(x) = 0$ implies

$$\frac{\partial}{\partial x}\left(d_T(x) + d_R(x)\right) = \left(\bar{\phi}_T - \bar{\phi}_R\right) \tag{3.23}$$

or, equivalently,

$$\frac{(x - x_T)}{d_T(x)} - \frac{(x_R - x)}{d_R(x)} = \left(\bar{\phi}_T - \bar{\phi}_R\right) \tag{3.24}$$

By definition, $(x - x_T)/d_T(x) = \sin\theta_T(x)$ and $(x_R - x)/d_R(x) = \sin\theta_R(x)$. Thus, the solution to (3.24), called x_s, satisfies the condition

$$\sin\theta_T(x_s) - \sin\theta_R(x_s) = \left(\bar{\phi}_T - \bar{\phi}_R\right) = \sin\theta_{T0} - \sin\theta_{R0}. \tag{3.25}$$

Due to the monotinicity of the left-hand side expression of (3.25), x_s is unique and thus it is straightforward to see that $x_s = 0$ is the only solution to (3.25). Using this fact, we write $y_T/d_{T0} = \cos\theta_{T0}$ and $y_R/d_{R0} = \cos\theta_{R0}$. Similarly, we have

$$\mathscr{P}''(x_s) = \frac{\partial^2}{\partial x^2}\left(d_T(x) + d_R(x)\right)_{|x=0} = -\frac{\cos^2\theta_{T0}}{d_{T0}} - \frac{\cos^2\theta_{R0}}{d_{R0}}. \tag{3.26}$$

Clearly, $\mathscr{P}''(x_s) < 0$. Therefore, using the stationary approximation method (introduced in (4.15)), the approximation of $E_x(x_R, y_R)$ is given as

$$E_x(x_R, y_R) \approx \frac{1}{4\sqrt{2\pi k}} \frac{\cos\theta_{T0} + \cos\theta_{R0}}{\sqrt{\cos^2\theta_{R0} d_{T0} + \cos^2\theta_{T0} d_{R0}}} \exp\left(-jk\left(d_{T0} + d_{R0} - \frac{\phi_0}{k}\right) - \frac{j\pi}{4}\right)$$

$$\overset{(a)}{=} \frac{1}{4\sqrt{2\pi k}} \frac{\sqrt{1-\bar{\phi}_T^2}+\sqrt{1-\bar{\phi}_R^2}}{\sqrt{\left(1-\bar{\phi}_R^2\right) d_{T0}+\left(1-\bar{\phi}_T^2\right) d_{R0}}} \exp\left(-jk\left(d_{T0}+d_{R0}-\frac{\phi_0}{k}\right)-\frac{j\pi}{4}\right) \tag{3.27}$$

where (a) is due to the fact that $\cos\theta_{T0} = \sqrt{1-\bar{\phi}_T^2}$ and $\cos\theta_{R0} = \sqrt{1-\bar{\phi}_R^2}$. The intensity of the electric field is then given as

$$\left|E_x(x_R, y_R)\right| \approx \frac{1}{4\sqrt{2\pi k}} \frac{\sqrt{1-\bar{\phi}_T^2}+\sqrt{1-\bar{\phi}_R^2}}{\sqrt{\left(1-\bar{\phi}_R^2\right) d_{T0}+\left(1-\bar{\phi}_T^2\right) d_{R0}}}. \tag{3.28}$$

3.6.3 Proof of Proposition 3.2

In the electrically-small regime, we approximate the distances $d_T(x)$ and $d_R(x)$ according to Definition 4.1, i.e.:

$$d_T(x) \approx d_{T0} + x\sin\theta_{T0}, \qquad d_R(x) \approx d_{R0} - x\sin\theta_{R0}, \tag{3.29}$$

with $d_{T0} = d_T(0)$, $d_{R0} = d_R(0)$, $\theta_{T0} = \theta_R(0)$ and $\theta_{R0} = \theta_R(0)$. Thus, (3.4) can be approximated as

$$\begin{aligned}
E_x\left(x_R, y_R\right) &\approx \frac{1}{8\pi}\int_{-L}^{L} \frac{\exp\left(-jk\left(d_{T0}+d_{R0}+x\left(\sin\theta_{T0}-\sin\theta_{R0}\right)-\Phi(x)\right)\right)}{\sqrt{d_{T0}d_{R0}+x\left(d_{R0}\sin\theta_{T0}-d_{T0}\sin\theta_{R0}\right)+x^2\sin\theta_{T0}\sin\theta_{R0}}} \\
&\left(\frac{y_T}{d_{T0}+x\sin\theta_{T0}}+\frac{y_R}{d_{R0}-x\sin\theta_{R0}}\right)dx. \\
&\overset{(a)}{\approx} \frac{\left(\frac{y_T}{d_{T0}}+\frac{y_R}{d_{R0}}\right)\exp\left(-jk\left(d_{T0}+d_{R0}\right)\right)}{8\pi\sqrt{d_{T0}d_{R0}}}\int_{-L}^{L}\exp\left(-jk\left(x\left(\sin\theta_{T0}-\sin\theta_{R0}\right)-\Phi(x)\right)\right)dx \\
&\overset{(b)}{\approx} \frac{\left(\cos\theta_{T0}+\cos\theta_{R0}\right)\exp\left(-jk\left(d_{T0}+d_{R0}\right)\right)}{8\pi\sqrt{d_{T0}d_{R0}}}\int_{-L}^{L}\exp\left(-jk\left(x\left(\sin\theta_{T0}-\sin\theta_{R0}\right)-\Phi(x)\right)\right)dx
\end{aligned} \tag{3.30}$$

In case of an RIS acting as anomalous reflector, $\Phi(x) = \left(\bar{\phi}_T-\bar{\phi}_R\right)x+\phi_0/k$ where $\bar{\phi}_T = \sin\theta_{T0}$ and $\bar{\phi}_R = \sin\theta_{T0}$. Therefore, at electrically-small regime the electric field can be approximated according to (3.30) as

$$\begin{aligned}
E_x\left(x_R, y_R\right) \approx\ & \frac{\left(\cos\theta_{T0}+\cos\theta_{R0}\right)\exp\left(-jk\left(d_{T0}+d_{R0}\right)\right)}{8\pi\sqrt{d_{T0}d_{R0}}} \\
& \int_{-L}^{L}\exp\left(-jk\left(x\left(\sin\theta_{T0}-\sin\theta_{R0}\right)-\left(\bar{\phi}_T-\bar{\phi}_R\right)x-\phi_0/k\right)\right)dx \\
=\ & \frac{\left(\cos\theta_{T0}+\cos\theta_{R0}\right)\exp\left(-jk\left(d_{T0}+d_{R0}-\phi_0/k\right)\right)}{8\pi\sqrt{d_{T0}d_{R0}}}
\end{aligned}$$

$$\begin{aligned}&\int_{-L}^{L} \exp\left(-jk\left(x\left(\sin\theta_{T0} - \sin\theta_{R0} + \bar{\phi}_R - \bar{\phi}_T\right)\right)\right) dx \\ &\overset{(a)}{=} \frac{L\left(\sqrt{1-\bar{\phi}_T^2} + \sqrt{1-\bar{\phi}_R^2}\right) \exp\left(-jk\left(d_{T0} + d_{R0} - \phi_0/k\right)\right)}{4\pi\sqrt{d_{T0}d_{R0}}}\end{aligned} \tag{3.31}$$

where (a) is due to the facts that $\sin\theta_{T0} - \sin\theta_{R0} + \bar{\phi}_R - \bar{\phi}_T = 0$ and $\cos\theta_{T0} + \cos\theta_{R0} = \sqrt{1-\bar{\phi}_T^2} + \sqrt{1-\bar{\phi}_R^2}$. The intensity of the electric field is then given as

$$\left|E_x(x_R, y_R)\right| \approx \frac{L}{4\pi} \frac{\sqrt{1-\bar{\phi}_T^2} + \sqrt{1-\bar{\phi}_R^2}}{\sqrt{d_{T0}d_{R0}}}. \tag{3.32}$$

4 On the Path-Loss of Reconfigurable Intelligent Surfaces: Physical Optics Analytical Formulation and Scaling Laws

In this chapter, we introduce a physics-based analytical characterization of the free-space path-loss of a wireless link in the presence of a reconfigurable intelligent surface. The proposed approach is based on the vector generalization of Green's theorem. The obtained path-loss model can be applied to two-dimensional homogenized metasurfaces, which are made of sub-wavelength scattering elements and that operate either in reflection or transmission mode. The path-loss is formulated in terms of a computable integral that depends on the transmission distances, the polarization of the radio waves, the size of the surface, and the desired surface transformation. Closed-form expressions are obtained in two asymptotic regimes that are representative of far-field and near-field deployments. Based on the proposed approach, the impact of several design parameters and operating regimes is unveiled.

4.1 Introduction 76

4.2 System Model 76

  4.2.1 Source Modeling 78

  4.2.2 Metasurface Modeling 79

4.3 Preliminaries 81

  4.3.1 Received Field at Rx 81

  4.3.2 Approximations and Asymptotic Regimes 85

  4.3.3 Application to Communication Systems 88

4.4 Electric Field In the Presence of a Reflecting Surface 89

  4.4.1 $\mathscr{S}$ is Configured for Specular Reflection 90

  4.4.2 $\mathscr{S}$ is Configured for Anomalous Reflection 91

  4.4.3 $\mathscr{S}$ is Configured for Focusing 93

  4.4.4 Comparison with other Path-Loss Models 95

4.5 Electric Field In the Presence of a Transmitting Surface 95

  4.5.1 $\mathscr{S}$ is Configured for Specular Transmission 96

  4.5.2 $\mathscr{S}$ is Configured for Anomalous Transmission 97

  4.5.3 $\mathscr{S}$ is Configured for Focusing 98

4.6 Numerical Results 100

  4.6.1 Anomalous Reflection and Focusing 100

4.6.2 Transmission Distance . . . 102

4.6.3 Surface Size . . . 104

4.6.4 Operating Frequency . . . 105

4.7 Conclusion . . . 106

4.8 Appendices . . . 106

4.8.1 Proof of Theorem 4.1 . . . 106

4.8.2 Proof of Theorem 4.2 . . . 107

4.8.3 Proof of Lemma 4.4 . . . 108

4.8.4 Proof of Lemma 4.5 . . . 108

4.8.5 Proof of Proposition 4.1 . . . 109

4.8.6 Proof of Corollary 4.1 . . . 110

4.8.7 Proof of Proposition 4.2 . . . 110

4.1 Introduction

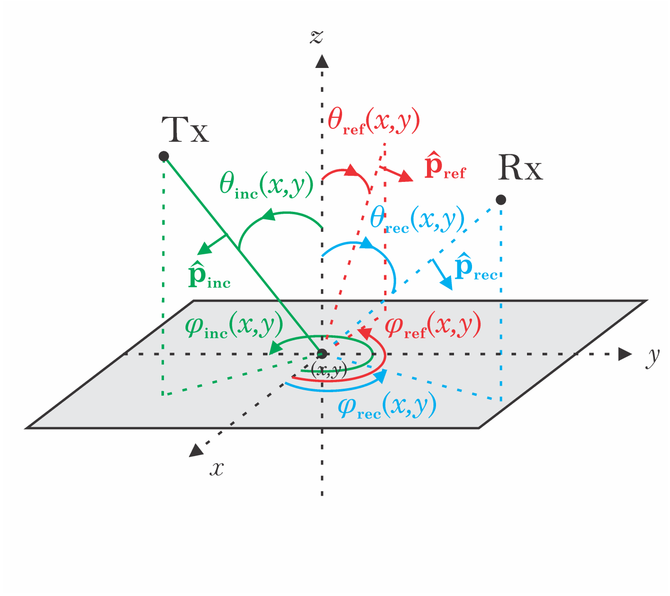
Motivated by recent experiments on the realization of unobtrusive transparent glasses that implement anomalous reflections and transmissions [53], we aim to characterize the free-space path-loss of a planar metamaterial-based RIS whose scattering elements have sizes and inter-distances much smaller than the wavelength. As we have seen in Chapter 1, under these conditions, the RIS is homogenizable and can be modeled as a continuous surface through appropriate functions, e.g., susceptibilities, impedances. To obtain an accurate but tractable path-loss models in order to quantify the performance of RISs in wireless networks, we propose an approach for calculating the free-space path-loss of an RIS-aided transmission link. The proposed approach adheres to the principles of physical optics, which overcome the limitations of geometric optics [54, Sec. 8.2.1], and leverages the vector generalization of Green's theorem [39]. Unlike the one in Chapter 3, the path-loss model proposed in this chapter leverages the vector generalization of Green's theorem [39], and it is formulated in terms of a computable integral that depends on the transmission distances, the polarization of the radio waves, the size of the RIS, and the desired surface transformations. Closed-form expressions are obtained in two asymptotic regimes that are representative of far-field and near-field transmission. Based on the proposed model, the impact of several design parameters is unveiled, and the differences and similarities between the far-field and near-field asymptotic regimes are discussed. Numerical results are illustrated and discussed in order to validate the accuracy and applicability of the asymptotic analytical formulations of the path-loss. Our study shows that the path-loss highly depends on the size of the RIS and the transmission distances, especially in the near-field regime.

4.2 System Model

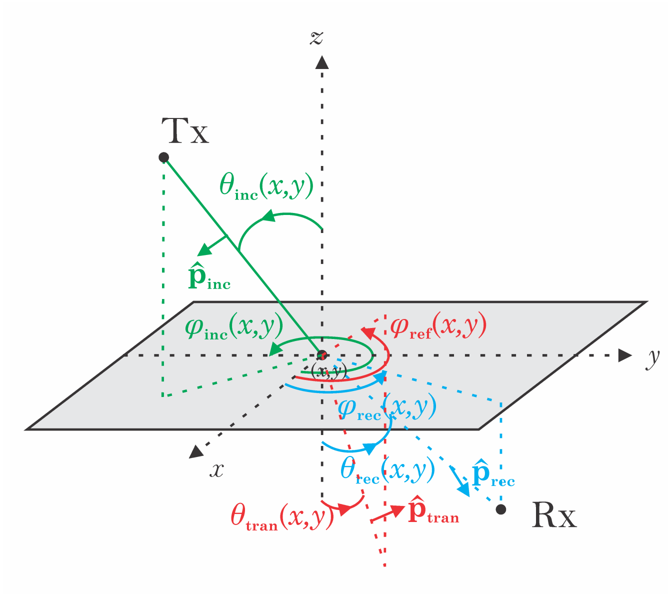
In a three-dimensional (3D) space, we consider a system that consists of a transmitter (Tx), a receiver (Rx), and a flat surface ($\mathscr{S}$) of zero-thickness. The surface $\mathscr{S}$ is a rectangle that lies on the xy-plane (i.e., $z = 0$) whose center is located at the origin. The sides of $\mathscr{S}$ are parallel to the x-axis and y-axis and have length $2L_x$ and $2L_y$, respectively. $\mathscr{S}$ is defined as follows:

$$\mathscr{S} = \left\{ \mathbf{s} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} : |x| \le L_x, |y| \le L_y \right\} \tag{4.1}$$

As shown in Fig. 4.1, Tx and Rx are located at $\mathbf{r}_{\mathrm{Tx}} = x_{\mathrm{Tx}}\hat{\mathbf{x}} + y_{\mathrm{Tx}}\hat{\mathbf{y}} + z_{\mathrm{Tx}}\hat{\mathbf{z}}$ and $\mathbf{r}_{\mathrm{Rx}} = x_{\mathrm{Rx}}\hat{\mathbf{x}} + y_{\mathrm{Rx}}\hat{\mathbf{y}} + z_{\mathrm{Rx}}\hat{\mathbf{z}}$, respectively. Without loss of generality, we assume $z_{\mathrm{Tx}} > 0$. As for z_{Rx}, we consider two cases: (i) $z_{\mathrm{Rx}} > 0$, i.e., Tx and Rx are located on the same side of $\mathscr{S}$; and (ii) $z_{\mathrm{Rx}} < 0$, i.e., Tx and Rx are located on opposite sides of $\mathscr{S}$. In the first case, the radio wave scattered by $\mathscr{S}$ towards Rx is referred to as the *reflected wave*, and, thus, $\mathscr{S}$ operates as a *reflecting surface*. In the second case, the radio wave scattered by $\mathscr{S}$ towards Rx is referred to as the *transmitted wave* and, thus, $\mathscr{S}$ operates as a *transmitting surface*. Tx emits electromagnetic (EM) waves through the vacuum whose permittivity and permeability are ϵ_0 and μ_0, respectively. The EM waves emitted by Tx travel at the speed of light $c = 1/\sqrt{\epsilon_0\mu_0}$. The carrier frequency, the wavelength,



(a) Tx and Rx are on the same side of the surface



(b) Tx and Rx are on opposite sides of the surface.

Figure 4.1 – System model.

and the wavenumber are denoted by f, $\lambda = c/f$, and $k = 2\pi/\lambda$, respectively.

For any point $\mathbf{s} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \in \mathscr{S}$, the Tx-to-$\mathscr{S}$ and $\mathscr{S}$-to-Rx distances are denoted by $d_{\mathrm{Tx}}(x,y) = \sqrt{(x - x_{\mathrm{Tx}})^2 + (y - y_{\mathrm{Tx}})^2 + z_{\mathrm{Tx}}{}^2}$ and $d_{\mathrm{Rx}}(x,y) = \sqrt{(x_{\mathrm{Rx}} - x)^2 + (y_{\mathrm{Rx}} - y)^2 + z_{\mathrm{Rx}}{}^2}$, respectively. More precisely, $d_{\mathrm{Tx}}(x,y)$ is the radius of the wavefront of the EM wave that is emitted by Tx and intersects $\mathscr{S}$ at $\mathbf{s}$, and $d_{\mathrm{Rx}}(x,y)$ is the radius of the wavefront of the EM wave that originates

Table 4.1 – Main operators ($G(x,y,z)$ is a scalar function, $\mathbf{F} = F_x\hat{\mathbf{x}}+F_y\hat{\mathbf{y}}+F_z\hat{\mathbf{z}}$ is a vector field with $\mathbf{F} = \mathbf{F}(x,y,z)$ and $F_a = F_a(x,y,z)$ for $a = x,y,z$). Symbols in bold denote vectors. Unit-norm vectors are denoted by $\hat{(\cdot)}$.

Operator	Definition
$\delta(\cdot,\cdot)$, $\mathrm{Hess}(\cdot)$, $\mod(\cdot)$	Dirac delta function, Hessian matrix, modulo operator
$\lvert C\rvert$, $\angle C$	Modulus and argument of complex number C
$\cdot$, $\times$	Scalar product and vector product
$\nabla^2 G(x,y,z)$	Laplacian of $G(x,y,z)$
$\nabla G(x,y,z)$	Gradient of $G(x,y,z)$
$\nabla\times\mathbf{F}$	Curl of $\mathbf{F}$
$\nabla\cdot\mathbf{F} = \frac{\partial F_x}{\partial x}+\frac{\partial F_y}{\partial y}+\frac{\partial F_z}{\partial z}$	Divergence of $\mathbf{F}$
$\vec{\nabla}^2\mathbf{F} = \nabla^2 F_x\hat{\mathbf{x}}+\nabla^2 F_y\hat{\mathbf{y}}+\nabla^2 F_z\hat{\mathbf{z}}$	Vector Laplacian of $\mathbf{F}$
$\nabla^2_{\mathbf{r}}G(x,y,z)$, $\nabla_{\mathbf{r}}G(x,y,z)$	Laplacian and gradient of $G(x,y,z)$ evaluated at $\mathbf{r}$
$\vec{\nabla}^2_{\mathbf{r}}\mathbf{F}$, $\nabla_{\mathbf{r}}\cdot\mathbf{F}$	Vector Laplacian and divergence of $\mathbf{F}$ evaluated at $\mathbf{r}$
$G(\mathbf{r}_1,\mathbf{r}_2) = \frac{\exp(-jk\lvert\mathbf{r}_1-\mathbf{r}_2\rvert)}{4\pi\lvert\mathbf{r}_1-\mathbf{r}_2\rvert}$	Green's function solution of (4.8)

from $\mathcal{S}$ at $\mathbf{s}$ and is observed at Rx. We define $d_{\mathrm{Tx}0} = d_{\mathrm{Tx}}(0,0)$ and $d_{\mathrm{Rx}0} = d_{\mathrm{Rx}}(0,0)$, i.e., $d_{\mathrm{Tx}0}$ and $d_{\mathrm{Rx}0}$ are the distances of Tx and Rx with respect to the center of $\mathcal{S}$, respectively. The polar angle of the incident wave at $\mathbf{s}$ is denoted by $\theta_{\mathrm{inc}}(x,y) = \cos^{-1}\left(z_{\mathrm{Tx}}/d_{\mathrm{Tx}}(x,y)\right)$. It represents the smallest angle formed by the z-axis and the wavefront of the EM wave that originates from Tx and intersects $\mathcal{S}$ at $\mathbf{s}$. The polar angle of the received wave at $\mathbf{r}_{\mathrm{Rx}}$ is denoted by $\theta_{\mathrm{rec}}(x,y) = \cos^{-1}\left(\lvert z_{\mathrm{Rx}}\rvert/d_{\mathrm{Rx}}(x,y)\right)$. It represents the smallest angle formed by the z-axis and the wavefront of the EM wave that is emitted by $\mathcal{S}$ at $\mathbf{s}$ and is observed at Rx. The azimuth angle of incidence and reflection at $\mathbf{s}$ are denoted by $\varphi_{\mathrm{inc}}(x,y)$ and $\varphi_{\mathrm{rec}}(x,y)$, respectively. In particular, $\varphi_{\mathrm{inc}}(x,y)$ represents the angle formed by the x-axis and the projection of the EM wavefront emitted from Tx towards $\mathcal{S}$ onto the xy-plane, and $\varphi_{\mathrm{rec}}(x,y)$ represents the angle formed by the x-axis and the projection of the EM wavefront emitted from $\mathcal{S}$ towards Rx onto the xy-plane:

$$\sin\varphi_{\mathrm{inc}}(x,y) = \frac{y_{\mathrm{Tx}}-y}{\sqrt{(x_{\mathrm{Tx}}-x)^2+(y_{\mathrm{Tx}}-y)^2}}, \qquad \cos\varphi_{\mathrm{inc}}(x,y) = \frac{x_{\mathrm{Tx}}-x}{\sqrt{(x_{\mathrm{Tx}}-x)^2+(y_{\mathrm{Tx}}-y)^2}},$$
$$\sin\varphi_{\mathrm{rec}}(x,y) = \frac{y_{\mathrm{Rx}}-y}{\sqrt{(x_{\mathrm{Rx}}-x)^2+(y_{\mathrm{Rx}}-y)^2}}, \qquad \cos\varphi_{\mathrm{rec}}(x,y) = \frac{x_{\mathrm{Rx}}-x}{\sqrt{(x_{\mathrm{Rx}}-x)^2+(y_{\mathrm{Rx}}-y)^2}}.$$

The polar and azimuth angles of the incident and received waves with respect to the center of $\mathcal{S}$ are denoted by $\theta_{Q0} = \theta_Q(0,0)$ and $\varphi_{Q0} = \varphi_Q(0,0)$, where $Q = \mathrm{inc}$ for the incident wave and $Q = \mathrm{rec}$ for the reflected/transmitted wave, respectively. Further notation is given in Table 4.1.

4.2.1 Source Modeling

Tx is characterized by the charge density $\rho(\mathbf{r},\mathbf{r}_{\mathrm{Tx}})$ and the current density $\mathbf{J}(\mathbf{r},\mathbf{r}_{\mathrm{Tx}})$, where $\mathbf{r}_{\mathrm{Tx}}$ is the center location of Tx and $\mathbf{r}$ is a generic location in the 3D space. We assume that $\rho(\mathbf{r},\mathbf{r}_{\mathrm{Tx}})$ and $\mathbf{J}(\mathbf{r},\mathbf{r}_{\mathrm{Tx}})$ are non-zero within a volume V_{Tx} that contains $\mathbf{r}_{\mathrm{Tx}}$ and are zero elsewhere. In particular, $\rho(\mathbf{r},\mathbf{r}_{\mathrm{Tx}})$ and $\mathbf{J}(\mathbf{r},\mathbf{r}_{\mathrm{Tx}})$ are not independent and fulfill the charge density continuity equation [55, Sec. IV], i.e., $\nabla_{\mathbf{r}}\cdot\mathbf{J}(\mathbf{r},\mathbf{r}_{\mathrm{Tx}}) + j\omega\rho(\mathbf{r},\mathbf{r}_{\mathrm{Tx}}) = 0$, where $\omega = 2\pi f$ and the universal time-dependency $e^{j\omega t}$ is assumed. Our proposed analytical framework can be applied to general EM sources, but, to obtain concrete results, we model Tx as a dipole antenna (e.g., short dipole). In this case, $\rho(\mathbf{r},\mathbf{r}_{\text{Tx}})$ and $\mathbf{J}(\mathbf{r},\mathbf{r}_{\text{Tx}})$ are [48, Eq. (15.5.2)]:

$$\rho(\mathbf{r},\mathbf{r}_{\text{Tx}}) = -\mathbf{p}\cdot\nabla_{\mathbf{r}}\delta(\mathbf{r},\mathbf{r}_{\text{Tx}}), \quad \mathbf{J}(\mathbf{r},\mathbf{r}_{\text{Tx}}) = j\omega\mathbf{p}\delta(\mathbf{r},\mathbf{r}_{\text{Tx}}) \tag{4.2}$$

where $\mathbf{p} = p_{\text{dm}}\hat{\mathbf{p}}_{\text{inc}}$ is the electric dipole moment, $p_{\text{dm}} = |\mathbf{p}|$ is the modulus of the dipole moment, and $\hat{\mathbf{p}}_{\text{inc}} = \tilde{\mathbf{p}}_{\text{inc}}e^{j\phi_{\text{inc}}}$ is the (complex) transmit polarization vector with $\tilde{\mathbf{p}}_{\text{inc}}$ being a real unit-norm vector and $\phi_{\text{inc}} \in [0,2\pi)$ being the phase of each component of $\hat{\mathbf{p}}_{\text{inc}}$. Similar results can be obtained for other source models, e.g., small linear wire antennas [48, Sec. 15.4].

4.2.2 Metasurface Modeling

We assume that the surface $\mathscr{S}$ is a metamaterial-based RIS, which is electrically-large and is made of sub-wavelength reconfigurable scattering elements whose inter-distances are much smaller than the wavelength. As detailed in [11, Sec. III-E], $\mathscr{S}$ is, therefore, homogenizable, i.e., it can be modeled through appropriate continuous surface-averaged functions (e.g., susceptibilities), even though the RIS is made of discrete elements. Based on [56, Sec. 2.1], more specifically, two conditions need to be fulfilled to make an RIS homogenizable: (i) the first homogenization condition requires that the incident field radiated by a point source varies little over one spatial period (i.e., the largest inter-distance among the scattering elements) of the RIS; and (ii) the second homogenization condition requires that the evanescent field scattered by the RIS is negligible at the observation point. Both conditions hold in this paper. The second condition, in particular, is fulfilled in the far-field of the RIS microstructure [11, Fig. 29], as elaborated next. The impact of homogenization on the radiation pattern of an RIS can be found in [57].

