Spring 2022

Resonances and Fundamental Bounds in Wave Scattering

Hanwen Zhang
Yale University Graduate School of Arts and Sciences, hanwen.zhang@yale.edu

Abstract

Resonances and Fundamental Bounds in Wave Scattering

Hanwen Zhang

2022

This thesis presents three results in wave scattering, centered around scattering matrices and fundamental limits in scattering responses.

First, we develop a new resonance-based construction of scattering matrices in open electromagnetic systems. We use quasinormal modes to develop an exact, *ab initio* generalized coupled-mode theory from Maxwell's equations. This quasinormal coupled-mode theory, which we denote "QCMT", enables a direct, mode-based construction of scattering matrices without resorting to external solvers or data. We consider canonical scattering bodies, for which we show that a conventional coupled-mode theory model will necessarily be highly inaccurate, whereas QCMT exhibits near-perfect accuracy.

Second, for arbitrary scattering matrices, we obtain power-concentration bounds for wave scattering by generalizing classical brightness theorem to wave scattering. We show that power per scattering channel generalizes brightness, and the rank of an appropriate density matrix generalizes étendue to states with arbitrary coherence. The bounds apply to nonreciprocal systems that are of increasing interest, and we demonstrate their applicability to maximal control in nanophotonics for metasurfaces and waveguide junctions. Through inverse design, we discover metasurface elements operating near the theoretical limits.

Finally, adapting recently developed techniques for electromagnetic-response bounds to quantum dynamics, we develop a general framework for identifying fundamental bounds in quantum control. We show that an integral-equation-based formulation of conservation laws in quantum dynamics leads to fundamental limits for quantum control scenarios. We demonstrate the utility of our bounds in three prototype systems– three-level driving, decoherence suppression, and maximum-fidelity gate implementations – and show that our bounds are tight or nearly so in each case. Global bounds complement local-optimization-based designs, illuminating performance levels that may be possible and those that cannot be surpassed.

Resonances and Fundamental Bounds in Wave Scattering

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Hanwen Zhang

Dissertation Director: Owen D. Miller

May, 2022

To my parents

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank Owen Miller for his mentorship over the past five years. You have taught me numerous lessons about physics and research. You always encourage me to explore outside of physics and make me appreciate the importance of computations. I would also like to thank Vladimir Rokhlin and John Schotland for many inspiring lectures on mathematics and cheerful conversations. I also thank Douglas Stone for being on my committee and Peter Rakich for being my outside reader.

None of this would be possible without Berg Englert at NUS, who demonstrated to me the craftsmanship of theoretical physics and stimulated my interests in it.

I thank my labmates, Lang, Zeyu, Wenjin, Shai, Haejung, and Andrew, for your support, friendship, and many delightful chats.

To my friends, Xiansong, Yiqi, Junlin, Bingjun, and many others, thank you for your support along the way, especially Xiansong and Yiqi for many stimulating discussions about physics.

I am grateful to Wen Xiong, who introduced me to the internship at Facebook Reality Lab. I thank my colleagues there, Marc Aurele Gilles, Yang Yang, and Jianbo Zhao, for teaching me many things about optical design and imaging.

To my parents, Guiqing Zhang and Yifen Deng: words cannot express my gratitude. I would never have made it this far in my life without your guidance, unconditional love, and support. Thank you!

To Qinglian, thank you for everything. You turned two years of isolation during the pandemic into pleasant memories. I cannot imagine my life without you.

Chapter 1

Introduction

Nature has developed various nanoscale structures that possess unique optical properties. Lord Rayleigh in 1917 [9] pointed out that nanoscale structures create special reflection profiles via optical interference and cause iridescent colors of many birds, butterflies, and beetles. However, it is only until recently that engineering of artificial nanophotonic structures became possible due to the advancement of nanoscale fabrication techniques. That opens the door for nanophotonics, where light is manipulated by designing optical structures on a scale comparable to the wavelength. As a prominent example, by designing the spatial periodicity of the dielectric constant, photonic crystals can inhibit the propagation of certain light frequencies. It provides a way to guide light through sharp bends seems impossible in waveguides. In plasmonics, the interaction between light and metallic nanostructures can sustain the oscillation of free electron gas in metals. This creates highly localized resonances much smaller than the wavelength, and significant field enhancement in the near field that has been utilized for amplifying near field heat transfer [10, 11, 12], spontaneous emission [13, 14, 15], and Raman scattering [16, 17, 18]. More recently, many efforts have been made to create multifunctional metasurfaces by patterning a piece of material subwavelength in thickness. Metasurfaces [19, 20, 21] hold the promise to significantly reduce the size of traditional optical elements such as gratings and lenses.

A critical distinction between nanophotonics and ray optics is that the interference between waves is fully harnessed to create scattering responses much larger than those in geometric optics . Resonances formed by wave interference play a significant role in creating

large scattering responses. In nanophotonics, coupled-mode theory (CMT) exploits this fact and decomposes the scattering process into power-carrying channels interacting with resonances. However, CMT is perturbative in nature and can only be applied to cases of well-separated resonances with high quality factors. It has been a successful theory in explaining mode couplings in photonic crystals. However, with the growing interests in plasmonics and metasurfaces where resonances are often of low quality factors, CMT breaks down, and alternative theories for such modal analysis are needed.

Another challenge in nanophotonics is how to design optimal structures to achieve the desired scattering responses. Computational inverse design [5, 22, 23, 24] has become a popular strategy for finding local optimal designs through the rapid computation of gradients of design variables. It often produces highly complex designs with superior performance compared with intuition-based design methods. However, this local optimization approach does not provide physical insights on why certain designs are better, and this local method is often trapped by numerous low-quality local optima. An emerging need is a framework that identifies fundamental limits to what is possible for these design problems.

1.1 Overview of this thesis

This thesis aims to address important problems related to the modal analysis of wave scattering and fundamental bounds in nanophotonics design. In addition, we also apply a bound framework developed in nanophotonics to quantum control problems, obtaining the limit on which quantum states can be controlled. In Chapter 2, we provide an overview of the scattering framework of Maxwell equations, introducing important concepts such as definitions of passivity, reciprocity, scattering channels, and scattering matrices. The remainder of this thesis is divided into three parts, each addressing a separate question related to wave scattering. Chapters 3, 5, and 7 serve respectively as the background information for Chapters 4, 6, and 8, which contain original contributions of this thesis that can be found in Ref. [25, 26, 27].

For the first part, we use quasinormal modes to develop an exact coupled-mode theory for analyzing scattering responses, decomposing scattering responses into contributions from

each resonance. In Chapter 3, we introduce the coupled-mode point of view of scattering. In the weak-coupling limit, the coupled-mode theory (CMT) naturally emerges, and we summarize its essential ingredients. In the strong-coupling limit, CMT breaks down, and we review a conceptually related approach originated from nuclear physics that formally include resonant information. We then introduce quasinormal modes, how they are related to resonances, and how they can incorporate strong coupling in scattering processes.

In Chapter 4, we start the construction of an exact coupled-mode theory from quasinormal modes. We use the quasinormal mode information inside the scatterers only, making it possible to expand the scattering matrix in terms of each resonance. We derive exact equations analogous to conventional CMT ones that describe the interaction between scattering channels and resonances. We show that the breakdown of conventional CMT is primarily due to the presence of low quality factor modes. These modes contribute to a smooth varying nonresonant background that is absent in conventional CMT. We show that conventional CMT is recovered in the weak coupling and high-quality-factor limit from our exact theory. In addition, we demonstrate the equivalence of a few different quasinormal mode formulae by pole expansions of Green's functions.

For the second part, we generalize the brightness theorem of ray optics to nanophotonics. In Chapter 5, we introduce the classical étendue and brightness theorem in ray optics with a simple derivation from Hamilton mechanics. We take a phase-space approach to illustrate these two concepts: the étendue is the phase space volume of a ray bundle, and the brightness is related to the intensity of rays. We demonstrate how these concepts lead to concentration bounds in non-imaging optics and how they reveal difficulties when designing imaging systems.

In Chapter 6, we generalize the brightness and étendue concepts into wave scattering and prove that concentrating incoherent power is an impossible task. First, we introduce the density matrix framework to incorporate coherence information between scattering channels. We generalize the brightness to be the power each scattering channel carries and the wave étendue to be the rank of the relevant density matrix. We prove the brightness theorem for wave scattering: it is impossible to concentrate incoherent power. The argument only involves the passivity of the system, so this result can be applied to lossy, nonreciprocal

systems. We derive analogous channel concentration bounds as in ray optics and inverse designed grating structures that approach these bounds. As a ramification of the concentration bound, we demonstrate the concept of wave étendue governs the robustness of power transports. Combining coherent power from multiple channels is inherently unstable and susceptible to incoherence noise.

For the third part, we apply a local-conservation-law-based framework developed in nanophotonics for identifying fundamental bounds to quantum control problems. In Chapter 7, we introduce the optimization frameworks in nanophotonics as finding optimal scatterer distributions in space. Then we summarize inverse design as a local optimization approach to nanophotonics design. We also provide the derivation of a framework for identifying global bounds originally developed in nanophotonics. Since quantum control can be seen as designing scattering responses in time, we naturally translate this framework and related concepts from nanophotonics to quantum control in Chapter 8. Although this nanophotonic framework is not the contribution of this thesis, we include it here because many concepts are more naturally discussed in the nanophotonics setting.

In Chapter 8, we repeat the derivation of the bound framework in the quantum control setting, turning the pulse design problem into a quadratically constraint quadratic program. We then relax it into a convex problem, which can be solved efficiently for global bounds to original design problems. We apply this framework to find minimum transition time, maximum decoherence suppression, and maximum gate fidelity. Compared with designed driving pulses, we find the bounds identified are almost tight and thus significantly improve results on quantum speed limits from previous works. Our framework possesses the unique advantage that constraints of practical importance, such as leakage prevention, can be easily incorporated.

Chapter 2

Preliminaries

In this chapter, we introduce basic equations and notions that will be useful in this thesis. Since most of this work is related to the scattering of electromagnetic waves, we first introduce Maxwell's equation and essential elements of scattering theory. Then we introduce scattering channels as the basis for far-field radiations that carry power away from scatterers. This enables us to represent scattering problems by scattering matrices. Scattering matrices represent both a conceptual and notational simplification when analyzing scattering problems and will be of central importance from Chapter 3 to Chapter 6. It is also implicitly related to Chapter 8, where many concepts in electromagnetic scattering are the same as those in driven quantum systems, and the propagator of a quantum system plays the same role in time as scattering matrices do in space.

2.1 Maxwell equations

We start by introducing time-harmonic Maxwell equations at frequency ω in SI convention with dimensionless units ($\epsilon_0 = \mu_0 = c = 1$):

$$\begin{aligned} \boldsymbol{\nabla} \cdot \boldsymbol{D}(\boldsymbol{r},\omega) &= \rho_e(\boldsymbol{r},\omega)\,, & \boldsymbol{\nabla} \times \boldsymbol{E}(\boldsymbol{r},\omega) &= \mathrm{i}\omega \boldsymbol{B}(\boldsymbol{r},\omega) - \boldsymbol{J}_m(\boldsymbol{r},\omega)\,, \\ \boldsymbol{\nabla} \cdot \boldsymbol{B}(\boldsymbol{r},\omega) &= \rho_m(\boldsymbol{r},\omega)\,, & \boldsymbol{\nabla} \times \boldsymbol{H}(\boldsymbol{r},\omega) &= -\mathrm{i}\omega \boldsymbol{D}(\boldsymbol{r},\omega) + \boldsymbol{J}_e(\boldsymbol{r},\omega)\,. \end{aligned} \tag{2.1}$$

In the above equations, $\boldsymbol{E}, \boldsymbol{D}$ and $\boldsymbol{H}, \boldsymbol{B}$ are respectively the electric and magnetic fields, ρ_e, ρ_m and $\boldsymbol{J}_e, \boldsymbol{J}_m$ are respectively the electric and magnetic charge densities, electric and magnetic current densities The magnetic charge and current are included here for mathematical convenience. When we include continuity equations for both electric and magnetic charges and currents $\boldsymbol{\nabla} \cdot \boldsymbol{J}_{e/m} = \mathrm{i}\omega\rho_{e/m}$, the two divergence equations on the left side of Eq. (2.1) are automatically satisfied by the two curl equations on the right side.

The constitutive relation between $\boldsymbol{D}, \boldsymbol{B}$ and $\boldsymbol{E}, \boldsymbol{H}$ works as a phenomenological theory to incorporate the macroscopic interaction between materials and electromagnetic waves. Throughout this work, we are interested light-matter interaction in the linear and time-invariant regime, so the most general constitutive relations are given by

$$\begin{aligned} \boldsymbol{D}(\boldsymbol{r},\omega) &= \epsilon(\boldsymbol{r},\omega)\boldsymbol{E}(\boldsymbol{r},\omega) + \zeta(\boldsymbol{r},\omega)\boldsymbol{H}(\boldsymbol{r},\omega)\,, \\ \boldsymbol{B}(\boldsymbol{r},\omega) &= \eta(\boldsymbol{r},\omega)\boldsymbol{E}(\boldsymbol{r},\omega) + \mu(\boldsymbol{r},\omega)\boldsymbol{H}(\boldsymbol{r},\omega)\,, \end{aligned} \tag{2.2}$$

where $\epsilon, \mu, \zeta, \eta$ are all 3 by 3 matrix variables that specify the nature of materials interacting with electromagnetic fields $\boldsymbol{E}, \boldsymbol{H}$.

Combining the two curl equation in Eq. (2.1) and the constitutive relation in Eq. (2.2), we can write the equation between $\boldsymbol{E}, \boldsymbol{H}$ and material quantities in compact six-vector notation:

$$\underbrace{\begin{pmatrix} & -\boldsymbol{\nabla}\times \\ -\boldsymbol{\nabla}\times & \end{pmatrix}}_{=\Theta} \underbrace{\begin{pmatrix} \boldsymbol{E} \\ \boldsymbol{H} \end{pmatrix}}_{=\psi} - \mathrm{i}\omega \underbrace{\begin{pmatrix} \epsilon & \zeta \\ -\eta & -\mu \end{pmatrix}}_{=B} \begin{pmatrix} \boldsymbol{E} \\ \boldsymbol{H} \end{pmatrix} = -\underbrace{\begin{pmatrix} \boldsymbol{J}_e \\ -\boldsymbol{J}_m \end{pmatrix}}_{=\xi}. \tag{2.3}$$

The seemingly strange sign convention in Eq. (2.3) is to ensure that the operator $\Theta - \mathrm{i}\omega B$ is complex symmetric when $B^T = B$. The reason for such convention will be clear when we clarify a few important cases of B. In the following, we may often hide the position $\boldsymbol{r}$ and frequency ω dependence whenever there is no confusion.

2.2 Important classes of material B

2.2.1 Passivity

We call a system *passive* if it cannot generate energy by itself, so the output power will never be greater than the input one. By a simple calculation with the Poynting flux and the fact the energy within a closed surface cannot increase, we obtain the relation

$$\operatorname{Im}\begin{pmatrix}\epsilon(\omega) & \zeta(\omega)\\ \eta(\omega) & \mu(\omega)\end{pmatrix} \geq 0\,, \quad \text{for real } \omega. \tag{2.4}$$

Note that Im of a matrix A is given by $\frac{1}{2\mathrm{i}}(A - A^\dagger)$ and $\dagger$ is the Hermitian conjugation, and $A \geq 0$ with matrix variables means the matrix is positive semidefinite. In particular, when ω is real, the material is lossless if the equality holds, which means $\begin{pmatrix}\epsilon(\omega) & \zeta(\omega)\\ \eta(\omega) & \mu(\omega)\end{pmatrix}$ is Hermitian.

2.2.2 Reciprocity

We call a system *reciprocal* if there is symmetry between the detection of fields and the source that generates them. This is true when B is complex symmetric:

$$B = \begin{pmatrix}\epsilon & \zeta\\ -\eta & -\mu\end{pmatrix} = \begin{pmatrix}\epsilon & \zeta\\ -\eta & -\mu\end{pmatrix}^T = B^T\,, \tag{2.5}$$

where T is the unconjugated transpose of B. This ensures that $\Theta - \mathrm{i}\omega B$ is complex symmetric because the sign convention of Θ makes it already complex symmetric. This ultimately results in the system Green's function being complex symmetric $\Gamma^T(\boldsymbol{r}, \boldsymbol{r}') = \Gamma(\boldsymbol{r}', \boldsymbol{r})$, since its Green's function can be formally understood as the inverse of $\Gamma = (\Theta - \mathrm{i}\omega B)^{-1}$.

In most practical situations, we only need an isotropic permittivity tensor ϵ, a constant multiple of a 3 by 3 identity matrix. When proving general results, we only rely on passivity and in some instance reciprocity, so the compact notation in Eq. (2.3) allows us to include general material tensors at no extra cost.

2.3 Scattering and volume integral equations

We first consider an incident wave ψ_{inc} in the presence of background $B_{\text{bg}}(\boldsymbol{r})$ and assume we have complete knowledge of the system under such incident excitation. Now we introduce inhomogeneity $\Delta B(\boldsymbol{r})$ in addition to $B_{\text{bg}}(\boldsymbol{r})$, so that $B(\boldsymbol{r}) = B_{\text{bg}}(\boldsymbol{r}) + \Delta B(\boldsymbol{r})$. We refer to the inhomogeneity as a collection of scattering bodies or scatterers. The scatterers are polarized by the existing field ψ_{inc}, producing a scattered field ψ_{scat} generated by the polarization current. Naturally, ψ_{scat} satisfies the outgoing radiation boundary condition.

The above description can be captured by the integral representation of the solution $\psi(\boldsymbol{r})$ to Eq. (2.3) under ψ_{inc} excitation:

$$\psi(\boldsymbol{r}) = \psi_{\text{inc}}(\boldsymbol{r}) + \underbrace{\mathrm{i}\omega \int \Gamma_{\text{bg}}(\boldsymbol{r}, \boldsymbol{r}')\Delta B(\boldsymbol{r}')\psi(\boldsymbol{r}')}_{=\psi_{\text{scat}}(\boldsymbol{r})}, \tag{2.6}$$

where Γ_{bg} is the background Green's function without any scatters ($B = B_{\text{bg}}$):

$$(\Theta - \mathrm{i}\omega B_{\text{bg}})\Gamma_{\text{bg}}(\boldsymbol{r}, \boldsymbol{r}') = \delta(\boldsymbol{r} - \boldsymbol{r}')I\,. \tag{2.7}$$

where I is the 6 by 6 identity matrix. We note that ψ_{inc} satisfies $(\Theta - \mathrm{i}\omega B_{\text{bg}})\psi_{\text{inc}} = 0$, subject to appropriate boundary conditions. The total field ψ consists of the incident field ψ_{inc} and the scattered field ψ_{scat}.

2.4 Scattering channels

Although ψ_{inc} and ψ_{scat} include all the information of a scattering problem, we often only care about a specific set of incoming and outgoing waves. The details near or inside scatterers may be irrelevant. For example, scattering can happen inside junctions of waveguides. Far away from the scatterers, only guided modes can carry energy in and out of the junction. For the scattering of a periodic structure, such as gratings or photonic crystal slabs, only the coupling between diffraction modes (Bloch modes) is of interest. For scattering in vacuum with localized scatterers, we may primarily be interested in certain angular distributions of far-field patterns. In general, there is always a physically important basis set in the far

field. We naturally want to use the basis to represent our scattered fields.