More specifically, the RIS is regarded as an EM discontinuity, i.e., the total tangential components of the EM fields at the two sides ($z = 0^+$ and $z = 0^-$) of $\mathscr{S}$ are discontinuous, and their difference is dictated by constituent equations that are referred to as generalized sheet transition conditions [11, Fig. 17]. For a homogenizable metamaterial-based RIS, the relation between the reflected (transmitted) tangential components of the EM fields can be formulated in terms of inhomogeneous functions as stated in [11, Eq. (50)]. Each Cartesian component of the reflected (transmitted) EM field may be formulated as a weighted linear combination of all the Cartesian components of the incident EM field. By virtue of linearity, we consider, without loss of generality, one term of the linear combination, whose corresponding inhomogeneous function is referred to as (field) local reflection or transmission coefficient if $\mathscr{S}$ operates as a reflecting surface or as a transmitting surface, respectively. It is worth noting that the generalized sheet transition conditions formulate the scattered field only on $\mathscr{S}$ [11], [58]. They do not yield the scattered field in any point of the volume of interest, which is the main objective of this paper.

In particular, the reflection (transmission) coefficient is denoted by $\widetilde{\Gamma}_{\text{ref}}(\mathbf{s})$ ($\widetilde{\Gamma}_{\text{tran}}(\mathbf{s})$), which is a complex function that is appropriately engineered (through the design of surface-averaged susceptibilities) in order to apply specified transformations to the impinging EM waves. Specific examples are provided later. As elaborated in [11, Fig. 14] and detailed alter, the surface equivalent theorem dictates that the EM field scattered by $\mathscr{S}$ at any point in a 3D space can be formulated in terms of only the incident fields, $\widetilde{\Gamma}_{\text{ref}}(\mathbf{s})$, and $\widetilde{\Gamma}_{\text{tran}}(\mathbf{s})$ at $\mathbf{s} \in \mathscr{S}$.

For generality, the RIS is assumed to be capable of modifying the polarization of the impinging radio waves. More precisely, given an incident signal with polarization $\hat{\mathbf{p}}_{\text{inc}}$, the polarizations of the reflected and transmitted signals are denoted by $\hat{\mathbf{p}}_{\text{ref}} = \tilde{\mathbf{p}}_{\text{ref}} e^{j\phi_{\text{ref}}}$ and $\hat{\mathbf{p}}_{\text{tran}} = \tilde{\mathbf{p}}_{\text{tran}} e^{j\phi_{\text{tran}}}$, respectively. Similar to the definition of $\hat{\mathbf{p}}_{\text{inc}}$, $\tilde{\mathbf{p}}_{\text{ref}}$ and $\tilde{\mathbf{p}}_{\text{tran}}$ are real unit-norm vectors and $\phi_{\text{ref}} \in [0, 2\pi)$ and $\phi_{\text{tran}} \in [0, 2\pi)$ are the phases of each component of $\hat{\mathbf{p}}_{\text{ref}}$ and $\hat{\mathbf{p}}_{\text{tran}}$, respectively.

Based on these modeling assumptions, the electric field at any point $\mathbf{s} \in \mathscr{S}$ on the reflection side of the RIS (i.e., $z = 0^+$) can be formulated as follows:

$$\begin{aligned} \mathbf{E}_{\mathscr{S}}(\mathbf{s}) &= \mathbf{E}_{\mathscr{S}}(\mathbf{s}, z = 0^+) \\ &= \mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{inc}}) + \widetilde{\Gamma}_{\text{ref}}(\mathbf{s}) \mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{ref}}) \end{aligned} \tag{4.3}$$

where $\mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{inc}})$ is the incident field at $\mathbf{s}$ with polarization $\hat{\mathbf{p}}_{\text{inc}}$ and $\widetilde{\Gamma}_{\text{ref}}(\mathbf{s}) = \Gamma_{\text{ref}}(\mathbf{s}) \mathscr{E}_{\text{ref}}(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{ref}})$ is the reflection coefficient. To make explicit the impact of the change of polarization introduced by $\mathscr{S}$, $\widetilde{\Gamma}_{\text{ref}}(\mathbf{s})$ is formulated as the product of two terms: (i) $\Gamma_{\text{ref}}(\mathbf{s})$ that is polarization-independent; and (ii) $\mathscr{E}_{\text{ref}}(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{ref}})$ that denotes the efficiency of the change of polarization from $\hat{\mathbf{p}}_{\text{inc}}$ to $\hat{\mathbf{p}}_{\text{ref}}$. In addition, $\mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{ref}})$ denotes the reflected electric field whose polarization is $\hat{\mathbf{p}}_{\text{ref}}$. We emphasize that the incident field $\mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{ref}})$ is formally equal to the reflected field $\mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{inc}})$ (in terms of amplitude and phase) except for the change of polarization.

Remark 4.1 *The polarization model in* (4.3) *accounts only for the impact of the desired radio wave while the analysis of the change of polarization on unwanted radio waves (e.g., multipath or interference) is postponed to future research. Also, the proposed approach can be used to analyze more general polarization models. The formulation in* (4.3) *is considered only as an example.*

Along the same lines and with a similar meaning of the symbols, the electric field at any point $\mathbf{s} \in \mathscr{S}$ on the transmission side of the RIS (i.e., $z = 0^-$) can be formulated as follows:

$$\mathbf{E}_{\mathscr{S}}(\mathbf{s}) = \mathbf{E}_{\mathscr{S}}(\mathbf{s}, z = 0^-) = \widetilde{\Gamma}_{\text{tran}}(\mathbf{s}) \mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{tran}}) = \Gamma_{\text{tran}}(\mathbf{s}) \mathscr{E}_{\text{tran}}(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{tran}}) \mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{tran}}) \tag{4.4}$$

where we have taken into account that at $z = 0^-$ there is no incident field [11, Eq. (6)].

We emphasize, as detailed in [11, Fig. 29], that (4.3) and (4.4) are applicable in the far-field of the RIS microstructure, i.e., at distances from $\mathscr{S}$ at which the presence of possible evanescent fields that are excited to realize RISs with high reflection and transmission efficiency can be safely ignored. In the next sections, we assume $k \gg 1/d_{\text{Tx}}(x, y)$ and $k \gg 1/d_{\text{Rx}}(x, y)$ that are typically fulfilled for wireless applications and allow us to ignore the presence of possible evanescent fields. The far-field of the RIS microstructure encompasses the near-field and the far-field of the RIS. These two regimes are analyzed later in detail.

Remark 4.2 *The period (i.e., the largest inter-distance among the scattering elements) of the metasurface determines the scattering properties of an RIS and has an important impact on the angular response of periodic RISs. Based on Floquet's theory, in fact, periodical variations of the metasurface properties lead to the existence of multiple diffracted modes, which depend on the angle of illumination, the period of the metasurface, and the operating frequency. In this paper, we do not consider higher-order modes because our main objective is the characterization of the scaling laws of the electromagnetic field scattered by an RIS as a function of its size, the transmission distances, and the wave transformations. Preliminary results on the analysis and impact of higher diffracted modes can be found in [59].*

4.3 Preliminaries

In this section, we introduce a general formulation of the received EM field at Rx in the presence of $\mathscr{S}$. The proposed approach adheres to the principles of physical optics and overcomes the limitations of geometric optics. Also, we introduce methods for computing recurrent integrals that characterize the EM field scattered by reflecting and transmitting surfaces.

4.3.1 Received Field at Rx

Assuming the universal time-dependency $e^{j\omega t}$, the electric field, $\mathbf{E}(\mathbf{r})$, and magnetic field, $\mathbf{H}(\mathbf{r})$, at any location $\mathbf{r} \in \mathbb{R}^3$ in vacuum satisfy the differential equations [55, Eqs. (6), (7)]:

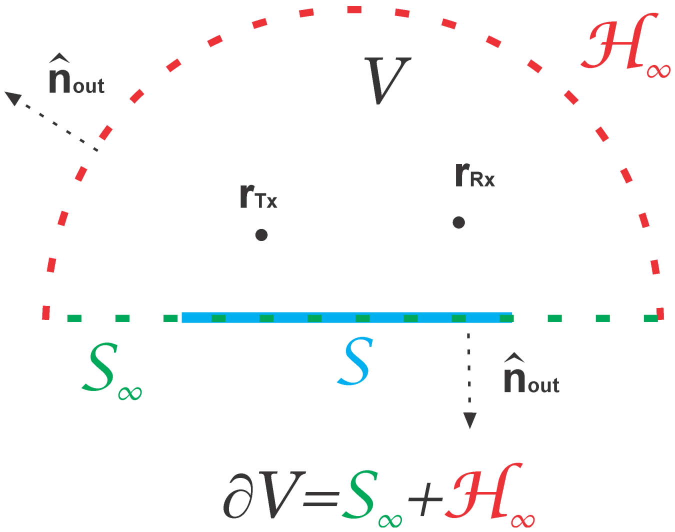
$$\nabla_{\mathbf{r}} \times (\nabla_{\mathbf{r}} \times \mathbf{E}(\mathbf{r})) = k^2 \mathbf{E}(\mathbf{r}) - j\omega\mu_0 \mathbf{J}(\mathbf{r}, \mathbf{r}_{\mathrm{Tx}}) \tag{4.5}$$

$$\nabla_{\mathbf{r}} \times (\nabla_{\mathbf{r}} \times \mathbf{H}(\mathbf{r})) = k^2 \mathbf{H}(\mathbf{r}) + \nabla_{\mathbf{r}} \times \mathbf{J}(\mathbf{r}, \mathbf{r}_{\mathrm{Tx}}) \tag{4.6}$$

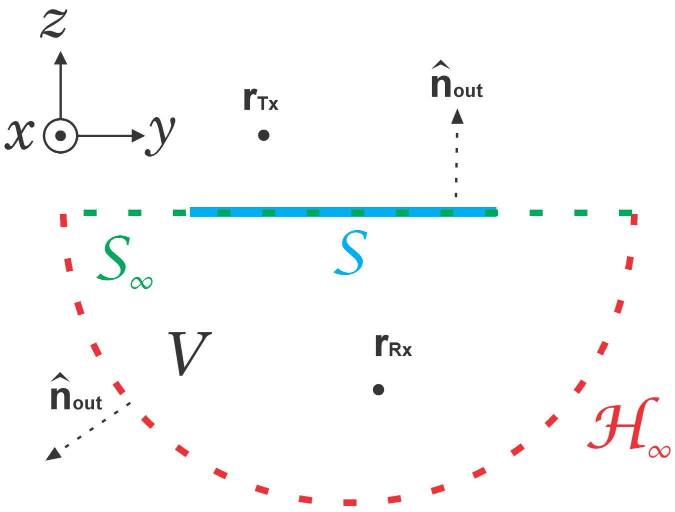
The solutions of (4.5) and (4.6) are related through the relation $\mathbf{H}(\mathbf{r}) = -\nabla_{\mathbf{r}} \times \mathbf{E}(\mathbf{r})/(j\omega\mu_0)$. Therefore, the complete characterization of the EM field can be given only through $\mathbf{E}(\mathbf{r})$.

In the absence of the RIS, the solution of (4.5), i.e., $\mathbf{E}(\mathbf{r})$, observed at $\mathbf{r}_{\mathrm{Rx}}$ boils down, by definition, to the incident electric field with polarization $\hat{\mathbf{p}}_{\mathrm{inc}}$. This latter electric field is denoted by $\mathbf{E}_{\mathrm{inc}}(\mathbf{r}_{\mathrm{Rx}}; \hat{\mathbf{p}}_{\mathrm{inc}})$. Using the notation in Sec. 4.2.1, it can be formulated as [48, Eq. (15.3.10)]:

$$\mathbf{E}_{\mathrm{inc}}(\mathbf{r}_{\mathrm{Rx}}; \hat{\mathbf{p}}_{\mathrm{inc}}) = \int_{V_{\mathrm{Tx}}} \left(-j\omega\mu_0 \mathbf{J}(\mathbf{r}, \mathbf{r}_{\mathrm{Tx}}) G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) + \frac{\rho(\mathbf{r}, \mathbf{r}_{\mathrm{Tx}})}{\epsilon_0} \nabla_{\mathbf{r}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \right) d\mathbf{r} \tag{4.7}$$



(a) reflecting surface.



(b) transmitting surface.

Figure 4.2 – Volume V and closed boundary ∂V.

where $G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r})$ is the Green function defined as follows [48, Eq. (18.10.2)]:

$$\nabla^2_{\mathbf{r}_{\mathrm{Rx}}} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) + k^2 G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) = -\delta(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}) \tag{4.8}$$

In the presence of $\mathscr{S}$, $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$, at any point $\mathbf{r}_{\mathrm{Rx}}$ in a volume $V \subseteq \mathbb{R}^3$, does not have a simple formulation as in (4.7). Under the assumptions of physical optics [54, Sec. 8.2.1], the field $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ solution of (4.5) in the presence of $\mathscr{S}$ can be characterized by using the Stratton-Chu formula [55].

Lemma 4.1 *Let* $\mathbf{r}_{\mathrm{Rx}}$ *be the observation point of interest in a generic volume* $V \subseteq \mathbb{R}^3$. *Let* ∂V *be a* *generic closed boundary of* V *such that (see Fig. 4.2): (i) Rx is always located inside the volume, i.e.,* $\mathbf{r}_{\mathrm{Rx}} \in V$*; (ii)* $\mathscr{S}$ *is part of the boundary, i.e.,* $\mathscr{S} \in \partial V$*; and (iii) Tx is located inside the volume, i.e.,* $\mathbf{r}_{\mathrm{Tx}} \in V$*, in the reflection case and outside the volume, i.e.,* $\mathbf{r}_{\mathrm{Tx}} \notin V$ *in the transmission case, respectively. Let* $\mathbf{r}'$ *be a generic point on the closed boundary* ∂V*, i.e.,* $\mathbf{r}' \in \partial V$*, and let* $\mathbf{E}_{\partial V}(\mathbf{r}')$ *and* $\mathbf{H}_{\partial V}(\mathbf{r}')$ *denote the total electric and magnetic fields at* $\mathbf{r}'$*, respectively. Then,* $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ *solution of* (4.5) *in the presence of* $\mathscr{S}$ *can be formulated as follows:*

$$\begin{aligned}\mathbf{E}(\mathbf{r}_{\mathrm{Rx}}) = \mathbb{1}_{(\mathbf{r}_{\mathrm{Tx}} \in V)} \mathbf{E}_{\mathrm{inc}}(\mathbf{r}_{\mathrm{Rx}}; \hat{\mathbf{p}}_{\mathrm{inc}}) - \int_{\partial V} \Big[& -j\omega\mu_0 \left(\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}_{\partial V}(\mathbf{r}')\right) G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \\ & + \left(\hat{\mathbf{n}}_{\mathrm{out}} \cdot \mathbf{E}_{\partial V}(\mathbf{r}')\right) \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') + (\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{E}_{\partial V}(\mathbf{r}')) \times \nabla_{\mathbf{r}'} G(\mathbf{r}_{\mathrm{Rx}}, \mathbf{r}') \Big] d\mathbf{r}' \end{aligned} \tag{4.9}$$

where $\hat{\mathbf{n}}_{\mathrm{out}}$ *is the normal vector pointing outwards the volume and* $\mathbb{1}_{(\cdot)}$ *is the indicator function.*

Proof 4.1 *See [55, Eq. (14)].*

Remark 4.3 *In [55, Sec. V], a contour integral that accounts for the impact of the boundary of* $\mathscr{S}$ *is added into* (4.9)*. In this chapter, we omit it since it can be proved to be negligible compared to the surface integral in* (4.9) *for* $k \gg 1$.

Remark 4.4 *The choice of* V *and* ∂V *is not unique. For convenience, Fig. 4.2 shows an example in which* V *is the upper or lower half-plane of the 3D space and* $\partial V = \mathscr{H}_\infty + \mathscr{S}_\infty$*, where* $\mathscr{H}_\infty$ *is the hemisphere for* $z > 0$ *or* $z < 0$ *with an infinite radius for a reflecting or transmitting surface, respectively, and* $\mathscr{S}_\infty$ *is the entire* xy*-plane (including* $\mathscr{S}$*).*

Remark 4.5 *There exist alternative integral expressions for the solution of* (4.5) *in the presence of* $\mathscr{S}$*, e.g., Franz's formula [60, Eq. (3)]. We choose* (4.9) *as the basis of our analysis for two reasons: (i) the incident field* $\mathbf{E}_{\mathrm{inc}}(\mathbf{r}_{\mathrm{Rx}}; \hat{\mathbf{p}}_{\mathrm{inc}})$ *explicitly appears in* (4.7)*, which leads to simple interpretations as elaborated later; and (ii) the two terms* $\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{E}_{\partial V}(\mathbf{r}')$ *and* $\hat{\mathbf{n}}_{\mathrm{out}} \times \mathbf{H}_{\partial V}(\mathbf{r}')$ *are directly related to the magnetic and electric currents, respectively, that are induced by the incident signal in the scattering elements (i.e., the inclusions) of the metasurface [11, Eq. (1)], which provides us with explicit evidence of the physics-based phenomena that govern the operation of RISs. In particular,* (4.9) *can be viewed as an instance of the surface equivalent theorem [11, Fig. 14].*

Even though (4.9) provides us with a computable integral for $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$, it does not offer an explicit analytical expression that depends on $\mathscr{S}$ and that yields insights. In the sequel, we analyze $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ in detail and compute equivalent explicit expressions for (4.9) that are useful in wireless applications. To this end, we assume that Rx is equipped with an antenna whose polarization is $\hat{\mathbf{p}}_{\mathrm{rec}} = \tilde{\mathbf{p}}_{\mathrm{rec}} e^{j\phi_{\mathrm{rec}}}$ [61], where $\tilde{\mathbf{p}}_{\mathrm{rec}}$ is a real unit-norm vector and $\phi_{\mathrm{rec}} \in [0, 2\pi)$ is the common phase of the three components of $\hat{\mathbf{p}}_{\mathrm{rec}}$. In general, $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ depends on $\hat{\mathbf{p}}_{\mathrm{inc}}$, $\hat{\mathbf{p}}_{\mathrm{ref}}$ or $\hat{\mathbf{p}}_{\mathrm{tran}}$, and $\hat{\mathbf{p}}_{\mathrm{rec}}$. To explicitly highlight the impact of $\hat{\mathbf{p}}_{\mathrm{rec}}$, we reformulate (4.9) as follows.

Theorem 4.1 *The projection of* $\mathbf{E}(\mathbf{r}_{\text{Rx}})$ *in* (4.9) *onto the receive polarization vector* $\hat{\mathbf{p}}_{\text{rec}}$ *is:*

$$\begin{aligned}\mathbf{E}(\mathbf{r}_{\text{Rx}}) \cdot \hat{\mathbf{p}}_{\text{rec}} &= \mathbb{1}_{(\mathbf{r}_{\text{Tx}} \in V)} \mathbf{E}_{\text{inc}}(\mathbf{r}_{\text{Rx}}; \hat{\mathbf{p}}_{\text{inc}}) \cdot \hat{\mathbf{p}}_{\text{rec}} \\ &\quad - \int_{\partial V} \left[(\mathbf{E}_{\partial V}(\mathbf{r}') \cdot \hat{\mathbf{p}}_{\text{rec}}) \nabla_{\mathbf{r}'} G(\mathbf{r}_{\text{Rx}}, \mathbf{r}') - G(\mathbf{r}_{\text{Rx}}, \mathbf{r}') \nabla_{\mathbf{r}'} \left(\mathbf{E}_{\partial V}(\mathbf{r}') \cdot \hat{\mathbf{p}}_{\text{rec}} \right) \right] \cdot \hat{\mathbf{n}}_{\text{out}} d\mathbf{r}'\end{aligned} \tag{4.10}$$

Proof 4.2 *See Appendix 4.8.1*

By appropriately choosing $\hat{\mathbf{p}}_{\text{rec}}$, $\mathbf{E}(\mathbf{r}_{\text{Rx}})$ along any directions can be retrieved, e.g., $\hat{\mathbf{p}}_{\text{rec}} = \hat{\mathbf{x}}$, $\hat{\mathbf{p}}_{\text{rec}} = \hat{\mathbf{y}}$, and $\hat{\mathbf{p}}_{\text{rec}} = \hat{\mathbf{z}}$. The Stratton-Chu formula in (4.10), however, does not explicitly reveal the impact of $\mathscr{S}$. Thus, we reformulate (4.10) such that $\mathscr{S}$, instead of V and ∂V, appears explicitly.

Theorem 4.2 *Let* $\mathbf{E}_{\mathscr{S}}(\mathbf{s})$ *be the surface electric field at point* $\mathbf{s} \in \mathscr{S}$ *in* (4.3) *and* (4.4) *for a reflecting and transmitting surface, respectively. Then,* (4.10) *can be equivalently reformulated as follows:*

$$\begin{aligned}\mathbf{E}(\mathbf{r}_{\text{Rx}}) \cdot \hat{\mathbf{p}}_{\text{rec}} = \mathbf{E}_{\text{inc}}(\mathbf{r}_{\text{Rx}}; \hat{\mathbf{p}}_{\text{inc}}) \cdot \hat{\mathbf{p}}_{\text{rec}} - \int_{\mathscr{S}} \big[&((\mathbf{E}_{\mathscr{S}}(\mathbf{s}) - \mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{inc}})) \cdot \hat{\mathbf{p}}_{\text{rec}}) \nabla_{\mathbf{s}} G(\mathbf{r}_{\text{Rx}}, \mathbf{s}) \\ &- G(\mathbf{r}_{\text{Rx}}, \mathbf{s}) \nabla_{\mathbf{s}} ((\mathbf{E}_{\mathscr{S}}(\mathbf{s}) - \mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{inc}})) \cdot \hat{\mathbf{p}}_{\text{rec}}) \big] \cdot \hat{\mathbf{n}}_{\text{out}} d\mathbf{s}\end{aligned} \tag{4.11}$$

Proof 4.3 *See Appendix 4.8.2*

Remark 4.6 *By direct inspection of the integrand in* (4.11), *we note that the electric field at an observation point is determined by both the EM field and its first-order derivative on* $\mathscr{S}$, *which are appropriately weighted by the first order-derivative of the Green function and by the Green function itself that operate as integral kernels. This formulation is similar to its scalar counterpart, i.e., the Fresnel-Kirchhoff integral theorem and the Huygens-Fresnel equation [1].*

The reformulation in (4.11) can be applied to any physical source at Tx, i.e., $\rho(\mathbf{r}, \mathbf{r}_{\text{Tx}})$ and $\mathbf{J}(\mathbf{r}, \mathbf{r}_{\text{Tx}})$, which determine the incident fields $\mathbf{E}_{\text{inc}}(\mathbf{s}; \hat{\mathbf{p}}_{\text{inc}})$ and $\mathbf{E}_{\text{inc}}(\mathbf{r}_{\text{Rx}}; \hat{\mathbf{p}}_{\text{inc}})$, and to any field transformations applied by the RIS, i.e., $\mathbf{E}_{\mathscr{S}}(\mathbf{s})$. In the following, as a concrete example, we focus our attention on a physical source that corresponds to a dipole antenna [48, Sec. (15.5)].

Lemma 4.2 *Let* $\hat{\mathbf{r}}_{\text{Rx-Tx}}$ *be the unit-norm propagation vector from* $\mathbf{r}_{\text{Tx}}$ *to* $\mathbf{r}_{\text{Rx}}$. *The incident electric field at* $\mathbf{r}_{\text{Rx}}$ *generated by a dipole antenna is* $\mathbf{E}_{\text{inc}}(\mathbf{r}_{\text{Rx}}; \hat{\mathbf{p}}_{\text{inc}}) \approx \mathbf{E}_{0,\text{inc}}\left(\mathbf{r}_{\text{Rx}}; \hat{\mathbf{p}}_{\text{inc}}\right) G\left(\mathbf{r}_{\text{Rx}}, \mathbf{r}_{\text{Tx}}\right)$ *where* $\mathbf{E}_{0,\text{inc}}\left(\mathbf{r}_{\text{Rx}}; \hat{\mathbf{p}}_{\text{inc}}\right) = \frac{k^2 p_{\text{dm}}}{\epsilon_0} \left(\hat{\mathbf{p}}_{\text{inc}} - (\hat{\mathbf{r}}_{\text{Rx-Tx}} \cdot \hat{\mathbf{p}}_{\text{inc}}) \hat{\mathbf{r}}_{\text{Rx-Tx}}\right) = \frac{k^2 p_{\text{dm}}}{\epsilon_0} \left(\tilde{\mathbf{p}}_{\text{inc}} - (\hat{\mathbf{r}}_{\text{Rx-Tx}} \cdot \tilde{\mathbf{p}}_{\text{inc}}) \hat{\mathbf{r}}_{\text{Rx-Tx}}\right) e^{j\phi_{\text{inc}}}$.

Proof 4.4 *The electric field radiated by a dipole antenna is [48, Eq. (15.5.5)]. The approximation follows from* $k \gg 1/|\mathbf{r}_{\text{Rx}} - \mathbf{r}_{\text{Tx}}|$. *The proof follows by simplifying the triple vector product.*

Remark 4.7 *The first addend in* (4.11), *i.e., the incident field at* $\mathbf{r}_{\mathrm{Rx}}$*, and the integral that yields the contribution from the RIS sum up, in general, incoherently and, thus, interfere with each other. The phase terms* $\angle\Gamma_{\mathrm{ref}}(x,y)$ *and* $\angle\Gamma_{\mathrm{tran}}(x,y)$ *of* $\mathscr{S}$ *can, however, be optimized in order to make sure that both contributions (incident field and scattered field) add up coherently at* $\mathbf{r}_{\mathrm{Rx}}$.

4.3.2 Approximations and Asymptotic Regimes

In Secs. 4.4 and 4.5, we capitalize on (4.11) in order to derive explicit expressions for the electric field reflected and transmitted by an RIS, and to unveil scaling laws as a function of relevant design parameters. To this end, some recurrent integrals need to be computed and some asymptotic approximations are exploited. In this section, we introduce methods to compute these integrals and we formally introduce the asymptotic operating regimes of interest.

Type-1 Integral

Consider the following type of integral:

$$I_1 = \int_{-L_y}^{L_y}\int_{-L_x}^{L_x} \mathscr{A}_1(d_{\mathrm{Tx}}(x,y), d_{\mathrm{Rx}}(x,y))\mathscr{B}_1(x,y)e^{-jk(d_{\mathrm{Tx}}(x,y)+d_{\mathrm{Rx}}(x,y)-\mathscr{C}(x,y))}dxdy \tag{4.12}$$

where $\mathscr{A}_1(d_{\mathrm{Tx}}(x,y), d_{\mathrm{Rx}}(x,y))$, $\mathscr{B}_1(x,y)$, and $\mathscr{C}(x,y)$ are real-valued functions. In particular, we assume that: (i) $\mathscr{A}_1(d_{\mathrm{Tx}}(x,y), d_{\mathrm{Rx}}(x,y))$ depends on $d_{\mathrm{Tx}}(x,y)$ and $d_{\mathrm{Rx}}(x,y)$; (ii) $\mathscr{B}_1(x,y)$ is independent of $d_{\mathrm{Tx}}(x,y)$ and $d_{\mathrm{Rx}}(x,y)$; and (iii) $\mathscr{C}(x,y)$ is a linear function in x and y.