We can enclose the scatterer region with a large enough bounding surface Σ, on which we define our basis functions for incoming and outgoing waves as illustrated in Fig. 4.1. We call the relevant basis functions *scattering channels*. The surface normal of the bounding surface is parallel to the propagation direction. Depending on the propagation direction, we can define the in and out channel bases

$$\Phi_{\text{out}} = \begin{pmatrix} \boldsymbol{E}_{\text{out},1} & \boldsymbol{E}_{\text{out},2} & \dots \\ \boldsymbol{H}_{\text{out},1} & \boldsymbol{H}_{\text{out},2} & \dots \end{pmatrix}, \Phi_{\text{in}} = \begin{pmatrix} \boldsymbol{E}_{\text{in},1} & \boldsymbol{E}_{\text{in},2} & \dots \\ \boldsymbol{H}_{\text{in},1} & \boldsymbol{H}_{\text{in},2} & \dots \end{pmatrix}. \tag{2.8}$$

One feature of the aforementioned examples of scattering channels is that they all have constant wave front on the surface whose normal is parallel to the wave propagation. For a waveguide invariant in z direction, the modes always have $\mathrm{e}^{\pm \mathrm{i} k_z z}$ dependence. Bloch modes with scatterers periodic in x and y direction are very similar to the waveguide case, except their waveform in the x, y is also related to the scatterer's periodicity. The natural bounding surfaces are planes across which the modes propagate for these two cases. For vacuum scattering with localized scatterers, the natural bounding surface is an enclosing sphere, so the basis can be chosen propagating in the radial direction with different angular patterns. Such a choice leads to vector spherical waves (VSWs) with asymptotic radial dependence $\frac{\mathrm{e}^{\pm \mathrm{i} k r}}{r}$.

2.4.1 Power orthonormal basis

An important consequence of channels' translational or spherical symmetry is that a suitable inner product can be defined on the bounding surface so that the scattering channel basis can be chosen as orthonormal in terms of power-flow-based inner product. Due to the particular functional form of the waveguide and Bloch channel basis, the $\boldsymbol{\nabla}\times$ operator in the splits into two parts

$$\boldsymbol{\nabla}\times = \boldsymbol{\nabla}_{\parallel}\times + \mathrm{i}\boldsymbol{k}_{\perp}\times, \tag{2.9}$$

where $\boldsymbol{k}_{\perp} = k_{\perp}\boldsymbol{n}$ and $\boldsymbol{n}$ is the surface normal of the bounding surface and $k_{\perp}$ is the wave vector magnitude in the propagation direction. Such separation indicates that these channel

functions are eigenfunctions to a generalized Hermitian eigenproblem with "weight" defined by matrix $Q = \begin{pmatrix} & -\boldsymbol{n}\times \\ \boldsymbol{n}\times & \end{pmatrix}$ with eigenvalue $k_\perp$. We can define a power inner product with Q between these channels basis as

$$\int_\Sigma \psi^\dagger \frac{Q}{4} \psi = \frac{1}{2} \operatorname{Re} \int_\Sigma \mathrm{d}S \boldsymbol{n} \cdot \boldsymbol{E}^* \times \boldsymbol{H} \,, \tag{2.10}$$

which is the integration of the Poynting vector over the bounding surface. The Hermitian nature of channel function eigenproblems immediately shows different channel functions are power orthogonal among incoming and outgoing sets. In addition, incoming and outgoing channels are automatically power orthogonal since their $k_\perp$ are in opposite signs. With spherical symmetry and direction calculation, this orthogonality relation also holds for VSWs. In other words, we can choose the channels basis to be orthonormal:

$$\begin{aligned} \langle \Phi_{\text{out}}, \Phi_{\text{out}} \rangle &= \int_\Sigma \Phi_{\text{out}}^\dagger \frac{Q}{4} \Phi_{\text{out}} \, \mathrm{d}S = I, \\ \langle \Phi_{\text{in}}, \Phi_{\text{in}} \rangle &= \int_\Sigma \Phi_{\text{in}}^\dagger \frac{Q}{4} \Phi_{\text{in}} \, \mathrm{d}S = -I \\ \langle \Phi_{\text{in}}, \Phi_{\text{out}} \rangle &= \int_\Sigma \Phi_{\text{in}}^\dagger \frac{Q}{4} \Phi_{\text{out}} = 0 \,, \end{aligned} \tag{2.11}$$

where I represents the identity matrix. In particular, the channels mentioned, including VSWs, Bloch, and waveguide modes, all satisfy the above relations.

2.4.2 Representation of in and out fields

A finite number of basis functions can give arbitrarily good accuracy because the integral operator in Eq. (2.6) is compact [28]. Physically speaking, this means the total interaction strength of a finite scatterer is also finite. Consequently, we can simply represent the incoming and outgoing fields with the help of column vectors $\boldsymbol{c}_{\text{in}}$ and $\boldsymbol{c}_{\text{out}}$ containing the coefficients of expansions incoming and outgoing fields:

$$\psi_{\text{in}} = \Phi_{\text{in}} \boldsymbol{c}_{\text{in}} \,, \quad \psi_{\text{out}} = \Phi_{\text{out}} \boldsymbol{c}_{\text{out}} \,. \tag{2.12}$$

A special and sometimes convenient choice is to make Φ_{out} the time-reversal of Φ_{in}.

Incoming and outgoing fields carry energy into and out of the surface Σ, so their spatial dependence contains singular points as sinks or sources. However, the incident wave is always a traveling wave, so the power integral for incident fields over a closed surface is always zero, which must be a combination of incoming and outgoing ones so that the singularities are canceled. After all, it is a solution to the homogeneous background Maxwell equations. As a result, the incident channel functions can always be written as

$$\Phi_{\text{inc}} = \alpha\Phi_{\text{in}} + \beta\Phi_{\text{out}}, \tag{2.13}$$

where α and β are constants. Up to arbitrary phase factors, for vector spherical waves, $\alpha = \beta = \frac{1}{2}$; for plane waves, $\alpha = \beta = 1$. (See Appendix A for details of the plane wave basis definition.) Any given incident wave can be represented in the incoming/outgoing-wave basis by the expression

$$\psi_{\text{inc}} = \Phi_{\text{inc}}\boldsymbol{c}_{\text{inc}} = \Phi_{\text{in}}\boldsymbol{c}_{\text{in}} + \Phi_{\text{out}}S_{\text{bg}}\boldsymbol{c}_{\text{in}}. \tag{2.14}$$

By matching the vectors multiplying Φ_{in} in Eqs. (2.13,2.14), we can see that $\boldsymbol{c}_{\text{inc}} = \frac{1}{\alpha}\boldsymbol{c}_{\text{in}}$. Here S_{bg} is a matrix that connects $\boldsymbol{c}_{\text{in}}$ and $\boldsymbol{c}_{\text{out}}$ without scatterers and is simply identity times β by Eq. (2.13) when scattering background is vacuum. We call S_{bg} the background scattering matrix. In the next section, we will introduce the concept of the scattering matrix in detail.

2.5 Scattering matrix representation

Since scattering problems we are considering here are linear, $\boldsymbol{c}_{\text{out}}$ must depend linearly on $\boldsymbol{c}_{\text{in}}$. We can represent this linear dependence by a matrix S, called the *scattering matrix*,

$$\boldsymbol{c}_{\text{out}} = S\boldsymbol{c}_{\text{in}}\,. \tag{2.15}$$

The scattering matrix S not only encapsulates the solution to the scattering problem for any relevant incident fields but also shows how different channels are coupled. Because of Eq. (2.11), the incoming/outgoing power into/out of the bounding surface Σ are given by $\boldsymbol{c}_{\text{in}}^{\dagger}\boldsymbol{c}_{\text{in}}$ and $\boldsymbol{c}_{\text{out}}^{\dagger}\boldsymbol{c}_{\text{out}}$ respectively.

In addition, properties and symmetries in the original problem are reflected as simple algebraic relations of S. If the scattering process is passive as in Sec. 2.2.1, we must have power out $\boldsymbol{c}_{\text{out}}^{\dagger}\boldsymbol{c}_{\text{out}} = \boldsymbol{c}_{\text{in}}^{\dagger}S^{\dagger}S\boldsymbol{c}_{\text{in}}$ no greater than power in $\boldsymbol{c}_{\text{in}}^{\dagger}\boldsymbol{c}_{\text{in}}$ for any $\boldsymbol{c}_{\text{in}}$, so we conclude passivity leads to the positive semidefiniteness of $I - S^{\dagger}S$ with I being the identity matrix:

$$S^{\dagger}S \leq I\,. \tag{2.16}$$

A useful restatement of the above relation is that the eigenvalues of $S^{\dagger}S$ are no greater than 1. If the scattering problem is reciprocal as in Sec. 2.2.2, the scattering matrix is complex symmetric, a property already reflected in Green's functions:

$$S^{T} = S\,, \tag{2.17}$$

when the incoming and outgoing channels are chosen to be time-reversal pairs. (A proof of this fact is a byproduct of results in Chapter 4.) The complex symmetric property makes sense physically since we can think of obtaining $\boldsymbol{c}_{\text{out}}$ as a detection process from source $\boldsymbol{c}_{\text{in}}$. S^{T} represents the switch between sources and detections, and reciprocity states the switch is physically equivalent.

2.6 Summary

In this chapter, we introduced basic equations and the concept of passive and reciprocal systems. We defined scattering channels as the natural basis for representing scattering far fields. This leads to the scattering matrix framework, where passivity and reciprocity are encoded as algebraic properties of scattering matrices. In Chapter 3 and 4, we will view scattering as a mode coupling process and represent S in terms of resonant modes. The scattering matrix gives a general and powerful framework to represent scattering problems

in general; it is the critical tool in generalizing the classical brightness theorem to wave scattering in Chapter 6. In Chapter 8, we switch from Maxwell equations to the Schrödinger equation, in which propagators U are conceptually equivalent to the scattering matrices we introduced here.

Chapter 3

Resonant Scattering

The presence of resonances is the singular cause of large scattering responses. For physical intuitions into scatterings, it is highly desirable to decompose a scattering matrix S into contributions of individual resonances. This chapter surveys resonant scattering from the coupled mode point of view. In Sec. 3.1, we first partition between a scatterer and a channel region. In the weak-coupling limit, the interaction between the two can be treated perturbatively, resulting in a simple yet powerful framework widely used in nanophotonics for representing the scattering matrix S by resonances. Beyond the weak coupling limit, this perturbative approach fails and one approach that originated in nuclear scattering is to invert the channel block directly to deal with the exact interaction. This process produces a representation of S by unperturbed modes in the scatterer region; this can be seen as an indirect way of incorporating resonant information, so it does not lead to a true extension of the coupled-mode theory. Then we show another way of achieving the same goal is to start from the resonances directly, formalized by the introduction of quasinormal modes. We introduce these modes and summarize important properties in Sec. 3.2 and pave the way for the construction of S , which comes in Chapter 4.

3.1 Coupled mode point of view

The scattering response is often noticeably larger near certain frequencies because of a build-up of energy inside a scatterer formed by interference between waves, creating rela-

tively long-lived resonant modes inside. Then the resonant modes leak out of the scatterer, producing large scattering responses. This two-fold process is called resonant scattering, and the frequency dependence of the extinction cross section σ_{ext} often takes the form of a Lorentzian peak

$$\sigma_{\text{ext}}(\omega) \sim \frac{1}{(\omega - \omega_0)^2 + \Gamma_0^2}\,, \tag{3.1}$$

where ω_0 is the resonance frequency with Γ_0 represents the half-width of the peak. A useful dimensionless parameter is the quality factor of the resonance given by $Q = \frac{\omega_0}{2\Gamma_0}$. The parameter Γ is the inverse lifetime of the resonance, as it dictates the time required for energy to be lost by the fraction $1/e$. The higher the quality factor, the sharper the peak.

Since resonant scattering can be understood as resonant modes forming inside scatterers under external excitation, it is advantageous to view such processes as external scattering channels being coupled to resonant modes in scatterers. As a result, one can partition the Maxwell operator into two spatially separated regions. We use "C" to denote the part where channel modes are defined, and "S" to denote the other where the scatterer modes are defined. After partitioning the operator into the scattering "S" and the channel "C" parts, an appropriate ordering of spatial variables bring the Maxwell operator into the following block form

$$\Theta = \begin{pmatrix} \Theta_S & \Theta_{SC} \\ \Theta_{CS} & \Theta_C \end{pmatrix}\,, \tag{3.2}$$

where Θ_{SC} and Θ_{CS} couple the Θ_S and Θ_C parts. We carry out the same partitioning procedure for B, so Eq. (2.3) is divided into the "C" and "S" blocks with couplings. Scatterer modes and channel modes can be respectively defined as the generalized eigenfunctions of Θ_S and Θ_C:

$$\Theta_S \psi_S = \mathrm{i}\omega_S B_S \psi_S, \quad \Theta_C \psi_C = \mathrm{i}\omega_C B_C \psi_C \tag{3.3}$$

The channel defined here is technically different from those defined on the bounding surface in Sec. 2.4, but they are conceptually the same. The definition here coincides with the one for the waveguide resonator interaction setup but not for scattering in free space. It is because, in the waveguide and resonant cavity setup, channel and resonant modes do live in disjoint spatial regions, but the separation becomes artificial for free-space scattering since

channels such as VSWs are defined everywhere in space. Despite this difference, these two definitions are conceptually the same and thus can help us understand a few coupled-mode frameworks.

We note here that this partition approach originates from nuclear physics [29, 30, 31, 32]. In nuclear scattering, it is also reasonable to consider the states in the nuclei coupled to a continuum scattering states (channels). This framework is usually referred to as the Feshbach projection.

In the following, we discuss how this partitioning leads to a formal representation of the scattering matrix.

3.1.1 Weak coupling

When the coupling between scatterer and channel modes is weak, indicated by properties of Θ_{SC} and Θ_{CS}, we can treat the coupling via perturbations. The decay rate of individual resonances comes from the change of eigenvalues ω_S caused by the overlap between ψ_S and ψ_C. Those scatterer modes are essentially *closed* modes with weak coupling to external channels. These closed modes can be normalized in terms of the stored energy by Hermitian inner products. Since the coupling is assumed to be weak, we will see in Sec. 3.1.2 that the resulting coupling constants (related to the decay rate) are also small required by energy conservation. This shows that high Q feature is an equivalent way of stating weak coupling between modes and channels.

The intuition discussed above forms the basis of a simple yet powerful framework: coupled-mode theory (CMT). CMT attempts to describe all scattering processes purely in terms of high Q resonances interacting with scattering channels and uses physical arguments such as passivity and reciprocity to reduce the number of undetermined parameters.

3.1.2 Coupled mode theory

In the weak-coupling limit, scattering response is dominated by sharp Lorentzian peaks associated with high Q resonances. When these resonances are well separated in frequency, it is natural to describe the scattering problem purely in terms of a phenomenological theory, called the coupled-mode theory (CMT). We simply adopt the coupled-mode point view, re-

placing the scatterer with a set of resonant modes represented by a vector $\boldsymbol{a}$, whose elements represents their modal amplitudes. Coupled mode theory has been applied extensively to resonant phenomena across the nanophotonics landscape [33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44]. There are four key matrices that appear in coupled-mode theory. Their resonant frequencies are denoted by ω_n with decay rate Γ_n, which are summarized into a diagonal matrix Ω with $\Omega_{ii} = \omega_i - \mathrm{i}\Gamma_i$. These resonant modes are coupled to incoming and outgoing channels represented by vectors $\boldsymbol{c}_{\text{in}}$ and $\boldsymbol{c}_{\text{out}}$ respectively. Their coupling strengths are elements in matrices K and D. Finally, without any scatterers, there is a "direct" scattering process of incoming waves to outgoing waves captured by the matrix S_{bg}. Often this matrix is denoted as a "C" matrix [45], but we use S_{bg} to emphasize it is the background scattering matrix. These relations are summarized in the CMT equations, which at frequency ω ($e^{-i\omega t}$ harmonic time dependence) read [45]:

$$\mathrm{i}(\Omega - \omega)\boldsymbol{a} = D^T \boldsymbol{c}_{\text{in}}\,, \tag{3.4}$$

$$\boldsymbol{c}_{\text{out}} = S_{\text{bg}}\boldsymbol{c}_{\text{in}} + K\boldsymbol{a}\,, \tag{3.5}$$

where the first one describes how the incoming waves are coupled to resonances and the second relates the outgoing wave to the resonance amplitude $\boldsymbol{a}$. By reciprocity, passivity, and time-reversal arguments, all parameters are constrained by the following relations:

$$K = D \quad \text{(reciprocity)}\,, \tag{3.6}$$

$$K^\dagger K = -2\,\mathrm{Im}\,\Omega \quad \text{(passivity)}\,, \tag{3.7}$$

$$S_{\text{bg}} D^* = -D \quad \text{(time-reversal invariance)}\,, \tag{3.8}$$

where D^* is the elementwise complex conjugate of D. By solving $\boldsymbol{a}$ in Eq. (3.4) and substituting it into Eq. (3.5), we obtain the scattering matrix S as

$$S = S_{\text{bg}} - \mathrm{i}K(\Omega - \omega)^{-1} D^T\,, \tag{3.9}$$

which is a generalization of the response from as single Lorentzian peak in Eq. (3.1).

As an artifact of CMT's phenomenological origin, the second constraint implicitly determines the number of channels that can be coupled to the number of resonances since the rank of $K^\dagger K$ must equal the rank of Ω. This constraint also shows that weak coupling (small coupling constants in K) is equivalent to the presence of high Q resonances. We remark that another phenomenological feature of this theory is that energy losses are accounted for by adding one physically inaccessible channel. That is why the overall theory always appears to have time-reversal symmetry, leading to the third constraint above.

3.1.3 Strong coupling

When the coupling is strong, the perturbative approach fails, so does CMT as well. This is expected because when the coupling represented by the off-diagonal blocks is large, the new modes formed via coupling can be entirely different from their original counterparts. In this case, a simple perturbative treatment certainly cannot capture this dramatic change.

Although CMT has been a successful theory for applications ranging from waveguide filters [46] to photovoltaic absorbers [42] to transparent displays [47] and enables predictions of high-performance designs and fundamental limits, such predictions are only valid within the limits of CMT itself, which requires a key assumption: weak coupling between high-quality-factor (Q) resonances [33]. This assumption is violated in many scenarios of emerging interest (including photovoltaic absorbers). For example, plasmonic structures [48, 49] have significant material losses, large-area metasurfaces [50, 51, 52] comprise large numbers of low-Q, highly coupled resonances, and random media that are encountered in wavefront-shaping applications [53] have significant nonresonant contributions to their response. There is a need for a CMT-like framework without the assumption of weakly-coupled high-Q resonances.

In the nuclear physics approach, to deal with strong coupling in the partitioned representation, the Θ_S block is inverted in the process of solving the scattering problem, retaining the coupled-mode view while accounting for the off-diagonal coupling exactly. This leads to a formal construction of scattering matrices in terms of the *unperturbed* scatterer modes with the coupling modification hidden in the self-energy term. As a result, the resonance information is not used directly, so it is not the resonant expansion that we seek. However,

this representation is still useful for formal manipulations and leads to few-mode expansions useful for nuclear scattering [32] and optics [54, 55].

Next, we introduce how to incorporate strong coupling by directly using resonant modes, also known as quasinormal modes, unlike in the partitioned approach above.

3.2 Quasinormal mode

Another way of incorporating the strong coupling is to start with the exact resonant modes, also known as the quasinormal modes (QNMs), defined via the non-Hermitian eigenvalue problem of the Maxwell operator in Eq. (2.3) with radiation boundary conditions [56]:

$$\Theta\psi_m = \mathrm{i}\widetilde{\omega}_m B(\widetilde{\omega}_m)\psi_m \quad \text{with outgoing radiation boundary conditions}\,, \tag{3.10}$$

where the whole Maxwell operator is diagonalized and these modes already contain the exact coupling between scatterer modes with channel modes. Whereas *normal* modes are solutions of Hermitian eigenproblems and are guaranteed to form a complete basis for the fields in a scatterer, the solutions of a non-Hermitian eigenproblem do not necessarily form a complete basis, and these singularities are the "exceptional points" that have garnered tremendous recent interest [57, 58, 59, 60]. Yet these are, necessarily, singular points in the space of all possible field and material distributions, and at any generic perturbation away from an exceptional point, the quasinormal modes will comprise a complete basis [59]. In fact, QNMs are likely to be over-complete [61, 62]—only a subset are likely necessary to form the expansion basis—and this leads to multiple equivalent expansion formulae and various sum rule relations [63].

3.2.1 Normalization issue

Because of the radiation boundary condition, the eigenproblem for finding QNM is open, and thus non-Hermitian, whose eigenvalues are complex with a negative imaginary part. This negative imaginary part indicates that these modes generate energy that propagates to infinity. Although such behavior is the signature of resonances, it causes normalization

difficulties mathematically; the radiation condition requires the far field to take the form $e^{i\omega r}$ with a negative $\mathrm{Im}\,\omega = -\Gamma$, so we have

$$\left|e^{i\omega r}\right| = e^{\Gamma r}\,. \tag{3.11}$$

This shows the far field amplitude diverges exponentially for large r as illustrated in Fig. 3.1(a). Any conventional normalization methods for closed systems are going to fail when dealing with QNMs.

The divergence, however, does not pose any physical inconsistency and we can understand why it happens. Our intuition about resonances comes from the time domain analysis, regarding them as transient modes inside the resonator. However, the definition in Eq. (3.10) is in the frequency domain. The origin of the divergence comes from us taking a stationary view of such transient phenomenon. Consequently, there is no physical inconsistency and the divergence can be saved.

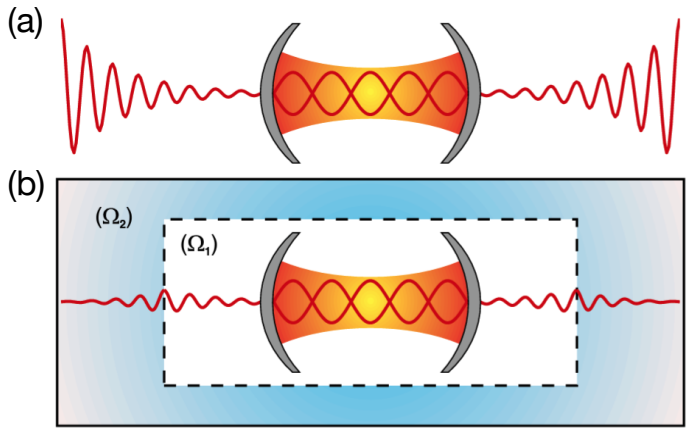

Figure 3.1: (a) Illustration of a quasinormal mode diverging outside of the resonant cavity. (b) By introducing perfect absorbers in the PML region Ω_2, the previous divergent mode in (a) decays. Then the unconjugated product in Eq. (3.17) can be used to normalize the qusinormal mode. The normalization is independent of parameters of PML implementations. (Reproduced from Ref. [1].)

The normalization of QNMs is still an active field of research. Here we adopt the perfectly matched layer (PML) based normalization method together with an unconjugated inner product as in Ref. [1, 56]. As shown in Fig. 3.1(b), the perfect absorbering material in the PML region squashes the divergence outside the cavity/scatterer. This method can be regarded as a generalization of the complex coordinate stretching method for 1D

systems [61]. Over different normalization methods, our choice here is convenient since it utilizes the complex symmetric nature of the operator in Eq. (3.10), so this normalization method is a manifestation of the complex symmetric properties of Θ. The open radiation boundary condition destroys any hope for a Hermitian problem, so we can only rely on reciprocity for general results. When the material is not reciprocal ($B \neq B^T$), we can use generalized reciprocity and two sets of basis, one for B and one for B^T.

In the following section, we will summarize results about the orthogonality and normalization of QNMs based on unconjugated inner products. We will not assume B is reciprocal, and use generalized reciprocity instead. With a slight burden on the notation, the results will form a coupled-mode theory for nonreciprocal systems. This sheds light on how nonreciprocity will change the coupling between channels and resonances.

3.2.2 Unconjugated orthogonality and normalization

Since we do not assume reciprocity, for any eigenfrequency $\widetilde{\omega}_m$, indexed by m, we have to distinguish between right quasinormal modes $\psi_{\mathrm{R},m}$ and left quasinormal modes $\psi_{\mathrm{L},m}$:

$$\Theta\psi_{\mathrm{R},m} = \mathrm{i}\widetilde{\omega}_m B(\widetilde{\omega}_m)\psi_{\mathrm{R},m}\,, \tag{3.12}$$

$$\Theta\psi_{\mathrm{L},n} = \mathrm{i}\widetilde{\omega}_n B^T(\widetilde{\omega}_n)\psi_{\mathrm{L},n}\,. \tag{3.13}$$

where the transpose in the material B operator arises from $\psi_{\mathrm{L},n}$ being defined on the left of the operators Θ and B, transposing the entire equation, and utilizing the complex-symmetry of the Θ operator ($\Theta^T = \Theta$). For reciprocal materials, $B = B^T$ and the left QNMs (comprising the "dual basis" [59]) coincide with the right QNMs. We assume any standard computational discretization of the problem (with sufficiently high accuracy) [64], and write the QNMs as columns in a basis matrix:

$$\Phi_{\mathrm{Rqnm}} = \begin{pmatrix} \psi_{\mathrm{R},1} & \psi_{\mathrm{R},2} & \dots \end{pmatrix}, \tag{3.14}$$

with corresponding eigenvalues

$$\Omega = \begin{pmatrix} \widetilde{\omega}_1 & & \\ & \widetilde{\omega}_2 & \\ & & \ddots \end{pmatrix}, \tag{3.15}$$

and similarly define Φ_{Lqnm} for the left QNMs, which share the same eigenvalues. One can multiply Eq. (3.12) and Eq. (3.13) on the left by $\psi_{\mathrm{L},n}^T$ and $\psi_{\mathrm{R},m}^T$, respectively; integrating over all space V yields the QNM orthogonality relation [1]:

$$\int_V \psi_{\mathrm{L},n}^T \left(\widetilde{\omega}_n B(\widetilde{\omega}_n) - \widetilde{\omega}_m B(\widetilde{\omega}_m)\right) \psi_{\mathrm{R},m} = 0\,. \tag{3.16}$$

To normalize the individual QNMs, we consider Eq. (3.16) with the left and right QNMs indexed by the same value n; the material-dependent term in the middle goes to zero, but instead, one can divide by $\widetilde{\omega}_n - \omega$ and take the limit as the difference goes to zero:

$$\begin{aligned} &\lim_{\omega\to\widetilde{\omega}_n} \int_V \psi_{\mathrm{L},n}^T \left(\frac{\widetilde{\omega}_n B(\widetilde{\omega}_n) - \omega B(\omega)}{\widetilde{\omega}_n - \omega}\right) \psi_{\mathrm{R},n} \\ = &\int_V \psi_{\mathrm{L},n}^T \frac{\partial}{\partial\omega} \left[\omega B(\omega)\right]_{\omega=\widetilde{\omega}_n} \psi_{\mathrm{R},n} = 1. \end{aligned} \tag{3.17}$$

Equation (3.17) cannot be used to independently normalize both $\psi_{\mathrm{L},n}$ and $\psi_{\mathrm{R},n}$ for a given n, but in any scattering computation one only needs overlap integrals between them, for which Eq. (3.17) is sufficient. For nondispersive materials, Eq. (3.16) reduces to an orthogonality relation

$$(\widetilde{\omega}_n - \widetilde{\omega}_m) \int_V \psi_{\mathrm{L},n}^T B \psi_{\mathrm{R},m} = 0\,, \tag{3.18}$$

and the normalization is simply $\int_V \psi_{\mathrm{L},n}^T B \psi_{\mathrm{R},n} = 1$. Once all relevant modes are solved for, they become the natural basis for the fields inside of scatterers.

3.3 Previous works

In the Sec. 3.2, we have introduced the PML-based normalization of QNMs. We will start the expansion of scattered fields by QNMs. There are multiple approaches to such

QNM expansions, e.g., via orthogonality decompositions [1, 65] or complex-analysis-based Mittag–Leffler expansions [63, 66, 67]. These techniques have been used to successfully apply modal analysis to plasmonics [1, 68, 69, 65] and diffraction gratings [70], where a normal-mode approximation (as used in CMT) would necessarily be inaccurate. Yet while these approaches can decompose scattered fields into modes, they have not successfully captured the full interactions of the incident and scattered waves with the incoming and outgoing channels, which is the second key component to a CMT-like theory. References [71, 72] construct scattering matrices but require least-squares procedures and alternative full-Maxwell solvers to do so. Consequently, none of the previous approaches yield all relevant CMT equations, nor can they construct the full scattering matrices of the system.

3.4 Summary

In this chapter, we introduced the coupled-mode point of view of scattering, where spatial regions are partitioned into scattering and channel parts. In the weak coupling limit the framework of coupled-mode theory emerges, where resonances are well separated and have high quality factors. QNMs are the natural basis for a nonperturbative coupled-mode theory. We also summarized the normalization technique we use for QNMs and previous unsuccessful attempts to construct the exact scattering matrices from these modes. In Chapter 4 we will start the first-principle construction of the scattering matrices from QNMs.

Chapter 4

Quasinormal Coupled Mode Theory (QCMT)

In this chapter, we develop a complete CMT-like theory with quasinormal modes. Conventional CMT, briefly summarized in Sec. 3.1.2, comprises two linear matrix equations, the first of which connects the excited-mode amplitudes to the incoming-wave amplitudes, and the second of which connects the outgoing-wave amplitudes to the incoming waves and the resonant excitations. The key result of in this chapter is the derivation of a set of two analogous equations based on quasinormal modes. To derive these equations, one first needs a formal description of both QNMs in Sec. 3.2 and scattering channels in Sec. 2.4. We then derive the two key QCMT equations through integral-equation identities that connect the fields external to the scatterer to fields within the scatterer, where they can be expanded in QNMs. The first equation has appeared in various (not obviously identical) forms in previous works [1, 65, 63, 56], whereas the second QCMT equation has to our knowledge not appeared previously. We use these two equations to derive the QNM-based full scattering matrix of any system (Sec. 4.1), which comprises frequency-dependent coupling matrices (unlike their frequency-independent conventional counterparts). Our key results are summarized in Sec. 4.2, representing the foundational components of QCMT. In Sec. 4.3, we show that Mittag–Leffler pole expansions of our derived expressions, when certain asymptotic conditions are satisfied, can lead to simplifications of the coupling matrices. We consider

two canonical scattering problems as test examples of QCMT: one-dimensional Fabry-Perot scattering, and three-dimensional Mie scattering, where we demonstrate the accuracy of QCMT, and the necessary *inaccuracy* of conventional CMT (Sec. 4.4). A more general discussion of the valid regimes of conventional CMT follows in Sec. 4.5; interestingly, we find that QCMT shows a similar time-domain equation structure as its conventional CMT counterpart, except that there is an additional direct-scattering term and all coupling matrices are convolution operators in time, arising from the inherent frequency dependencies of the underlying Maxwell operators.

4.1 From Maxwell to QCMT

A coupled-mode-theory representation decomposes a scattering problem into three sets of degrees of freedom: the incoming-wave amplitudes, the outgoing-wave amplitudes, and the amplitudes of the resonances, i.e., the quasinormal modes as illustrated in Fig. 4.1. The incoming-wave amplitudes are specified by the problem of interest. This leaves two sets of amplitudes to solve for: the QNM amplitudes (Sec. 4.1.1) and the outgoing-wave amplitudes (Sec. 4.1.2). We derive the two corresponding QCMT equations in the following subsections.

4.1.1 QCMT Eqn 1: QNM amplitudes

The first QCMT equation should relate the quasinormal-mode amplitudes, $\boldsymbol{a}$, to the incoming-wave amplitudes $\boldsymbol{c}_{\text{in}}$ (like its CMT counterpart, Eq. (3.4)). To do this we need to find the QNM response for a given incident field. In the region of the scatterer, the scattered field ψ_{scat} can be decomposed into the QNMs:

$$\psi_{\text{scat}} = \begin{pmatrix} & & \\ \psi_{\text{R},1} & \psi_{\text{R},2} & \cdots \\ & & \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} = \Phi_{\text{Rqnm}} \boldsymbol{a}, \tag{4.1}$$

where Φ_{Rqnm} is the basis of QNM resonance fields and $\boldsymbol{a}$ are the unknown expansion coefficients. To relate the expansion coefficients to the incoming-wave amplitudes, it might be possible to use the typical differential form of Maxwell's equations, but the fields have

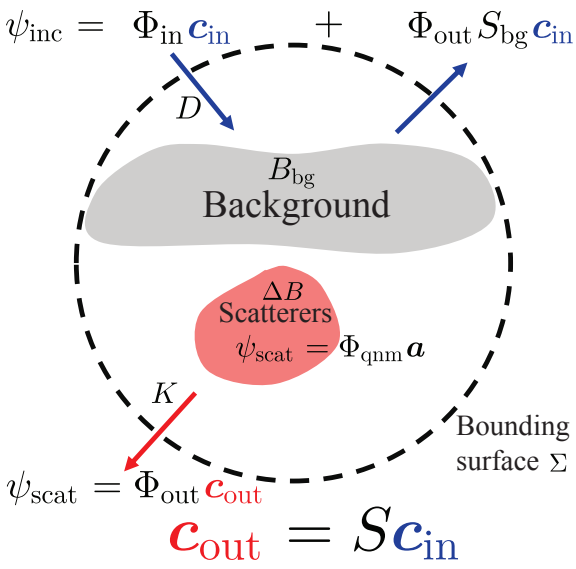

Figure 4.1: An incident field ψ_{inc} in a background of material tensor B_{bg} impinges upon scatterers, with susceptibility ΔB, exciting a scattered field ψ_{scat}. The coupling of the incident field to the resonances is described by a coupling matrix $D(\omega)$ in a relation that comprises the first QCMT equation, Eq. (4.14). Within the scatterers, the scattered field can be decomposed into quasinormal modes Φ_{qnm}, with modal amplitudes $\boldsymbol{a}$, which leak out into radiation amplitudes determined by a coupling matrix $K(\omega)$, as dictated by the second QCMT equation, Eq. (4.15). The "channels" carrying power into and out of the scattering bodies are defined on a bounding surface Σ, where they satisfy power-orthonormality relations, Eq. (2.11). The S-matrix connecting the incoming-wave coefficients, $\boldsymbol{c}_{\text{in}}$, to the outgoing-wave coefficients, $\boldsymbol{c}_{\text{out}}$, is determined by the QCMT equation of Eq. (4.16).

independent degrees of freedom everywhere in space, including outside the scatterer. Instead, all of the degrees of freedom can be brought inside the scatterer by using the volume equivalence principle [64]. The material tensor B can be separated into background and scatterer constituents as discussed in Sec. 2.3,

$$B = B_{\text{bg}} + \Delta B\,. \tag{4.2}$$

The incident field ψ_{inc} is the solution of Maxwell's equations in the absence of the scatterer (i.e. with $B = B_{\text{bg}}$ everywhere), while the scattered field ψ_{scat} is the additional field excited when ΔB is introduced, which is given by the difference between the total and incident fields, $\psi_{\text{scat}} = \psi - \psi_{\text{inc}}$. Then straightforward manipulation of Maxwell's equations yields a differential equation for ψ_{scatt} with all degrees of freedom within the scatterer [64]:

$$\Theta\psi_{\text{scatt}} - \mathrm{i}\omega B\psi_{\text{scatt}} = \mathrm{i}\omega\Delta B\psi_{\text{inc}}\,, \tag{4.3}$$

One can see that Eq. (4.3) relates ψ_{scat} to ψ_{inc}, and thus should also define the connection from $\boldsymbol{a}$ to $\boldsymbol{c}_{\text{in}}$. Inserting the expression relating ψ_{inc} to $\boldsymbol{c}_{\text{in}}$, Eq. (2.13), into the right-hand side of Eq. (4.3), and the decomposition of ψ_{scat} into QNMs, Eq. (4.1), on the left-hand side, and utilizing generalized reciprocity relations as well as the normalization properties of the QNMs, Eqs. (3.16,3.17), leads directly to the first QCMT equation:

$$N(\omega)\mathrm{i}(\Omega-\omega)\boldsymbol{a} = \frac{\mathrm{i}\omega}{\alpha}(\Phi_{\text{Lqnm}}, \Delta B\Phi_{\text{inc}})\boldsymbol{c}_{\text{in}}\,, \tag{4.4}$$

where $N(\omega)$ is a nonsingular matrix with entries

$$N_{nm}(\omega) = \int_V \psi_{\text{L},n}^T \left(\frac{\widetilde{\omega}_n B(\widetilde{\omega}_n) - \omega B(\omega)}{\widetilde{\omega}_m - \omega}\right)\psi_{\text{R},m}\,. \tag{4.5}$$

If B is nondispersive, the normalization condition of Eq. (3.17) implies that N is the identity matrix.

Equation (4.4) is our first QCMT equation. It is not new, having been derived in Refs. [1, 56], but we included a brief derivation for completeness. The intuition behind Eq. (4.4) is similar to that of the first CMT equation, Eq. (3.4): the QNM amplitudes $\boldsymbol{a}$ are large for the modes (within Φ_{Lqnm}) that have large overlap with the incident field over the scatterer volume, and/or whose resonant frequencies have small imaginary parts and real parts close to the excitation frequency ω. The matrix $N(\omega)$ accounts for modal overlaps in dispersive media, though as we show in Sec. 4.3, this matrix can be dropped from the equation when a Mittag–Leffler expansion is valid.