Definition 4.1 *Define* $r_{ES} = 8(L_x^2 + L_y^2)/\lambda$*. Assume that* (4.11) *is formulated in terms of type-1 integrals. An RIS is said to operate in the electrically-small regime if* $d_{\mathrm{Tx0}} > r_{ES}$ *and* $d_{\mathrm{Rx0}} > r_{ES}$.

The electrically-small regime in Definition 4.1 is analogous to the Fraunhofer far-field [62, Sec. 4.4.1]. This can be shown from the Taylor expansion of, e.g., $d_{\mathrm{Tx}}(x,y)$ around the origin:

$$d_{\mathrm{Tx}}(x,y) = d_{\mathrm{Tx0}} - x\sin\theta_{\mathrm{inc0}}\cos\varphi_{\mathrm{inc0}} - y\sin\theta_{\mathrm{inc0}}\sin\varphi_{\mathrm{inc0}} + R_2(x,y) \tag{4.13}$$

where $R_2(x,y)$ collects the terms of higher order than the first degree. In general, the Fraunhofer distance is calculated for linear structures, e.g., by assuming $L_y \ll L_x$, and by then replacing the length of the structure ($L = L_x$) with the largest dimension of $\mathscr{S}$ [62, Eq. (4.41)]. Based on (4.1), the largest dimension of $\mathscr{S}$ is its diagonal $D = 2\sqrt{L_x^2 + L_y^2}$. For linear structures, (4.13) reduces to $d_{\mathrm{Tx}}(x) = d_{\mathrm{Tx0}} - x\sin\theta_{\mathrm{inc0}} + R_2(x)$. By definition, the Fraunhofer far-field is the distance r_F at which the identity $\max\{R_2(x)\} = \pi/8$ holds true, which gives $r_F = 2D^2/\lambda$. Thus, we obtain $r_F = r_{ES}$. Notably, r_{ES} can be formulated in terms of the ratio between the surface area and the wavelength, i.e., $r_{ES} = 2\frac{A_{\mathscr{S}}}{\lambda}\frac{a_x^2+a_y^2}{a_x a_y}$, where $A_{\mathscr{S}}$ is the area of $\mathscr{S}$ and $L_x = a_x L$, $L_y = a_y L$.

Lemma 4.3 *In the electrically-small regime, the integral in* (4.12) *can be approximated as:*

$$I_1 \approx \mathscr{A}_1(d_{\mathrm{Tx0}}, d_{\mathrm{Rx0}}) e^{-jk(d_{\mathrm{Tx0}}+d_{\mathrm{Rx0}})} \int_{-L_y}^{L_y} \int_{-L_x}^{L_x} \mathscr{B}_1(x,y) e^{-jk(\mathscr{D}_x x + \mathscr{D}_y y - \mathscr{C}(x,y))} dx dy \tag{4.14}$$

where $\mathscr{D}_x = \sin\theta_{\mathrm{inc0}} \cos\varphi_{\mathrm{inc0}} + \sin\theta_{\mathrm{rec0}} \cos\varphi_{\mathrm{rec0}}$ *and* $\mathscr{D}_y = \sin\theta_{\mathrm{inc0}} \sin\varphi_{\mathrm{inc0}} + \sin\theta_{\mathrm{rec0}} \sin\varphi_{\mathrm{rec0}}$.

Proof 4.5 *It follows from* (4.13) *by ignoring* $R_2(x,y)$ *and noting that* $\mathscr{A}_1(\cdot,\cdot) \to \mathscr{A}_1(d_{\mathrm{Tx0}}, d_{\mathrm{Rx0}})$.

Definition 4.2 *Define* $\mathscr{P}(x,y) = d_{\mathrm{Tx}}(x,y) + d_{\mathrm{Rx}}(x,y) - \mathscr{C}(x,y)$. *The stationary points of* $\mathscr{P}(x,y)$ *are the points* (x_s, y_s) *such that* $\frac{\partial}{\partial x}\mathscr{P}(x,y)|_{(x,y)=(x_s,y_s)} = \frac{\partial}{\partial y}\mathscr{P}(x,y)|_{(x,y)=(x_s,y_s)} = 0$.

Definition 4.3 *Define* $\mathscr{P}(x,y) = d_{\mathrm{Tx}}(x,y) + d_{\mathrm{Rx}}(x,y) - \mathscr{C}(x,y)$ *and let* Ψ *be the set of its stationary points. Let* $D = 2\sqrt{L_x^2 + L_y^2}$ *be the diagonal of* $\mathscr{S}$. *Assume that* (4.11) *is formulated in terms of type-1 integrals. An RIS is said to operate in the electrically-large regime if*

$$(2D^2/\lambda)(z_{\mathrm{Tx}}^2/[d_{\mathrm{Tx}}(x_s,y_s)]^3 + z_{\mathrm{Rx}}^2/[d_{\mathrm{Rx}}(x_s,y_s)]^3) \gg 1$$

for all stationary points $(x_s, y_s) \in \Psi$.

Similar to the Fraunhofer far-field [62, Sec. 4.4.1], Definition 4.3 can be justified by starting from a line surface, e.g., by assuming $L_y \ll L_x$ and by then replacing the length of the line ($L = L_x$) with the diagonal D of $\mathscr{S}$. Consider the line integral $I_\ell = \int_{-L}^{L} \mathscr{M}(\ell) e^{-jk\mathscr{P}(\ell)} d\ell$ corresponding to (4.12), where $\mathscr{M}(\ell)$ is a slowly-varying function in $[-L, L]$, $\mathscr{P}(\ell) = d_{\mathrm{Tx}}(\ell) + d_{\mathrm{Rx}}(\ell) - \mathscr{C}(\ell)$, $d_{\mathrm{Tx}}(\ell) = \sqrt{\ell^2 + z_{\mathrm{Tx}}^2}$, $d_{\mathrm{Rx}}(\ell) = \sqrt{\ell^2 + z_{\mathrm{Rx}}^2}$, and $\mathscr{C}(\ell)$ is a linear function in ℓ. Let $\ell_s \in [-L, L]$ be a stationary point (assuming that it exists) of $\mathscr{P}(\ell)$, i.e., $\frac{\partial}{\partial x}\mathscr{P}(\ell)|_{\ell=\ell_s} = 0$. Definition 4.3 can be justified by invoking the stationary phase method to compute I_ℓ [63]. In particular, the integrand of I_ℓ oscillates very quickly outside a small region centered at ℓ_s, and, thus, the contributions outside the small region around ℓ_s cancel out when computing the integral. Under these conditions, I_ℓ can be well approximated by replacing $\mathscr{P}(\ell)$ with its Taylor approximation at $\ell_s \in [-L, L]$, i.e., $\mathscr{P}(\ell) \approx \mathscr{P}(\ell_s) + \frac{1}{2}\left({z_{\mathrm{Tx}}}^2/[d_{\mathrm{Tx}}(\ell_s)]^3 + {z_{\mathrm{Rx}}}^2/[d_{\mathrm{Rx}}(\ell_s)]^3\right)(\ell - \ell_s)^2$, and by letting the extremes of integration go to infinity, provided that the region around ℓ_s that dominates I_ℓ is well contained in $[-L, L]$. This is usually true when the minimum of the integrand of I_ℓ falls within $[-L, L]$, which occurs if the condition in Definition 4.3 is fulfilled. Notably, the latter condition can be formulated in terms of ratio between the area of the surface and the wavelength, i.e., $d_s \ll r_{EL}$ with $r_{EL} = \frac{2D^2}{\lambda}\sqrt{\frac{z_{\mathrm{Tx}}^2}{b_{\mathrm{Tx}}^3} + \frac{z_{\mathrm{Rx}}^2}{b_{\mathrm{Rx}}^3}} = \frac{2A_{\mathscr{S}}}{\lambda}\frac{a_x^2 + a_y^2}{a_x a_y}\sqrt{\frac{z_{\mathrm{Tx}}^2}{b_{\mathrm{Tx}}^3} + \frac{z_{\mathrm{Rx}}^2}{b_{\mathrm{Rx}}^3}}$, $d_{\mathrm{Tx}}\left(x_s, y_s\right) = b_{\mathrm{Tx}} d_s$, $d_{\mathrm{Rx}}\left(x_s, y_s\right) = b_{\mathrm{Rx}} d_s$.

Remark 4.8 *Based on Definition 4.1 and Definition 4.3, the following comments can be made: (i) the terminology electrically-small RIS originates from the conditions* $d_{\mathrm{Tx0}} > r_{ES}$ *and* $d_{\mathrm{Rx0}} > r_{ES}$, *i.e., the transmission distances (computed with respect to the center of* $\mathscr{S}$*) are larger than* *the electrical size of $\mathscr{S}$, which is $A_{\mathscr{S}}/\lambda$; (ii) the terminology electrically-large RIS originates from the condition $d_s \ll r_{EL}$, i.e., the transmission distances (computed with respect to the stationary point of the phase term) are smaller than the electrical size, $A_{\mathscr{S}}/\lambda$, of $\mathscr{S}$; and (iii) since, in general, $\sqrt{\frac{z_{\rm Tx}^2}{b_{\rm Tx}^3}+\frac{z_{\rm Rx}^2}{b_{\rm Rx}^3}} < 1$, then $r_{EL} = r_{ES}\sqrt{\frac{z_{\rm Tx}^2}{b_{\rm Tx}^3}+\frac{z_{\rm Rx}^2}{b_{\rm Rx}^3}} < r_{ES}$. This implies that the electrically-large regime holds for shorter distances than the radiating near-field regime [62, Sec. (4.4.2)].*

Lemma 4.4 *Define $\mathbf{A}(x_s, y_s) = \text{Hess}(\mathscr{P}(x,y))|_{(x,y)=(x_s,y_s)}$. Assume that Ψ is not empty and, for $(x_s, y_s) \in \Psi$, $\det(\mathbf{A}(x_s, y_s)) \neq 0$. In the electrically-large regime,* (4.12) *can be approximated as:*

$$I_1 \approx (2\pi/k)\sum\nolimits_{(x_s,y_s)\in\Psi}\mathscr{A}_1(d_{\rm Tx}(x_s,y_s), d_{\rm Rx}(x_s,y_s))\mathscr{B}_1(x_s,y_s)\left|\det(\mathbf{A}(x_s,y_s))\right|^{-1/2} \exp\left(-jk\mathscr{P}(x_s,y_s) - j\pi\text{sign}\left(\mathbf{A}(x_s,y_s)\right)/4\right) \tag{4.15}$$

where $\text{sign}\left(\mathbf{A}(x_s,y_s)\right) = N^+(\mathbf{A}(x_s,y_s)) - N^-(\mathbf{A}(x_s,y_s))$ *is the signature of* $\mathbf{A}(x_s,y_s)$*, with* $N^+(\mathbf{A}(x_s,y_s))$ *and* $N^-(\mathbf{A}(x_s,y_s))$ *the number of positive and negative eigenvalues of* $\mathbf{A}(x_s,y_s)$.

Proof 4.6 *See Appendix 4.8.3.*

Lemma 4.5 *Let Ψ be empty. In the electrically-large regime,* (4.12) *can be approximated as:*

$$I_1 \approx \frac{1}{(-jk)^2}\left[\frac{\mathscr{A}_1(d_{\rm Tx}(x,y), d_{\rm Rx}(x,y))\mathscr{B}_1(x,y)e^{-jk\mathscr{P}(x,y)}}{\mathscr{P}_x(x,y)\mathscr{P}_y(x,y)}\right]\Bigg|_{x=-L_x}^{x=L_x}\Bigg|_{y=-L_y}^{y=L_y} \tag{4.16}$$

where $\mathscr{P}_x(x,y) = \frac{\partial}{\partial x}\mathscr{P}(x,y)$ *and* $\mathscr{P}_y(x,y) = \frac{\partial}{\partial y}\mathscr{P}(x,y)$.

Proof 4.7 *See Appendix 4.8.4.*

By comparing Lemmas 4.4 and 4.5, we evince that, since $\det(\mathbf{A}(x_s, y_s))$ is independent of k, $|I_1|$ is inversely proportional to k if at least one stationary point is contained in $\mathscr{S}$, and is inversely proportional to k^2 if no stationary point lies in $\mathscr{S}$. For $k \gg 1$, thus, $|I_1|$ is dominated by the contributions from the stationary points. In the rest of this chapter, therefore, we focus our attention on the case studies (in the electrically-large regime) in which at least one stationary point exists.

Type-2 Integral

Consider the following type of integral:

$$I_2 = \int_{-L_y}^{L_y}\int_{-L_x}^{L_x}\mathscr{A}_2(d_{\rm Tx}(x,y), d_{\rm Rx}(x,y))\mathscr{B}_2(x,y)\,dx\,dy \tag{4.17}$$

where $\mathscr{A}_2(d_{\rm Tx}(x,y), d_{\rm Rx}(x,y))$ is a real-valued function of the distances $d_{\rm Tx}(x,y)$ and $d_{\rm Rx}(x,y)$, and $\mathscr{B}_2(x,y)$ is a real-valued function that is independent of $d_{\rm Tx}(x,y)$ and $d_{\rm Rx}(x,y)$.

Table 4.2 – Approximate conditions for which the electrically-large and electrically-small asymptotic regimes hold.

Integral Type	Electrically-Large Regime (Lemma 4.5)	Electrically-Small Regime (Lemmas 4.3, 4.6)
Type-1 integral in (4.12)	$(2D^2/\lambda)(z_{\mathrm{Tx}}^2/[d_{\mathrm{Tx}}(x_s,y_s)]^3 + z_{\mathrm{Rx}}^2/[d_{\mathrm{Rx}}(x_s,y_s)]^3) \gg 1$	$d_{\mathrm{Tx0}}, d_{\mathrm{Rx0}} > 8(L_x^2+L_y^2)/\lambda$
Type-2 integral in (4.17)	see Sections 4.4.3 and 4.5.3 for an example	$d_{\mathrm{Tx0}}, d_{\mathrm{Rx0}} \gg 2\sqrt{L_x^2+L_y^2}$

Lemma 4.6 *Assume $d_{\mathrm{Tx0}} \gg D$ and $d_{\mathrm{Rx0}} \gg D$, where $D = 2\sqrt{L_x^2+L_y^2}$ is the diagonal of $\mathscr{S}$. The integral in* (4.17) *can be approximated as follows:*

$$I_2 \approx \mathscr{A}_2(d_{\mathrm{Tx0}}, d_{\mathrm{Rx0}}) \int_{-L_y}^{L_y} \int_{-L_x}^{L_x} \mathscr{B}_2(x,y)\,dx\,dy \tag{4.18}$$

Proof 4.8 *It follows from* (4.13) *noting that $d_{\mathrm{Tx}}(x,y) \approx d_{\mathrm{Tx0}}$, $d_{\mathrm{Rx}}(x,y) \approx d_{\mathrm{Rx0}}$ if $d_{\mathrm{Tx0}}, d_{\mathrm{Rx0}} \gg D$.*

If $d_{\mathrm{Tx0}} \ll D$ and $d_{\mathrm{Rx0}} \ll D$, it is not straightforward to compute (4.17) in general. This case study is analyzed in Secs. 4.4.3 and 4.5.3 for the specific $\mathscr{A}_2(x,y)$ and $\mathscr{B}_2(x,y)$ of interest.

Remark 4.9 *The asymptotic regime in Lemma 4.6 is independent of λ and is, in general, different from the asymptotic regime in Definition 4.1 that depends on λ. We still refer to it as electrically-small regime, however, since $d_{\mathrm{Tx0}} \gg D$ and $d_{\mathrm{Rx0}} \gg D$ implies $D/\lambda \ll d_{\mathrm{Tx0}}/\lambda$ and $D/\lambda \ll d_{\mathrm{Rx0}}/\lambda$. Likewise, the regime $d_{\mathrm{Tx0}} \ll D$ and $d_{\mathrm{Rx0}} \ll D$ is referred to as electrically-large regime.*

Remark 4.10 *The closed-form approximations obtained for type-1 and type-2 integrals are used in the next two sections to quantify the scaling laws of the EM field scattered by reflecting and transmitting RISs in the electrically-large and electrically-small (asymptotic) regimes, respectively. More precisely, the conditions (system setups) for which the obtained closed-form expressions can be considered to be accurate are summarized in Table 4.2.*

4.3.3 Application to Communication Systems

Before analyzing reflecting and transmitting RISs, we briefly discuss the application of the proposed approach in communication systems. The electric field in Theorem 4.2 can be used in free-space, either when the direct link is available or when it is not by retaining only the integral term in the right-hand side of (4.11). $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ in (4.11) can be used to compute the Poynting vector at any locations $\mathbf{r}_{\mathrm{Rx}}$ of interest and to evaluate the received power density [48, Eq. (1.9.6)]. From the Poynting vector, the link budget of RIS-assisted communications can be assessed quantitatively.

4.4 Electric Field In the Presence of a Reflecting Surface

In this section, we analyze $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ under the assumption that $\mathscr{S}$ is a reflecting surface according to the definitions and assumptions given in Sec. 4.2 (see Fig. 4.1a and Fig. 4.2a).

Proposition 4.1 *Let* $\hat{\mathbf{s}}_{(x,y)} = \sin\theta_{\mathrm{inc}}(x,y)\cos\varphi_{\mathrm{inc}}(x,y)\hat{\mathbf{x}} + \sin\theta_{\mathrm{inc}}(x,y)\sin\varphi_{\mathrm{inc}}(x,y)\hat{\mathbf{y}} + \cos\theta_{\mathrm{inc}}(x,y)\hat{\mathbf{z}}$, *be the unit-norm propagation vector from* $\mathbf{r}_{\mathrm{Tx}}$ *to* $\mathbf{s} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \in \mathscr{S}$. *Define* $\Omega_{\mathrm{ref}}(x,y;\hat{\mathbf{p}}_{\mathrm{ref}},\hat{\mathbf{p}}_{\mathrm{rec}}) = (k^2/\epsilon_0)p_{\mathrm{dm}}\left(\tilde{\mathbf{p}}_{\mathrm{rec}}\cdot\tilde{\mathbf{p}}_{\mathrm{ref}} - \left(\hat{\mathbf{s}}_{(x,y)}\cdot\tilde{\mathbf{p}}_{\mathrm{rec}}\right)\left(\hat{\mathbf{s}}_{(x,y)}\cdot\tilde{\mathbf{p}}_{\mathrm{ref}}\right)\right)\mathscr{E}\left(\hat{\mathbf{p}}_{\mathrm{inc}},\hat{\mathbf{p}}_{\mathrm{ref}}\right)$. *Under the assumptions stated in Lemma 4.2 (i.e., dipole transmit antenna), the electric field* $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ *projected onto* $\hat{\mathbf{p}}_{\mathrm{rec}}$ *can be formulated as follows:*

$$\begin{aligned}
&\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})\cdot\hat{\mathbf{p}}_{\mathrm{rec}} \approx \hat{\mathbf{p}}_{\mathrm{rec}}\cdot\mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{r}_{\mathrm{Rx}};\hat{\mathbf{p}}_{\mathrm{inc}}\right)G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}_{\mathrm{Tx}}) \\
&+ jke^{j(\phi_{\mathrm{ref}}+\phi_{\mathrm{rec}})}\int_{\mathscr{S}}\Gamma_{\mathrm{ref}}(\mathbf{s})\Omega_{\mathrm{ref}}(x,y;\hat{\mathbf{p}}_{\mathrm{ref}},\hat{\mathbf{p}}_{\mathrm{rec}})G(\mathbf{s},\mathbf{r}_{\mathrm{Tx}})\,G(\mathbf{r}_{\mathrm{Rx}},\mathbf{s})\left[\frac{z_{\mathrm{Rx}}}{|\mathbf{s}-\mathbf{r}_{\mathrm{Rx}}|}+\frac{z_{\mathrm{Tx}}}{|\mathbf{s}-\mathbf{r}_{\mathrm{Tx}}|}\right]d\mathbf{s} \\
&= \hat{\mathbf{p}}_{\mathrm{rec}}\cdot\mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{r}_{\mathrm{Rx}};\hat{\mathbf{p}}_{\mathrm{inc}}\right)G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}_{\mathrm{Tx}}) + \mathscr{I}_0\int_{-L_y}^{L_y}\int_{-L_x}^{L_x}\mathscr{I}_R(x,y)e^{-jk\mathscr{P}_R(x,y)}\,dx\,dy
\end{aligned} \tag{4.19}$$

where $\mathscr{I}_0 = jk/(16\pi^2)$, *and the following shorthand notation is used:*

$$\mathscr{P}_R(x,y) = d_{\mathrm{Tx}}(x,y) + d_{\mathrm{Rx}}(x,y) - (\phi_{\mathrm{rec}} + \phi_{\mathrm{ref}} + \angle\Gamma_{\mathrm{ref}}(x,y))/k \tag{4.20}$$

$$\mathscr{I}_R(x,y) = \frac{\left|\Gamma_{\mathrm{ref}}(x,y)\right|\Omega_{\mathrm{ref}}(x,y;\hat{\mathbf{p}}_{\mathrm{ref}},\hat{\mathbf{p}}_{\mathrm{rec}})}{d_{\mathrm{Tx}}(x,y)d_{\mathrm{Rx}}(x,y)}\left(\cos\theta_{\mathrm{inc}}(x,y)+\cos\theta_{\mathrm{rec}}(x,y)\right) \tag{4.21}$$

Proof 4.9 *See Appendix 4.8.5.*

Remark 4.11 *The approximation in* (4.19) *originates only from the assumptions* $k \gg 1/d_{\mathrm{Tx}}(x,y)$, $k \gg 1/d_{\mathrm{Rx}}(x,y)$ *(see Sec. 4.2). This is apparent from the proof in Appendix 4.8.5. The proof in Appendix 4.8.5 can, however, be readily generalized in order to avoid these assumptions.*

Remark 4.12 *In* (4.21), *we note that the scattered electric field depends on the sum of the cosines (known as obliquity factors) of the (local) angles of incidence and reflection. This implies that the scattered power computed from the Poynting vector would be proportional to* $\left(\cos\theta_{\mathrm{inc}}(x,y)+\cos\theta_{\mathrm{rec}}(x,y)\right)^2$. *The sum of cosines originates from the gradient of the Green functions associated with the incident and reflected waves, as shown in* (4.61) *and* (4.62).

The electric field in (4.19) is formulated as the sum of the incident electric field in the absence of $\mathscr{S}$ and the contribution due to the reflection from $\mathscr{S}$. This latter term is denoted by $F_R(\mathbf{r}_{\mathrm{Rx}}) = \mathscr{I}_0\int_{-L_y}^{L_y}\int_{-L_x}^{L_x}\mathscr{I}_R(x,y)e^{-jk\mathscr{P}_R(x,y)}\,dx\,dy$ and is analyzed next to better understand the performance of RISs as a function of important design parameters and configurations for $\mathscr{S}$, e.g., $\angle\Gamma_{\mathrm{ref}}(x,y)$. As illustrations, we consider case studies that correspond to using phase gradient metasurfaces, which are known to be approximated implementations of perfect anomalous reflectors [11]. This choice is motivated by analytical convenience and to reveal the impact of important design parameters. Proposition 4.1 has, in fact, general applicability.

4.4.1 $\mathscr{S}$ is Configured for Specular Reflection

This setup is obtained if $\angle\Gamma_{\text{ref}}(x,y) = \phi_0$ for $(x,y) \in \mathscr{S}$, where $\phi_0 \in [0,2\pi)$ is a fixed phase.

Corollary 4.1 *Let* $(x_s, y_s) \in \mathscr{S}$ *be the solution of the following system of equations:*

$$\frac{(x_s - x_{\text{Tx}})}{d_{\text{Tx}}(x_s,y_s)} + \frac{(x_s - x_{\text{Rx}})}{d_{\text{Rx}}(x_s,y_s)} = 0, \qquad \frac{(y_s - x_{\text{Tx}})}{d_{\text{Tx}}(x_s,y_s)} + \frac{(y_s - x_{\text{Rx}})}{d_{\text{Rx}}(x_s,y_s)} = 0 \tag{4.22}$$

In the electrically-large regime, $F_R(\mathbf{r}_{\text{Rx}})$ *can be approximated as follows:*

$$F_R(\mathbf{r}_{\text{Rx}}) \approx \frac{\left|\Gamma_{\text{ref}}(x_s,y_s)\right| \Omega_{\text{ref}}(x_s,y_s;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}})}{4\pi(d_{\text{Tx}}(x_s,y_s) + d_{\text{Rx}}(x_s,y_s))} e^{-jk\left(d_{\text{Tx}}(x_s,y_s)+d_{\text{Rx}}(x_s,y_s)-(\phi_0+\phi_{\text{ref}}+\phi_{\text{rec}})/k\right)} \tag{4.23}$$

Proof 4.10 *See Appendix 4.8.6 .*

Remark 4.13 *Assume that Tx and Rx move along directions such that* (x_s, y_s)*, and* $\theta_{\text{inc}}(x_s,y_s)$, $\theta_{\text{rec}}(x_s,y_s)$, $\varphi_{\text{inc}}(x_s,y_s)$, $\varphi_{\text{rec}}(x_s,y_s)$ *are kept fixed. From Corollary 4.1, we evince the following.*

- *Since* $\Gamma_{\text{ref}}(x_s,y_s)$ *depends only on* (x_s,y_s) *and* $\Omega_{\text{ref}}(x_s,y_s;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}})$ *depends only on* $\tilde{\mathbf{p}}_{\text{inc}}$, $\tilde{\mathbf{p}}_{\text{rec}}$, $\theta_{\text{inc}}(x_s,y_s)$, $\theta_{\text{rec}}(x_s,y_s)$, $\varphi_{\text{inc}}(x_s,y_s)$*, and* $\varphi_{\text{rec}}(x_s,y_s)$*, they are both independent of the Tx-to-*(x_s,y_s) *and* (x_s,y_s)*-to-Rx distances. In the electrically-large regime, therefore,* $|F_R(\mathbf{r}_{\text{Rx}})|$ *decays as a function of the sum of the Tx-to-*(x_s,y_s) *and* (x_s,y_s)*-to-Rx distances.*
- *In the electrically-large regime,* $|F_R(\mathbf{r}_{\text{Rx}})|$ *is independent of the size of* $\mathscr{S}$*. This implies that the received power is bounded, even though the size of* $\mathscr{S}$ *grows large (tending to infinity).*
- *The system of equations in* (4.22) *is equivalent to* $\varphi_{\text{inc}}(x_s,y_s) = (\varphi_{\text{rec}}(x_s,y_s) + \pi) \mod 2\pi$ *and* $\theta_{\text{inc}}(x_s,y_s) = \theta_{\text{rec}}(x_s,y_s)$*. These conditions correspond to the law of reflection.*

Corollary 4.2 *In the electrically-small regime,* $F_R(\mathbf{r}_{\text{Rx}})$ *can be approximated as follows:*

$$\begin{aligned} F_R(\mathbf{r}_{\text{Rx}}) \approx{}& \frac{jk\Omega_{\text{ref}}(0,0;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}})\left(\cos\theta_{\text{inc0}} + \cos\theta_{\text{rec0}}\right)}{16\pi^2\left(d_{\text{Tx0}}d_{\text{Rx0}}\right)} e^{-jk\left(d_{\text{Tx0}}+d_{\text{Rx0}}-(\phi_0+\phi_{\text{ref}}+\phi_{\text{rec}})/k\right)} \\ & \int_{-L_y}^{L_y}\int_{-L_x}^{L_x} |\Gamma_{\text{ref}}(x,y)| e^{jk(\mathscr{D}_x x+\mathscr{D}_y y)} dx dy \end{aligned} \tag{4.24}$$

where $\mathscr{D}_x$ *and* $\mathscr{D}_y$ *are defined in Lemma 4.3. Let* $\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$ *be the sinc function. If* $|\Gamma_{\text{ref}}(x,y)| = \Gamma_{\text{ref}} > 0$ *for* $(x,y) \in \mathscr{S}$*, then* $F_R(\mathbf{r}_{\text{Rx}})$ *can be further simplified as follows:*

$$\begin{aligned} F_R(\mathbf{r}_{\text{Rx}}) \approx{}& \frac{jk\Gamma_{\text{ref}}\Omega_{\text{ref}}(0,0;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}})L_x L_y\left(\cos\theta_{\text{inc0}} + \cos\theta_{\text{rec0}}\right)}{4\pi^2 d_{\text{Tx0}}d_{\text{Rx0}}} \operatorname{sinc}(kL_x\mathscr{D}_x)\operatorname{sinc}\left(kL_y\mathscr{D}_y\right) \\ & e^{-jk\left(d_{\text{Tx0}}+d_{\text{Rx0}}-(\phi_0+\phi_{\text{ref}}+\phi_{\text{rec}})/k\right)} \end{aligned} \tag{4.25}$$

Proof 4.11 *If follows directly from* (4.14).