4.1.2 QCMT Eqn 2: Outgoing-channel wave amplitudes

The second QCMT equation, analogous to Eq. (3.5) of conventional CMT, should determine the outgoing-wave amplitudes from the incoming-wave and quasinormal-mode amplitudes. First, we recognize that the outgoing-wave amplitudes are given by an overlap integral on the bounding region Σ of the outgoing-wave basis functions with the total field ψ:

$$\boldsymbol{c}_{\text{out}} = \langle \Phi_{\text{out}}, \psi\rangle = \int_\Sigma \Phi_{\text{out}}^\dagger \frac{Q}{4}\psi. \tag{4.6}$$

We can write the total field as the incident field plus the scattered field, with the scattered field itself being the field radiated from the polarization field inside the scatterer via a convolution with the background Green's function, Γ_{bg}:

$$\psi_{\text{scat}}(\boldsymbol{r}) = \mathrm{i}\omega \int_{V'} \Gamma_{\text{bg}}(\boldsymbol{r}, \boldsymbol{r}') \Delta B(\boldsymbol{r}') \psi(\boldsymbol{r}'). \tag{4.7}$$

The polarization field itself must be separated into incident and scattered fields, the latter of which comprise the QNMs, per Eq. (4.1). Making these substitutions, we have

$$\begin{aligned} \boldsymbol{c}_{\text{out}} = & \int_{\Sigma} \Phi_{\text{out}}^{\dagger} \frac{Q}{4} \psi_{\text{inc}}(\boldsymbol{r}) \\ & + \mathrm{i}\omega \int_{\Sigma} \Phi_{\text{out}}^{\dagger} \frac{Q}{4} \int_{V} \Gamma_{\text{bg}}(\boldsymbol{r}, \boldsymbol{r}') \Delta B(\boldsymbol{r}') \psi_{\text{inc}}(\boldsymbol{r}') \\ & + \mathrm{i}\omega \int_{\Sigma} \Phi_{\text{out}}^{\dagger} \frac{Q}{4} \int_{V} \Gamma_{\text{bg}}(\boldsymbol{r}, \boldsymbol{r}') \Delta B(\boldsymbol{r}') \Phi_{\text{Rqnm}}(\boldsymbol{r}') \boldsymbol{a} \,. \end{aligned} \tag{4.8}$$

The first term simply isolates the outgoing components of the incident field; from Eq. (2.14), this term equals $S_{\text{bg}} \boldsymbol{c}_{\text{in}}$.

The second and third terms start with polarization-field terms inside the scatterer, $\Delta B \psi_{\text{inc}}$ and $\Delta B \Phi_{\text{Rqnm}} \boldsymbol{a}$, respectively, convolve with the background Green's function to yield radiated (outgoing) fields on the bounds surface Σ, and then computes the power-normalized overlap with the outgoing-wave basis functions, Φ_{out}. We can simplify these expressions via reciprocity. We assume that the background medium is a reciprocal material. (If it is not, then one can simply use generalized reciprocity [73] and slightly modified versions of the expressions below.) Reciprocity in the background means that the background Green's function has a symmetry under reversal of its position arguments, i.e., transposition of source and receiver locations. Typically that would be written $\Gamma_{\text{bg}}^{T}(\boldsymbol{r}, \boldsymbol{r}') = \Gamma_{\text{bg}}(\boldsymbol{r}', \boldsymbol{r})$. However, this exact expression relies on either free-space or periodic boundary conditions and does not apply, for example, in scattering scenarios with Bloch-periodic boundary conditions. With full generality, for *any* boundary conditions, the reciprocal scenario represented by the transpose and argument reversal of the Green's function is given by *time-reversing*

the scattering channels, which we can encode in the Green's function as the relation

$$\Gamma^T_{\mathrm{bg}}(\boldsymbol{r},\boldsymbol{r}') = \Gamma^{\mathrm{TR}}_{\mathrm{bg}}(\boldsymbol{r}',\boldsymbol{r})\,, \tag{4.9}$$

where the "TR" superscript implies time-reversal of the channels. This relation does not require any time-reversal symmetry of the scatterer itself. This definition includes both the usual symmetry of the Green's function for free-space or periodic boundary conditions, as well as the special inner product defined in the Bloch-periodic case [74].

We now simplify the third expression on the right-hand side of Eq. (4.8), which will imply a similar simplification for the second expression as well. First, we observe that adding in any constant multiple of Φ_{in} to Φ_{out} in the overlap integrals will not change the value of that integral, due to the orthogonality $\langle \Phi_{\mathrm{in}}, \Phi_{\mathrm{out}} \rangle = 0$ and the scattered field being always outgoing. As a result , we add $\frac{\alpha}{\beta}\Phi_{\mathrm{in}}$ to Φ_{out}, and this sum becomes $\frac{1}{\beta}\Phi_{\mathrm{inc}}$ by Eq. (2.13), so we can replace $\Phi^\dagger_{\mathrm{out}}$ with $\left(\frac{1}{\beta}\Phi_{\mathrm{inc}}\right)^\dagger$. Now we can work with a regular field Φ_{inc}, instead of the singular field Φ_{out}, which will enable a simplification below. The basis functions in Φ^*_{inc} are related to their time-reversed partners by $\Phi^{\mathrm{TR}}_{\mathrm{inc}} = P\Phi^*_{\mathrm{inc}}$, where $P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Using the fact that $P^2 = \mathbb{I}$, we can write $\left(\frac{1}{\beta}\Phi_{\mathrm{inc}}\right)^\dagger$ as $\frac{1}{\beta^*}(P\Phi^{\mathrm{TR}}_{\mathrm{inc}})^T$. To use reciprocity in the Green's function part of the expression in Eq. (4.8), we want to take its transpose. Upon using the reciprocity relation, Eq. (4.9), the position arguments of the Green's function can be reversed, in which case one can interpret the expression in a new way: fields on the bounding surface Σ are transported *into* the scatterer, by the background Green's function, where they are overlapped with the fields inside the scatterer. Working through this intuition mathematically, we find:

$$\begin{aligned}
&\left\{\int_\Sigma \frac{\mathrm{i}\omega}{\beta^*}(P\Phi^{\mathrm{TR}}_{\mathrm{inc}})^T(\boldsymbol{r})\frac{Q}{4}\int_V \Gamma_{\mathrm{bg}}(\boldsymbol{r},\boldsymbol{r}')\Delta B(\boldsymbol{r}')\Phi_{\mathrm{Rqnm}}(\boldsymbol{r}')\boldsymbol{a}\right\}^T \\
&= \frac{\mathrm{i}\omega\boldsymbol{a}^T}{4\beta^*}\int_V \Phi^T_{\mathrm{Rqnm}}(\boldsymbol{r}')\Delta B^T(\boldsymbol{r}')\int_\Sigma \Gamma^{\mathrm{TR}}_{\mathrm{bg}}(\boldsymbol{r}',\boldsymbol{r})QP\Phi^{\mathrm{TR}}_{\mathrm{inc}}(\boldsymbol{r}) \\
&= \frac{\mathrm{i}\omega\boldsymbol{a}^T}{4\beta^*}\int_V \Phi^T_{\mathrm{Rqnm}}(\boldsymbol{r}')\Delta B^T(\boldsymbol{r}')\Phi^{\mathrm{TR}}_{\mathrm{inc}}(\boldsymbol{r}') \\
&= \frac{\mathrm{i}\omega}{4\beta^*}(\Phi^{\mathrm{TR}}_{\mathrm{inc}}, \Delta B\Phi_{\mathrm{Rqnm}})\boldsymbol{a}\,,
\end{aligned} \tag{4.10}$$

The initial expression is the third right-hand side expression of Eq. (4.8), with $\frac{1}{\beta^*}(P\Phi_{\rm inc}^{\rm TR})^T$ replacing $\Phi_{\rm out}^\dagger$. The first equality expression is the transpose of the initial with $\Gamma_{\rm bg}^{\rm TR}(\boldsymbol{r}',\boldsymbol{r})$ replacing $\Gamma_{\rm bg}^T(\boldsymbol{r},\boldsymbol{r}')$ via reciprocity, and Q^T replacing Q because it is complex symmetric. The second equality expression simplifies the integral on the right-hand side of the previous expression, $\int_\Sigma \Gamma_{\rm b}^{\rm TR}(\boldsymbol{r}',\boldsymbol{r})QP\Phi_{\rm inc}^{\rm TR}(\boldsymbol{r}) = \Phi_{\rm inc}^{\rm TR}(\boldsymbol{r}')$, through the surface-equivalence principle [73, 64]. The fact that $\Phi_{\rm inc}^{\rm TR}$ is a traveling wave and free of singularities is crucial, otherwise the surface-equivalence principle cannot be applied. The term $QP\Phi_{\rm inc}^{\rm TR}(\boldsymbol{r})$ are the equivalent surface currents that generate the incoming fields $\Phi_{\rm inc}^{\rm TR}$, and the convolution with the Green's function $\Gamma_{\rm bg}^{\rm TR}(\boldsymbol{r}',\boldsymbol{r})$ produces the fields $\Phi_{\rm inc}^{\rm TR}$ at points $\boldsymbol{r}'$ in the scattering body. The final equality expression is simply the transpose of the previous one, written in inner product notation, where we now clearly see that the lengthy expressions on the right-hand side of Eq. (4.8) are proportional to simple unconjugated overlap integrals of the time-reversed incident fields with the fields $\Delta B\Phi_{\rm Rqnm}$, which are nonzero only in the scatterer. The specific nature of $\Phi_{\rm Rqnm}$ played no role in the derivation of Eq. (4.10), and thus the simplification of the second term in Eq. (4.8) is of exactly the same form but with the replacement $\Phi_{\rm Rqnm}\boldsymbol{a} \to \psi_{\rm inc}$.

Having simplified each of the three terms in Eq. (4.8), we now have the second QCMT equation:

$$\boldsymbol{c}_{\rm out} = \left\{ S_{\rm bg} + \frac{1}{4\alpha\beta^*}\mathrm{i}\omega(\Phi_{\rm inc}^{\rm TR}, \Delta B\Phi_{\rm inc}) \right\} \boldsymbol{c}_{\rm in} + \frac{1}{4\beta^*}\mathrm{i}\omega(\Phi_{\rm inc}^{\rm TR}, \Delta B\Phi_{\rm Rqnm})\boldsymbol{a}\,. \tag{4.11}$$

Equation (4.11), to our knowledge, has not been derived before. (See Appendix A.2 for a more direct direction, without using the equivalence principle, for structures in vacuum.) Intuitively, the first and third terms represent the direct background process from incoming to outgoing waves, and the radiation from QNMs to outgoing waves, respectively. Interestingly, the second term represents a Born-like scattering term (it is the first term in a Born scattering series expansion) that apparently is not captured in the resonant response of the third term. Equation (4.11) is the crucial QCMT equation that enables the solution of the outgoing fields for a given input, and it will be the key to enabling an expression for the scattering matrix.

4.2 Quasinormal coupled mode theory

We can synthesize the two key results of the previous sections into our quasinormal coupled-mode theory. To simplify comparisons with conventional CMT, we define frequency-dependent matrices $D(\omega)$ and $K(\omega)$ that play similar roles to D and K in conventional CMT:

$$D(\omega) = \frac{\mathrm{i}\omega}{\alpha}(\Phi_{\mathrm{Lqnm}}, \Delta B \Phi_{\mathrm{inc}})^T\,, \tag{4.12}$$

$$K(\omega) = \frac{\mathrm{i}\omega}{4\beta^*}(\Phi_{\mathrm{inc}}^{\mathrm{TR}}, \Delta B \Phi_{\mathrm{Rqnm}})\,. \tag{4.13}$$

Eqs. (4.12,4.13) are spatial overlap integrals between resonant modes and scattering channel functions weighted by the scatterer susceptibility ΔB, representing the coupling strengths between modes and channels. With these matrices, we can write the key QCMT equations, Eqs. (4.4,4.11), as:

$$\mathrm{i}N(\omega)(\Omega - \omega)\boldsymbol{a} = D^T(\omega)\boldsymbol{c}_{\mathrm{in}}\,, \tag{4.14}$$

$$\boldsymbol{c}_{\mathrm{out}} = \left\{ S_{\mathrm{bg}} + \frac{\mathrm{i}\omega}{4\alpha\beta^*}(\Phi_{\mathrm{inc}}^{\mathrm{TR}}, \Delta B \Phi_{\mathrm{inc}}) \right\} \boldsymbol{c}_{\mathrm{in}} + K(\omega)\boldsymbol{a}\,. \tag{4.15}$$

Additionally, we can solve the first QCMT equation, Eq. (4.14), for the quasinormal-mode amplitudes $\boldsymbol{a}$, insert the result into the second QCMT equation, Eq. (4.15), and extract the QCMT scattering matrix:

$$S = S_{\mathrm{bg}} + \frac{\mathrm{i}\omega}{4\alpha\beta^*}(\Phi_{\mathrm{inc}}^{\mathrm{TR}}, \Delta B \Phi_{\mathrm{inc}}) - \mathrm{i}K(\omega)\left[N(\omega)(\Omega - \omega)\right]^{-1} D^T(\omega)\,. \tag{4.16}$$

Eq. (4.16) shows that, once all QNMs are computed, S can be constructed at any frequency from overlap integrals for the coupling strengths.

We see that Eqs. (4.14)–(4.16) show a similar functional form to the analogous CMT equations, Eqs. (3.4,3.5,3.9). There are two key differences. First, the coupling matrices $K(\omega)$ and $D(\omega)$ are now frequency-dependent, with a possible additional frequency dependence arising in dispersive media from the matrix $N(\omega)$. This frequency dependence is critical to accurate simulations, as we show in Sec. 4.4; in the time domain, they indicate that the coupling operators are convolutions, as we discuss in Sec. 4.5. The second

key difference is the appearance of the second term on the right-hand sides of Eq. (4.15) and Eq. (4.16), which is proportional to the overlap of the time-reversal-generated incident waves with the incident field, in the domain of the scatterer. This Born-scattering term arises only in the presence of a scatterer (i.e. $\Delta B \neq 0$ everywhere), and yet is part of the "direct" scattering process, a term that has no counterpart in conventional CMT.

One can similarly ask whether the QCMT equations satisfy conservation laws similar to those of Eqs. (3.6)–(3.8) of conventional CMT. With reciprocal materials and outgoing channel functions that are time-reversed partners of the incoming channel functions, one can see from Eqs. (4.12,4.13) that $D(\omega)$ and $K(\omega)$ will be identical,

$$D(\omega) = K(\omega) \qquad \text{(reciprocal)} \tag{4.17}$$

up to the numerical factor $\alpha/4\beta^*$. However, that is as far as one can go with simple QCMT conservation laws. The analog of Eq. (3.7), $K^\dagger K = 2\,\mathrm{Im}\,\Omega$, does not hold. Intuitively, that equality is a statement that the mode-energy decay rate equals the power in the outgoing channels, and certainly one could codify such a statement in a Poynting-flux evaluation of the CMT quantities. However, $\mathrm{Im}\,\Omega$ does not determine the mode-energy decay rate at arbitrary frequency ω in the general scenario when the coupling matrices are frequency-dependent. Similarly, there is no simple analog of Eq. (3.8), $S_{\mathrm{bg}} D^* = -D$, which enforces a relation between the background scattering matrix and the coupling matrix D. A key impediment is the presence of the Born term in the second CMT equation, Eq. (4.15), which augments the background with a scatterer- and frequency-dependent matrix.

4.3 Pole expansion representations

The key equations derived to this point are Eqs. (4.14)–(4.16), which are the two QCMT equations and the corresponding scattering matrix, in order. A key distinction between conventional CMT and these equations is that the QCMT matrices are frequency-dependent, and require the use of an overlap matrix $N(\omega)$ and its inverse. In this section, we show how Mittag–Leffler pole expansions allow for removal of the N matrix and simplification of the

frequency-dependent matrices in Eq. (4.23). When such pole expansions are applied to the S matrix, it is also revealed in Eq. (4.24) that the dominant contribution of the nonresonant scattering background comes from low-Q modes.

For our purposes, we can use the following form of a Mittag–Leffler expansion [75, 76]: given a meromorphic function $f(z)$ with simple poles $z_1, z_2, \ldots, z_n, \ldots$, no pole at 0, and $\lim_{n\to\infty} z_n = \infty$, one can expand $f(z)$ around $z = 0$ as

$$f(z) = f(0) + h(z) + \sum_{n=1}^{\infty} \frac{\mathrm{Res}(f(z_n))}{z_n} + \sum_{n=1}^{\infty} \frac{\mathrm{Res}(f(z_n))}{z - z_n}\,, \tag{4.18}$$

where $\mathrm{Res}(f(z_n))$ is the residue of $f(z)$ at z_n and $h(z)$ is an entire function. By definition, $h(0) = 0$. Furthermore, if $f(z)$ is bounded as z goes to complex infinity, then $h(z)$ is zero everywhere. In this section we emphasize only new expressions; in Appendix A.3 we show that Mittag–Leffler expansions at various stages of the QCMT formulation unify many seemingly different expressions that have previously appeared across the literature.

4.3.1 Simplified QCMT equation

A first use of the Mittag–Leffler expansion is to simplify the first QCMT equation, Eq. (4.14). We can start with the volume-integral expression relating the scattered field at a point $\boldsymbol{r}$ in the scatterer to the incident field at a point $\boldsymbol{r}'$ in the scatterer:

$$\psi_{\mathrm{scat}}(\boldsymbol{r}) = \mathrm{i}\omega \int \Gamma(\boldsymbol{r}, \boldsymbol{r}', \omega) \Delta B(\boldsymbol{r}', \omega) \psi_{\mathrm{inc}}(\boldsymbol{r}')\,. \tag{4.19}$$

To identify all possible poles in the response, we require knowledge of the frequency dependence of the permittivity, which by the Kramers–Kronig relations [77] (or a pole-expansion representation [78, 79]), can be written

$$B(\omega) = B_\infty - \sum_n \frac{\sigma_n}{\omega - \omega_n}, \tag{4.20}$$

where the ω_n are the complex-frequency material poles (to be distinguished from the quasinormal-mode frequencies $\widetilde{\omega}_i$) and the σ_n are matrix-valued residues. The Green's function can be decomposed into a summation over the quasinormal-mode resonances [66,

56, 67],

$$\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega)=\sum_i \frac{\psi_{R,i}(\boldsymbol{r})\psi_{L,i}^T(\boldsymbol{r}')}{\mathrm{i}(\widetilde{\omega}_i-\omega)}. \tag{4.21}$$

From Eq. (4.21) and a known sum rule [63], one can show that Γ evaluated at any material pole is 0, i.e. $\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega_n)=0$, still considering $\boldsymbol{r}$ and $\boldsymbol{r}'$ at points inside the scatterer. Physically this must be true because at any material pole the permittivity diverges and the materials acts as a perfect conductor, such that the field within the scatterer is 0. The term $\Gamma\Delta B$ in Eq. (4.19) contains the frequency-dependent term $\Gamma\sigma_n/(\omega-\omega_n)$; applying ML to that term yields:

$$\begin{aligned}
&\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega)\sum_n \frac{\sigma_n}{\omega-\omega_n}\\
=\;&\sum_{i,n}\frac{\psi_{R,i}(\boldsymbol{r})\psi_{L,i}^T(\boldsymbol{r}')}{\mathrm{i}(\widetilde{\omega}_i-\omega)}\frac{\sigma_n}{(\widetilde{\omega}_i-\omega_n)}\\
&-\sum_{i,n}\frac{1}{\mathrm{i}}\psi_{R,i}(\boldsymbol{r})\psi_{L,i}^T(\boldsymbol{r}')\frac{\sigma_n}{\widetilde{\omega}_i(\widetilde{\omega}_i-\omega_n)}\\
&+\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega_n)\sum_n\frac{\sigma_n}{\omega-\omega_n}-\Gamma(\boldsymbol{r},\boldsymbol{r}',0)\sum_n\frac{\sigma_n}{\omega_n}\\
=\;&\sum_{i,n}\frac{\psi_{R,i}(\boldsymbol{r})\psi_{L,i}^T(\boldsymbol{r}')}{\mathrm{i}(\widetilde{\omega}_i-\omega)}\sum_n\frac{\sigma_n}{(\widetilde{\omega}_i-\omega_n)}\,,\\
=\;&\sum_{i,n}\frac{\psi_{R,i}(\boldsymbol{r})\psi_{L,i}^T(\boldsymbol{r}')}{\mathrm{i}(\widetilde{\omega}_i-\omega)}(B_\infty-B(\widetilde{\omega}_i))
\end{aligned} \tag{4.22}$$

where only the first term remains in the expansion because the fourth term contains $\Gamma(\boldsymbol{r},\boldsymbol{r}',0)$, the third term contains $\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega_n)$, and the second term is proportional (via partial-fraction expansion, cf. Appendix A.3) to the difference $\Gamma(\boldsymbol{r},\boldsymbol{r}',0)-\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega_n)$, all of which are zero due to the identities discussed above, $\Gamma(\boldsymbol{r},\boldsymbol{r}',0)=\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega_n)=0$. We define a basis matrix $\widetilde{\Phi}_{\mathrm{Lqnm}}$ with elements $[\Delta B(\widetilde{\omega}_i)]^T\,\psi_{L,i}$. Inserting this expansion back into Eq. (4.19), we find a simple expansion expression:

$$(\Omega-\omega)\boldsymbol{a}=\frac{\omega}{\alpha}(\widetilde{\Phi}_{\mathrm{Lqnm}},\Phi_{\mathrm{inc}})\boldsymbol{c}_{\mathrm{in}}\,, \tag{4.23}$$

This expression simplifies the frequency dependencies of Eq. (4.4), as it does not contain the frequency-dependent overlap matrix $N(\omega)$, nor does it require re-evaluating the material constant B at every frequency ω, and it is very similar to the first CMT equation, albeit with a frequency-dependent prefactor. To our knowledge, Eq. (4.23) has not been derived before. Similarly, substituting Γ in Eq. (4.21) into Eq. (4.19) produces a version of Eq. (4.4) without $N(\omega)$, so we can similarly drop $N(\omega)$ in the expansions to follow. Furthermore, applying the same procedures to $\omega\Gamma\Delta B$, we obtain the expansion expression in Refs.[63, 65] (cf. Appendix A.3). This method also makes the equivalence evident among other expansion expressions highlighted in Ref. [56].

4.3.2 Resonant expansion of S matrices

The S matrix expression in Eq. (4.16) is similar to a pole expansion via resonance frequencies, except for the presence of the N matrix and the frequency dependencies of the K and D matrices. We can apply the Mittag–Leffler expansion to Eq. (4.16) to obtain a simpler form. As shown in Appendix A.3, the result utilizes frequency-independent matrices $\widetilde{K}$ and $\widetilde{D}$, defined by $\widetilde{K}_{ij} = K_{ij}(\widetilde{\omega}_j)$ and $\widetilde{D}_{ij} = D_{ij}(\widetilde{\omega}_i)$. Then Eq. (4.16) becomes

$$S = \underbrace{S_{\mathrm{bg}}(\omega = 0) + H(\omega) + \mathrm{i}\widetilde{K}\Omega^{-1}\widetilde{D}^T}_{=\text{“background part”}} - \underbrace{\mathrm{i}\widetilde{K}(\Omega - \omega)^{-1}\widetilde{D}^T}_{=\text{“resonant part”}}, \tag{4.24}$$

where $H(\omega)$ is a frequency dependent background term (generalizing h from Eq. (4.18) to a matrix), whose elements are all entire functions of ω, containing the first Born scattering term and the difference between $\widetilde{K}(\Omega - \omega)^{-1}\widetilde{D}^T$ and $K(\Omega - \omega)^{-1}D^T$. (cf. Appendix A.3) The difference grows smaller as the Q factor of a mode becomes larger. For a mode with large Q factor, the relevant range of ω around the resonant frequency $\widetilde{\omega}_i$ is very small, so replacing ω by $\widetilde{\omega}_i$ in K and D makes little difference. This implies that $H(\omega)$ mostly consists of contributions from low-Q modes and the first Born scattering term. This fact forms the basis on determining when the conventional CMT is valid in Sec. 4.5.

Because Eq. (4.24) is derived from Eq. (4.16), it is not useful for computing S. After all, one can only determine $H(\omega)$ from Eq. (4.16). However, Eq. (4.24) separates the background

and resonant scattering contributions. The "resonant" term with the inverse of $(\Omega - \omega)$ will be the dominant contributor near resonances, while the remaining terms can be considered a non-resonant "background." With the frequency dependencies of K and D removed (and N removed altogether), the resonant term is now purely a complex Lorentzian form. For simplicity, we assumed the scattering poles do not exactly coincide with the material poles. For choices of channel basis functions such that $S(\omega)$ goes to zero at infinity everywhere in the complex plane (as is possible in the Fabry–Perot example in Sec. 4.4), $H(\omega) = 0$ and Eq. (4.24) represents a dramatic simplification of the QCMT scattering matrix.

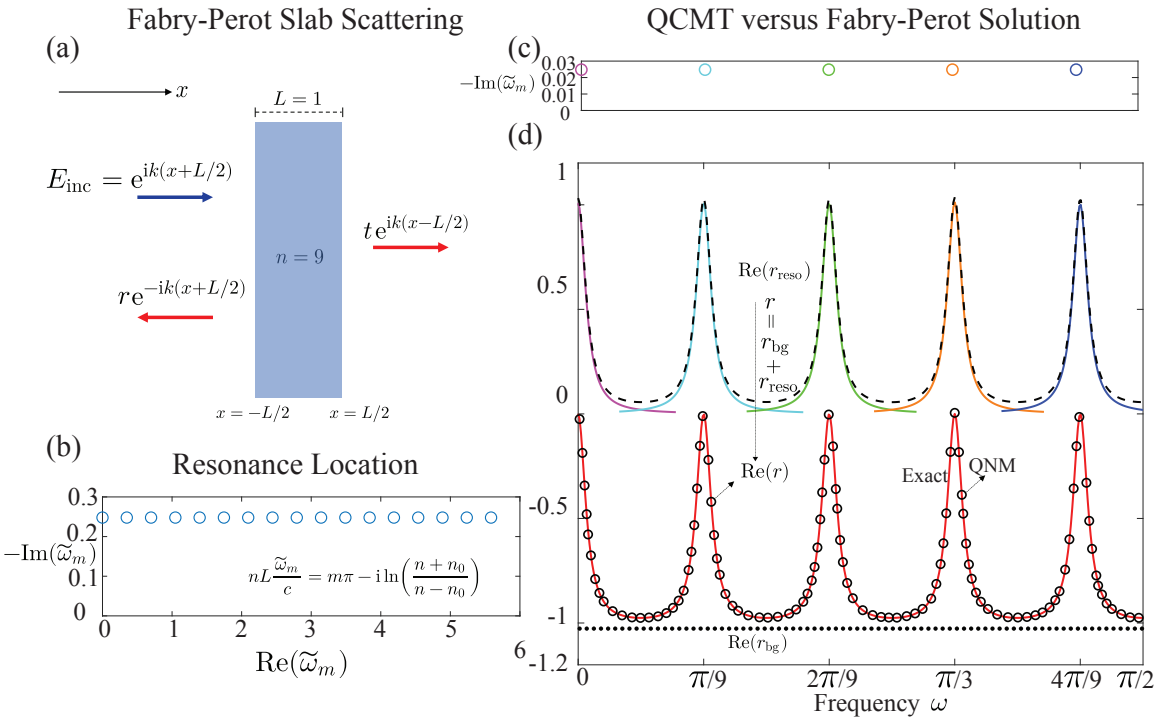

Figure 4.2: (a) A plane wave in vacuum is incident upon a Fabry-Perot slab of thickness $L = 1$ (alternatively, all frequencies are scaled by $1/L$ for arbitrary L) and refractive index $n = 9$. The transmission and reflection coefficients of the slab are t and r, respectively. The thickness-dependent phases of the channel functions ensure $H(\omega) = 0$ in Eq. (4.24). (b) Resonances $\widetilde{\omega}_m$ in the complex frequency plane, computed analytically (inset). (c,d) Five resonances (c) and the real parts of the background (black dotted), resonant (colored solid, black dashed, top), and total (black circles) reflection coefficients computed by QCMT, as well as the exact expression (red solid).

4.4 QCMT Computations

In this section, we demonstrate the accuracy of the QCMT theory. We compute wave-scattering solutions by Eq. (4.16) from a Fabry-Perot slab as well as from a three-dimensional

Mie sphere, two cases where one can easily compare against exact solutions. In each example, we demonstrate that it is critical to accurately model the contribution of the background term in Eq. (4.24) (including the Born scattering term), which was missing from previous quasinormal-mode descriptions of the scattering matrix [71, 72]. In the sphere case, we go a step further and compare the exact and QCMT results with the best possible CMT model of the scattering process. We show that CMT *cannot* accurately model the scattering response, as even the best CMT approximation is highly inaccurate.

The Fabry–Perot example is depicted in Fig. 4.2(a). The basis functions for the incoming and outgoing waves are plane waves. The basis-function phases are defined relative to each interface, ensuring that the scattering matrix decays to zero at infinity everywhere in the complex plane, allowing us to use the simpler pole-expansion expressions of Sec. 4.3. In Fig. 4.2(b) we plot the resonances of the slab (with refractive index $n = 9$), which can be found analytically [56] by the inset expression. (The high value of n makes the resonances more distinct; one can expect equivalent or even higher accuracy for smaller refractive indices.) We consider normal incidence where both s and p polarizations exhibit identical responses. The reflection coefficients computed by QCMT are shown in Fig. 4.2(d), where each of the resonant contribution r_{reso}, the background contribution r_{bg}, and the total reflection coefficient r are shown. For the resonant-only reflection coefficient, the dashed black line shows the total computed r_{reso}, while the solid lines depict the isolated contributions of each resonance. The background reflection coefficient, which our method predicts exactly, provides an important shift in the reflection coefficient that is critical to getting the correct final answer. The total reflection coefficient is depicted by the open circles, agreeing very well with the exact solution (red solid line). See Appendix A for further details of the channel function definitions and quasinormal modes. An accurate CMT model of the form of Eqs. (3.4)–(3.9) for this example is impossible. Since the direct process here is perfect transmission, no CMT model can produce a reflection coefficient r as in Fig. 4.2(d), with a few dips in a background of unity (total reflection).

Fig. 4.3 depicts the scattering by a sphere with refractive index $n = 4.5$. Parts (a–e) of the figure are analogous to their counterparts in Fig. 4.2, and again show very high accuracy in the QCMT computations. The channel basis functions are vector spherical

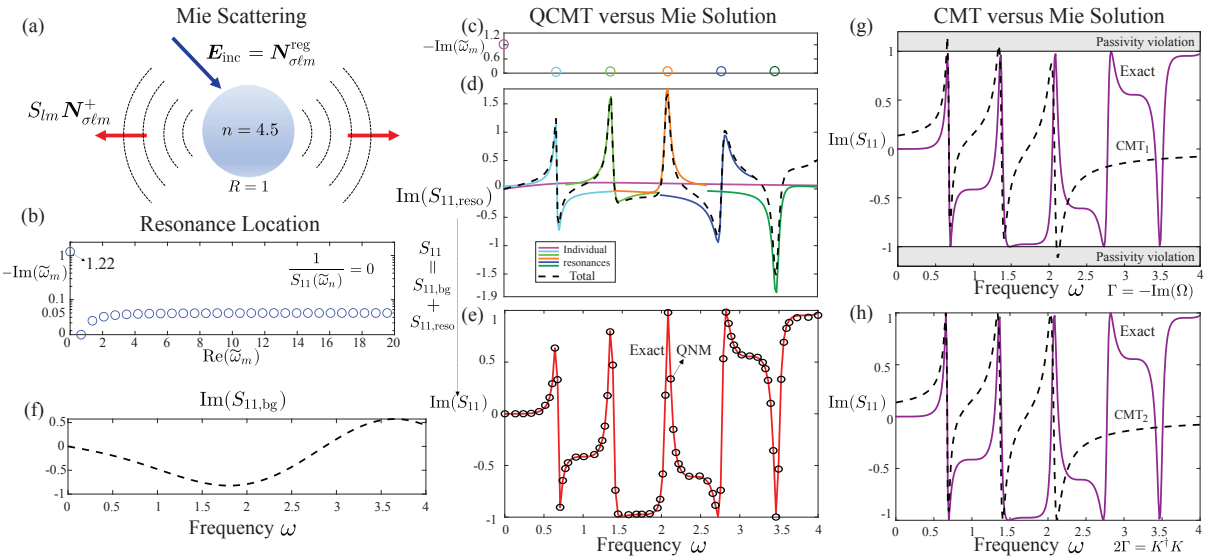

Figure 4.3: (a) An incoming spherical wave in vacuum scatters from a sphere of refractive index $n = 4.5$ and radius $R = 1$. (b) Computed resonances of the Mie sphere. (c) Six highlighted resonances. (d) Resonant contributions to the real part of the S_{11} scattering-matrix coefficient from each resonance (solid colors), as well as in total (black dashed). (e) Total scattering-matrix coefficient by QCMT (black circles) and the exact solution (red solid lines). (f) The frequency-dependent background contribution. (g,h) CMT models of the Mie-scattering process, selecting the matrix Γ equal to either $-\operatorname{Im}\Omega$ (g, CMT_1), in which case passivity can be violated in the CMT model, or $K^{\dagger}K/2$ (h, CMT_2). Neither CMT model can accurately capture the response, even in the vicinity of the resonant peaks.

waves $\boldsymbol{N}_{\sigma\ell m}$ (Ref. [80]). The indices ℓ and m are the usual angular-momentum indices, while σ separates the angular components into even and odd constituents. For a sphere the different angular momentum channels decouple, and here we consider the $\ell = 1$ channel for the incident wave and scattering-matrix calculations. Unlike the Fabry–Perot case, the resonance locations cannot be found analytically. Instead, they are found as the poles of the sub-block of the S-matrix corresponding to the $\ell = 1$ channel, S_{11}, i.e., the solution of the equation $(S_{11}(\widetilde{\omega}_m))^{-1} = 0$. However, their QNM fields *can* be obtained analytically using an appropriate rotation of the outgoing fields in the complex plane (cf. Appendix A). In Fig. 4.3(a) we compare the imaginary part of S_{11} constructed from Eq. (4.16) with the exact value in Fig. 4.3(e). Unlike the Fabry-Perot case, there does not appear to be a choice of basis functions that makes the background part frequency-independent, and Fig. 4.3(f) shows the oscillatory background term $S_{11,\mathrm{bg}}$. Furthermore, we try to construct a CMT model of the form of Eqs. (3.4)–(3.9) to the highest possible accuracy. In this CMT model, a single channel is coupled to multiple resonances represented by Ω as in Eq. (3.15), so K and D are row vectors, with the number of elements set by the number of coupled

resonances. Since the system here is reciprocal and lossless with $S_{\text{bg}} = 1$, Eqs. (3.6,3.8) must hold in the CMT model, requiring $K = D$ with purely imaginary elements. This fixes the phase of elements in K. Hence, we choose $K_m = \text{i}\sqrt{-2\,\text{Im}\,\widetilde{\omega}_m}$ to accurately model the decay of the resonances. The choices of Γ and Ω are not as straightforward as K and D because we have more modes than channels. This means $\Gamma = -\text{Im}(\Omega)$ and $\Gamma = \frac{1}{2}K^\dagger K$ cannot be satisfied simultaneously. ($K^\dagger K$ is rank 1 with the single incoming/outgoing channel.) In Fig. 4.3(g), with model "CMT$_1$," we choose $\Gamma = -\text{Im}(\Omega)$, in which case $K^\dagger K \neq 2\Gamma$ and energy conservation is violated. In Fig. 4.3(h), for model "CMT$_2$," we choose $\Gamma = \frac{1}{2}K^\dagger K$, in which case $\Gamma \neq -\text{Im}(\Omega)$ (which is the choice in Ref. [43]). In both cases, the agreement is poor between the CMT model and the exact soution, demonstrating the inability of the CMT model to accurately capture the multi-resonant scattering response.

4.5 When is CMT accurate?

The results of the previous section prompt a more general question about the validity of conventional CMT. Our exact QCMT theory allows us to uniquely answer this question. The QCMT theory simplifies to conventional CMT when the following conditions hold: (1) the Born-scattering background term is small, and (2) the coupling strengths $D(\omega)$ and $K(\omega)$ must be approximately frequency-independent over the bandwidth of interest. The second condition is a more precise statement of the well-understood requirement [81, 33] that CMT requires high-quality-factor, well-separated resonances.

The time-domain versions of the CMT equations of Eqs. (3.4,3.5) are [45]

$$\frac{\text{d}}{\text{d}t}\boldsymbol{a}(t) = -\text{i}\Omega\boldsymbol{a}(t) + D^T\boldsymbol{c}_{\text{in}}(t) \tag{4.25}$$

$$\boldsymbol{c}_{\text{out}}(t) = S_{\text{bg}}\boldsymbol{c}_{\text{in}}(t) + K\boldsymbol{a}(t)\,, \tag{4.26}$$

which can be interpreted as the inverse Fourier transform of the frequency-domain equations, Eqs. (3.4,3.5). To identify the approximations inherent to the CMT equations, we can find the inverse Fourier transform of the frequency-domain QCMT equations, Eqs. (4.14,4.15);

for nondispersive media (for which $N(\omega)$ is the identity), analogous manipulations yield

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}(t) = -\,\mathrm{i}\Omega\boldsymbol{a}(t) + \int \mathrm{d}t' D^T(t-t')\boldsymbol{c}_{\mathrm{in}}(t')\,, \tag{4.27}$$

$$\boldsymbol{c}_{\mathrm{out}}(t) = \int \mathrm{d}t' \left\{S_{\mathrm{bg}}(t-t') + E(t-t')\right\}\boldsymbol{c}_{\mathrm{in}}(t') + \int \mathrm{d}t' K(t-t')\boldsymbol{a}(t')\,, \tag{4.28}$$

where $E(t)$ denotes the inverse Fourier transform of the Born scattering term. There are two prominent differences between the QCMT time-domain equations, Eqs. (4.27,4.28), and the CMT time-domain equations, Eqs. (4.25,4.26). First, the Born-scattering $E(t)$ term is a background contribution that is not accounted for in conventional CMT. (It cannot simply be lumped into S_{bg}, which by definition is defined in the absence of any scatterer, and thus is scatterer-independent.) Second, one can see that all of the linear relations between the mode amplitudes and the incoming- and outgoing-wave coefficients are convolutions in time, which is required due to the frequency-dependence of the relevant matrices in the frequency domain.

Consequently, one can conclude that CMT will be valid when two conditions are met: when the Born term E is negligible (i.e., single-pass scattering excites a much smaller amplitude than the incident wave), and when the coupling matrices such as $D(t)$ and $K(t)$ are sharply peaked in time; mathematically, one recovers the conventional CMT equations by neglecting $E(t)$ and assuming all other matrices are delta functions in time, or dispersionless.

In the frequency domain, we can more precisely identify the condition in which the coupling matrices can be treated as constants. Consider for example the matrix $D(\omega)$, whose i^{th} row $D_i(\omega)$ corresponds to the coupling strengths between resonance i, with resonant frequency $\widetilde{\omega}_i = \omega_i - \mathrm{i}\Gamma_i$, and all incoming-wave amplitudes. This row can be approximated as a constant row $D_i(\widetilde{\omega}_i)$ if the next term in the Taylor expansion is sufficiently small. That term will be proportional to $\omega - \widetilde{\omega}_i$ (and the first derivative of D_i); for a mode with high quality factor Q, the frequencies of interest will occupy a bandwidth proportional to $1/Q$ that will reduce the size of $|\omega - \widetilde{\omega}_i|$. For high-quality-factor, well-separated modes, conventional CMT will apply. Beyond this limit, QCMT is required.