Remark 4.14 *Assume that Tx and Rx move along directions such that* $\theta_{\text{inc}0}, \theta_{\text{rec}0}, \varphi_{\text{inc}0}, \varphi_{\text{rec}0}$ *are kept fixed. From Corollary 4.2, the following conclusions can be drawn.*

- *In the electrically-small regime,* $|F_R(\mathbf{r}_{\text{Rx}})|$ *decays as a function of the product of the Tx-to-*$(0,0)$ *and* $(0,0)$*-to-Rx distances, where* $(0,0)$ *is the center of* $\mathscr{S}$*.*

- *In the electrically-small regime,* $|F_R(\mathbf{r}_{\text{Rx}})|$ *grows linearly with the area of* $\mathscr{S}$*, i.e.,* $A_{\mathscr{S}} = 4L_xL_y$*. This does not imply that* $|F_R(\mathbf{r}_{\text{Rx}})|$ *grows unbounded if the size of* $\mathscr{S}$ *tends to infinity. If* $A_{\mathscr{S}} \to \infty$*, in fact, the RIS does not operate in the electrically-small regime anymore, but in the electrically-large regime. Therefore, the approximation in Corollary 4.2 needs to be replaced with the approximation in Corollary 4.1, which does not depend on the size of* $\mathscr{S}$*.*

- *In the electrically-small regime,* $|F_R(\mathbf{r}_{\text{Rx}})|$ *attains its maximum for* $\mathscr{D}_x = \mathscr{D}_y = 0$*. If the angle of incidence* $\theta_{\text{inc}0}$ *is fixed, this is fulfilled in correspondence of the angles of observation* $\theta_{\text{inc}0} = \theta_{\text{rec}0}$ *and* $\varphi_{\text{inc}0} = (\varphi_{\text{rec}0} + \pi) \mod 2\pi$*, which can be interpreted as the law of reflection. Also, the main lobe of* $\text{sinc}(kL_x\mathscr{D}_x)$ *and* $\text{sinc}(kL_y\mathscr{D}_y)$ *gets narrower if* L_x *and* L_y *increase.*

4.4.2 $\mathscr{S}$ is Configured for Anomalous Reflection

This setup is obtained by setting $\angle\Gamma_{\text{ref}}(x,y) = k(\alpha_R x + \beta_R y) + \phi_0$ for $(x,y) \in \mathscr{S}$, where $\alpha_R \in \mathbb{R}$, $\beta_R \in \mathbb{R}$ are design parameters, and $\phi_0 \in [0, 2\pi)$ is a fixed phase. As will be detailed later, the direction of anomalous reflection is determined by the specific choice of α_R and β_R.

Corollary 4.3 *Let* $(x_s, y_s) \in \mathscr{S}$ *be the solution of the following system of equations:*

$$\frac{(x_s - x_{\text{Tx}})}{d_{\text{Tx}}(x_s,y_s)} + \frac{(x_s - x_{\text{Rx}})}{d_{\text{Rx}}(x_s,y_s)} = \alpha_R, \qquad \frac{(y_s - x_{\text{Tx}})}{d_{\text{Tx}}(x_s,y_s)} + \frac{(y_s - x_{\text{Rx}})}{d_{\text{Rx}}(x_s,y_s)} = \beta_R \tag{4.26}$$

Define the shorthand notation $\Theta_{\text{Q}} = \theta_{\text{Q}}(x_s,y_s)$ *and* $\Phi_{\text{Q}} = \varphi_{\text{Q}}(x_s,y_s)$ *for* $\text{Q} \in \{\text{inc}, \text{rec}\}$*. In the electrically-large regime,* $F_R(\mathbf{r}_{\text{Rx}})$ *can be approximated as follows:*

$$F_R(\mathbf{r}_{\text{Rx}}) \approx \frac{\left|\Gamma_{\text{ref}}(x_s,y_s)\right| \Omega_{\text{ref}}(x_s,y_s;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}}) e^{-jk(d_{\text{Tx}}(x_s,y_s)+d_{\text{Rx}}(x_s,y_s)-(\alpha_R x_s+\beta_R y_s)-(\phi_0+\phi_{\text{ref}}+\phi_{\text{rec}})/k)}}{8\pi\sqrt{\mathscr{R}_1(d_{\text{Tx}}(x_s,y_s))^2 + \mathscr{R}_2(d_{\text{Rx}}(x_s,y_s))^2 + \mathscr{R}_3 d_{\text{Tx}}(x_s,y_s) d_{\text{Rx}}(x_s,y_s)}} \tag{4.27}$$

where $\mathscr{R}_1 = \cos^2\Theta_{\text{rec}}/(\cos\Theta_{\text{inc}} + \cos\Theta_{\text{rec}})^2$*,* $\mathscr{R}_2 = \cos^2\Theta_{\text{inc}}/(\cos\Theta_{\text{inc}} + \cos\Theta_{\text{rec}})^2$*, and* $\mathscr{R}_3 = \left(\cos^2\Theta_{\text{inc}} + \cos^2\Theta_{\text{rec}} + \sin^2\Theta_{\text{inc}}\sin^2\Theta_{\text{rec}}\sin^2(\Phi_{\text{inc}} - \Phi_{\text{rec}})\right)/(\cos\Theta_{\text{inc}} + \cos\Theta_{\text{rec}})^2$.

Proof 4.12 *It follows from Lemma 4.4 along the same lines as the proof of Corollary 4.1. The only difference is that* $\mathscr{P}(x,y) = \mathscr{P}_R(x,y)$*,* $\det(\mathbf{A}(x_s,y_s))$*, and* $\text{sign}(\mathbf{A}(x_s,y_s))$ *depend on* α_R, β_R*.*

The analytical formulation in (4.27) does not provide direct design insights. To this end, we introduce an approximation for (4.27) in order to unveil scaling laws and performance trends.

Corollary 4.4 *Consider* $\zeta_1' > 0, \zeta_2' > 0$. *Define* $K_1 = (\mathscr{R}_1\zeta_1' + \frac{1}{2}\mathscr{R}_3\zeta_2')/\sqrt{\mathscr{R}_1{\zeta_1'}^2 + \mathscr{R}_2{\zeta_2'}^2 + \mathscr{R}_3\zeta_1'\zeta_2'}$, $K_2 = (\mathscr{R}_2\zeta_2' + \frac{1}{2}\mathscr{R}_3\zeta_1')/\sqrt{\mathscr{R}_1{\zeta_1'}^2 + \mathscr{R}_2{\zeta_2'}^2 + \mathscr{R}_3\zeta_1'\zeta_2'}$. *Then,* (4.27) *can be approximated as follows:*

$$F_R(\mathbf{r}_{\mathrm{Rx}}) \approx \frac{\left|\Gamma_{\mathrm{ref}}(x_s, y_s)\right| \Omega_{\mathrm{ref}}(x_s, y_s; \hat{\mathbf{p}}_{\mathrm{ref}}, \hat{\mathbf{p}}_{\mathrm{rec}})}{8\pi(K_1 d_{\mathrm{Tx}}(x_s, y_s) + K_2 d_{\mathrm{Rx}}(x_s, y_s))} e^{-jk(d_{\mathrm{Tx}}(x_s,y_s)+d_{\mathrm{Rx}}(x_s,y_s)-(\alpha_R x_s+\beta_R y_s)-(\phi_0+\phi_{\mathrm{ref}}+\phi_{\mathrm{rec}})/k)} \tag{4.28}$$

Proof 4.13 *For simplicity, let us denote* $\zeta_1 = d_{\mathrm{Tx}}(x_s, y_s)$ *and* $\zeta_2 = d_{\mathrm{Rx}}(x_s, y_s)$. *Define* $f(\zeta_1,\zeta_2) = \sqrt{\mathscr{R}_1\zeta_1^2 + \mathscr{R}_2\zeta_2^2 + \mathscr{R}_3\zeta_1\zeta_2}$. *Consider a generic pair of points* (ζ_1',ζ_2'). *The function* $f(\zeta_1,\zeta_2)$ *can be approximated at* (ζ_1',ζ_2') *by using the Taylor approximation, which yields* $f(\zeta_1,\zeta_2) \approx \sqrt{\mathscr{R}_1{\zeta_1'}^2 + \mathscr{R}_2{\zeta_2'}^2 + \mathscr{R}_3\zeta_1'\zeta_2'} + (\mathscr{R}_{1,3}(\zeta_1 - \zeta_1') + \mathscr{R}_{2,3}(\zeta_2 - \zeta_2'))/\sqrt{\mathscr{R}_1{\zeta_1'}^2 + \mathscr{R}_2{\zeta_2'}^2 + \mathscr{R}_3\zeta_1'\zeta_2'}$, *where* $\mathscr{R}_{1,3} = (\mathscr{R}_1\zeta_1' + \mathscr{R}_3\zeta_2'/2)$ *and* $\mathscr{R}_{2,3} = (\mathscr{R}_2\zeta_2' + \mathscr{R}_3\zeta_1'/2)$. *The proof follows with the aid of algebraic steps. The parameters* K_1 *and* K_2 *are independent of the pair* (ζ_1',ζ_2') *if* $\zeta_1' = \zeta_2'$.

Given Θ_{inc}, Θ_{rec}, Φ_{inc}, and Φ_{rec}, (4.27) and (4.28) coincide only if $K_1 = (\cos\Theta_{\mathrm{rec}})/(\cos\Theta_{\mathrm{inc}} + \cos\Theta_{\mathrm{rec}})$, $K_2 = (\cos\Theta_{\mathrm{inc}})/(\cos\Theta_{\mathrm{inc}} + \cos\Theta_{\mathrm{rec}})$, and $2K_1K_2 = [\cos^2\Theta_{\mathrm{inc}} + \cos^2\Theta_{\mathrm{rec}} + \sin^2\Theta_{\mathrm{inc}} \sin^2\Theta_{\mathrm{rec}} \sin^2(\Phi_{\mathrm{inc}} - \Phi_{\mathrm{rec}})]/(\cos\Theta_{\mathrm{inc}} + \cos\Theta_{\mathrm{rec}})^2$ are satisfied simultaneously. This holds true only if $\mathscr{S}$ is a uniform surface, i.e., $\alpha_R = \beta_R = 0$, which corresponds to specular reflection. As for anomalous reflection, (4.28) is an approximation for (4.27) because Taylor's approximation is used. The approximation in (4.28) depends, in general, on ζ_1' and ζ_2'. A convenient choice for these parameters is $\zeta_1' = \zeta_2'$, since (4.28) is independent of ζ_1' and ζ_2' (i.e., ζ_1' and ζ_2' cancel out in (4.28)) if $\zeta_1' = \zeta_2'$. With the aid of (4.28), the impact and scaling laws of key parameters can be unveiled.

Remark 4.15 *Assume that Tx and Rx move along directions such that* (x_s, y_s), *and* $\theta_{\mathrm{inc}}(x_s, y_s)$, $\theta_{\mathrm{rec}}(x_s, y_s)$, $\varphi_{\mathrm{inc}}(x_s, y_s)$, $\varphi_{\mathrm{rec}}(x_s, y_s)$ *are kept fixed. From* (4.28), *we evince the following.*

- *In the electrically-large regime,* $|F_R(\mathbf{r}_{\mathrm{Rx}})|$ *decays as a function of the weighted sum of the Tx-to-*(x_s, y_s) *and* (x_s, y_s)*-to-Rx distances. Also,* $|F_R(\mathbf{r}_{\mathrm{Rx}})|$ *is independent of the size of* $\mathscr{S}$.
- *From* (4.26), *we have* $\sin\theta_{\mathrm{inc}}(x_s, y_s)\cos\varphi_{\mathrm{inc}}(x_s, y_s) + \sin\theta_{\mathrm{rec}}(x_s, y_s)\cos\varphi_{\mathrm{rec}}(x_s, y_s) = -\alpha_R$ *and* $\sin\theta_{\mathrm{inc}}(x_s, y_s)\sin\varphi_{\mathrm{inc}}(x_s, y_s) + \sin\theta_{\mathrm{rec}}(x_s, y_s)\sin\varphi_{\mathrm{rec}}(x_s, y_s) = -\beta_R$. *This implies that, in general, the polar and azimuthal angles of incidence and reflection in correspondence of the stationary point* (x_s, y_s) *are different and depend on* α_R *and* β_R. *This corresponds to the generalized law of reflection. By using* (4.26), *in particular,* α_R *and* β_R *can be optimized in order to obtain the desired angle of reflection for a given angle of incidence.*
- *If* $\alpha_R = \beta_R = 0$, (4.27) *and* (4.28) *reduce, as expected, to* (4.23).

Corollary 4.5 *In the electrically-small regime, $F_R(\mathbf{r}_{\text{Rx}})$ can be approximated as follows:*

$$
\begin{aligned}
F_R(\mathbf{r}_{\text{Rx}}) \approx & \frac{jk\Omega_{\text{ref}}(0,0;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}})\left(\cos\theta_{\text{inc}0}+\cos\theta_{\text{rec}0}\right)}{16\pi^2 d_{\text{Tx}0}d_{\text{Rx}0}} e^{-jk(d_{\text{Tx}0}+d_{\text{Rx}0}-(\phi_0+\phi_{\text{ref}}+\phi_{\text{rec}})/k)} \\
& \int_{-L_y}^{L_y}\int_{-L_x}^{L_x} |\Gamma_{\text{ref}}(x,y)| e^{jk(\mathscr{D}_{\alpha_R}x+\mathscr{D}_{\beta_R}y)}\,dxdy
\end{aligned}
\tag{4.29}
$$

where the shorthand notation $\mathscr{D}_{\alpha_R}=\alpha_R+\mathscr{D}_x$ and $\mathscr{D}_{\beta_R}=\beta_R+\mathscr{D}_y$ is used. If $|\Gamma_{\text{ref}}(x,y)|=\Gamma_{\text{ref}}>0$ for $(x,y)\in\mathscr{S}$, then $F_R(\mathbf{r}_{\text{Rx}})$ can be further simplified as follows:

$$
\begin{aligned}
F_R(\mathbf{r}_{\text{Rx}}) \approx & \frac{jk\Gamma_{\text{ref}}\Omega_{\text{ref}}(0,0;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}})L_xL_y\left(\cos\theta_{\text{inc}0}+\cos\theta_{\text{rec}0}\right)}{4\pi^2 d_{\text{Tx}0}d_{\text{Rx}0}} \\
& \operatorname{sinc}\left(kL_x\mathscr{D}_{\alpha_R}\right)\operatorname{sinc}\left(kL_y\mathscr{D}_{\beta_R}\right)e^{-jk(d_{\text{Tx}0}+d_{\text{Rx}0}-(\phi_0+\phi_{\text{ref}}+\phi_{\text{rec}})/k)}
\end{aligned}
\tag{4.30}
$$

Proof 4.14 *It follows by substituting $\mathscr{C}(x,y)=k(\alpha_Rx+\beta_Ry)+\phi_0$ in* (4.14).

Remark 4.16 *From* (4.30)*, conclusions similar to Remark 4.14 can be drawn with one exception.* $|F_R(\mathbf{r}_{\text{Rx}})|$ *in* (4.30) *attains its maximum in correspondence of angles of incidence and reflection that fulfill the equalities $\alpha_R=-(\sin\theta_{\text{inc}0}\cos\theta_{\text{inc}0}+\sin\theta_{\text{rec}0}\cos\theta_{\text{rec}0})$ and $\beta_R=-(\sin\theta_{\text{inc}0}\sin\theta_{\text{inc}0}+\sin\theta_{\text{rec}0}\sin\theta_{\text{rec}0})$. Thus, α_R and β_R can be appropriately optimized for maximizing the reflected signal towards a desired direction, given the angle of incidence with respect to the center of $\mathscr{S}$.*

4.4.3 $\mathscr{S}$ is Configured for Focusing

This setup is obtained by setting $\angle\Gamma_{\text{ref}}(x,y)=k\left(d_{\text{Tx}}(x,y)+d_{\text{Rx}}(x,y)\right)+\phi_0$ for $(x,y)\in\mathscr{S}$, where $\phi_0\in[0,2\pi)$ is a fixed phase. With this setup, $F_R(\mathbf{r}_{\text{Rx}})$ in (4.19) simplifies as follows:

$$
\begin{aligned}
F_R(\mathbf{r}_{\text{Rx}}) \approx & \frac{jke^{j(\phi_0+\phi_{\text{rec}}+\phi_{\text{ref}})}}{16\pi^2}\int_{-L_y}^{L_y}\int_{-L_x}^{L_x}\Omega_{\text{ref}}(x,y;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}}) \\
& \frac{|\Gamma_{\text{ref}}(x,y)|\left(\cos\theta_{\text{inc}}(x,y)+\cos\theta_{\text{rec}}(x,y)\right)}{d_{\text{Tx}}(x,y)d_{\text{Rx}}(x,y)}\,dxdy
\end{aligned}
\tag{4.31}
$$

In the electrically-large regime, (4.31) cannot be, in general, further simplified, since no fast oscillating term is present in the integrand function, and, hence, the stationary phase method cannot be applied. In this case, therefore, we focus our attention on analyzing an upper-bound for $|F_R(\mathbf{r}_{\text{Rx}})|$, in order to unveil the impact of the size of $\mathscr{S}$ (e.g., if it tends to infinity).

Corollary 4.6 *Assume $d_{\text{P1}}(x,y)\le d_{\text{P2}}(x,y)$, where* (P1, P2) = (Tx, Rx) *or* (P1, P2) = (Rx, Tx),

$|\Gamma_{\text{ref}}(x,y)| = \Gamma_{\text{ref}} > 0$ for $(x,y) \in \mathscr{S}$, and $z_{\text{P1}} \neq 0$. Define $C_{\text{ref}} = \frac{2k^3 p_{\text{dm}} \Gamma_{\text{ref}} \mathscr{E}(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{ref}})}{16\pi^2 \epsilon_0}$. Then:

$$|F_R(\mathbf{r}_{\text{Rx}})| \leq C_{\text{ref}} \left(1 + \frac{z_{\text{P2}}}{z_{\text{P1}}}\right) \tan^{-1} \left[\frac{(x_{\text{P1}} - x)(y_{\text{P1}} - y)}{z_{\text{P1}} \sqrt{(x_{\text{P1}} - x)^2 + (y_{\text{P1}} - y)^2 + z_{\text{P1}}^2}} \right] \Bigg|_{x=-L_x}^{x=L_x} \Bigg|_{y=-L_y}^{y=L_y} \tag{4.32}$$

Proof 4.15 *Consider $\Omega_{\text{ref}}(x,y;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}})$ in Proposition 4.1, where $\tilde{\mathbf{p}}_{\text{rec}}$, $\tilde{\mathbf{p}}_{\text{ref}}$, and $\hat{\mathbf{s}}_{(x,y)}$ are real unit-norm vectors. By virtue of Cauchy-Schwarz's inequality (i.e., $-|\mathbf{u}|\,|\mathbf{v}| \leq \mathbf{u} \cdot \mathbf{v} \leq |\mathbf{u}|\,|\mathbf{v}|$ for any $\mathbf{u}$ and $\mathbf{v}$), we have $-1 \leq \tilde{\mathbf{p}}_{\text{rec}} \cdot \tilde{\mathbf{p}}_{\text{ref}} \leq 1$, $-1 \leq \hat{\mathbf{s}}_{(x,y)} \cdot \tilde{\mathbf{p}}_{\text{rec}} \leq 1$, and $-1 \leq \hat{\mathbf{s}}_{(x,y)} \cdot \tilde{\mathbf{p}}_{\text{ref}} \leq 1$. Hence, we obtain $-2\frac{k^2 p_{\text{dm}} \mathscr{E}(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{tran}})}{\epsilon_0} \leq \Omega_{\text{ref}}(x,y;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}}) \leq 2\frac{k^2 p_{\text{dm}} \mathscr{E}(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{tran}})}{\epsilon_0}$ for $(x,y) \in \mathscr{S}$. Thus:*

$$|F_R(\mathbf{r}_{\text{Rx}})| \leq \frac{2k^3 p_{\text{dm}} \Gamma_{\text{ref}} \mathscr{E}\left(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{ref}}\right)}{16\pi^2 \epsilon_0} \int_{-L_y}^{L_y} \int_{-L_x}^{L_x} \frac{\left(\cos\theta_{\text{inc}}(x,y) + \cos\theta_{\text{rec}}(x,y)\right)}{d_{\text{Tx}}(x,y) d_{\text{Rx}}(x,y)} dx dy \tag{4.33}$$

Since $d_{\text{P1}}(x,y) \leq d_{\text{P2}}(x,y)$, we have $\frac{(\cos\theta_{\text{inc}}(x,y)+\cos\theta_{\text{rec}}(x,y))}{d_{\text{P1}}(x,y) d_{\text{P2}}(x,y)} \leq \frac{z_{\text{Tx}}+z_{\text{Rx}}}{(d_{\text{P1}}(x,y))^3}$. Using the notable integral $\int_{-L_y}^{L_y} \int_{-L_x}^{L_x} \frac{1}{(d_{\text{P1}}(x,y))^3} dx dy = z_{\text{P1}}^{-1} \tan^{-1} \left[\frac{(x_{\text{P1}}-x)(y_{\text{P1}}-y)}{z_{\text{P1}} \sqrt{(x_{\text{P1}}-x)^2+(y_{\text{P1}}-y)^2+z_{\text{P1}}^2}} \right] \Big|_{x=-L_x}^{x=L_x} \Big|_{y=-L_y}^{y=L_y}$, the proof follows.

Remark 4.17 *From Corollary 4.6, we observe that $|F_R(\mathbf{r}_{\text{Rx}})| \lessapprox \left(1 + \frac{z_{\text{P2}}}{z_{\text{P1}}}\right) \frac{2k^3 p_{\text{dm}} \Gamma_{\text{ref}} \mathscr{E}(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{ref}})}{8\pi\epsilon_0}$ for $L_x, L_y \to \infty$. This implies that $|F_R(\mathbf{r}_{\text{Rx}})|$ is upper-bounded if the size of $\mathscr{S}$ increases without bound. Thus, the received power is bounded even for an infinitely large RIS. The scaling law as a function of the transmission distances is, in general, different from the weighted-sum distance obtained in* (4.27)*. This is because of the different optimization of $\angle\Gamma_{\text{ref}}(x,y)$. In the electrically-large regime, an anomalous reflecting RIS and a focusing RIS behave, in general, differently.*

Corollary 4.7 *In the electrically-small regime, $F_R(\mathbf{r}_{\text{Rx}})$ in* (4.31) *can be approximated as follows:*

$$F_R(\mathbf{r}_{\text{Rx}}) \approx \frac{jke^{j(\phi_0+\phi_{\text{rec}}+\phi_{\text{ref}})}}{16\pi^2 d_{\text{Tx0}} d_{\text{Rx0}}} \Omega_{\text{ref}}(0,0;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}}) \left(\cos\theta_{\text{inc0}} + \cos\theta_{\text{rec0}}\right) \int_{-L_y}^{L_y} \int_{-L_x}^{L_x} |\Gamma_{\text{ref}}(x,y)| dx dy \tag{4.34}$$

If $|\Gamma_{\text{ref}}(x,y)| = \Gamma_{\text{ref}} > 0$ for $(x,y) \in \mathscr{S}$, then $F_R(\mathbf{r}_{\text{Rx}})$ can be further approximated as follows:

$$F_R(\mathbf{r}_{\text{Rx}}) \approx \frac{jk\Gamma_{\text{ref}} \Omega_{\text{ref}}(0,0;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}}) L_x L_y \left(\cos\theta_{\text{inc0}} + \cos\theta_{\text{rec0}}\right)}{4\pi^2 d_{\text{Tx0}} d_{\text{Rx0}}} e^{j(\phi_0+\phi_{\text{rec}}+\phi_{\text{ref}})} \tag{4.35}$$

Proof 4.16 *It follows directly from* (4.18).

Remark 4.18 *The scaling laws of* $|F_R(\mathbf{r}_{\mathrm{Rx}})|$ *in* (4.35) *as a function of the distances and the size of* $\mathscr{S}$ *are the same as in* (4.30) *for anomalous reflection. This can be justified by analyzing* $\angle\Gamma_{\mathrm{ref}}(x,y)$ *for focusing and anomalous reflection. As for focusing, we have*

$$\angle\Gamma_{\mathrm{ref}}(x,y) = k\left(d_{\mathrm{Tx}}(x,y) + d_{\mathrm{Rx}}(x,y)\right) + \phi_0.$$

In the electrically-small regime, $d_{\mathrm{Tx}}(x,y)$ *and* $d_{\mathrm{Rx}}(x,y)$ *can be approximated by using* (4.13) *and ignoring* $R_2(x,y)$, *which yields* $\angle\Gamma_{\mathrm{ref}}(x,y) = k(\alpha_R + \beta_R) + (d_{\mathrm{Tx0}} + d_{\mathrm{Rx0}} + \phi_0)$ *with* α_R *and* β_R *as given in Remark 4.16. The obtained* $\angle\Gamma_{\mathrm{ref}}(x,y)$ *coincides with that of a surface that operates as an anomalous reflector towards the same direction as the focusing spot of a surface that operates as a focusing lens. In the electrically-small regime, hence, anomalous reflectors and focusing lenses are almost equivalent. This does not apply in the electrically-large regime.*

4.4.4 Comparison with other Path-Loss Models

Compared with other path-loss models recently reported in the literature, e.g., [33], [64], [36], [37], the distinctive features of the path-loss model in Proposition 4.1 can be summarized as follows: (i) the electric and magnetic currents induced on the RIS according to the surface equivalence principle are explicitly taken into account; (ii) the impact of the physical sources and the polarization of the radio waves are explicitly considered; (iii) the scaling laws and performance trends as a function of the transmission distances and the size of the RIS are unveiled in the electrically-large and electrically-small regimes. For example, the scaling law of the scattered power when the size of the RIS tends to infinity is evaluated analytically in Corollaries 4.1, 4.3, and 4.6, and it is proved to be compliant with the underlying assumptions of physical optics, geometric optics, and Fresnel zones [54, Sec. 8.3.2, Eq. (8.75)]; and (iv) although based on different electromagnetic methods, the obtained path-loss in the far-field region, i.e., (4.30) and (4.35), is consistent with those reported in [33, Eq. (4)] and [37, Eq. (10)].