4.6 Conclusions

We have developed a QCMT framework as an exact generalization of CMT. Compared with CMT, the coupling constants between resonances and channels are frequency-dependent and an extra non-resonant term, previously missing in CMT, appears. This framework reveals the underlying structure of scattering matrices and enables the exact decomposition and analysis of the response due to individual resonances in a complex scattering problem. It also provides guidelines for the usage of CMT, allowing a systematic approximation from the exact theory. Looking forward, this work opens multiple avenues. QCMT can serve as a modeling paradigm for the design of complex nanophotonic structures, for applications ranging from metasurfaces [82, 19, 50, 51, 83, 24, 52] to random media [53, 55] to energy harvesting devices [84, 85]. There is an emerging interest in identifying fundamental limits to response in such structures [26, 86], and the QCMT framework could be ideal for identifying new bounds via the convenient mode/channel structure of the underlying equations. The QCMT framework could be paired with known sum rules on modal densities [87, 88], or used in tandem with energy-conservation constraints [89, 90, 91] to identify the extreme limits of what is possible.

Chapter 5

Brightness Theorem in Ray Optics

In the last two chapters, we have explored the internal resonance structure of scattering matrices. In Chapter 5 and 6, we characterize bounds to scattering-matrix properties, generalizing the brightness theorem of ray optics to wave scattering. In ray optics, two conserved quantities, étendue and brightness, play a fundamental role in designing both imaging [92] and non-imaging optical systems [93]. Étendue and brightness are also intimately related to the second law of thermodynamics since the black body radiation connects the temperature of a thermal body with its power radiated over frequencies. In this chapter, we review the classical ray optics brightness theorem in Sec. 5.1, including a simple derivation from Hamiltonian mechanics and its connection to thermodynamics. In Sec. 5.3 we show its applications in optical designs to demonstrate its importance. The purpose of this chapter is to provide the background for Chapter 6, where we attempt to generalize the concepts of étendue and brightness to wave scattering.

5.1 Classical brightness theorem and étendue

In ray optics, the central objects are propagating rays in the direction perpendicular to the wavefront of the wave they represent. Each ray can be specified by its position and direction of propagation. The power flux it carries, usually called the radiance L, is proportional to the intensity of the electric field intensity associated with the ray. In the pure ray optics setting (without physical optics correction), all rays are implicitly assumed to be *incoherent*.

5.1.1 Rays in phase space

We are often interested in the flow of a bundle of rays extending over some spatial and angular range. It is then useful to view the flow of ray bundles in a phase space with a ray's transverse position and angular tilt as variables. The phase space volume of a ray bundle is called the étendue. As rays flow, the étendue representing them moves around in the phase space as illustrated in Fig. 6.1(a). In non-imaging optics, it is often desirable to concentrate rays of all angles to a surface area as small as possible as in Fig. 5.1. The phase space representation is equivalent to compressing the spatial extension of the étendue. It is natural to ask the limit of the concentration. Can we focus a bundle of rays into a single bright arrow to create the ultimate focusing in the extreme case? The answer is undoubtedly no as a consequence of the brightness theorem and étendue conservation.

In an ideal passive optical system, the étendue is conserved, and the brightness of the ray (proportional to L) is non-increasing. We will formulate these two concepts more precisely in the next section, and now let us understand them intuitively. Roughly speaking, étendue conservation means the number of rays does not change. In order words, combining rays by focusing is impossible! Consequently, the radiance L of rays will not increase (at least this is true in homogeneous material) if its power is not amplified by an active medium. Hence, if we want to focus a ray bundle to a small area by a passive optical device, the only hope is to expand its angular extension to maintain the étendue. This is precisely the picture of a lens focusing rays to its focal point.

5.1.2 Connection to thermodynamics

Thermodynamics provides a different perspective to understand combining incoherent rays is impossible. We consider a heat reservoir A transferring heat to a body B in the form of far-field radiation. We assume the surrounding has a homogeneous refraction index. Imagine a bundle of rays is emitted by a thermal reservoir as body A thermally at temperature T_A, and we are trying to focus it onto B. A thermal equilibrium between A and B is reached. Then if we can concentrate rays to increase radiance, body B will emit rays thermally with the increased radiance by the condition of thermal equilibrium and energy balance.

By the law of black body radiation, this is equivalent to stating that B now has a higher temperature T_B than T_A. Then we can connect body B with another thermal reservoir C at temperature T_A ($T_A < T_B$) to form a Carnot cycle that converts heat into work. Thus we have constructed a thermal engine between A and C with the same temperature that converts heat into work with perfect efficiency. This is a direct violation of the second law of thermodynamics formulated by Lord Kelvin.

The above argument shows that the brightness theorem and étendue conservation is also a direct consequence of the second law of thermodynamics. This fact continues to hold after we generalize everything into waves. We will not pursue this line of argument anymore but provide a clearer proof of the brightness theorem and étendue conservation from a Hamiltonian mechanics point of view.

5.2 Derivation of the brightness theorem

Analogous to the principle of stationary action in classical mechanics, Fermat's principle states the path of a ray follows the minimum optical path. The optical path is given by

$$\int \mathcal{L}(x,y,z)\mathrm{d}s = \int n(x,y,z)\mathrm{d}s = \int n(x,y,z)\sqrt{1+\dot{x}^2+\dot{y}^2}\mathrm{d}z \tag{5.1}$$

where n is the refractive index and $\mathrm{d}s$ is the length of an infinitesimal line segment. Note that in the second equation in Eq. (5.1) we have chosen a local parametrization of s in z. Following the process of classical mechanics, conjugate variables to the spatial variable x, y can be found as

$$p_x = \frac{\partial \mathcal{L}}{\partial \dot{x}} = \frac{n\dot{x}}{\sqrt{1+\dot{x}^2+\dot{y}^2}} = n\cos\alpha, \quad p_y = \frac{\partial \mathcal{L}}{\partial \dot{y}} = \frac{n\dot{y}}{\sqrt{1+\dot{x}^2+\dot{y}^2}} = n\cos\beta, \tag{5.2}$$

where $\cos\alpha$ and $\cos\beta$ are the direction cosines between the ray direction and the x and$y$ axes respectively. They can be converted into and spherical coordinates with z axis pointing in the north pole. By Liouville's theorem, the phase space volume elements, the étendue

element, is given by

$$\mathrm{d}x\mathrm{d}y\mathrm{d}p_x\mathrm{d}p_y = n^2 \underbrace{\cos\theta \mathrm{d}x\mathrm{d}y}_{=\mathrm{d}A_\mathrm{p}} \,\mathrm{d}\Omega \tag{5.3}$$

is conserved along the path of ray propagation. Here $\cos\theta \mathrm{d}x\mathrm{d}y$ is the projected area onto the xy plane, denoted by $\mathrm{d}A_\mathrm{p}$.

The radiance L of a ray is defined as the power per unit project area per unit solid angle, so $L\mathrm{d}A_\mathrm{p}\mathrm{d}\Omega$ is the power flow into a small area and angle range. In the absence of sources and losses, local energy conservation requires the local power flow

$$L\mathrm{d}A_\mathrm{p}\mathrm{d}\Omega = \frac{L}{n^2} \underbrace{n^2\mathrm{d}A_\mathrm{p}\mathrm{d}\Omega}_{\text{étendue}} \tag{5.4}$$

to be conserved throughout. Since étendue is also conserved, then we conclude that the basic radiance or the brightness of the ray $\frac{L}{n^2}$ is also conserved. If there is loss present, then the brightness can decrease. Thus we have derived the brightness theorem: brightness cannot increase in a passive ray-optical system.

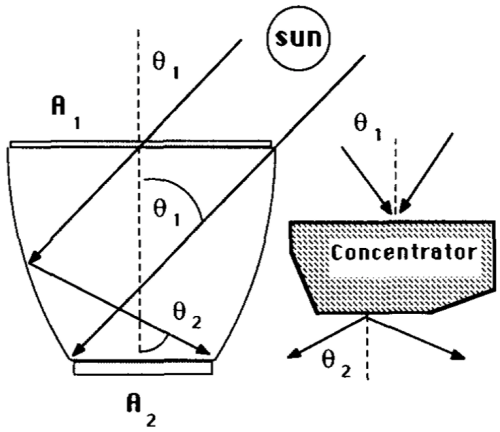

Figure 5.1: Solar concentrators focusing sunlight over angular spread θ_1 and area A_1 onto a smaller area A_2. (Reproduced from Ref. [2])

5.3 Applications to optical designs

As already mentioned in Sec. 5.1, the brightness theorem is fundamental in nonimaging system designs, where one often wishes to focus all rays into a small area. In the design

of a passive solar concentration, the brightness theorem leads to a fundamental focusing limit [2]. Consider a situation in Fig. 5.1 where we want to concentrate sunlight and we need to design a concentration with entrance opening area A_1 and acceptance cone of θ_1 and exit opening area A_2 and acceptance cone of θ_2. Suppose that the incoming rays have uniform radiance L_1 both the entrance and exit rays are in the air, so the refractive $n_1 = n_2 = 1$. The power entering the concentrator is given by integrating Eq. (5.4), which is $\Phi_1 = A_1 \int L_1 \cos\theta \mathrm{d}\Omega = \pi A_1 L_1 \sin^2\theta_1$. Similarly, we have the exiting power by $\pi A_2 L_2 \sin\theta_2$ if the rays have radiance L_2 after going through the concentration. Then the concentration ratio C defined below has a limit:

$$C = \frac{\Phi_2/A_2}{\Phi_1/A_1} = \frac{L_2 \sin^2\theta_2}{L_1 \sin^2\theta_1} \leq \frac{\sin^2\theta_2}{\sin^2\theta_1} \leq \frac{1}{\sin^2\theta_1}\,, \tag{5.5}$$

where the first inequality comes from the brightness theorem so that $L_2 \leq L_1$ and the second one by $\sin^2\theta_2 \leq 1$, a statement of the maximum angular spread gives the best spatial concentration by the conservation of étendue.

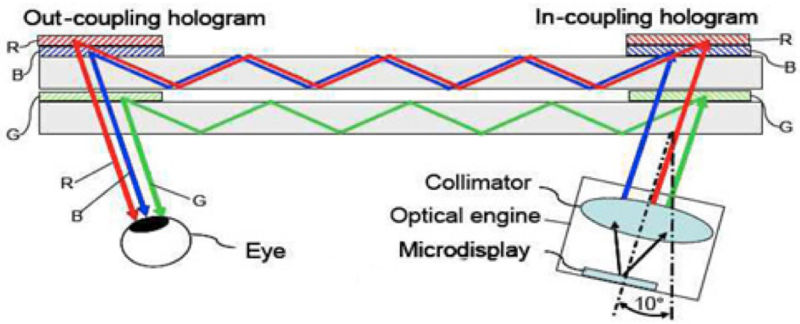

Figure 5.2: Glasses of the augmented reality goggle is a waveguide that carries lights from the microdisplay to the user's eye. The input and output grating couplers need to be designed as small as possible in order not to block the user's view while maintaining the image quality from the microdisplay. (Reproduced from Ref. [3])

For imaging system design, the conservation of étendue can be applied to understand the feasibility of certain system designs. In an imaging system, various apertures of different sizes will be present. These apertures set the spatial extension for which rays can pass. To achieve high efficiency, we need most rays to travel through these apertures unblocked. By the conservation of étendue, for a ray bundle to pass a smaller aperture, its angular

extension must be widened, but such an angular manipulation could be difficult and could lead to degradation of image qualities. This fact reveals the fundamental difficulties in head-mounted display, such as augmented or virtual reality headsets as in Fig. 5.2, where a small form factor is essential. To achieve compact designs, lights can only be coupled in and out of the overhead display via gratings with very small spatial extensions while keeping the image information intact. This imposes a big limitation on how the étendue can be squeezed and the optical design tasks involved are thus very challenging.

5.4 Difficulties in the extension to wave optics

Given the importance of the brightness theorem and étendue conservation, a natural question is whether Maxwell's equations, and more general wave-scattering physics, exhibit related conservation laws? There are two fundamental difficulties to overcome if any such generalization is possible. First, coherent waves can interfere to create focused fields and the concept of coherence is absent in ray optics. Second, in wave physics, the concept of rays ceases to be valid and we only have modes and channels.

Extending radiometric concepts such as brightness into wave systems with coherence, beyond ray optics, has been the subject of considerable study [94, 95, 96, 97, 98, 99, 100, 101, 102, 103]. Wigner functions can represent generalized phase-space distributions in such settings, and are particularly useful for "first-order optics," i.e. paraxial approximations, spherical waves, etc. Yet generalizations to the brightness theorem have not been identified because Wigner-function and similar approaches cannot simultaneously satisfy all necessary properties of a generalized radiance [96, 100, 102]. We circumvent these challenges by recognizing power transported on scattering channels as the "brightness" constrained in general wave-scattering systems in Chapter 6.

Chapter 6

Brightness Theorem for Nanophotonics

In the previous chapter, we summarized the ray-optical brightness theorem and illustrate its utility in ray optics designs. In this chapter, we develop analogous conservation laws for power flow through the scattering channels that comprise the bases of linear scattering matrices. By a density-matrix framework more familiar to quantum settings, in Sec. 6.1 we derive bounds on power concentration in scattering channels, determined by the coherence of the incident field. In Sec. 6.2, we show the ranks of the density matrices for the incoming and outgoing fields play the role of étendue, and maximal eigenvalues dictate maximum possible power concentration. For the specific case of a purely incoherent excitation of N incoming channels, we show power cannot be concentrated onto fewer than N outgoing channels, which in the ray-optical limit simplifies to the classical brightness theorem. In resonant systems described by temporal coupled-mode theory, we show in Sec. 6.4 the number of coupled resonant modes additionally restricts the flow of wave étendue through the system. The bounds require only passivity and apply to nonreciprocal systems. We discuss their ramifications in nanophotonics—for the design of metasurfaces, waveguide multiplexers, random-media transmission, and more——while noting that the bounds apply more generally to scattering in acoustic, quantum, and other wave systems.

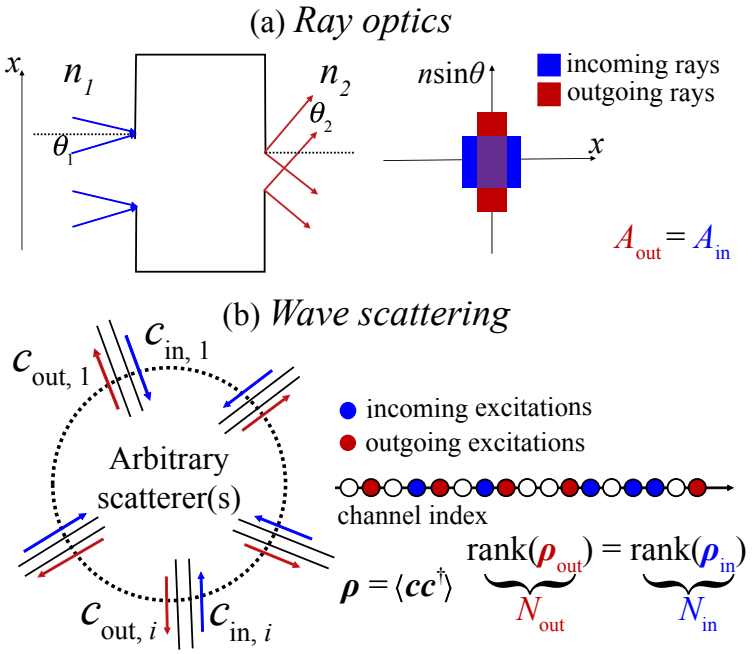

Figure 6.1: (a) In ray optics, there is a tradeoff in spatial and angular concentration of rays, by virtue of étendue conservation and the brightness theorem. (b) For general wave scattering, the scattering channels comprise the phase space. In ideal systems, the phase-space volumes are conserved: $A_{\text{out}} = A_{\text{in}}$ in (a), and $N_{\text{out}} = N_{\text{in}}$ in (b), where N denotes the number of excited channels (filled circles) or, more generally, the rank of the respective density matrix $\boldsymbol{\rho}$.

6.1 Concentration bounds

Consider generic linear wave scattering as shown in Fig. 6.1(b), in which the scattering process is described with a finite number of input and output "channels" (i.e., scattering states represented by power-normalized basis functions). We assume the scattering process is not amplifying, but does not have to be reciprocal or unitary. The channel states form a basis for incoming and outgoing fields, whose decomposition onto this basis is described by vectors $\boldsymbol{c}_{\text{in}}$ and $\boldsymbol{c}_{\text{out}}$, respectively. A scattering matrix S relates the two,

$$\boldsymbol{c}_{\text{out}} = S\boldsymbol{c}_{\text{in}}, \tag{6.1}$$

and we take the channels to be power-orthogonal, such that $\boldsymbol{c}_{\text{in}}^{\dagger}\boldsymbol{c}_{\text{in}}$ and $\boldsymbol{c}_{\text{out}}^{\dagger}\boldsymbol{c}_{\text{out}}$ represent the total incoming and outgoing powers, respectively. Perfectly coherent excitations allow for arbitrarily large concentration (e.g., through phase-conjugate optics [104, 105], perfect coherent absorbers [106, 54]), but the introduction of incoherence incurs restrictions. To describe the coherence of incoming waves, we use a density matrix $\boldsymbol{\rho}_{\text{in}}$ [107] that is the

ensemble average (denoted $\langle\cdot\rangle$, over the source of incoherence) of the outer product of the incoming wave amplitudes:

$$\boldsymbol{\rho}_{\text{in}} = \langle \boldsymbol{c}_{\text{in}} \boldsymbol{c}_{\text{in}}^{\dagger} \rangle. \tag{6.2}$$

The incoherence of the outgoing channels is represented in the corresponding outgoing-wave density matrix, $\boldsymbol{\rho}_{\text{out}} = \langle \boldsymbol{c}_{\text{out}} \boldsymbol{c}_{\text{out}}^{\dagger} \rangle = S \boldsymbol{\rho}_{\text{in}} S^{\dagger}$. Both matrices are Hermitian and positive semidefinite.