4.5 Electric Field In the Presence of a Transmitting Surface

In this section, we analyze $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ under the assumption that $\mathscr{S}$ is a transmitting surface according to the definitions and assumptions given in Sec. 4.2 (see Fig. 4.1b and Fig. 4.2b). Some analytical steps are similar to the setup of reflecting surfaces. Thus, only the final results and the most important steps of the analysis are reported. The same applies to the performance trends.

Proposition 4.2 *Let* $\hat{\mathbf{s}}_{(x,y)} = \sin\theta_{\mathrm{inc}}(x,y)\cos\varphi_{\mathrm{inc}}(x,y)\hat{\mathbf{x}} + \sin\theta_{\mathrm{inc}}(x,y)\sin\varphi_{\mathrm{inc}}(x,y)\hat{\mathbf{y}} + \cos\theta_{\mathrm{inc}}(x,y)\hat{\mathbf{z}}$, *be the unit-norm propagation vector from* $\mathbf{r}_{\mathrm{Tx}}$ *to* $\mathbf{s} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \in \mathscr{S}$. *Define* $\Omega_{\mathrm{inc}}(x,y;\hat{\mathbf{p}}_{\mathrm{inc}},\hat{\mathbf{p}}_{\mathrm{rec}}) = (k^2/\epsilon_0)p_{\mathrm{dm}}\left(\tilde{\mathbf{p}}_{\mathrm{rec}}\cdot\tilde{\mathbf{p}}_{\mathrm{inc}} - \left(\hat{\mathbf{s}}_{(x,y)}\cdot\tilde{\mathbf{p}}_{\mathrm{rec}}\right)\left(\hat{\mathbf{s}}_{(x,y)}\cdot\tilde{\mathbf{p}}_{\mathrm{inc}}\right)\right)$ *and* $\Omega_{\mathrm{tran}}(x,y;\hat{\mathbf{p}}_{\mathrm{tran}},\hat{\mathbf{p}}_{\mathrm{rec}}) = (k^2/\epsilon_0)p_{\mathrm{dm}}(\tilde{\mathbf{p}}_{\mathrm{rec}}\cdot\tilde{\mathbf{p}}_{\mathrm{tran}} - \left(\hat{\mathbf{s}}_{(x,y)}\cdot\tilde{\mathbf{p}}_{\mathrm{rec}}\right)\left(\hat{\mathbf{s}}_{(x,y)}\cdot\tilde{\mathbf{p}}_{\mathrm{tran}}\right))\mathscr{E}\left(\hat{\mathbf{p}}_{\mathrm{inc}},\hat{\mathbf{p}}_{\mathrm{tran}}\right)$. *Under the assumptions stated in Lemma 4.2, the* *electric field* $\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})$ *projected onto* $\hat{\mathbf{p}}_{\mathrm{rec}}$ *can be formulated as follows:*

$$\begin{aligned}
\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})\cdot\hat{\mathbf{p}}_{\mathrm{rec}} &\approx \hat{\mathbf{p}}_{\mathrm{rec}}\cdot\mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{r}_{\mathrm{Rx}};\hat{\mathbf{p}}_{\mathrm{inc}}\right)G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}_{\mathrm{Tx}}) + jk\int_{\mathscr{S}}\Big[\Gamma_{\mathrm{tran}}(x,y)\Omega_{\mathrm{tran}}(x,y;\hat{\mathbf{p}}_{\mathrm{tran}},\hat{\mathbf{p}}_{\mathrm{rec}})e^{j(\phi_{\mathrm{tran}}+\phi_{\mathrm{rec}})}\\
&\quad -\Omega_{\mathrm{inc}}(x,y;\hat{\mathbf{p}}_{\mathrm{inc}},\hat{\mathbf{p}}_{\mathrm{rec}})e^{j(\phi_{\mathrm{inc}}+\phi_{\mathrm{rec}})}\Big]\,G(\mathbf{s},\mathbf{r}_{\mathrm{Tx}})\,G(\mathbf{r}_{\mathrm{Rx}},\mathbf{s})\left[\frac{z_{\mathrm{Rx}}}{|\mathbf{s}-\mathbf{r}_{\mathrm{Rx}}|}-\frac{z_{\mathrm{Tx}}}{|\mathbf{s}-\mathbf{r}_{\mathrm{Tx}}|}\right]d\mathbf{s}\\
&= \hat{\mathbf{p}}_{\mathrm{rec}}\cdot\mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{r}_{\mathrm{Rx}};\hat{\mathbf{p}}_{\mathrm{inc}}\right)G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}_{\mathrm{Tx}}) \qquad (4.36)\\
&\quad -\mathscr{I}_0\int_{-L_y}^{L_y}\int_{-L_x}^{L_x}\mathscr{I}_D(x,y)e^{-jk\mathscr{P}_D(x,y)}\,dx\,dy+\mathscr{I}_0\int_{-L_y}^{L_y}\int_{-L_x}^{L_x}\mathscr{I}_T(x,y)e^{-jk\mathscr{P}_T(x,y)}\,dx\,dy \qquad (4.37)
\end{aligned}$$

where $\mathscr{I}_0 = jk/(16\pi^2)$, $\mathscr{P}_D(x,y) = d_{\mathrm{Tx}}(x,y)+d_{\mathrm{Rx}}(x,y)-(\angle\phi_{\mathrm{inc}}+\angle\phi_{\mathrm{rec}})/k$, $\mathscr{P}_T(x,y) = d_{\mathrm{Tx}}(x,y)+d_{\mathrm{Rx}}(x,y)-(\angle\Gamma_{\mathrm{tran}}(x,y)+\angle\phi_{\mathrm{tran}}+\angle\phi_{\mathrm{rec}})/k$, *and the following shorthands are used:*

$$\mathscr{I}_D(x,y) = \frac{\Omega_{\mathrm{inc}}(x,y;\hat{\mathbf{p}}_{\mathrm{inc}},\hat{\mathbf{p}}_{\mathrm{rec}})\left(\cos\theta_{\mathrm{inc}}(x,y)+\cos\theta_{\mathrm{rec}}(x,y)\right)}{d_{\mathrm{Tx}}(x,y)d_{\mathrm{Rx}}(x,y)} \qquad (4.38)$$

$$\mathscr{I}_T(x,y) = \frac{\left|\Gamma_{\mathrm{tran}}(x,y)\right|\Omega_{\mathrm{inc}}(x,y;\hat{\mathbf{p}}_{\mathrm{tran}},\hat{\mathbf{p}}_{\mathrm{rec}})\left(\cos\theta_{\mathrm{inc}}(x,y)+\cos\theta_{\mathrm{rec}}(x,y)\right)}{d_{\mathrm{Tx}}(x,y)d_{\mathrm{Rx}}(x,y)} \qquad (4.39)$$

Proof 4.17 *See Appendix 4.8.7.*

Remark 4.19 *Consider* $\hat{\mathbf{p}}_{\mathrm{tran}} = \hat{\mathbf{p}}_{\mathrm{inc}}$, $\angle\Gamma_{\mathrm{tran}}(x,y) = 0$, *and* $|\Gamma_{\mathrm{tran}}(x,y)| = 1$. *By definition, we have* $\mathscr{E}(\hat{\mathbf{p}}_{\mathrm{inc}},\hat{\mathbf{p}}_{\mathrm{inc}}) = 1$. *Then, the integral terms in Proposition 4.2 coincide and their difference vanishes. This result is consistent with the fact that, under the considered special setup, the surface transmits the impinging wave without any modifications. Thus, we retrieve the setup in the absence of* $\mathscr{S}$, *and the received field coincides with the incident field in the absence of* $\mathscr{S}$.

In (4.36), the only term that depends on the design and properties of $\mathscr{S}$ is the last one, which we denote by $F_T(\mathbf{r}_{\mathrm{Rx}}) = \mathscr{I}_0\int_{-L_y}^{L_y}\int_{-L_x}^{L_x}\mathscr{I}_T(x,y)e^{-jk\mathscr{P}_T(x,y)}dx\,dy$. In the next sub-sections, therefore, we focus our attention only on the analysis of $F_T(\mathbf{r}_{\mathrm{Rx}})$. Similar to Sec. 4.4, $F_T(\mathbf{r}_{\mathrm{Rx}})$ is analyzed as a function of important design parameters and configurations for $\mathscr{S}$, e.g., $\angle\Gamma_{\mathrm{tran}}(x,y)$. As illustrative examples, similar to reflecting surfaces, we consider phase gradient metasurfaces [11].

4.5.1 $\mathscr{S}$ is Configured for Specular Transmission

This setup is obtained if $\angle\Gamma_{\mathrm{tran}}(x,y) = \phi_0$ for $(x,y)\in\mathscr{S}$, where $\phi_0\in[0,2\pi)$ is a fixed phase.

Corollary 4.8 *Let* $(x_s,y_s)\in\mathscr{S}$ *be the solution of the following system of equations:*

$$\frac{(x_s-x_{\mathrm{Tx}})}{d_{\mathrm{Tx}}(x_s,y_s)}+\frac{(x_s-x_{\mathrm{Rx}})}{d_{\mathrm{Rx}}(x_s,y_s)}=0,\qquad \frac{(y_s-x_{\mathrm{Tx}})}{d_{\mathrm{Tx}}(x_s,y_s)}+\frac{(y_s-x_{\mathrm{Rx}})}{d_{\mathrm{Rx}}(x_s,y_s)}=0 \qquad (4.40)$$

In the electrically-large regime, $F_T(\mathbf{r}_{\mathrm{Rx}})$ can be approximated as follows:

$$F_T(\mathbf{r}_{\mathrm{Rx}}) \approx \frac{\left|\Gamma_{\mathrm{tran}}(x_s, y_s)\right| \Omega_{\mathrm{tran}}(x_s, y_s; \hat{\mathbf{p}}_{\mathrm{tran}}, \hat{\mathbf{p}}_{\mathrm{rec}})}{4\pi(d_{\mathrm{Tx}}(x_s, y_s) + d_{\mathrm{Rx}}(x_s, y_s))} e^{-jk\left(d_{\mathrm{Tx}}(x_s,y_s)+d_{\mathrm{Rx}}(x_s,y_s)-(\phi_0+\phi_{\mathrm{tran}}+\phi_{\mathrm{rec}})/k\right)} \tag{4.41}$$

Proof 4.18 *It is similar to the proof of Corollary 4.1.*

Corollary 4.9 *In the electrically-small regime, $F_T(\mathbf{r}_{\mathrm{Rx}})$ can be approximated as follows:*

$$\begin{aligned} F_T(\mathbf{r}_{\mathrm{Rx}}) \approx & \frac{jk\Omega_{\mathrm{tran}}(0,0; \hat{\mathbf{p}}_{\mathrm{tran}}, \hat{\mathbf{p}}_{\mathrm{rec}}) (\cos\theta_{\mathrm{inc}0} + \cos\theta_{\mathrm{rec}0})}{16\pi^2 d_{\mathrm{Tx}0} d_{\mathrm{Rx}0}} e^{-jk\left(d_{\mathrm{Tx}0}+d_{\mathrm{Rx}0}-(\phi_0+\phi_{\mathrm{tran}}+\phi_{\mathrm{rec}})/k\right)} \\ & \int_{-L_y}^{L_y} \int_{-L_x}^{L_x} |\Gamma_{\mathrm{tran}}(x, y)| e^{jk(\mathscr{D}_x x+\mathscr{D}_y y)} dx dy \end{aligned} \tag{4.42}$$

where definitions and notation similar to Corollary 4.2 are employed. If $|\Gamma_{\mathrm{tran}}(x, y)| = \Gamma_{\mathrm{tran}} > 0$ for $(x, y) \in \mathscr{S}$, then $F_T(\mathbf{r}_{\mathrm{Rx}})$ can be simplified as follows:

$$\begin{aligned} F_R(\mathbf{r}_{\mathrm{Rx}}) \approx & \frac{jk\Gamma_{\mathrm{tran}}\Omega_{\mathrm{tran}}(0,0; \hat{\mathbf{p}}_{\mathrm{tran}}, \hat{\mathbf{p}}_{\mathrm{rec}}) L_x L_y (\cos\theta_{\mathrm{inc}0} + \cos\theta_{\mathrm{rec}0})}{4\pi^2 d_{\mathrm{Tx}0} d_{\mathrm{Rx}0}} \operatorname{sinc}(kL_x\mathscr{D}_x) \operatorname{sinc}\left(kL_y\mathscr{D}_y\right) \\ & e^{-jk\left(d_{\mathrm{Tx}0}+d_{\mathrm{Rx}0}-(\phi_0+\phi_{\mathrm{tran}}+\phi_{\mathrm{rec}})/k\right)} \end{aligned} \tag{4.43}$$

Proof 4.19 *It follows by direct application of* (4.14).

The intensity of $F_T(\mathbf{r}_{\mathrm{Rx}})$ in Corollaries 4.8 and 4.9 is similar to that in Corollaries 4.1 and 4.2, respectively. Therefore, similar scaling laws and performance trends are obtained. In the electrically-large regime, in particular, the law of transmission, i.e., $\varphi_{\mathrm{inc}}(x_s, y_s) = (\varphi_{\mathrm{rec}}(x_s, y_s) + \pi) \mod 2\pi$ and $\theta_{\mathrm{inc}}(x_s, y_s) = \theta_{\mathrm{rec}}(x_s, s_y)$ can be retrieved by direct inspection of (4.40).

4.5.2 $\mathscr{S}$ is Configured for Anomalous Transmission

This setup is obtained by setting $\angle\Gamma_{\mathrm{tran}}(x, y) = k(\alpha_T x + \beta_T y) + \phi_0$ for $(x, y) \in \mathscr{S}$, where $\alpha_T \in \mathbb{R}$ and $\beta_T \in \mathbb{R}$ are design parameters, and $\phi_0 \in [0, 2\pi)$ is a fixed phase. Similar to reflecting surfaces, the direction of anomalous transmission is determined by the setup of α_T and β_T.

Corollary 4.10 *Let $(x_s, y_s) \in \mathscr{S}$ be the solution of the following system of equations:*

$$\frac{(x_s - x_{\mathrm{Tx}})}{d_{\mathrm{Tx}}(x_s, y_s)} + \frac{(x_s - x_{\mathrm{Rx}})}{d_{\mathrm{Rx}}(x_s, y_s)} = \alpha_T, \qquad \frac{(y_s - x_{\mathrm{Tx}})}{d_{\mathrm{Tx}}(x_s, y_s)} + \frac{(y_s - x_{\mathrm{Rx}})}{d_{\mathrm{Rx}}(x_s, y_s)} = \beta_T \tag{4.44}$$

Assume the same notation and definitions as in Corollary 4.3 and Corollary 4.4. In the electrically-large regime, $F_T(\mathbf{r}_{\mathrm{Rx}})$ can be approximated as follows:

$$F_T(\mathbf{r}_{\mathrm{Rx}}) \approx \frac{\left|\Gamma_{\mathrm{tran}}(x_s, y_s)\right| \Omega_{\mathrm{tran}}(x_s, y_s; \hat{\mathbf{p}}_{\mathrm{tran}}, \hat{\mathbf{p}}_{\mathrm{rec}}) e^{-jk\left(d_{\mathrm{Tx}}(x_s,y_s)+d_{\mathrm{Rx}}(x_s,y_s)-(\alpha_T x_s+\beta_T y_s)-(\phi_0+\phi_{\mathrm{tran}}+\phi_{\mathrm{rec}})/k\right)}}{8\pi\sqrt{\mathscr{R}_1(d_{\mathrm{Tx}}(x_s, y_s))^2 + \mathscr{R}_2(d_{\mathrm{Rx}}(x_s, y_s))^2 + \mathscr{R}_3 d_{\mathrm{Tx}}(x_s, y_s) d_{\mathrm{Rx}}(x_s, y_s)}}$$

$$\approx \frac{\left|\Gamma_{\text{tran}}(x_s, y_s)\right| \Omega_{\text{tran}}(x_s, y_s; \hat{\mathbf{p}}_{\text{tran}}, \hat{\mathbf{p}}_{\text{rec}})}{8\pi(K_1 d_{\text{Tx}}(x_s, y_s) + K_2 d_{\text{Rx}}(x_s, y_s))} e^{-jk(d_{\text{Tx}}(x_s,y_s)+d_{\text{Rx}}(x_s,y_s)-(\alpha_T x_s+\beta_T y_s)-(\phi_0+\phi_{\text{tran}}+\phi_{\text{rec}})/k)} \tag{4.45}$$

Proof 4.20 *It is similar to the proofs of Corollary 4.3 and Corollary 4.4.*

Corollary 4.11 *In the electrically-small regime, $F_T(\mathbf{r}_{\text{Rx}})$ can be approximated as follows:*

$$F_T(\mathbf{r}_{\text{Rx}}) \approx \frac{jk\Omega_{\text{tran}}(0,0; \hat{\mathbf{p}}_{\text{tran}}, \hat{\mathbf{p}}_{\text{rec}})(\cos\theta_{\text{inc0}} + \cos\theta_{\text{rec0}})}{16\pi^2 d_{\text{Tx0}} d_{\text{Rx0}}} e^{-jk(d_{\text{Tx0}}+d_{\text{Rx0}}-(\phi_0+\phi_{\text{tran}})/k)} \int_{-L_y}^{L_y} \int_{-L_x}^{L_x} |\Gamma_{\text{tran}}(x,y)| e^{jk(\mathscr{D}_{\alpha_T} x + \mathscr{D}_{\beta_T} y)} dx dy \tag{4.46}$$

where the shorthand notation $\mathscr{D}_{\alpha_T} = \alpha_T + \mathscr{D}_x$ and $\mathscr{D}_{\beta_T} = \beta_T + \mathscr{D}_y$ is used. If $\left|\Gamma_{\text{tran}}(x,y)\right| = \Gamma_{\text{tran}}$ for $(x,y) \in \mathscr{S}$, then $F_T(\mathbf{r}_{\text{Rx}})$ can further be approximated as follows:

$$F_T(\mathbf{r}_{\text{Rx}}) \approx \frac{jk\Gamma_{\text{tran}}\Omega_{\text{tran}}(0,0; \hat{\mathbf{p}}_{\text{tran}}, \hat{\mathbf{p}}_{\text{rec}}) L_x L_y (\cos\theta_{\text{inc0}} + \cos\theta_{\text{rec0}})}{4\pi^2 d_{\text{Tx0}} d_{\text{Rx0}}} \operatorname{sinc}\left(kL_x \mathscr{D}_{\alpha_T}\right) \operatorname{sinc}\left(kL_y \mathscr{D}_{\beta_T}\right) e^{-jk(d_{\text{Tx0}}+d_{\text{Rx0}}-(\phi_0+\phi_{\text{tran}})/k)} \tag{4.47}$$

Proof 4.21 *It is similar to the proof of Corollary 4.5.*

Once again, we observe that the intensity of $F_T(\mathbf{r}_{\text{Rx}})$ in Corollaries 4.10 and 4.11 is similar to that in Corollaries 4.3 and 4.5, respectively. In particular, the angles of transmission can be optimized through the setup of α_T and β_T, similar to the optimization of α_R and β_R for reflecting surfaces.

4.5.3 $\mathscr{S}$ is Configured for Focusing

This setup is obtained by setting $\angle\Gamma_{\text{tran}}(x,y) = k\left(d_{\text{Tx}}(x,y) + d_{\text{Rx}}(x,y)\right) + \phi_0$ for $(x,y) \in \mathscr{S}$, where $\phi_0 \in [0, 2\pi)$ is a fixed phase. With this setup, $F_T(\mathbf{r}_{\text{Rx}})$ in (4.36) simplifies as follows:

$$F_T(\mathbf{r}_{\text{Rx}}) \approx \frac{jke^{j(\phi_0+\phi_{\text{rec}}+\phi_{\text{tran}})}}{16\pi^2} \int_{-L_y}^{L_y} \int_{-L_x}^{L_x} \Omega_{\text{tran}}(x, y; \hat{\mathbf{p}}_{\text{tran}}, \hat{\mathbf{p}}_{\text{rec}}) \frac{|\Gamma_{\text{tran}}(x,y)| \left(\cos\theta_{\text{inc}}(x,y) + \cos\theta_{\text{rec}}(x,y)\right)}{d_{\text{Tx}}(x,y) d_{\text{Rx}}(x,y)} dx dy \tag{4.48}$$

Similar to reflecting surfaces, the following corollaries provide an upper-bound and an asymptotic approximation for (4.48) in the electrically-large and electrically-small regimes, respectively.

Table 4.3 – Simulation setup

Settings
$f = 28$ GHz, $\lambda = 10.71$ mm
$\epsilon_0 = 8.85 \cdot 10^{-12}$ Farad/meter
$p_{\text{dm}} = (k^2/\epsilon_0)^{-1}$
$\tilde{\mathbf{p}}_{\text{inc}} = \tilde{\mathbf{p}}_{\text{ref}} = \tilde{\mathbf{p}}_{\text{tran}} = \tilde{\mathbf{p}}_{\text{rec}} = \hat{\mathbf{y}}$ (transverse electric)
$\phi_0 = \phi_{\text{inc}} = \phi_{\text{ref}} = \phi_{\text{tran}} = \phi_{\text{rec}} = 0$
$\mathscr{E}\left(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{ref}}\right) = \mathscr{E}\left(\hat{\mathbf{p}}_{\text{inc}}, \hat{\mathbf{p}}_{\text{tran}}\right) = 1$
$\lvert\Gamma_{\text{ref}}(x,y)\rvert = \lvert\Gamma_{\text{tran}}(x,y)\rvert = 1 \ \forall (x,y) \in \mathscr{S}$
$\theta_{\text{inc0}} = \pi/4$, $\varphi_{\text{inc0}} = \pi/3$
$\theta_{\text{rec0}} = \pi/6$, $\varphi_{\text{rec0}} = \pi$ (reflecting $\mathscr{S}$)
$\theta_{\text{rec0}} = \pi/3$, $\varphi_{\text{rec0}} = 5\pi/4$ (transmitting $\mathscr{S}$)
$\alpha_R = \alpha_T = -\sin\theta_{\text{inc0}}\cos\varphi_{\text{inc0}} - \sin\theta_{\text{rec0}}\cos\varphi_{\text{rec0}}$
$\beta_R = \beta_T = -\sin\theta_{\text{inc0}}\sin\varphi_{\text{inc0}} - \sin\theta_{\text{rec0}}\sin\varphi_{\text{rec0}}$

Corollary 4.12 *Assume* $d_{\text{P1}}(x,y) \le d_{\text{P2}}(x,y)$*, where* $(\text{P1},\text{P2}) = (\text{Tx},\text{Rx})$ *or* $(\text{P1},\text{P2}) = (\text{Rx},\text{Tx})$, $\lvert\Gamma_{\text{tran}}(x,y)\rvert = \Gamma_{\text{tran}} > 0$ *for* $(x,y) \in \mathscr{S}$*, and* $z_{\text{P1}} \neq 0$*. Define* $C_{\text{tran}} = \frac{2k^3 p_{\text{dm}}\Gamma_{\text{tran}}\mathscr{E}\left(\hat{\mathbf{p}}_{\text{inc}},\hat{\mathbf{p}}_{\text{tran}}\right)}{16\pi^2\epsilon_0}$*. Then:*

$$\lvert F_T(\mathbf{r}_{\text{Rx}})\rvert \le C_{\text{tran}}\left(1 - \frac{z_{\text{P2}}}{z_{\text{P1}}}\right)\tan^{-1}\left[\frac{(x_{\text{P1}}-x)(y_{\text{P1}}-y)}{\lvert z_{\text{P1}}\rvert\sqrt{(x_{\text{P1}}-x)^2+(y_{\text{P1}}-y)^2+z_{\text{P1}}^2}}\right]\Bigg|_{x=-L_x}^{x=L_x}\Bigg|_{y=-L_y}^{y=L_y} \quad (4.49)$$

Proof 4.22 *It is the same as for Corollary 4.6.*

Corollary 4.13 *In the electrically-small regime,* $F_T(\mathbf{r}_{\text{Rx}})$ *can be approximated as follows:*

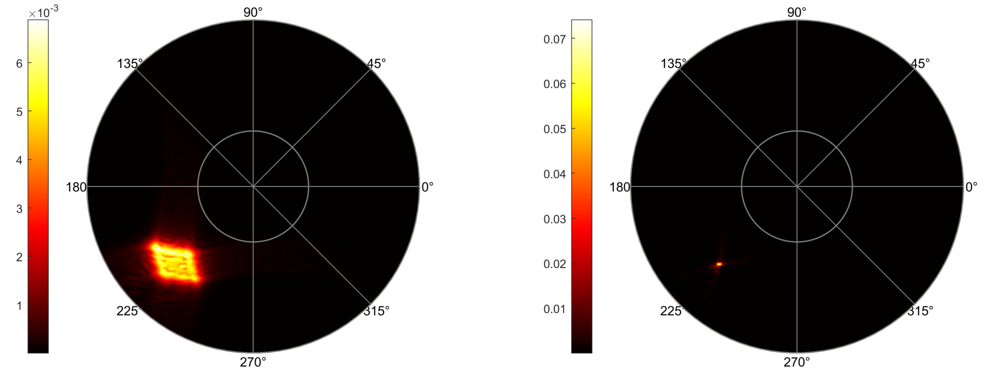
$$F_T(\mathbf{r}_{\text{Rx}}) \approx \frac{jke^{j(\phi_0+\phi_{\text{rec}}+\phi_{\text{tran}})}}{16\pi^2}\Omega_{\text{tran}}(0,0;\hat{\mathbf{p}}_{\text{tran}},\hat{\mathbf{p}}_{\text{rec}})\left(\cos\theta_{\text{inc0}}+\cos\theta_{\text{rec0}}\right)$$
$$\int_{-L_y}^{L_y}\int_{-L_x}^{L_x}\lvert\Gamma_{\text{tran}}(x,y)\rvert dxdy \quad (4.50)$$

If $\lvert\Gamma_{\text{tran}}(x,y)\rvert = \Gamma_{\text{tran}} > 0$ *for* $(x,y)\in\mathscr{S}$*, then* $F_T(\mathbf{r}_{\text{Rx}})$ *can be further approximated as follows:*

$$F_T(\mathbf{r}_{\text{Rx}}) \approx \frac{jk\Gamma_{\text{tran}}\lvert\Omega_{\text{tran}}(0,0;\hat{\mathbf{p}}_{\text{tran}},\hat{\mathbf{p}}_{\text{rec}})\rvert L_x L_y\left(\cos\theta_{\text{inc0}}+\cos\theta_{\text{rec0}}\right)}{4\pi^2 d_{\text{Tx0}}d_{\text{Rx0}}}e^{j(\phi_0+\phi_{\text{rec}}+\phi_{\text{tran}})} \quad (4.51)$$

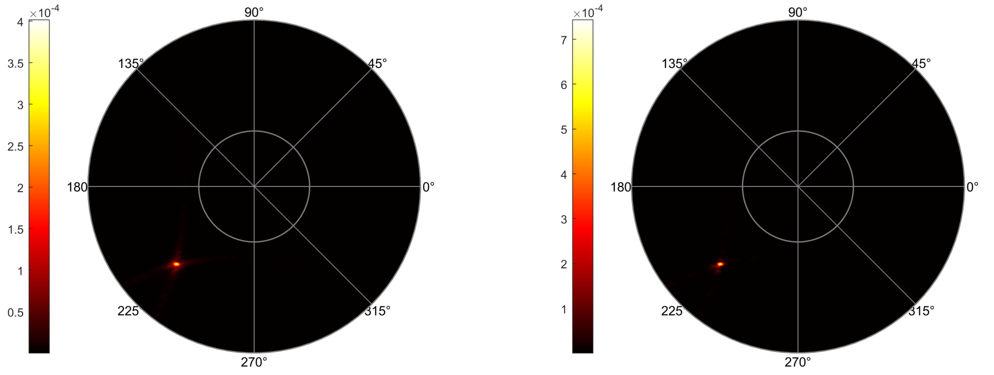
Proof 4.23 *It follows by direct application of* (4.18).

In conclusion, we show that $\lvert F_T(\mathbf{r}_{\text{Rx}})\rvert$ in Corollaries 4.12 and 4.13 is similar to $\lvert F_R(\mathbf{r}_{\text{Rx}})\rvert$ in Corollaries 4.6 and 4.7, respectively. As for the performance trends as a function of the size of $\mathscr{S}$ and the transmission distances, reflecting and transmitting surfaces have a similar behavior.



(a) Anomalous transmission, $d_{\text{Tx}0} = d_{\text{Rx}0} = 5$ m.

(b) Focusing lens, $d_{\text{Tx}0} = d_{\text{Rx}0} = 5$ m.



(c) Anomalous transmission, $d_{\text{Tx}0} = d_{\text{Rx}0} = 50$ m.

(d) Focusing lens, $d_{\text{Tx}0} = d_{\text{Rx}0} = 50$ m.

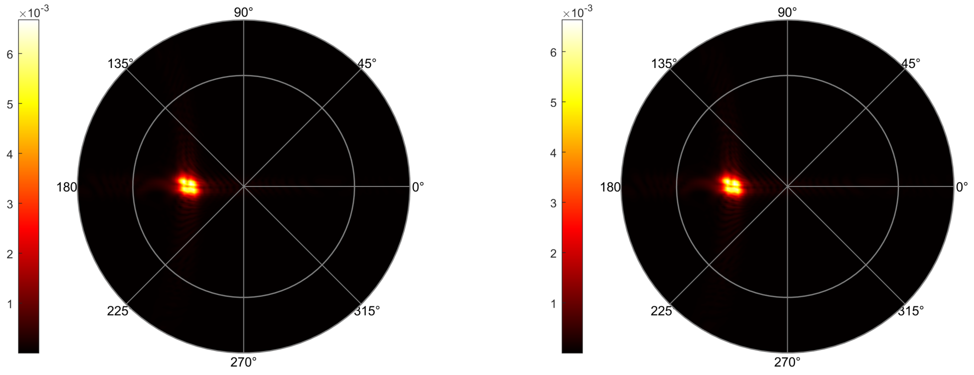
Figure 4.3 – Anomalous transmission vs. focusing lens (transmitting surface). Setup: $2L_x = 2L_y = 1$ m.

4.6 Numerical Results

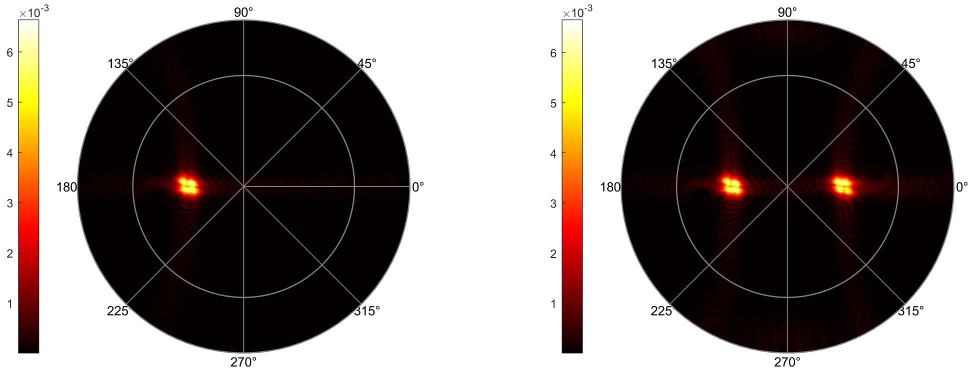
In this section, we illustrate some numerical examples in order to shed light on the behavior of the path-loss in the presence of RISs. In addition, we aim to analyze the conditions under which the considered asymptotic regimes hold true, and whether the considered phase gradient metasurfaces allow us to realize anomalous reflection/transmission and focusing as elaborated in Secs. 4.4 and 4.5. Unless otherwise stated, we use the simulation setup in Table 4.3. The simulation results illustrate $F_R(\mathbf{r}_{\text{Rx}})$ and $F_T(\mathbf{r}_{\text{Rx}})$ obtained in Secs. 4.4 and 4.5, respectively.

4.6.1 Anomalous Reflection and Focusing

In Fig. 4.3, we analyze anomalous transmission and focusing (transmitting surface) by using Proposition 4.2. The radial lines spaced by 45 degrees denote the angle φ and the three inner



(a) Anomalous reflection, no discretization. (b) Anomalous reflection, discretization step = 0.25λ

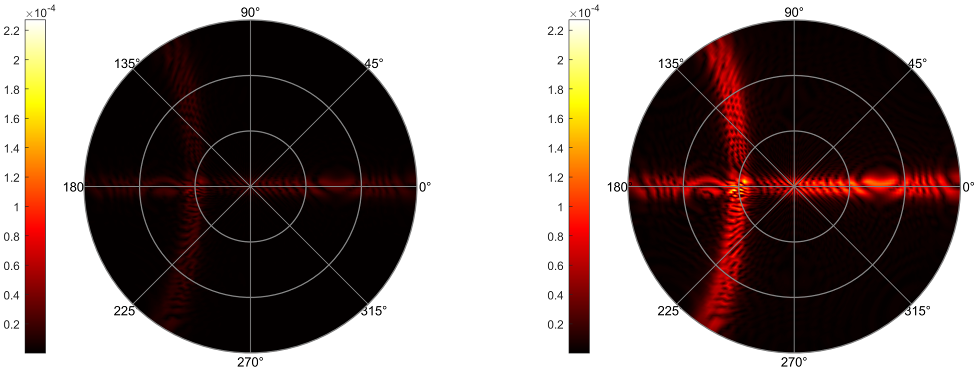


(c) Anomalous reflection, discretization step = 0.5λ. (d) Anomalous reflection, discretization step = λ.

Figure 4.4 – Anomalous reflection: Impact of discretization. Setup: $2L_x = 2L_y = 0.5$ m; $d_{\text{Tx0}} = d_{\text{Rx0}} = 5$ m.

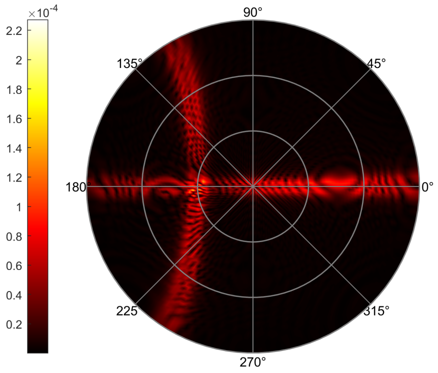
circles spaced by 30 degrees denote the angle θ (some lines and circles are removed for clarity). We observe that the correct angles of transmission are obtained. We note that anomalous transmitting surfaces yield a larger coverage area than focusing lenses. This is obtained, however, only for short transmission distances (near-field of the RIS). On the other hand, the larger coverage area is not apparent for long transmission distances (far-field of the RIS). This confirms Remark 4.18.

In Fig. 4.4, we analyze anomalous reflection and focusing (reflecting surface) by using Proposition 4.1. In particular, we consider a discretized version of the integral in (4.19), which corresponds to a practical implementation of the RIS based on (discrete) scattering elements. Provided that the scattering elements are spaced less than half of the wavelength apart (i.e., the discretization step is $\lambda/2$), we show that no significant differences can be observed at the naked eye. If the discretization step is greater than $\lambda/2$, e.g., it is less than λ, we observe the presence of grating lobes (spurious reflections) in unwanted directions. To better appreciate the impact of discretization, Fig. 4.5 reports the absolute error difference that corresponds



(a) Absolute error of Figs. 4.4(b), 4.4(a).

(b) Absolute error of Figs. 4.4(c), 4.4(a).



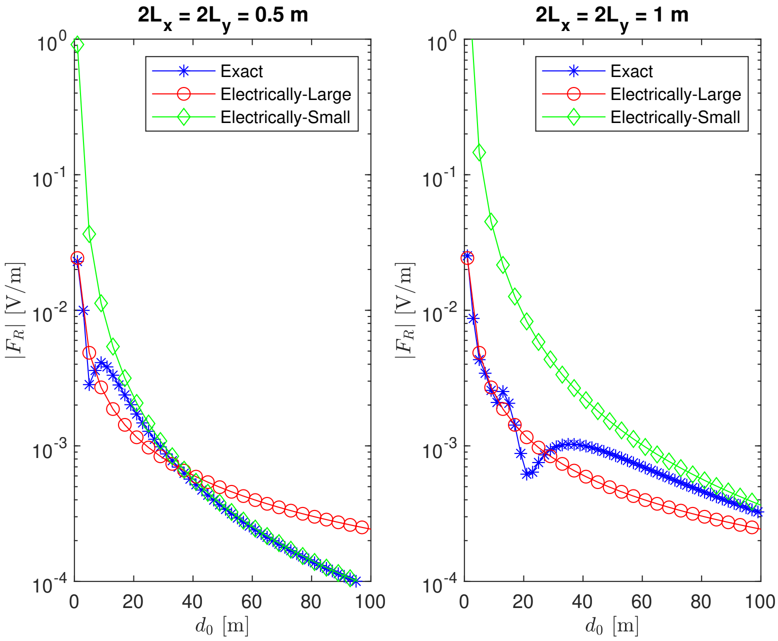
(c) Absolute error of Figs. 4.4(c), 4.4(b).

Figure 4.5 – Absolute error difference corresponding to Fig. 4.4.

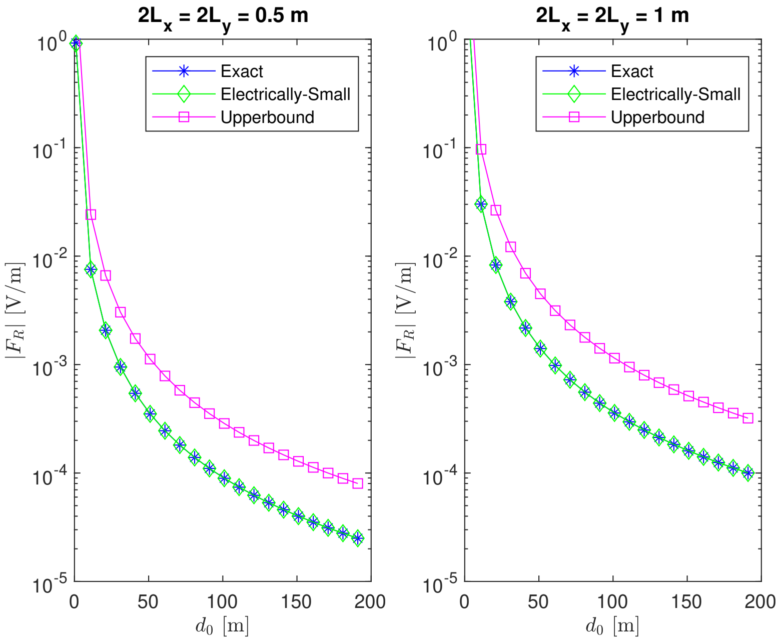
to the setups in Fig. 4.4. We observe that some differences are indeed apparent and that more closely spaced scattering elements yield more accurate estimates of the electric field (especially in the considered near-field regime).

4.6.2 Transmission Distance

In Fig. 4.6, we analyze the impact of the transmission distance in the context of anomalous reflection and focusing. In particular, the distances from the transmitter to the center of the RIS, and from the center of the RIS to the receiver are denoted by $d_{\mathrm{Tx}0} = d_{\mathrm{Rx}0} = d_0$. The angles of observation computed with respect to the center of the RIS are kept fixed as d_0 increases or decreases. We observe that the analytical frameworks obtained in the electrically-large and electrically-small asymptotic regimes well overlap, in the regions of interest, with the integral representation of the electric field. In particular, we note a major difference between anomalous reflectors and focusing lenses. As for anomalous reflectors, we observe two scaling laws as a function of the distance: (i) the weighted-sum path-loss model for short distances



(a) Anomalous reflection.



(b) Focusing lens.

Figure 4.6 – Anomalous reflection vs. focusing lens: Impact of transmission distance.

and (ii) the product path-loss model for long distances. As for focusing lenses, on the other hand, we observe a single scaling law: the product path-loss model that is sufficiently accurate for short and long distances. This highlights that the impact of the distance depends on the

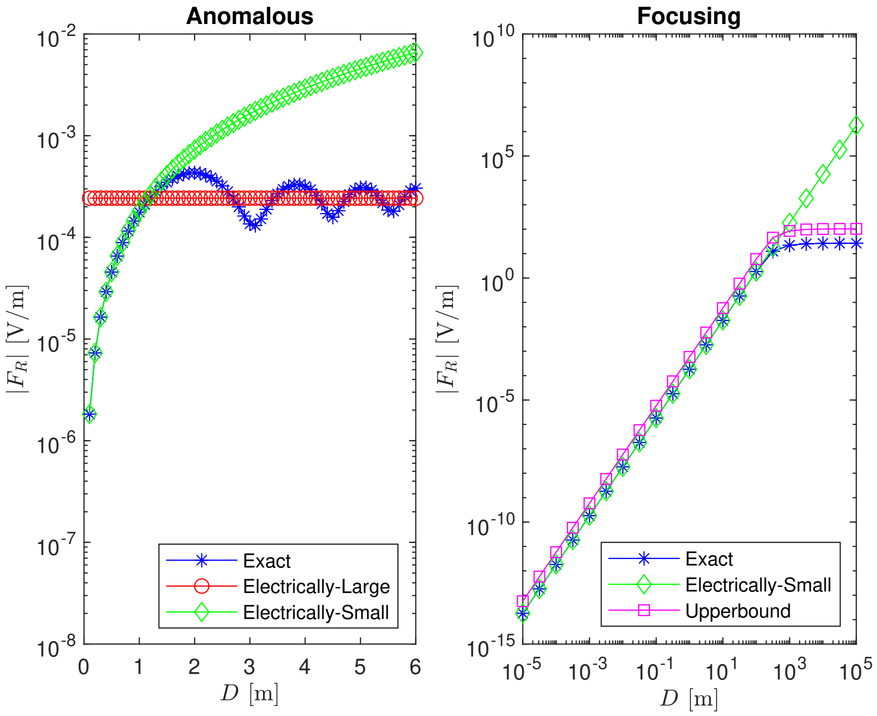


Figure 4.7 – Impact of surface size.

setup of the RIS. A focusing lens co-phases all the contributions scattered from the RIS and this yields a different scaling law as compared with an anomalous reflector. As for anomalous reflectors, it is worth noting that the weighted-sum path-loss model may be accurate up to a few tens of meters, which may be important in indoor scenarios and for local coverage enhancement in outdoor scenarios.

4.6.3 Surface Size

In Fig. 4.7, we analyze the impact of the size of the RIS for anomalous reflection and focusing (reflecting surface). In particular, the figure reports the intensity of the electric field as a function of the diagonal, D, of the RIS. We observe that the intensity of the electric field is bounded even if the size of the surface increases without bound. The closed-form analytical frameworks and the bounds obtained in the electrically-large and electrically-small regimes well predict the scaling law. It is worth noting that the electrically-small approximation may significantly overestimate the intensity of the electric field even for relatively small surfaces and for long transmission distances (100 meters in the figure). These results confirm that the proposed path-loss model is compliant with the power conservation law and that the received power is always bounded, regardless of the size of the surface, in the far-field of the RIS microstructure.

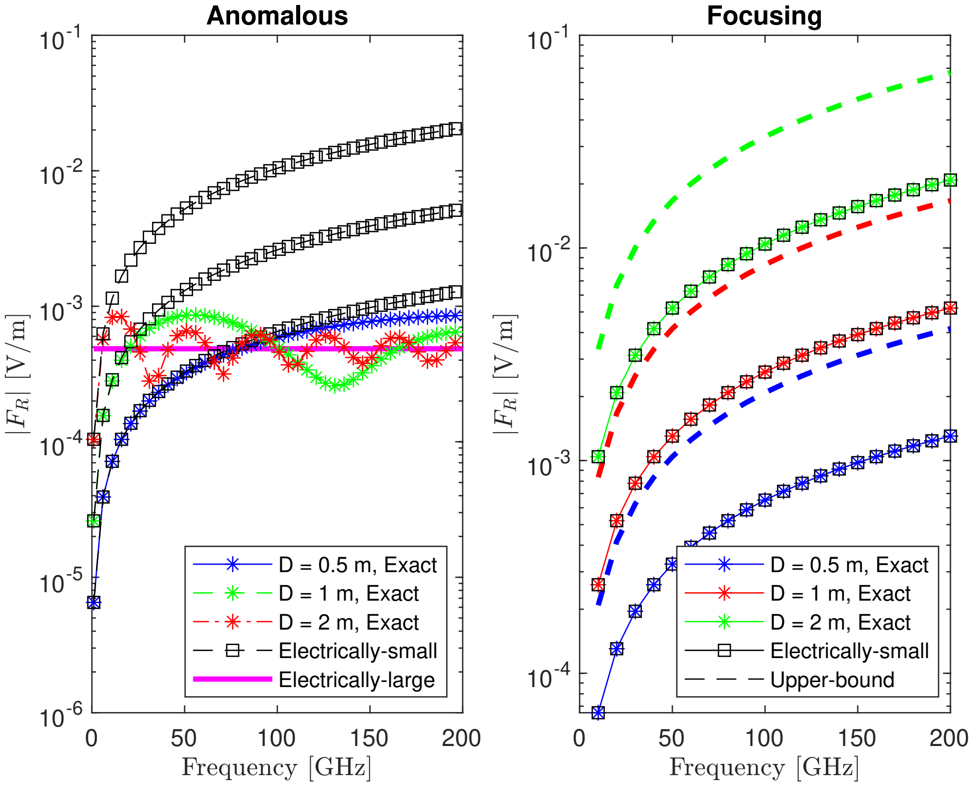


Figure 4.8 – Impact of transmission frequency. Reflecting surface. Setup: $d_{\mathrm{Tx}0} = d_{\mathrm{Rx}0} = 50$ m.

4.6.4 Operating Frequency

In Fig. 4.8, we consider anomalous reflection and focusing (reflecting surface), and analyze the impact of the transmission frequency on the scaling laws of the scattered field. It is worth nothing that the results are normalized by imposing the relation $(k^2/\epsilon_0)p_{\mathrm{dm}} = 1$ (see Table 4.3) that makes Corollaries 4.3 and 4.4 independent of the frequency. The numerical results are consistent with the analytical derivations. The higher the transmission frequency is, in particular, the more likely an RIS whose size is kept fixed with the frequency is viewed as electrically-large by the transmitter and receiver. Similar to other figures, the boundary between the electrically-large and electrically-small region is different for anomalous reflection and focusing. It is worth mentioning that the higher the transmission frequency is, the more scattering elements and tuning circuits are needed if the size of the RIS is kept fixed. Thus, as remarked in [64], the deployment of RISs in high frequency bands requires a careful system design that accounts for the tradeoff between the fine controllability of the EM field and the cost and power consumption of the RIS.

4.7 Conclusion

We have introduced a physics-compliant path-loss model for RIS-aided wireless transmission. The proposed path-loss model is general enough for application to various operating regimes, which include near-field and far-field asymptotic regimes. The far-field approximations agree with existing results available in e.g., [36, 37]. Our work advances those results by also taking into account the near-field approximations. The impact of several design parameters has been analyzed. In particular, we have proved that the scaling laws of the received power as a function of the transmission distance and the size of the RIS are different in the near-field and far-field regimes, and they depend on the wave transformations applied by the RIS. Notably, the received power scattered by an RIS is bounded as its size increases without bound.

In the context of wireless communications, in general terms, one should always use the integral representation of the path-loss in order to make sure that the received power is physically meaningful as a function of every design parameter, e.g., the surface size and the transmission distance. The simple analytical expressions obtained in the near-field and far-field asymptotic regimes can be employed provided that the considered system setup is compliant with their regime of validity. For application to the performance evaluation and optimization of wireless networks, one may consider the use of a two-law path-loss model (in analogy with two-slope path-loss models), which combines together the closed-form analytical expressions obtained in the near-field and far-field regimes. This approach may avoid the analytical intractability of using two-fold integrals while ensuring compliance with physics-based constraints.

4.8 Appendices

4.8.1 Proof of Theorem 4.1

By inserting $\mathbf{H}_{\partial V}(\mathbf{r}') = -\nabla_{\mathbf{r}'} \times \mathbf{E}_{\partial V}(\mathbf{r}')/(j\omega\mu_0)$ in (11) and by scalar multiplying both sides of the resulting equations with $\hat{\mathbf{p}}_{\text{rec}}$, we obtain:

$$\begin{aligned}\mathbf{E}(\mathbf{r}_{\text{Rx}}) \cdot \hat{\mathbf{p}}_{\text{rec}} = \mathbb{1}_{(\mathbf{r}_{\text{Tx}} \in V)} \mathbf{E}_{\text{inc}}(\mathbf{r}_{\text{Rx}}; \hat{\mathbf{p}}_{\text{inc}}) \cdot \hat{\mathbf{p}}_{\text{rec}} - \int_{\partial V} \Bigg\{ \Bigg[G(\mathbf{r}_{\text{Rx}}, \mathbf{r}') \Big(\hat{\mathbf{n}}_{\text{out}} \times \left(\nabla_{\mathbf{r}'} \times \mathbf{E}_{\partial V}(\mathbf{r}') \right) \Big) \Bigg] \cdot \hat{\mathbf{p}}_{\text{rec}} \\ + \left(\hat{\mathbf{n}}_{\text{out}} \cdot \mathbf{E}_{\partial V}(\mathbf{r}') \right) \left(\nabla_{\mathbf{r}'} G(\mathbf{r}_{\text{Rx}}, \mathbf{r}') \cdot \hat{\mathbf{p}}_{\text{rec}} \right) + \Bigg[(\hat{\mathbf{n}}_{\text{out}} \times \mathbf{E}_{\partial V}(\mathbf{r}')) \times \nabla_{\mathbf{r}'} G(\mathbf{r}_{\text{Rx}}, \mathbf{r}') \Bigg] \cdot \hat{\mathbf{p}}_{\text{rec}} \Bigg\} d\mathbf{r}' \quad (4.52)\end{aligned}$$

By using the divergence theorem and identities in [61, Eqs. (20), (21), (23)], the first and third integrands in (4.52) can be transformed as follows:

$$\begin{aligned}&- \int_{\partial V} \Bigg\{ \Bigg[G(\mathbf{r}_{\text{Rx}}, \mathbf{r}') \Big(\hat{\mathbf{n}}_{\text{out}} \times \left(\nabla_{\mathbf{r}'} \times \mathbf{E}_{\partial V}(\mathbf{r}') \right) \Big) \Bigg] \cdot \hat{\mathbf{p}}_{\text{rec}} \Bigg\} d\mathbf{r}' \\ &\qquad = - \int_{V} \Bigg[\nabla_{\mathbf{r}} G(\mathbf{r}_{\text{Rx}}, \mathbf{r}) \cdot \Big(\left(\hat{\mathbf{p}}_{\text{rec}} \cdot \nabla_{\mathbf{r}} \right) \mathbf{E}(\mathbf{r}) \Big) \Bigg] d\mathbf{r} + \int_{V} \Bigg[\left(\nabla_{\mathbf{r}} G(\mathbf{r}_{\text{Rx}}, \mathbf{r}) \cdot \hat{\mathbf{p}}_{\text{rec}} \right) \left(\nabla_{\mathbf{r}} \cdot \mathbf{E}(\mathbf{r}) \right) \Bigg] d\mathbf{r} \\ &\qquad\quad + \int_{\partial V} \Bigg[G(\mathbf{r}_{\text{Rx}}, \mathbf{r}') \hat{\mathbf{n}}_{\text{out}} \cdot \Big(\nabla_{\mathbf{r}'} \left(\mathbf{E}_{\partial V}(\mathbf{r}') \cdot \hat{\mathbf{p}}_{\text{rec}} \right) \Big) \Bigg] d\mathbf{r}' \qquad (4.53)\end{aligned}$$

$$
\begin{aligned}
&-\int_{\partial V}\left\{\left[(\hat{\mathbf{n}}_{\text{out}}\times\mathbf{E}_{\partial V}(\mathbf{r}'))\times\nabla_{\mathbf{r}'}G(\mathbf{r}_{\text{Rx}},\mathbf{r}')\right]\cdot\hat{\mathbf{p}}_{\text{rec}}\right\}d\mathbf{r}'\\
&\quad=\int_{\partial V}\left[-\left(\hat{\mathbf{n}}_{\text{out}}\cdot\nabla_{\mathbf{r}'}G(\mathbf{r}_{\text{Rx}},\mathbf{r}')\right)\left(\mathbf{E}_{\partial V}(\mathbf{r}')\cdot\hat{\mathbf{p}}_{\text{rec}}\right)+\left(\hat{\mathbf{p}}_{\text{rec}}\cdot\nabla_{\mathbf{r}'}G(\mathbf{r}_{\text{Rx}},\mathbf{r}')\right)\left(\mathbf{E}_{\partial V}(\mathbf{r}')\cdot\hat{\mathbf{n}}_{\text{out}}\right)\right]d\mathbf{r}'\\
&\quad+\int_{V}\left[-(\nabla_{\mathbf{r}}\cdot\mathbf{E}(\mathbf{r}))\left(\hat{\mathbf{p}}_{\text{rec}}\cdot\nabla_{\mathbf{r}}G(\mathbf{r}_{\text{Rx}},\mathbf{r})\right)+\nabla_{\mathbf{r}}G(\mathbf{r}_{\text{Rx}},\mathbf{r})\cdot\left(\hat{\mathbf{p}}_{\text{rec}}\cdot\nabla_{\mathbf{r}}\right)\mathbf{E}(\mathbf{r})\right]d\mathbf{r} \qquad (4.54)
\end{aligned}
$$

where $\mathbf{E}(\mathbf{r})$ is the total electric field at a point $\mathbf{r}\in V$. The proof follows from substituting (4.53) and (4.54) into (4.52) and canceling out some terms.