For inputs defined by some $\boldsymbol{\rho}_{\text{in}}$, how much power can flow into a single output channel, or more generally into a linear combination given by a unit vector $\hat{\boldsymbol{u}}$? If we denote $\hat{\boldsymbol{u}}^{\dagger} \boldsymbol{c}_{\text{out}}$ as $\boldsymbol{c}_{\text{out},\hat{\boldsymbol{u}}}$, then the power through $\hat{\boldsymbol{u}}$ is $\langle |\boldsymbol{c}_{\text{out},\hat{\boldsymbol{u}}}|^2 \rangle = \hat{\boldsymbol{u}}^{\dagger} \boldsymbol{\rho}_{\text{out}} \hat{\boldsymbol{u}} = \hat{\boldsymbol{u}}^{\dagger} S \boldsymbol{\rho}_{\text{in}} S^{\dagger} \hat{\boldsymbol{u}}$. The quantity $\langle |\boldsymbol{c}_{\text{out},\hat{\boldsymbol{u}}}|^2 \rangle$ is a quadratic form in $\boldsymbol{\rho}_{\text{in}}$, such that its maximum value is dictated by its largest eigenvalue [108], $\lambda_{\max}$, leading to the inequality

$$\langle |\boldsymbol{c}_{\text{out},\hat{\boldsymbol{u}}}|^2 \rangle \leq \lambda_{\max}(\boldsymbol{\rho}_{\text{in}}) \left(\hat{\boldsymbol{u}}^{\dagger} S S^{\dagger} \hat{\boldsymbol{u}} \right). \tag{6.3}$$

To bound the term in parentheses, $\hat{\boldsymbol{u}}^{\dagger} S S^{\dagger} \hat{\boldsymbol{u}}$, we consider *coherent* scattering for a new input: $\boldsymbol{c}_{\text{in}} = S^{\dagger} \hat{\boldsymbol{u}}$. For this input field, the incoming power is $\hat{\boldsymbol{u}}^{\dagger} S S^{\dagger} \hat{\boldsymbol{u}}$, while the outgoing power in unit vector $\hat{\boldsymbol{u}}$ is $|\hat{\boldsymbol{u}}^{\dagger} \boldsymbol{c}_{\text{out}}|^2 = (\hat{\boldsymbol{u}}^{\dagger} S S^{\dagger} \hat{\boldsymbol{u}})^2$. Enforcing the inequality that the outgoing power in $\hat{\boldsymbol{u}}$ must be no larger than the (coherent) total incoming power, we immediately have the identity $\hat{\boldsymbol{u}}^{\dagger} S S^{\dagger} \hat{\boldsymbol{u}} \leq 1$. (We provide an alternative proof in the Appendix B.2.) Inserting into Eq. (6.3), we arrive at the bound

$$\langle |\boldsymbol{c}_{\text{out},\hat{\boldsymbol{u}}}|^2 \rangle \leq \lambda_{\max}(\boldsymbol{\rho}_{\text{in}}). \tag{6.4}$$

Equation (6.4) is a key theoretical result of this chapter. It states that for a system whose incoming power flow and coherence are described by a density matrix $\boldsymbol{\rho}_{\text{in}}$, the maximum concentration of power is the largest eigenvalue of that density matrix. For a coherent input (akin to quantum-mechanical "pure states" [109]), there is a single nonzero eigenvalue, equal to 1, such that all of the power can be concentrated into a single channel. For equal incoherent excitation of N independent incoming states, the density matrix is diagonal with

all nonzero eigenvalues equal to $1/N$, in which case

$$\langle |\boldsymbol{c}_{\text{out},\hat{\boldsymbol{u}}}|^2 \rangle \leq \frac{1}{N}. \tag{6.5}$$

Equation (6.5) is less general than Eq. (6.4) but provides intuition and is a closer generalization of the ray-optical brightness theorem. Since the average output power per independent state must be less than or equal to $1/N$, at least N independent outgoing states must be excited, or a commensurate amount of power must be lost to dissipation. (In reciprocal systems, this bound follows from reversibility.) In Appendix B.1, we prove that Eq. (6.5) simplifies to the ray-optical brightness theorem for continuous plane-wave channels in homogeneous media.

6.2 Wave étendue

Eqs. (6.4,6.5) imply that the incoherent excitation of N inputs cannot be fully concentrated to fewer than N outputs. This motivates the identification of "wave étendue" as the number of incoherent excitations on any subset of channels (incoming, outgoing, etc.). For a density matrix $\boldsymbol{\rho}$, wave étendue can be defined as its *rank*: étendue = rank($\boldsymbol{\rho}$).

Wave étendue satisfies a conservation law similar to its ray-optical counterpart. Any square S matrix without a zero eigenvalue is full rank. The rank of $\boldsymbol{\rho}_{\text{out}} = S\boldsymbol{\rho}_{\text{in}}S^\dagger$ is then the rank of $\boldsymbol{\rho}_{\text{in}}$ (Ref. [110]), giving:

$$\operatorname{rank} \boldsymbol{\rho}_{\text{out}} = \operatorname{rank} \boldsymbol{\rho}_{\text{in}}. \tag{6.6}$$

Equation (6.6) defines conservation of wave étendue in linear scattering systems and simplifies to the classical result in the ray-optics limit (Appendix B.1). Fig. 6.1(b) depicts the rank-defined (channel-counting) definition of wave étendue. In wave-scattering systems, phase space is defined by distinct scattering channels, without recourse to the position and momentum unique to free-space states.

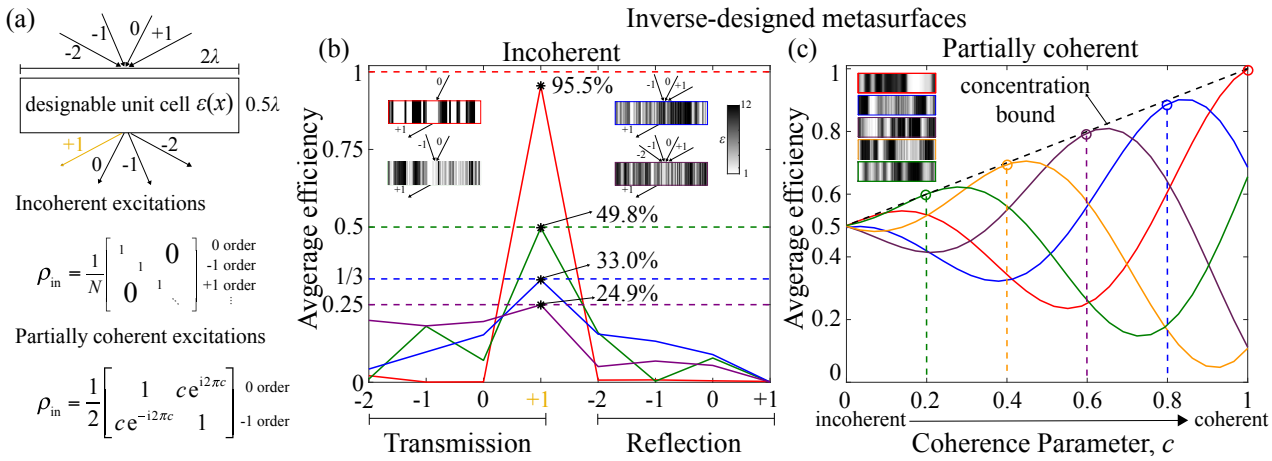

Figure 6.2: (a) A periodic metasurface element to be designed for maximal power in the +1 transmission diffraction order (yellow). We consider incoherent excitations among the four incident orders, with a diagonal density matrix, as well as partially coherent excitations between the 0 and -1 order, represented by an off-diagonal term with coherence parameter c. Inverse-designed metasurfaces closely approaching the coherence- and channel-dependent bounds are shown in (b) for incoherent excitations among up to four channels, and in (c) for partially coherent excitations between two channels. (Designs in (c) are all optimal for the fully incoherent case because $\boldsymbol{\rho}_{\text{in}}$ is a constant multiple of the identity matrix. This should not be considered a generic phenomenon when excitation powers are unevenly distributed.)

6.3 Metasurface design

To probe the channel-concentration bounds, we consider control of diffraction orders through complex metasurfaces, for potential applications such as augmented-reality optics [111, 112] and photovoltaic concentrators [113, 114, 20]. Fig. 6.2(a) depicts a designable gradient refractive-index profile with a period of 2λ and a thickness of 0.5λ. (Such an element could be one unit cell within a larger, non-periodic metasurface [115, 116, 21, 51].) For incoherent excitation of N diffraction orders, Eq. (6.5) dictates that the maximum average efficiency of concentrating light into a single output order (+1) cannot be greater than $1/N$ (dashed lines in Fig. 6.2(c)). For s-polarized light incoherently incident from orders 0 (red), -1, 0 (green), -1 ,0, +1 (blue), and -2 ,-1 ,0, +1 (purple) (20-degree angle of incidence for the 0^{th} order), we use adjoint-based "inverse design" [117, 118, 119, 120, 121, ?, 122, 123] (details in Sec. 7.2) to discover optimal refractive-index profiles of the four metasurfaces shown in Fig. 6.2(b). (Broader angular control and binary refractive-index profiles could be generated through standard optimization augmentations [120, 121], but here we emphasize the brightness-theorem consequences.) The transmission spectrum was computed by the Fourier modal method [124] with a freely available software package [125].

In Fig. 6.2(b), as the number of incoherent channels excited increases from 1 to 4, the average efficiency of the optimal structures decreases from 95.5% to 24.9%. We also probe the effects of partial coherence by varying the coherence between two input orders, per the density matrix in Fig. 6.2(a). By Eq. (6.4), maximum concentration is determined by the largest eigenvalue of $\boldsymbol{\rho}_{\text{in}}$, which is $1 - c/2$, where c is the coherence parameter. Fig. 6.2(c) shows inverse-designed structures for $c = 0.2, 0.4, 0.6, 0.8, 1$, with unique structures optimizing the response depending on the coherence of the excitation. All of the structures maximize efficiency in the incoherent $c = 0$ case, because the eigenvalues of the density matrix are degenerate and thus transmission of any state is optimal.

6.4 Étendue tranmission

A common scenario is to want to maximize power transmission from excitations on a set of incident channels (with density matrix $\boldsymbol{\rho}_{\text{inc}}$) to a distinct set of transmission channels, as depicted in Fig. 6.3. We define "étendue transmission" as the number of incoherent excitations that can propagate through the system without reflection. Eq. (6.5) dictates that at least N output channels are excited for N orthogonal inputs. If the number of transmission channels, N_{trans}, is less than $\text{rank}(\boldsymbol{\rho}_{\text{inc}})$, the incoherent excitations cannot all be concentrated onto the transmission channels, and some power must necessarily be back-scattered.

Resonance-assisted transmission, in which resonances couple the incident and transmission channels, introduces an additional constraint: the number of scatterer modes (resonances), M, coupled to the relevant channels. We consider systems that can be described by temporal coupled mode theory (CMT) as summarized in Sec. 3.1.2, wherein the scattering process is encoded in an $M \times M$ matrix Ω, comprising the real and imaginary parts of the resonant frequencies, and an $N \times M$ matrix K, denoting channel–mode coupling. In the transmission scenario of Fig. 6.3, the relevant matrix is the T-matrix ("transmission matrix"), which relates outputs on transmission channels, $\boldsymbol{c}_{\text{trans}}$, to inputs on incident channels, $\boldsymbol{c}_{\text{inc}}$: $\boldsymbol{c}_{\text{trans}} = T\boldsymbol{c}_{\text{inc}}$. In TCMT, the T-matrix for the resonance-assisted scattering component is (Appendix B.3): $T = -iK_{\text{trans}}\left(\Omega - \omega\right)^{-1} K_{\text{inc}}^{T}$, where K_{trans} and K_{inc} are the

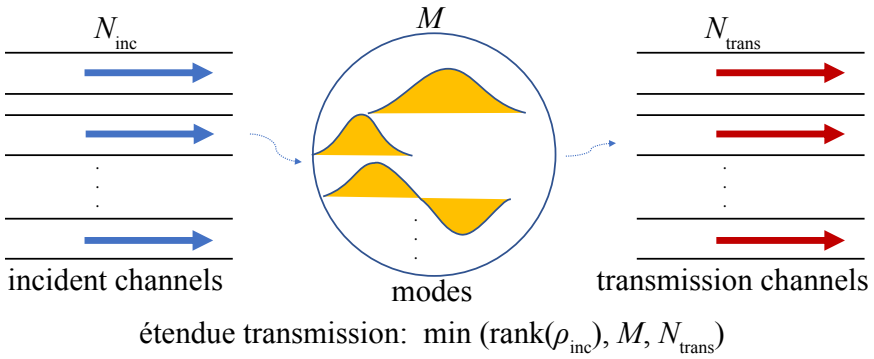

Figure 6.3: Étendue, defined as the rank of wave-scattering density matrices, is restricted in resonance-assisted transmission processes by the number of transmission channels and channel-coupled resonances in the process.

$N_{\text{trans}} \times M$ and $N_{\text{inc}} \times M$ submatrices of K denoting modal couplings to the transmission and incident scattering channels, respectively.

The maximum (average) power flow into a single transmission output is subject to the bounds of Eqs. (6.4,6.5), now in terms of a density matrix $\boldsymbol{\rho}_{\text{inc}}$ for the incident (not incoming) channels. The matrix $\boldsymbol{\rho}_{\text{trans}}$ equals $T\boldsymbol{\rho}_{\text{inc}}T^{\dagger}$. By recursive application of the matrix-rank inequality $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$ (Ref. [110]), we can see that

$$\text{rank}(\boldsymbol{\rho}_{\text{trans}}) \leq \min\left(\text{rank}(\boldsymbol{\rho}_{\text{inc}}), M, N_{\text{trans}}\right). \tag{6.7}$$

The number of orthogonal outputs is less than or equal to the *minimum* of the numbers of incident inputs, resonance modes, and transmission channels. Transmission channels and resonance modes act like apertures [126] in restricting the flow of étendue through a system.

We may also consider total transmission onto *all* N_{trans} transmission channels, i.e. $\sum_i \langle |\boldsymbol{c}_{\text{trans},i}|^2 \rangle$. Since the transmission onto a single output is bounded above by $\lambda_{\max}(\boldsymbol{\rho}_{\text{inc}})$, the total power is bounded above by the sum of the first $\text{rank}(\boldsymbol{\rho}_{\text{trans}})$ eigenvalues (Appendix B.3):

$$\sum_i \langle |\boldsymbol{c}_{\text{trans},i}|^2 \rangle \leq \sum_{i=1}^{\min(\text{rank}(\boldsymbol{\rho}_{\text{inc}}),M,N_{\text{trans}})} \lambda_i, \tag{6.8}$$

where the eigenvalues are indexed in descending order. For incoherent excitation of the N_{inc} channels, $\lambda_i(\boldsymbol{\rho}_{\text{inc}}) = 1/N_{\text{inc}}$ for all i, and the term on the right of Eq. (6.8) simplifies

to $\min\left(N_{\text{inc}}, M, N_{\text{trans}}\right)/N_{\text{inc}}$. In resonance-assisted transmission scenarios, Eq. (6.8) represents an additional constraint on power flow: in addition to the number of output channels, total power flow is further constrained by the number of distinct modes that interact with them. Étendue restrictions anywhere in the transmission process necessarily generate back reflections.

We apply the resonance-assisted-transmission bounds to CMT models of waveguide multiplexers as depicted in Fig. 6.4. There has been significant interest [127, 128, **?**, 129, 130, 131] in the design of compact junctions for routing light. In Fig. 6.4(a–c), we consider "input" waveguides and "output" waveguides coupled by a resonant scattering system. For two input waveguides, we consider three scenarios: (a) one output and two resonances, (b) one resonance and two outputs, and (c) two resonances and two outputs. In each case, a highly controlled coherent excitation can, through the appropriate design of the resonator, yield perfect transmission on resonance at the output port. But a *robust* design, impervious to noise or other incoherence, may be required, and such noise would introduce incoherence that is subject to the bound of Eq. (6.8). In each case, we optimize TCMT model parameters to maximize transmission for all phase differences between two inputs. Device (c) maintains perfect transmission, wheres devices (a) and (b) are highly sensitive to noise, as predicted by the restrictions to étendue flow.

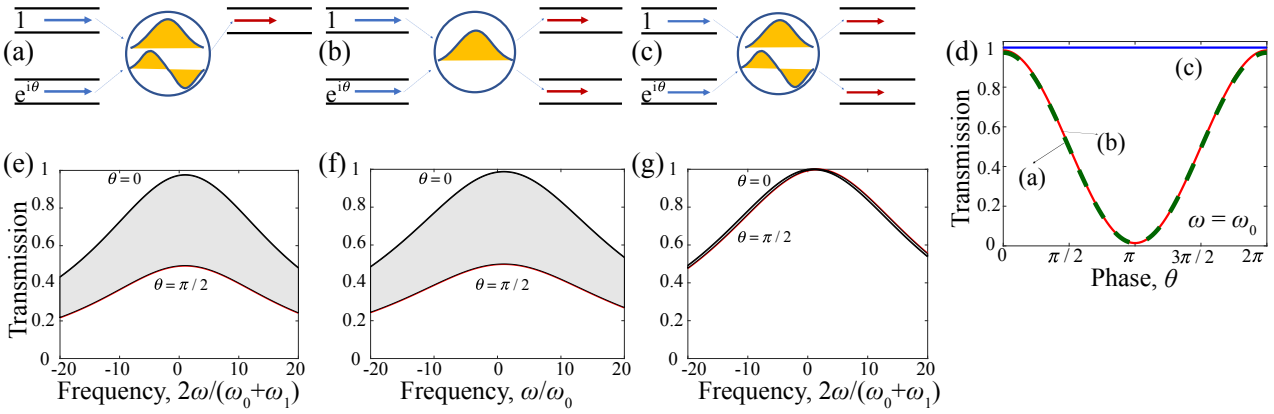

Figure 6.4: The robustness of waveguide junctions is susceptible to étendue restrictions. For two input channels, we consider (a) one output, (b) one mode, and (c) no restrictions. (e)-(g) Transmission for (a)-(c) with input phase angles in $\theta = [0, \pi/2]$. (d) Transmission as a function of phase, on resonance. Case (c) is designed to be almost perfectly insensitive to phase; such designs are impossible in cases (a) and (b).

6.5 Conclusion

The channel-concentration bounds and wave-étendue concept generalize classical ray-optical ideas to general wave scattering. In addition to the nanophotonic design problems considered, there are numerous potential applications. First, they resolve how to incorporate polarization into ray-optical étendue, showing unequivocally that polarizing unpolarized light requires doubling classical étendue, an uncertain conjecture in display design [132, 133]. Moreover, the natural incorporation of nonreciprocity into the bounds is of particular relevance given the emerging interest in nonreciprocal photonics [134, 135, 136] and acoustics [137], and places constraints on many of these systems (extensions to time-modulated CMT systems should be possible). Another additional application space is in random-scattering theory [138, 55, 53, 139]. There is significant interest in controlling the scattering channels of opaque optical media comprising random scatterers. The framework developed herein may lead to fundamental limits to control in such systems.

Chapter 7

Optimization Frameworks in Nanophotonics

The remainder of this thesis considers the design of scattering response from an optimization perspective. Design processes in nanophotonics can be viewed as optimizing the material distribution $B(\boldsymbol{r})$ iteratively; it is a design problem in space. Although physically completely different, the pulse design problem in quantum control, engineering the optimal pulse $\varepsilon(t)$ at each time t to drive the system into desired states, is a design problem in time. They are all scattering problems that share similar mathematical structures. As a result, techniques for nanophotonics designs [90, 140] can be applied to quantum control. In particular, the framework for identifying fundamental limits in scattering developed in nanophotonics is capable of obtaining "quantum speed limits" in quantum control.

In this chapter, we summarize the framework of nanophotonics design, in particular inverse design method for finding local optima (Sec. 7.2) and the convex relaxation for identifying global bounds (Sec. 7.3). Although the convex relaxation framework in nanophotonics is not the contribution of this thesis, many concepts, such as polarization current and optical theorem, are more naturally discussed in the Maxwell setting. Then in Chapter 8, we present a new application of computational bounds for the Schrödinger equation in quantum control.