4.8.2 Proof of Theorem 4.2

The total electric field at any point $\mathbf{r}'\in\partial V\setminus\mathscr{S}$ (i.e., not including $\mathscr{S}$) is equal to the incident field, i.e., $\mathbf{E}_{\partial V}(\mathbf{r}')=\mathbf{E}_{\text{inc}}(\mathbf{r}',\hat{\mathbf{p}}_{\text{inc}})$. On the other hand, the electric field at any point $\mathbf{s}\in\mathscr{S}$ is equal to $\mathbf{E}_{\partial V}(\mathbf{r}')=\mathbf{E}_{\mathscr{S}}(\mathbf{s})$, where $\mathbf{E}_{\mathscr{S}}(\mathbf{s})$ is given in (4.3) or (4.4). By denoting $M(\mathbf{r}')=\mathbf{E}_{\text{inc}}(\mathbf{r}',\hat{\mathbf{p}}_{\text{inc}})\cdot\hat{\mathbf{p}}_{\text{rec}}$ and $N(\mathbf{s})=\mathbf{E}_{\mathscr{S}}(\mathbf{s})\cdot\hat{\mathbf{p}}_{\text{rec}}$, (4.10) can be written, with the aid of some algebra, as follows:

$$
\begin{aligned}
\mathbf{E}(\mathbf{r}_{\text{Rx}})\cdot\hat{\mathbf{p}}_{\text{rec}}&\overset{(a)}{=}\mathbb{1}_{(\mathbf{r}_{\text{Tx}}\in V)}M(\mathbf{r}_{\text{Rx}})-\int_{\partial V}\left[M(\mathbf{r}')\nabla_{\mathbf{r}'}G(\mathbf{r}_{\text{Rx}},\mathbf{r}')-G(\mathbf{r}_{\text{Rx}},\mathbf{r}')\nabla_{\mathbf{r}'}M(\mathbf{r}')\right]\cdot\hat{\mathbf{n}}_{\text{out}}d\mathbf{r}'\\
&\quad-\int_{\mathscr{S}}\left[(N(\mathbf{s})-M(\mathbf{s}))\nabla_{\mathbf{s}}G(\mathbf{r}_{\text{Rx}},\mathbf{s})-G(\mathbf{r}_{\text{Rx}},\mathbf{s})\nabla_{\mathbf{s}}(N(\mathbf{s})-M(\mathbf{s}))\right]\cdot\hat{\mathbf{n}}_{\text{out}}d\mathbf{s} \qquad (4.55)
\end{aligned}
$$

where (a) is obtained by taking into account that: (i) $\int_{\partial V}=\int_{\partial V\setminus\mathscr{S}}+\int_{\mathscr{S}}$, and (ii) $N(\mathbf{s})-M(\mathbf{s})$ is the difference between the total electric field $\mathbf{E}_{\mathscr{S}}(\mathbf{s})$ and the incident field $\mathbf{E}_{\text{inc}}(\mathbf{s},\hat{\mathbf{p}}_{\text{inc}})$ on $\mathscr{S}$.

Let us consider $I_{\partial V}=-\int_{\partial V}\left[M(\mathbf{r}')\nabla_{\mathbf{r}'}G(\mathbf{r}_{\text{Rx}},\mathbf{r}')-G(\mathbf{r}_{\text{Rx}},\mathbf{r}')\nabla_{\mathbf{r}'}M(\mathbf{r}')\right]\cdot\hat{\mathbf{n}}_{\text{out}}d\mathbf{r}'$. By applying the divergence theorem [48, Eq. (C.25)] and the identity $\nabla_{\mathbf{r}}\cdot\left(f\nabla_{\mathbf{r}}g\right)=f\nabla^2_{\mathbf{r}}g+\nabla_{\mathbf{r}}f\cdot\nabla_{\mathbf{r}}g$ [62, Eq. (VII-46)] to generic scalar functions f and g, $I_{\partial V}$ can be simplified as follows:

$$
\begin{aligned}
I_{\partial V}&=-\int_{V}\left[M(\mathbf{r})(\nabla^2_{\mathbf{r}}+k^2)G(\mathbf{r}_{\text{Rx}},\mathbf{r})-G(\mathbf{r}_{\text{Rx}},\mathbf{r})(\nabla^2_{\mathbf{r}}+k^2)M(\mathbf{r})\right]d\mathbf{r} \qquad (4.56)\\
&\overset{(a)}{=}M(\mathbf{r}_{\text{Rx}})+\int_{V}G(\mathbf{r}_{\text{Rx}},\mathbf{r})(\nabla^2_{\mathbf{r}}+k^2)M(\mathbf{r})d\mathbf{r}\\
&\overset{(b)}{=}M(\mathbf{r}_{\text{Rx}})+\int_{V}G(\mathbf{r}_{\text{Rx}},\mathbf{r})\left((\vec{\nabla}^2_{\mathbf{r}}+k^2)\mathbf{E}_{\text{inc}}(\mathbf{r};\hat{\mathbf{p}}_{\text{inc}})\right)\cdot\hat{\mathbf{p}}_{\text{rec}}d\mathbf{r}
\end{aligned}
$$

where (a) follows from (4.8) and from the fact that $\mathbf{r}_{\text{Rx}}$ is always contained in V, and (b) follows from the identity $(\nabla^2_{\mathbf{r}}+k^2)M(\mathbf{r})=\left((\vec{\nabla}^2_{\mathbf{r}}+k^2)\mathbf{E}_{\text{inc}}(\mathbf{r};\hat{\mathbf{p}}_{\text{inc}})\right)\cdot\hat{\mathbf{p}}_{\text{rec}}$.

Consider the integral $I_V=\int_V G(\mathbf{r}_{\text{Rx}},\mathbf{r})\left((\vec{\nabla}^2_{\mathbf{r}}+k^2)\mathbf{E}_{\text{inc}}(\mathbf{r};\hat{\mathbf{p}}_{\text{inc}})\right)\cdot\hat{\mathbf{p}}_{\text{rec}}d\mathbf{r}$. From (4.5) and by virtue of the identities $\nabla_{\mathbf{r}}\times\nabla_{\mathbf{r}}\times\mathbf{E}_{\text{inc}}(\mathbf{r};\hat{\mathbf{p}}_{\text{inc}})=\nabla_{\mathbf{r}}(\nabla_{\mathbf{r}}\cdot\mathbf{E}_{\text{inc}}(\mathbf{r};\hat{\mathbf{p}}_{\text{inc}}))-\vec{\nabla}^2_{\mathbf{r}}\mathbf{E}_{\text{inc}}(\mathbf{r};\hat{\mathbf{p}}_{\text{inc}})$ [62, Eq. (VII-51)] and $\nabla\cdot\mathbf{E}_{\text{inc}}(\mathbf{r};\hat{\mathbf{p}}_{\text{inc}})=\rho(\mathbf{r},\mathbf{r}_{\text{Tx}})/\epsilon_0$ [48, Sec. 1.1], the integral I_V simplifies to $I_V=\int_V G(\mathbf{r}_{\text{Rx}},\mathbf{r})(\nabla_{\mathbf{r}}\rho(\mathbf{r},\mathbf{r}_{\text{Tx}})/\epsilon_0+j\omega\mu_0\mathbf{J}(\mathbf{r},\mathbf{r}_{\text{Tx}}))\cdot\hat{\mathbf{p}}_{\text{rec}}d\mathbf{r}$. By definition: (i) $I_V=0$ if Tx is not contained in V, and (ii) $I_V=-\mathbf{E}_{\text{inc}}(\mathbf{r}_{\text{Rx}};\hat{\mathbf{p}}_{\text{inc}})\cdot\hat{\mathbf{p}}_{\text{rec}}$ if Tx is contained in V [48, Eqs. (15.3.3), (15.3.6)]. Thus, we have $I_V=-\mathbb{1}_{(\mathbf{r}_{\text{Tx}}\in V)}M(\mathbf{r}_{\text{Rx}})$, and, from (4.56), $I_{\partial V}=M(\mathbf{r}_{\text{Rx}})-\mathbb{1}_{(\mathbf{r}_{\text{Tx}}\in V)}M(\mathbf{r}_{\text{Rx}})$. With the aid of some simplifications, the proof follows.

4.8.3 Proof of Lemma 4.4

The proof is based on the application of the stationary phase method [62, Appendix VIII], [63] to (4.12) under the assumption of operating in the electrically-large regime, as stated in Definition 4.3. Let $(x_s, y_s) \in \Psi$ be the stationary points of $\mathscr{P}(x, y) = d_{\mathrm{Tx}}(x, y) + d_{\mathrm{Rx}}(x, y) - \mathscr{C}(x, y)$ for $(x, y) \in \mathscr{S}$. By invoking the stationary phase method, the integral I_1 in (4.12) oscillates very quickly outside a small region centered at $(x_s, y_s) \in \Psi$, and, thus, the contributions outside the small region around the stationary points cancel out when computing the integral [62, pg. 923]. Under these conditions, I_1 can be well approximated by (i) replacing $\mathscr{P}(x, y)$ with its Taylor approximation evaluated at $(x_s, y_s) \in \Psi$, i.e., $\mathscr{P}(x, y) \approx \mathscr{P}(x_s, y_s) + A(x - x_s)^2 + B(y - y_s)^2 + C(x - x_s)(y - y_s)$ where $A = \frac{\partial^2}{\partial x^2}\mathscr{P}(x, y)|_{(x,y)=(x_s,y_s)}$, $B = \frac{\partial^2}{\partial y^2}\mathscr{P}(x, y)|_{(x,y)=(x_s,y_s)}$, and $C = \frac{\partial^2}{\partial x \partial y}\mathscr{P}(x, y)|_{(x,y)=(x_s,y_s)}$ and (ii) by letting the extremes of integration go to infinity, since I_1 is dominated by a small region around the stationary points and the contributions to the integral outside that small region cancel out. For simplicity, let us assume that a single stationary point exists. The case study with multiple stationary points is obtained by summing up the contributions from all the stationary points [63, Sec. 1.3, pg. 15]. Accordingly, I_1 can be approximated as follows:

$$
\begin{aligned}
I_1 \approx\ & \mathscr{A}_1(d_{\mathrm{Tx}}(x_s, y_s), d_{\mathrm{Rx}}(x_s, y_s))\mathscr{B}_1(x_s, y_s)e^{-jk\mathscr{P}(x_s,y_s)} \\
& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-jk\left(A(x-x_s)^2+C(x-x_s)(y-y_s)+B(x-x_s)^2\right)}dxdy
\end{aligned}
\tag{4.57}
$$

From (4.57), the proof follows from [62, Eqs. (VIII-10)-(VIII-22)].

4.8.4 Proof of Lemma 4.5

Define $\mathscr{Q}(x, y) = \mathscr{A}_1(d_{\mathrm{Tx}}(x, y), d_{\mathrm{Rx}}(x, y))\mathscr{B}_1(x, y)$. Since no stationary points lie in $\mathscr{S}$, we can divide and multiply the integrand of (4.12) by $\partial\mathscr{P}(x, y)/\partial x \neq 0$. Thus,(4.12) can be written as:

$$
\begin{aligned}
I_1 &= \int_{-L_y}^{L_y}\int_{-L_x}^{L_x} \frac{\mathscr{Q}(x, y)e^{-jk\mathscr{P}(x,y)}}{\partial\mathscr{P}(x, y)/\partial x}\frac{\partial\mathscr{P}(x, y)}{\partial x}dxdy \\
&\overset{(a)}{=} \frac{1}{(-jk)}\int_{-L_y}^{L_y}\left(\frac{\mathscr{Q}(x, y)e^{-jk\mathscr{P}(x,y)}}{\partial\mathscr{P}(x, y)/\partial x}\Bigg|_{x=-L_x}^{x=L_x} - \int_{-L_x}^{L_x}\frac{\partial}{\partial x}\left(\frac{\mathscr{Q}(x, y)}{\partial\mathscr{P}(x, y)/\partial x}\right)e^{-jk\mathscr{P}(x,y)}dx\right)dy \\
&\overset{(b)}{\approx} \frac{1}{(-jk)}\int_{-L_y}^{L_y}\left(\frac{\mathscr{Q}(L_x, y)e^{-jk\mathscr{P}(L_x,y)}}{\mathscr{P}_x(L_x, y)}\right)dy - \frac{1}{(-jk)}\int_{-L_y}^{L_y}\left(\frac{\mathscr{Q}(-L_x, y)e^{-jk\mathscr{P}(-L_x,y)}}{\mathscr{P}_x(-L_x, y)}\right)dy \\
&\overset{(c)}{\approx} \frac{1}{(-jk)^2}\left[\frac{\mathscr{Q}(L_x, y)e^{-jk\mathscr{P}(L_x,y)}}{\mathscr{P}_x(L_x, y)\mathscr{P}_y(L_x, y)} - \frac{\mathscr{Q}(-L_x, y)e^{-jk\mathscr{P}(-L_x,y)}}{\mathscr{P}_x(-L_x, y)\mathscr{P}_y(-L_x, y)}\right]\Bigg|_{y=-L_y}^{y=L_y}
\end{aligned}
\tag{4.58}
$$

where (a) is obtained by using integration by parts, (b) follows by virtue of Riemann-Lebesgue's lemma, which states that the integral over x decays with $1/k^2$, and, therefore, it can be ignored as compared with the first term [65, Eqs. (3.21), (3.22)], [66, Eq. (4.2)], and (c) follows by applying again the same procedure but by multiplying and dividing the two integrands in (b)

by $\partial\mathscr{P}(\pm L_x, y)/\partial y \neq 0$. The proof follows by iterating the same procedure once more.

4.8.5 Proof of Proposition 4.1

Consider (4.11). From Lemma 4.2, $\mathbf{E}_{\text{inc}}(\mathbf{r}_{\text{Rx}};\hat{\mathbf{p}}_{\text{inc}}) \approx \mathbf{E}_{0,\text{inc}}\left(\mathbf{r}_{\text{Rx}};\hat{\mathbf{p}}_{\text{inc}}\right) G(\mathbf{r}_{\text{Rx}},\mathbf{r}_{\text{Tx}})$. From (4.3) and Lemma 4.2, $\mathbf{E}_{\mathscr{S}}(\mathbf{s}) = \mathbf{E}_{\text{inc}}(\mathbf{s};\hat{\mathbf{p}}_{\text{inc}}) + \Gamma_{\text{ref}}(\mathbf{s})\mathscr{E}_{\text{ref}}(\hat{\mathbf{p}}_{\text{inc}},\hat{\mathbf{p}}_{\text{ref}})\mathbf{E}_{\text{inc}}(\mathbf{s};\hat{\mathbf{p}}_{\text{ref}})$, with $\mathbf{E}_{\text{inc}}(\mathbf{s};\hat{\mathbf{p}}_{\text{inc}}) \approx \mathbf{E}_{0,\text{inc}}\left(\mathbf{s};\hat{\mathbf{p}}_{\text{inc}}\right) G(\mathbf{s},\mathbf{r}_{\text{Tx}})$ and $\mathbf{E}_{\text{inc}}(\mathbf{s};\hat{\mathbf{p}}_{\text{ref}}) \approx \mathbf{E}_{0,\text{inc}}\left(\mathbf{s};\hat{\mathbf{p}}_{\text{ref}}\right) G(\mathbf{s},\mathbf{r}_{\text{Tx}})$. By inserting them in (4.11), we obtain:

$$\begin{aligned}\mathbf{E}(\mathbf{r}_{\text{Rx}})\cdot\hat{\mathbf{p}}_{\text{rec}} &\approx \hat{\mathbf{p}}_{\text{rec}}\cdot\mathbf{E}_{0,\text{inc}}\left(\mathbf{r}_{\text{Rx}};\hat{\mathbf{p}}_{\text{inc}}\right) G(\mathbf{r}_{\text{Rx}},\mathbf{r}_{\text{Tx}}) \\ &- \int_{\mathscr{S}} \left[\mathscr{F}_{\text{ref}}(\mathbf{s}) G(\mathbf{s},\mathbf{r}_{\text{Tx}})\nabla_{\mathbf{s}} G(\mathbf{r}_{\text{Rx}},\mathbf{s}) - G(\mathbf{r}_{\text{Rx}},\mathbf{s})\nabla_{\mathbf{s}}\left(\mathscr{F}_{\text{ref}}(\mathbf{s}) G(\mathbf{s},\mathbf{r}_{\text{Tx}})\right)\right]\cdot\hat{\mathbf{n}}_{\text{out}}\,d\mathbf{s}\end{aligned} \tag{4.59}$$

where $\mathscr{F}_{\text{ref}}(\mathbf{s}) = \Gamma_{\text{ref}}(\mathbf{s})\mathscr{E}_{\text{ref}}(\hat{\mathbf{p}}_{\text{inc}},\hat{\mathbf{p}}_{\text{ref}})\mathbf{E}_{0,\text{inc}}\left(\mathbf{s};\hat{\mathbf{p}}_{\text{ref}}\right)\cdot\hat{\mathbf{p}}_{\text{rec}} = \Gamma_{\text{ref}}(\mathbf{s})\Omega_{\text{ref}}(x,y;\hat{\mathbf{p}}_{\text{ref}},\hat{\mathbf{p}}_{\text{rec}})e^{j(\phi_{\text{ref}}+\phi_{\text{rec}})}$ for $(x,y)\in\mathscr{S}$, where $\Omega_{\text{ref}}(\cdot)$ is defined in the statement of Proposition 4.1.

In the reflection case, $\hat{\mathbf{n}}_{\text{out}} = -\hat{\mathbf{z}}$. Thus, by using the product rule of derivatives, we have $\hat{\mathbf{n}}_{\text{out}}\cdot\nabla_{\mathbf{s}}\left(\mathscr{F}_{\text{ref}}(\mathbf{s}) G(\mathbf{s},\mathbf{r}_{\text{Tx}})\right) = \mathscr{Z}_1 + \mathscr{Z}_2$, where $\mathscr{Z}_1 = -\mathscr{F}_{\text{ref}}(\mathbf{s})\frac{\partial}{\partial z} G(\mathbf{s},\mathbf{r}_{\text{Tx}})$ and $\mathscr{Z}_2 = -G(\mathbf{s},\mathbf{r}_{\text{Tx}})\frac{\partial}{\partial z}\mathscr{F}_{\text{ref}}(\mathbf{s})$. By computing the derivatives, it can be shown that $\mathscr{Z}_1 \propto \frac{k^3}{|\mathbf{s}-\mathbf{r}_{\text{Tx}}|} G(\mathbf{s},\mathbf{r}_{\text{Tx}})$ and $\mathscr{Z}_2 \propto \frac{k^2}{|\mathbf{s}-\mathbf{r}_{\text{Tx}}|} G(\mathbf{s},\mathbf{r}_{\text{Tx}})$. Under the assumption $k \gg 1/|\mathbf{s}-\mathbf{r}_{\text{Tx}}|$, $\mathscr{Z}_1$ dominates $\mathscr{Z}_2$ and hence $\hat{\mathbf{n}}_{\text{out}}\cdot\nabla\left(\mathscr{F}_{\text{ref}}(\mathbf{s}) G(\mathbf{s},\mathbf{r}_{\text{Tx}})\right) \approx \mathscr{Z}_1 = -\mathscr{F}_{\text{ref}}(\mathbf{s})\frac{\partial}{\partial z} G(\mathbf{s},\mathbf{r}_{\text{Tx}})$. Therefore, (4.59) can be simplified as follows:

$$\begin{aligned}\mathbf{E}(\mathbf{r}_{\text{Rx}})\cdot\hat{\mathbf{p}}_{\text{rec}} &\approx \hat{\mathbf{p}}_{\text{rec}}\cdot\mathbf{E}_{0,\text{inc}}\left(\mathbf{r}_{\text{Rx}};\hat{\mathbf{p}}_{\text{inc}}\right) G(\mathbf{r}_{\text{Rx}},\mathbf{r}_{\text{Tx}}) \\ &- \int_{\mathscr{S}} \mathscr{F}_{\text{ref}}(\mathbf{s})\left[G(\mathbf{s},\mathbf{r}_{\text{Tx}})\nabla_{\mathbf{s}} G(\mathbf{r}_{\text{Rx}},\mathbf{s}) - G(\mathbf{r}_{\text{Rx}},\mathbf{s})\nabla_{\mathbf{s}} G(\mathbf{s},\mathbf{r}_{\text{Tx}})\right]\cdot\hat{\mathbf{n}}_{\text{out}}\,d\mathbf{s} \\ &\overset{(a)}{\approx} \hat{\mathbf{p}}_{\text{rec}}\cdot\mathbf{E}_{0,\text{inc}}\left(\mathbf{r}_{\text{Rx}};\hat{\mathbf{p}}_{\text{inc}}\right) G(\mathbf{r}_{\text{Rx}},\mathbf{r}_{\text{Tx}}) + jk\int_{\mathscr{S}} \mathscr{F}_{\text{ref}}(\mathbf{s}) G(\mathbf{s},\mathbf{r}_{\text{Tx}}) G(\mathbf{r}_{\text{Rx}},\mathbf{s})\left[\frac{z_{\text{Rx}}}{|\mathbf{s}-\mathbf{r}_{\text{Rx}}|} + \frac{z_{\text{Tx}}}{|\mathbf{s}-\mathbf{r}_{\text{Tx}}|}\right] d\mathbf{s}\end{aligned} \tag{4.60}$$

where (a) is obtained by taking into account that $\hat{\mathbf{n}}_{\text{out}} = -\hat{\mathbf{z}}$, and, hence, by definition:

$$\nabla_{\mathbf{s}} G(\mathbf{s},\mathbf{r}_{\text{Tx}})\cdot\hat{\mathbf{n}}_{\text{out}} \overset{(b)}{=} +\frac{\partial}{\partial z} G(\mathbf{s},\mathbf{r}_{\text{Tx}})_{|z=0} \overset{(d)}{\approx} jkG(\mathbf{s},\mathbf{r}_{\text{Tx}})\frac{z_{\text{Tx}}}{|\mathbf{s}-\mathbf{r}_{\text{Tx}}|} \tag{4.61}$$

$$\nabla_{\mathbf{s}} G(\mathbf{r}_{\text{Rx}},\mathbf{s})\cdot\hat{\mathbf{n}}_{\text{out}} \overset{(c)}{=} -\frac{\partial}{\partial z} G(\mathbf{r}_{\text{Rx}},\mathbf{s})_{|z=0} \overset{(d)}{\approx} -jkG(\mathbf{r}_{\text{Rx}},\mathbf{s})\frac{z_{\text{Rx}}}{|\mathbf{r}_{\text{Rx}}-\mathbf{s}|} \tag{4.62}$$

where the "+" sign in (b) and the "-" sign in (c) take into account that the directions of propagation of the incident and reflected signals point towards the same and the opposite directions with respect to $\hat{\mathbf{n}}_{\text{out}}$, respectively, and the approximations in (d) take into account that $1/|\mathbf{s}-\mathbf{r}_{\text{Tx}}| \ll k$ and $1/|\mathbf{r}_{\text{Rx}}-\mathbf{s}| \ll k$. This completes the proof.

4.8.6 Proof of Corollary 4.1

Consider (4.19). The proof is based on the stationary phase method stated in Lemma 4.4. According to Definition 4.2, the stationary points of $\mathscr{P}(x,y) = \mathscr{P}_R(x,y)$ in (4.20) correspond to the solutions of (4.22). Due to the monotonicity of (4.22) with respect to x_s and y_s, either a single or no stationary point exists. More precisely, (4.22) can be equivalently re-written as follows:

$$\begin{aligned}\sin\theta_{\mathrm{inc}}(x_s,y_s)\cos\varphi_{\mathrm{inc}}(x_s,y_s) &= -\sin\theta_{\mathrm{rec}}(x_s,y_s)\cos\varphi_{\mathrm{rec}}(x_s,y_s) \\ \sin\theta_{\mathrm{inc}}(x_s,y_s)\sin\varphi_{\mathrm{inc}}(x_s,y_s) &= -\sin\theta_{\mathrm{rec}}(x_s,y_s)\sin\varphi_{\mathrm{rec}}(x_s,y_s)\end{aligned} \tag{4.63}$$

which, using some algebra, yields $\varphi_{\mathrm{inc}}(x_s,y_s) = (\varphi_{\mathrm{rec}}(x_s,y_s)+\pi) \mod 2\pi$ and $\theta_s = \theta_{\mathrm{inc}}(x_s,y_s) = \theta_{\mathrm{rec}}(x_s,y_s)$ (i.e., the law of reflection). Based on Lemma 4.4 with $\mathscr{P}(x,y) = \mathscr{P}_R(x,y)$, the determinant of $\mathbf{A}(x_s,y_s)$ is $\det(\mathbf{A}(x_s,y_s)) = P_{xx}P_{yy} - (P_{xy})^2 = \cos^2\theta_s \left(1/d_{\mathrm{Tx}}(x_s,y_s) + 1/d_{\mathrm{Rx}}(x_s,y_s)\right)^2$, where the derivatives are $P_{xx} = \frac{\partial^2}{\partial x^2}\mathscr{P}_R(x,y)|_{(x,y)=(x_s,y_s)}$, $P_{yy} = \frac{\partial^2}{\partial y^2}\mathscr{P}_R(x,y)|_{(x,y)=(x_s,y_s)}$, and $P_{xy} = \frac{\partial^2}{\partial x \partial y}\mathscr{P}_R(x,y)|_{(x,y)=(x_s,y_s)}$. In addition, it can be proved that $(P_{xy})^2 < P_{xx}P_{yy}$, which implies that the two eigenvalues of $\mathbf{A}(x_s,y_s)$ are positive and distinct. Therefore, we obtain $\mathrm{sign}(\mathbf{A}(x_s,y_s)) = 2$. The proof follows by inserting $\det(\mathbf{A}(x_s,y_s))$ and $\mathrm{sign}(\mathbf{A}(x_s,y_s))$ in 4.4.

4.8.7 Proof of Proposition 4.2

Consider (4.11). From Lemma 4.2, $\mathbf{E}_{\mathrm{inc}}(\mathbf{r}_{\mathrm{Rx}};\hat{\mathbf{p}}_{\mathrm{inc}}) \approx \mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{r}_{\mathrm{Rx}};\hat{\mathbf{p}}_{\mathrm{inc}}\right) G\left(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}_{\mathrm{Tx}}\right)$. From (4.4) and Lemma 4.2, $\mathbf{E}_{\mathscr{S}}(\mathbf{s}) = \Gamma_{\mathrm{tran}}(\mathbf{s})\mathscr{E}_{\mathrm{tran}}(\hat{\mathbf{p}}_{\mathrm{inc}},\hat{\mathbf{p}}_{\mathrm{tran}})\mathbf{E}_{\mathrm{inc}}(\mathbf{s};\hat{\mathbf{p}}_{\mathrm{tran}})$, with $\mathbf{E}_{\mathrm{inc}}(\mathbf{s};\hat{\mathbf{p}}_{\mathrm{tran}}) \approx \mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{s};\hat{\mathbf{p}}_{\mathrm{tran}}\right) G\left(\mathbf{s},\mathbf{r}_{\mathrm{Tx}}\right)$. By inserting them in (4.11), we obtain:

$$\begin{aligned}\mathbf{E}(\mathbf{r}_{\mathrm{Rx}})\cdot\hat{\mathbf{p}}_{\mathrm{rec}} \approx{}& \hat{\mathbf{p}}_{\mathrm{rec}}\cdot\mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{r}_{\mathrm{Rx}};\hat{\mathbf{p}}_{\mathrm{inc}}\right) G(\mathbf{r}_{\mathrm{Rx}},\mathbf{r}_{\mathrm{Tx}}) \\ &- \int_{\mathscr{S}}\left[\mathscr{F}_{\mathrm{tran}}(\mathbf{s})G\left(\mathbf{s},\mathbf{r}_{\mathrm{Tx}}\right)\nabla_{\mathbf{s}}G(\mathbf{r}_{\mathrm{Rx}},\mathbf{s}) - G(\mathbf{r}_{\mathrm{Rx}},\mathbf{s})\nabla_{\mathbf{s}}\left(\mathscr{F}_{\mathrm{tran}}(\mathbf{s})G\left(\mathbf{s},\mathbf{r}_{\mathrm{Tx}}\right)\right)\right]\cdot\hat{\mathbf{n}}_{\mathrm{out}}\,d\mathbf{s}\end{aligned} \tag{4.64}$$

where $\mathscr{F}_{\mathrm{tran}}(\mathbf{s}) = \left(\Gamma_{\mathrm{tran}}(\mathbf{s})\mathscr{E}_{\mathrm{tran}}(\hat{\mathbf{p}}_{\mathrm{inc}},\hat{\mathbf{p}}_{\mathrm{tran}})\mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{s};\hat{\mathbf{p}}_{\mathrm{tran}}\right) - \mathbf{E}_{0,\mathrm{inc}}\left(\mathbf{s};\hat{\mathbf{p}}_{\mathrm{inc}}\right)\right)\cdot\hat{\mathbf{p}}_{\mathrm{rec}}$, which can be formulated in terms of $\Omega_{\mathrm{inc}}(\cdot)$ and $\Omega_{\mathrm{tran}}(\cdot)$ as defined in the statement of Proposition 4.2.

The rest of the proof is similar to Appendix 4.8.5. The difference is that $\hat{\mathbf{n}}_{\mathrm{out}} = \hat{\mathbf{z}}$, and, hence, the signs in (b) and (c) that correspond to (4.61) and (4.62) are both negative because the direction of propagation of the incident and transmitted signals is opposite to $\hat{\mathbf{n}}_{\mathrm{out}}$.

5 Conclusions and Future Work

This chapter highlights general conclusions of the thesis and summaries possible directions for future work.

5.1 Conclusion 112

5.2 Future Work 113

  5.2.1 Physics-based Modeling 113

  5.2.2 Experimental Validation 114

  5.2.3 Constrained System Design 114

  5.2.4 Statistics of the RIS Channel Model 114

  5.2.5 Information and Communication Theory 115

5.1 Conclusion

In general, this thesis provides introduction and novel results on modeling reconfigurable inteligent surfaces (RISs), from the perspective of and using tools from theoretical physics, in the context of wireless communications applications. The thesis starts by introducing, in Chapter 1, the concept of SRE and RIS as its enabling technology, highlighting its properties, functionalities, and benefits compared to the existing technologies. In Chapter 2, it also introduces analytical methods and methodologies for modeling RISs that are made of metasurfaces, and for computing the EM field scattered by metasurfaces, which can be categorized into two parts: (i) modeling the metasurfaces, and (ii) modeling the radio waves.

In Chapter 3, leveraging the general scalar theory of diffraction, approximate analytical path-loss modelings for RISs is derived in closed-forms for different RIS functionalities (i.e., specular reflection, anomalous reflection, and beamformer) and for different regimes of operation (i.e., electrically-small and electrically-large regimes). To this end, we assume that the electromagnetic waves are cylindrical, i.e., it travels in 2D-space. From the analytical results, it is observed that different scaling laws exist in the two considered operating regimes. In particular, when the RIS is configured for specular or anomalous reflection, in the electrically-large regime, the intensity of the electric field at the receiver (Rx) scales with the *inverse of the sum* of the distance of the transmitter (Tx) and the Rx from *the stationary point* (i.e., the point on the RIS at which the gradient of the phase function of the received field is zero) and it is independent of the size of the RIS. On the other hand, in the electrically-small regime, the intensity of the electric field at Rx scales with the *inverse of the product* of the distance of Tx and Rx from *the center of the RIS* and linearly increases with the surface size. It is worth mentioning that, although the analytical closed-form formulation for the case where the RIS is used for beamforming is not explicitly derived in this chapter, it can be easily obtained using similar steps used for the case of specular and anomalous reflection. Still in Chapter 3, using the obtained analytical results, we compare the performance of RIS-aided communications (measured in terms of data rate) against that of relay-assisted one. In particular, we numerically compare the maximum achievable data rate in case of communications with the aid of a single RIS against that of half-duplex (HD) and full-duplex (FD) relays. Based on the presented comparison results, it can be concluded that RIS-aided transmission may outperform relay-aided transmission provided that the size of the RIS is sufficiently large.

In Chapter 4, a physics-compliant path-loss model for RIS-aided wireless transmission is derived by using the vector generalization of Green's theorem. The results presented in this chapter extend the path-loss models obtained in Chapter 2: a spherical wave propagating in 3D-space is considered and its vectorial form (including its polarization) is taken into account. We also consider two cases in which the RIS is used for either reflection (Tx and Rx in the same side of RIS) or transmission (Tx and Rx on different sides of RIS). In each case, the path-loss is formulated in terms of a computable integral that depends on the transmission distances, the polarization of the radio waves, the size of the surface, and the desired surface transformation. Furthermore, approximation techniques such as Taylor approximation and stationary phase method (SPM) are employed to obtain closed-form approximations in two asymptotic regimes that are representative of far-field and near-field deployments. Based on the obtained closed-form results, the impact of several design parameters is analyzed. In particular, it is proved that the scaling laws of the received power as a function of the transmission distance and the size of the RIS are different in the near-field and far-field regimes, and they depend on the wave transformations applied by the RIS (specular reflection, anomalous reflection, or beamforming/focusing). It is also important to note that it is also proved the the received power scattered by an RIS is bounded as its size increases without bound (regardless if the RIS is used for specular/anomalous reflection or beamforming). This implies that the path-loss model is consistent with the law of the energy conservation: the electric power reflected by the RIS and received by the Rx cannot exceed the transmitted power.

To summarize, in the general context of wireless communications, the results obtained in this thesis reveal that one should always use the exact (integral) representation of the path-loss in order to make sure that the received power is physically meaningful as a function of every design parameter, e.g., the surface size and the transmission distance. The simple analytical expressions obtained in the near-field and far-field asymptotic regimes can be employed provided that the considered system setup is compliant with their regime of validity. For application to the performance evaluation and optimization of wireless networks, one may consider the use of a two-law path-loss model (in analogy with two-slope path-loss models), which combines together the closed-form analytical expressions obtained in the near-field and far-field regimes. This approach may avoid the analytical intractability of using two-fold integrals while ensuring compliance with physics-based constraints.

5.2 Future Work

Regarding the road ahead, there are several fundamental and open research issues that deserve attention.

5.2.1 Physics-based Modeling

Current research on RISs relies on simplified models on how the meta-surfaces shape the impinging radio waves. Hence, there is a compelling need for developing sufficiently accurate but analytically tractable models for the meta-surfaces, whose foundation is to be built on the laws of electromagnetism and physics. The contribution of this thesis constitutes the fundamental step in RIS modeling for wireless communication system where, through physics-based approach which is based on Maxwell equations, the path-loss models obtained for given RIS modeling assumptions are very accurate and practical (due to their closed-form representations), as shown by extensive numerical results. However, at the same time, some modeling assumptions are not realistic or too simplistic for practical purposes, in exchange to obtain tractable results. For example, RISs are usually modeled as local structures, and, therefore, the spatial coupling among the meta-atoms is ignored. Therefore, in the future, these factors need to be taken into account to obtain practical and realistic RIS models.

5.2.2 Experimental Validation

To be accepted by the wireless community, the models obtained throughout this thesis need to be validated through hardware testbeds and empirical measurements. As has been mentioned throughout the thesis, as far as current research is concerned, there exist a few experimental results (e.g., [33]) that have validated these scaling laws as a function of the size of the RISs, the transmission distances involved, and the specified wave transformations applied by the RISs. However, more comprehensive tests are required, for example, to account for the spatial coupling among the meta-atoms, effect of polarization change, and the impact of incident and reflection angles towards the RIS's efficiency.

5.2.3 Constrained System Design

The potential gains and applications of RISs in wireless networks depend on their nearly passive implementation. This imposes stringent constraints on the development of efficient signal processing algorithms and communication protocols. The absence of power amplifiers and channel estimation units on the RISs implies, for example, that no channel estimation can be performed at the RISs, and new and efficient (low overhead) protocols need to be developed for acquiring the necessary environmental information for controlling and programming their operation [67].

5.2.4 Statistics of the RIS Channel Model

To accurately assess the potential of RIS wireless applications, a realistic channel model is indispensable. To this end, in addition to deterministic component of the path-loss modeling, which is the central subject of this thesis, the statistics of the channel between the transmitters and receivers also need to be incorporated during the system-level design and analysis for meaningful and practical applications. In most works on RIS-aided communications, an independent and identically distributed (i.i.d.) Rayleigh fading assumption is often assumed for modeling the channel between a RIS element and a user. While this greatly simplifies the analysis and leads to tractable analysis, the assumption is unrealistic since, unlike uniform linear array (ULA), the configuration of the RIS elements is rectangular in nature and thus there will always be correlation between channels from different elements (i.e., the channel fading in RIS-aided communications will always be spatially correlated [68]). Moreover, since RIS is typically deployed strategically to provide virtual line-of-sight (VLOS) between a transmitter-receiver pair through the RIS, one should always consider the existence of a dominant propagation path in the Tx-RIS and RIS-Rx links. Therefore, an assumption of (correlated) Rician fading should be considered instead of Rayleigh fading. Another important consideration in modeling the channel is the Doppler effect, i.e., the change of frequency observed by the RIS due to the mobility of the UAVs and the users.

5.2.5 Information and Communication Theory

Conventional information and communication theoretic models applied to wireless networks assume that the system, i.e., the environment, is represented by transition probabilities that are fixed and cannot be optimized. The concept of smart radio environments based on RISs challenges this assumption, allowing the channel states to be included among the degrees of freedom for encoding and modulation. This opens up new venues for system optimization that can provide a better channel capacity, as recently reported in [10].

Thanks to its low-cost, low-energy, and low-complexity of deployments, RIS has an advantage over its competing technologies to be implemented in a large-scale environments. Therefore, it is also interesting to understand the performance limit of RIS when deployed in a large-scale networks, in addition to point-to-point settings. However, unfortunately, most studies in the literature are limited to “small-size” system models where usually one or only a few RISs are considered. To investigate the potential of large-scale deployments of RISs, therefore, more studies that take into account large-scale networks with hundreds or possibly thousand of RISs are required. To this end, mathematical tools such as stochastic geometry is very powerful, due to its simplicity and tractability, to aid the analysis and to unveil the impact of RIS deployment rate, among others, towards the overall system-level performance. Indeed, there are several works that incorporate stochastic geometry in the analysis of RIS-aides systems (e.g., [69, 70]). Unfortunately, most works neglect the consideration of the height of the transmitters (i.e., base stations), which is an important factor in RIS performance evaluation, and limit the analysis to two-dimensional topology which leads to inappropriate network models [71]. This necessitates the use of three dimensional (3D) stochastic geometry frameworks to obtain accurate assessment of performance analysis.

Bibliography

[1] M. Di Renzo, F. Habibi Danufane, X. Xi, J. de Rosny, and S. Tretyakov, "Analytical modeling of the path-loss for reconfigurable intelligent surfaces – anomalous mirror or scatterer ?" in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020, pp. 1–5.

[2] F. Yang and Y. Rahmat-Samii, *Surface Electromagnetics: With Applications in Antenna, Microwave, and Optical Engineering.* Cambridge University Press, Jun. 2019.

[3] V. Arun and H. Balakrishnan, "Rfocus: Practical beamforming for small devices," *arXiv preprint arXiv:1905.05130*, 2019.

[4] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, 2018.

[5] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–20, 2019.

[6] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 7331–7376, 2019.

[7] C. Liaskos, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "Using any surface to realize a new paradigm for wireless communications," *Communications of the ACM*, vol. 61, no. 11, pp. 30–33, 2018.

[8] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[9] M. Di Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. De Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang *et al.*, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.

[10] R. Karasik, O. Simeone, M. Di Renzo, and S. S. Shitz, "Beyond max-SNR: Joint encoding for reconfigurable intelligent surfaces," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 2965–2970.

[11] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[12] N. Yu, P. Genevet, M. A. Kats, F. Aieta, J.-P. Tetienne, F. Capasso, and Z. Gaburro, "Light propagation with phase discontinuities: generalized laws of reflection and refraction," *science*, vol. 334, no. 6054, pp. 333–337, 2011.

[13] W. Tang, M. Z. Chen, J. Y. Dai, Y. Zeng, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with programmable metasurface: New paradigms, opportunities, and challenges on transceiver design."

[14] C. Caloz and Z.-L. Deck-Léger, "Spacetime metamaterials—part i: general concepts," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 3, pp. 1569–1582, 2019.

[15] V. G. Veselago, "Electrodynamics of substances with simultaneously negative ϵ and μ," *Usp. Fiz. Nauk*, vol. 92, p. 517, 1967.

[16] C. Caloz and T. Itoh, *Electromagnetic Metamaterials: Transmission Line Theory and Microwave Applications*. John Wiley & Sons, 2005.

[17] D. R. Smith, W. J. Padilla, D. Vier, S. C. Nemat-Nasser, and S. Schultz, "Composite medium with simultaneously negative permeability and permittivity," *Physical Review Letters*, vol. 84, no. 18, p. 4184, 2000.

[18] J. B. Pendry, "Negative refraction makes a perfect lens," *Physical Review Letters*, vol. 85, no. 18, p. 3966, 2000.

[19] D. Schurig, J. J. Mock, B. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, and D. R. Smith, "Metamaterial electromagnetic cloak at microwave frequencies," *Science*, vol. 314, no. 5801, pp. 977–980, 2006.

[20] U. Leonhardt, "Optical conformal mapping," *Science*, vol. 312, no. 5781, pp. 1777–1780, 2006.

[21] C. Caloz and A. Alù, "Guest editorial special cluster on magnetless nonreciprocity in electromagnetics," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 11, pp. 1931–1937, 2018.

[22] T. Kodera, D. L. Sounas, and C. Caloz, "Artificial Faraday rotation using a ring metamaterial structure without static magnetic field," *Applied Physics Letters*, vol. 99, no. 3, p. 031114, 2011.

[23] S. B. Glybovski, S. A. Tretyakov, P. A. Belov, Y. S. Kivshar, and C. R. Simovski, "Metasurfaces: From microwaves to visible," *Physics Reports*, vol. 634, pp. 1–72, 2016.

[24] E. F. Kuester, M. A. Mohamed, M. Piket-May, and C. L. Holloway, "Averaged transition conditions for electromagnetic fields at a metafilm," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 10, pp. 2641–2651, 2003.

[25] C. L. Holloway, M. A. Mohamed, E. F. Kuester, and A. Dienstfrey, "Reflection and transmission properties of a metafilm: With an application to a controllable surface composed of resonant particles," *IEEE Transactions on Electromagnetic Compatibility*, vol. 47, no. 4, pp. 853–865, 2005.

[26] L. Subrt and P. Pechac, "Controlling propagation environments using intelligent walls," in *2012 6th European Conference on Antennas and Propagation (EUCAP)*. IEEE, 2012, pp. 1–5.

[27] ——, "Intelligent walls as autonomous parts of smart indoor environments," *IET communications*, vol. 6, no. 8, pp. 1004–1010, 2012.

[28] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 1, pp. 183–198, 2013.

[29] P. Del Hougne, M. F. Imani, T. Sleasman, J. N. Gollub, M. Fink, G. Lerosey, and D. R. Smith, "Dynamic metasurface aperture as smart around-the-corner motion detector," *Scientific Reports*, vol. 8, no. 1, pp. 1–10, 2018.

[30] J. C. B. Garcia, M. Kamoun, and A. Sibille, "Reconfigurable passive relaying array for coverage enhancement," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–6.

[31] N. Kaina, M. Dupré, G. Lerosey, and M. Fink, "Shaping complex microwave fields in reverberating media with binary tunable metasurfaces," *Scientific reports*, vol. 4, no. 1, pp. 1–8, 2014.

[32] P. del Hougne, M. Fink, and G. Lerosey, "Optimally diverse communication channels in disordered environments with tuned randomness," *Nature Electronics*, vol. 2, no. 1, pp. 36–41, 2019.

[33] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 421–439, 2021.

[34] J. C. B. Garcia, A. Sibille, and M. Kamoun, "Reconfigurable intelligent surfaces: Bridging the gap between scattering and reflection," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2538–2547, 2020.

[35] W. Khawaja, O. Ozdemir, Y. Yapici, F. Erden, and I. Guvenc, “Coverage enhancement for NLOS mmwave links using passive reflectors,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 263–281, 2020.

[36] Ö. Özdogan, E. Björnson, and E. G. Larsson, “Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling,” *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, 2019.

[37] M. Najafi, V. Jamali, R. Schober, and H. V. Poor, “Physics-based modeling and scalable optimization of large intelligent reflecting surfaces,” *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2673–2691, 2021.

[38] D. Dardari, “Communicating with large intelligent surfaces: Fundamental limits and models,” *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2526–2537, 2020.

[39] G. Green, *An essay on the application of mathematical analysis to the theories of electricity and magnetism.* Wezäta-Melins Aktiebolag, Nottingham, 1828.

[40] E. F. Kuester, M. A. Mohamed, M. Piket-May, and C. L. Holloway, “Averaged transition conditions for electromagnetic fields at a metafilm,” *IEEE Trans. Antennas Propag.*, vol. 51, no. 10, pp. 2641–2651, Oct. 2003.

[41] C. L. Holloway, M. A. Mohamed, E. F. Kuester, and A. Dienstfrey, “Reflection and transmission properties of a metafilm: With an application to a controllable surface composed of resonant particles,” *IEEE Trans. Electromagn. Compat.*, vol. 47, no. 4, pp. 853–865, Nov. 2005.

[42] K. Achouri, M. A. Salem, and C. Caloz, “General metasurface synthesis based on susceptibility tensors,” *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 7, pp. 2977–2991, 2015.

[43] K. Achouri and C. Caloz, “Design, concepts, and applications of electromagnetic metasurfaces,” *De Gruyter Nanophotonics*, vol. 7, no. 6, pp. 1095–1116, Jun. 2018.

[44] Y. Vahabzadeh, N. Chamanara, K. Achouri, and C. Caloz, “Computational analysis of metasurfaces,” *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, vol. 3, pp. 37–49, 2018.

[45] M. Albooyeh, R. Alaee, C. Rockstuhl, and C. Simovski, “Revisiting substrate-induced bianisotropy in metasurfaces,” *Physical Review B*, vol. 91, no. 19, pp. 1–11, 195304, May 2015.

[46] F. H. Danufane, M. Di Renzo, J. de Rosny, and S. Tretyakov, “On the path-loss of reconfigurable intelligent surfaces: An approach based on the theory of diffraction,” 2020, (submitted to IEEE Transactions on Communications).

[47] D. J. Griffiths, “Introduction to electrodynamics,” 2005.

[48] S. J. Orfanidis, “Electromagnetic waves and antennas,” 2002.

[49] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, “Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond,” *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[50] W. Lu and M. Di Renzo, “Stochastic geometry modeling and system-level analysis & optimization of relay-aided downlink cellular networks,” *IEEE Transactions on Communications*, vol. 63, no. 11, pp. 4063–4085, 2015.

[51] Y. Vahabzadeh, N. Chamanara, K. Achouri, and C. Caloz, “Computational analysis of metasurfaces,” *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, vol. 3, pp. 37–49, 2018.

[52] E. Björnson, Ö. Özdogan, and E. G. Larsson, “Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?” *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2019.

[53] https://www.rcrwireless.com/20181205/5g/ntt-docomo-metawave-test-5g-mobile-system-tokyo, accessed: 2021-03-01.

[54] A. V. Osipov and S. A. Tretyakov, *Modern Electromagnetic Scattering Theory with Applications*. John Wiley & Sons, 2017.

[55] J. A. Stratton and L. Chu, “Diffraction theory of electromagnetic waves,” *Physical Review*, vol. 56, no. 1, p. 99, 1939.

[56] S. A. H. Gangaraj and F. Monticone, “Molding light with metasurfaces: from far-field to near-field interactions,” *Nanophotonics*, vol. 7, no. 6, pp. 1025–1040, 2018.

[57] Z. Wan, Z. Gao, F. Gao, M. Di Renzo, and M.-S. Alouini, “Terahertz massive mimo with holographic reconfigurable intelligent surfaces,” *IEEE Transactions on Communications*, 2021.

[58] N. M. Estakhri and A. Alu, “Wave-front transformation with gradient metasurfaces,” *Physical Review X*, vol. 6, no. 4, p. 041008, 2016.

[59] A. Diaz-Rubio and S. Tretyakov, “Macroscopic modeling of anomalously reflecting metasurfaces: Angular response and far-field scattering,” *arXiv preprint arXiv:2012.03727*, 2020.

[60] C.-T. Tai, “Kirchhoff theory: Scalar, vector, or dyadic?” *IEEE Transactions on Antennas and Propagation*, vol. 20, no. 1, pp. 114–115, 1972.

[61] R. N. Treuhaft, S. T. Lowe, and E. Cardellach, “Formulating a vector wave expression for polarimetric GNSS surface scattering,” *Progress In Electromagnetics Research*, vol. 33, pp. 257–276, 2011.

[62] C. A. Balanis, *Antenna Theory: Analysis and Design.* John wiley & sons, 2016.

[63] V. A. Borovikov, “Uniform stationary phase method,” *IEE Electromagnetic Waves Series*, vol. 40, 1994.

[64] W. Tang, X. Chen, M. Z. Chen, J. Y. Dai, Y. Han, M. Di Renzo, S. Jin, Q. Cheng, and T. J. Cui, “Path loss modeling and measurements for reconfigurable intelligent surfaces in the millimeter-wave frequency band,” *arXiv preprint arXiv:2101.08607*, 2021.

[65] R. Wong, *Asymptotic Approximations of Integrals.* SIAM, 2001.

[66] J. L. Lopez, E. Mainar, and N. M. Temme, “Asymptotic approximations of integrals: an introduction, with recent developments and applications to orthogonal polynomials,” *Electronic Transactions on Numerical Analysis*, vol. 19, pp. 58–83, 2005.

[67] Z.-Q. He and X. Yuan, “Cascaded channel estimation for large intelligent metasurface assisted massive mimo,” *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 210–214, 2019.

[68] E. Björnson and L. Sanguinetti, “Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces,” *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 830–834, 2021.

[69] M. A. Kishk and M. S. Alouini, “Exploiting randomly located blockages for large-scale deployment of intelligent surfaces,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1043–1056, 2021.

[70] M. Nemati, J. Park, and J. Choi, “RIS-assisted coverage enhancement in millimeter-wave cellular networks,” *IEEE Access*, vol. 8, pp. 188 171–188 185, 2020.

[71] T. Bai, R. Vaze, and R. W. Heath, “Analysis of blockage effects on urban cellular networks,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 5070–5083, 2014.

Synthèse en français

Les futurs réseaux sans fil devront prendre en charge des débits de données très élevés, car entre 2020 et 2030, il est prévu que le trafic de données du protocole Internet mondial (IP) augmente de 55% chaque année, atteignant à terme 5 016 exabytes. En outre, ils devraient offrir plusieurs autres services tels que le haut débit mobile amélioré (eMBB), les communications ultra-fiables et à faible latence (URLLC) et les communications massives de type machine (mMTC). Malheureusement, malgré les objectifs primordiaux que la 5G devrait atteindre, divers essais récents montrent que la technologie 5G actuelle n'est toujours pas en mesure de les atteindre tous. De plus, la plupart des tests sont encore loin d'utiliser un système complet, ce qui signifie que l'impact de l'interférence des autres cellules et les avantages des couches multiples sont largement ignorés. De plus, la plupart d'entre eux ne considèrent que les scénarios de liaison descendante (DL) mais pas les scénarios de liaison montante (UL). Afin de répondre à toutes les exigences de la 5G, qui n'ont pas encore été atteintes par les progrès actuels, une approche plausible consiste à repenser le paradigme actuel de conception du système de communication et à surmonter ses limites.

Au cours des cinq premières générations de réseaux sans fil, la conception des systèmes de communication sans fil a considéré les postulats selon lesquels l'environnement sans fil entre les appareils communicants (i) est contrôlé par nature, (ii) ne peut pas être modifié, (iii) ne peut être compensé que par la conception de systèmes d'émission et de réception sophistiqués. La sixième génération (6G) de réseaux mobile, quant à elle, est envisagée pour briser ces postulats en supposant que l'on puisse contrôler l'environnement sans fil pour obtenir des gains de performance majeurs. Par exemple, une station de base typique transmet des ondes radio dont l'amplitude est de l'ordre de grandeur du Watt tandis qu'un équipement utilisateur détecte des signaux de l'ordre de grandeur du μWatts. Le reste de l'énergie est, en général, perdu de différentes manières dans l'environnement, par exemple en générant des interférences avec d'autres éléments du réseau ou en créant des menaces pour la sécurité, car la propagation des ondes radio via le canal sans fil ne peut pas être contrôlée et modifiée une fois qu'elles sont émises par les émetteurs et avant qu'elles ne soient reçues par les récepteurs. Ce nouveau paradigme qui considère l'environnement comme un élément contrôlable dans la conception de systèmes sans fil est appelé environnement radio intelligent (SRE).

L'émergence de l'environnement radio intelligent (SRE) en tant que nouveau paradigme qui remet en question le statu quo des communications sans fil a encouragé l'utilisation de la

surface intelligente reconfigurable (RIS) basée sur des méta-surfaces, afin d’améliorer les limites de performance des systèmes de communication sans fil. L’objectif principal de cette thèse est la modélisation de systèmes de communication assistés par des surfaces intelligentes reconfigurables (RIS) à l’aide de méthodes électromagnétiques. Le Chapitre 1 présente le concept d’environnement radio intelligent. La définition d’une RIS et l‘utilisation d’une RIS dans le contexte d’un SRE sont également présentées. Le Chapitre 2 présente les concepts théoriques nécessaires tel que la modélisation des méta-surfaces et les approches analytiques qui nous permettent de calculer le champ l’électromagnétique dans un certain volume. Le Chapitre 3 fournit une comparaison des performances entre les RISs fonctionnant comme des réflecteurs anormaux et une méthode de relais basée sur le décodage et la transmission. Le Chapitre 4 propose une caractérisation analytique du champ électromagnétique en présence d’une RIS qui est modélisé comme une méta-surface homogène fonctionnant soit en réflexion ou transmission. Enfin, le Chapitre 5 résume la thèse et discute des perspectives futures qui méritent d’être étudiées pour mieux comprendre les avantages et les inconvénients des RISs pour application aux communications sans fil.