7.1 Nanophotonics design

In nanophotonics design, a desired figure of merit (FOM) is first identified, followed by maximizing it through changing the material distribution $B(\boldsymbol{r})$ using certain optimization procedures. The final design is given by the optimized $B(\boldsymbol{r})$ in space. We denote the general form of a FOM as a function $f(\psi(\boldsymbol{r}), B(\boldsymbol{r}))$ of the field $\psi(\boldsymbol{r})$ and the material distribution $B(\boldsymbol{r})$ in some design region V. These two quantities are related since they satisfy Maxwell equations Eq. (2.3), so a typical nanophotonics design problem can be written formally as

$$\max_{\psi(\boldsymbol{r}), B(\boldsymbol{r})} f(\psi(\boldsymbol{r}), B(\boldsymbol{r})) \tag{7.1}$$
$$\text{subject to Maxwell equations}\,.$$

Due to experimental or computational constraints, it is necessary to discretize V into smaller units (possibly as small cubes) with some minimum size, and the value $B(\boldsymbol{r})$ within each unit is constant. As a result, we only have control over a finite number n of design parameters/degrees of freedom as the value $B(\boldsymbol{r})$ within each unit. The variable n is often large so that we can ensure a large enough design parameter space for good outcomes.

As an example, in Sec. 6.3, the design region V is the unit cell of a periodic grating, and the design parameters are the $B(\boldsymbol{r})$ in the horizontal slices that the unit cell is being divided into. The target of the design is to maximize $f = \left|\psi_{\text{tar}}^{\dagger}\psi\right|^2$, where ψ_{tar} is a target field, so the FOM here measures the "overlap" between the designed response and the target. In particular, ψ_{tar} is the linear combination of different diffraction orders. In Fig. 7.1 we show a few more examples of inverse designed optical devices. Fig. 7.1(a) shows a Z-bend photonic crystal waveguide that achieves almost perfect transmission efficiency [4]. In Fig. 7.1(b) a two-channel splitter is able to demultiplex incoming waves of two different frequencies into separate channels. Fig. 7.1(c) shows a metasurface as a flat lens focusing a plane wave with 5 degree incidence angle into a diffraction-limited spot [6]. In general, the formulation in Eq. (7.1) covers almost everything of practical interest.

During the design process, we obtain approximate solutions to Maxwell equations via standard simulation techniques such as finite difference, finite element, or integral equation

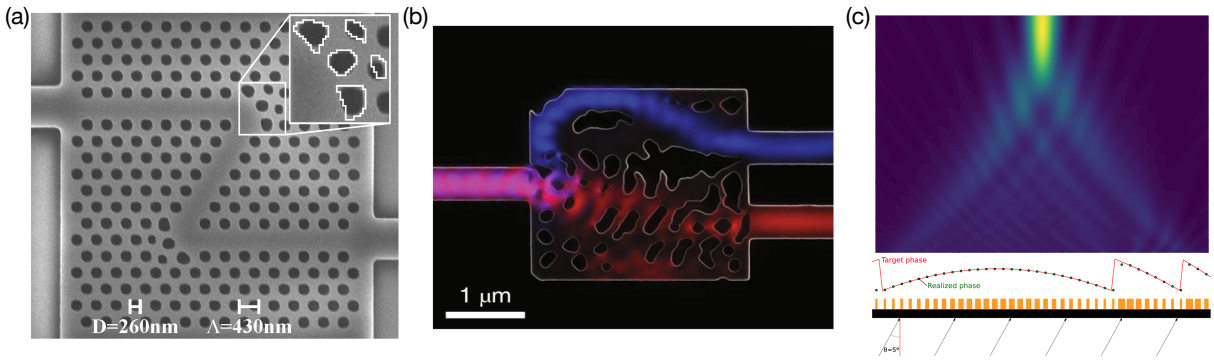

Figure 7.1: (a) A Z-bend of photonic crystal waveguides achieving perfect transmission efficiency [4]. (b) A two channel wavelength splitter [5]. (c) A metasurface that focuses 5 degree incident plane wave into a diffraction-limited spot [6].

approach. We will see that the "inverse design" method permits us to compute the gradients efficiently, *independent* of the number of design variables. As a result, the solution simulation is the most computationally expensive part of the design and thus limits the accuracy and size of the design.

7.1.1 Binary versus grayscale designs

It is common that the designer has the freedom to alter $B_{\text{bg}}(\boldsymbol{r})$ by a constant ΔB in each unit in V. As a result, $B(\boldsymbol{r})$ in each of those units can either be $B_{\text{bg}}(\boldsymbol{r})$ or $B_{\text{bg}}(\boldsymbol{r}) + \Delta B$. Such alteration can be caused by adding material of index ΔB into the unit or by removing existing material through etching. This scenario will be referred to as binary designs in the rest of this work. All three cases in Fig. 7.1 are binary designs.

Different from binary designs, greyscale designs allows $B(\boldsymbol{r})$ in each unit to vary continuously from $B_{\text{bg}}(\boldsymbol{r})$ to $B_{\text{bg}}(\boldsymbol{r}) + \Delta B$. Metasurface designs in Sec. 6.3 have continuously varying material index so they are greyscale designs. If V is divided fine enough so each unit is tiny, by the theory of homogenization, given any greyscale design, there always exits a binary design that creates the same scattering response, so these two different design cases are mathematically equivalent so we can pick whichever is convenient. We will see that binary designs are suitable for bounds, while greyscale methods are convenient for gradient-based local optimization.

7.2 Local optimization approach: inverse designs

Since the dependence of $\psi(\boldsymbol{r})$ on $B(\boldsymbol{r})$ is highly nonlinear and nonconvex, one implication of the binary design is that these optimization problems in nanophotonics are combinatorially hard with a large number of design parameters, so it is impossible to expect a globally optimal design in general. The local approach to nanophotonics design is to first obtain the gradient of FOM f with respect to $B(\boldsymbol{r})$ in the design region V; then a standard gradient-based optimization algorithm can produce a local optimal design. In the binary case, the gradient can only be defined with respect to boundary shifts between those two materials or with respect to topological changes (e.g. perforations). In the greyscale case, the value of $B(\boldsymbol{r})$ can be changed continuously, so gradient can be mathematical defined as usual. (Since binary and grayscale cases are equivalent, the greyscale results from this local optimization approach still reflect what to expect for the binary case. In addition, filtering or penalty techniques can force the greyscale results into binary ones if needed.)

The "inverse design" approach allows a large number of design parameters with almost no extra cost than intuition-based designs, where we only allow simple shape scatterers (such as circles or rods) with few geometric parameters to tune. Consequently, the inverse design method permits an efficient search in an enormous design space, so it generally produces superior results than intuition-based designs. Inverse designed structures sometimes approach the fundamental limits imposed by physical laws and thus can be applied to check the quality of those theoretical limits as in Sec. 6.3.

7.2.1 Direct gradient calculation with $2n$ simulations

The local optimization scheme only works provided we can obtain gradients efficiently. A naive way to find out the gradient f of $B(\boldsymbol{r})$ is to solve the scattering problem first to get the value of f, then perturbing the value of $B(\boldsymbol{r})$ in unit cell i by δB_i and solve the problem again to get a new f_i. Then the gradient is given obtained from the finite difference $\delta f/\delta B_i$. (Because δB_i is a vector, the gradient expression here is a also matrix whose elements are gradients of f with respect to each element in δB_i.) Since the number of design parameters n is often large, it seems that we need $2n$ simulation to get the full gradient for a *single*

optimization step, which is prohibitively expensive! Fortunately, the inverse design method reduces the cost of the task of finding gradients into solving the scattering problem twice and twice only, independently of the number n.

7.2.2 Adjoint gradient calculation with 2 simulations

The key to such cost reduction is to switch the point of view between "sources" and "detections" by reciprocity. The expensive method above for finding gradient can be understood physically as detecting the change in FOM δf generated by the source due to the perturbation δB_i for all i. However, by reciprocity, there is no distinction between "source" and "detection", and they can be switched. As illustrated in Fig. 7.2, we can view the process equivalently as detecting the adjoint field in each unit $\psi_{\text{adj}}(\boldsymbol{r})$ generated by current due to $\frac{\partial f}{\partial \psi(\boldsymbol{r})}$, which is related to δf above. In this switched view, $\psi_{\text{adj}}(\boldsymbol{r})$ in each unit can be found by a single simulation, and this is where the cost reduction comes from.

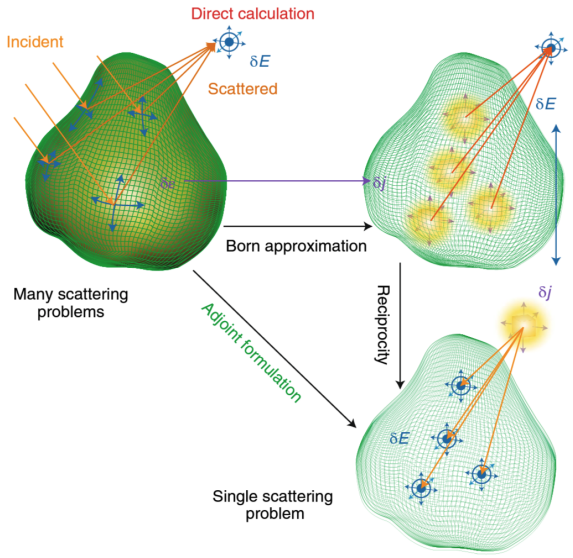

Figure 7.2: Reciprocity allows us to treat the gradient of the target as the adjoint source. This enable us to find the gradient of all design parameters with a extra adjoint simulation only, independent of the number of design parameters. (Reproduced from Ref. [5].)

The above explanation summarizes the main ingredients of inverse design and the detailed derivation can be found in Ref. [5, 23]. Here we only provide the formula for finding the gradient: given the current design $B(\boldsymbol{r})$, if we have solved the original scattering problem and obtained its solution $\psi(\boldsymbol{r})$, then the gradient can be obtained by doing another

adjoint simulation with source as $\frac{\partial f}{\partial \psi(\boldsymbol{r})}$ to obtain the adjoint field $\psi_{\text{adj}}(\boldsymbol{r})$, then

$$\delta f \propto \int_{\delta B_i} \text{Im}\left(\psi_{\text{adj}}^T(\boldsymbol{r})\delta B_i \psi(\boldsymbol{r})\right) , \tag{7.2}$$

where the integration is over the unit i where $B(\boldsymbol{r})$ is perturbed by δB_i. It is clear from Eq. (7.2) that we only need to solve two similar scattering problems: one for the original problem to get ψ and the other to get ψ_{adj} with source $\frac{\partial f}{\partial \psi(\boldsymbol{r})}$. With this method for computing the gradient efficiently for a large number of design parameters, gradient-based optimization method is able to converge quickly to a local optimal design.

7.2.3 Deficiency of local optimization

Despite the fact that locally optimized designs that are often superior to those from intuition-based methods, we are often stuck among numerous low quality local optima. It is not uncommon to restart from many different initial starting points many times to obtain a good local optimal design, especially when there is a large number of design parameters. In addition, the local optimization approach fails to provide insights whenever local optima designs are far from ideal. We want to explain why the target cannot be achieved, whether it is because the wrong material is used or simply the design region is too small for a complex design problem.

To remedy these drawbacks, a complementary approach is developed to relax the design problem in Eq. (7.1) into a convex optimization problem, which can be solved efficiently. In contrast to the local approach, this global approach can identity what is possible and what is not in a design setup, providing insights into what can be further improved. In the next section, we describe how such procedures are carried out.

7.3 Global bounds: convexification of nanophotonics designs

The convex relaxation is motivated by the observation that the Maxwell constraint in Eq. (7.1) can be turned into quadratic constraints of $\psi(\boldsymbol{r})$, and most important class of

FOM, such as scattering extinction, cross section, power flow, and field intensity, are also quadratic in $\psi(\boldsymbol{r})$. This makes it possible to transform the whole problem in Eq. (7.1) into a quadratically constraint quadratic program (QCQP), where standard relaxation method is available. We discuss this transformation into QCQP in Sec. 7.3.1, 7.3.2. Then we show how the resulting QCQP can be relaxed into a convex problem (Sec. 7.3.3) by standard semidefinite relaxation. Solving the convex problem gives us a bound to the original design problem in Eq. (7.1). We note that the application of this framework in nanophotonics is not the contribution of this thesis and can be found in Ref. [90, 140].

This framework for finding the fundamental bound of a nanophotonics design problem is first developed with the integral formulation of Eq. (2.6), equivalent to the differential version Maxwell's equations, so we start with Eq. (2.6) that contains the optimization variable $\Delta B(\boldsymbol{r}) = B(\boldsymbol{r}) - B_{\text{bg}}(\boldsymbol{r})$ and $\psi(\boldsymbol{r})$. Next, we first describe how to relax Eq. (7.1) into a QCQP with a single constraint representing the optical theorem, which states energy is conserved globally. This approach is conceptually important but is superseded in terms of the tightness of bounds produced by its improved version, the "D-matrix" approach, which adds in many more constraints from local energy conservation as a natural extension of the optical theorem constraint.

7.3.1 Optical theorem as a single quadratic constraint

To form quadratic constraints in ψ, we need to consider binary design problems and turn this binary nature into quadratic constraints. In this binary case, the product of these variables $\Delta B(\boldsymbol{r})\psi(\boldsymbol{r})$, the polarization current, is either $\Delta B\psi(\boldsymbol{r})$ or 0, where ΔB is a constant matrix as required by binary designs. We can equally represent this with the new variable $\phi(\boldsymbol{r}) = \Delta B(\boldsymbol{r})\psi(\boldsymbol{r})$ by the following two equations:

$$\begin{aligned} \psi_{\text{inc}}(\boldsymbol{r}) + \quad & \mathrm{i}\omega \textstyle\int \Gamma_{\text{bg}}(\boldsymbol{r},\boldsymbol{r}')\phi(\boldsymbol{r}') - \Delta B^{-1}\phi(\boldsymbol{r}) = 0\,, \quad \text{for } \boldsymbol{r} \text{ where } \Delta B(\boldsymbol{r}) = \Delta B\,, \\ & \phi(\boldsymbol{r}) = 0\,, \qquad\qquad\qquad\qquad \text{for } \boldsymbol{r} \text{ where } \Delta B(\boldsymbol{r}) = 0\,. \end{aligned} \tag{7.3}$$

At each spatial point $\boldsymbol{r}$, at least one of the above equations holds, so the product of these two will hold for any $\boldsymbol{r}$ in V. We form the product of these two equations with a constant

6 by 6 matrix D in between, obtaining a quadratic relation in ϕ at each $\boldsymbol{r}$:

$$\phi^\dagger(\boldsymbol{r})D\left(\psi_{\text{inc}}(\boldsymbol{r})+\mathrm{i}\omega\int\Gamma_{\text{bg}}(\boldsymbol{r},\boldsymbol{r}')\phi(\boldsymbol{r}')-\Delta B^{-1}\phi(\boldsymbol{r})\right)=0, \tag{7.4}$$

The variable $\phi^\dagger$ is the Hermitian conjugate of ϕ. Note that in Eq. (7.4) the spatial dependence of $B(\boldsymbol{r})$ disappears; it is embedded into the quadratic dependence of ϕ. Such "domain obliviousness" is crucial for obtaining bounds because the quadratic constraints of ϕ apply to any field and design that satisfies Maxwell equations. The extra D here is to capture the tensor nature of B so that we have included all possible physically permissible constraints.

Instead of using Eq. (7.4) as constraints at every $\boldsymbol{r}$, we first use the imaginary part of its integrated version over V with D the identity matrix, because the formed single constraint is the optical theorem:

$$\begin{aligned}&\underbrace{\frac{\omega}{2}\,\mathrm{Im}\int\phi^\dagger(\boldsymbol{r})\psi_{\text{inc}}(\boldsymbol{r})}_{=P_{\text{ext}}}\\ =\;&\underbrace{\frac{\omega}{2}\,\mathrm{Im}\iint\phi^\dagger(\boldsymbol{r})\mathrm{i}\omega\Gamma_{\text{bg}}(\boldsymbol{r},\boldsymbol{r}')\phi(\boldsymbol{r}')}_{=P_{\text{scat}}}+\underbrace{\frac{\omega}{2}\,\mathrm{Im}\int\phi^\dagger(\boldsymbol{r})(-\Delta B^{-1})\phi(\boldsymbol{r})}_{=P_{\text{abs}}},\end{aligned} \tag{7.5}$$

a statement of global energy conservation: the extinction power P_{ext} equals to the sum of scattered power P_{scat} and absorbed power P_{abs} .

Since energy conservation is automatically satisfied by Maxwell equations, any field that satisfies Maxwell equations will necessarily satisfy Eq. (7.5). Consequently, replacing the Maxwell constraint in Eq. (7.1) by the Eq. (7.5) is a relaxation to the original problem. Then if the FOM f is a quadratic function of ψ, thus of ϕ, we have succeeded in relaxing Eq. (7.1) into a QCQP:

$$\begin{gathered}\max_{\phi(\boldsymbol{r})} f(\phi(\boldsymbol{r}))\\ \text{subject to the optical theorem Eq. (7.5).}\end{gathered} \tag{7.6}$$

The Lagrangian dual of Eq. (7.6) can be found analytically and is always convex, so it can be solved efficiently to upper bound the original QCQP, providing a bound to the original

design problem Eq. (7.1). (In fact, strong duality holds even though the QCQP is not convex [141].) This is precisely the strategy in Ref. [142] for finding bounds. It turns out that we can keep all possible constraints in on top of the optical theorem to obtain even tighter bounds. This is called the "D-matrix" approach and will be explained in the next section.

7.3.2 Local conservation laws as infinite quadratic constraints

We have shown in Sec. 7.3.1 that the origin design problem Eq. (7.1) with quadratic FOM f can be relaxed into a QCQP, whose dual can be solved efficiently to obtain a bound to Eq. (7.1). The QCQP has a single constraint based on the optical theorem out of Eq. (7.4). It turns out that we can use all possible constraints in the form of Eq. (7.4) for all $\boldsymbol{r}$ in the design region V and all possible constant matrix D. Hence, we reformulate Eq. (7.1) into the "D-matrix" form:

$$\begin{gathered} \max_{\phi(\boldsymbol{r})} f(\phi(\boldsymbol{r})) \\ \text{subject to Eq. (7.4) for all } \boldsymbol{r} \text{ in } V \text{ and } D, \end{gathered} \tag{7.7}$$

where $f(\phi)$ is a quadratic function of ϕ. It turns out that the form Eq. (7.7) is equivalent to Eq. (7.1) [90]. This makes sense intuitively since constraints Eq. (7.7) comes from material placement at each point $\boldsymbol{r}$ and the tensor natural of the material is captured by D; the number of constraint in Eq. (7.7) is tantamount to all possible design parameters in Eq. (7.1).

Similar to solving the dual of Eq. (7.6), Eq. (7.7) also admits a standard convex relaxation that upper bounds Eq. (7.7). Since the optical theorem comes from the linear combination of a subset of constraints here, bounds from the convex relaxed version of Eq. (7.7) are expected to be tighter than those from the dual of Eq. (7.6). Next, we describe how the procedure of the convex relaxation of Eq. (7.7).

7.3.3 Semidefinite relaxation of QCQP

After standard computational discretization, the infinite set of quadratic constraints labeled $\boldsymbol{r}$ becomes finite and its size equals the number of computational grids m. After

discretization, Eq. (7.7) typically takes the following form:

$$\begin{aligned} &\max_{\boldsymbol{\phi}} \quad \boldsymbol{\phi}^\dagger A \boldsymbol{\phi} + 2\,\mathrm{Re}\,\boldsymbol{b}^\dagger \boldsymbol{\phi} \\ \text{subject to} \quad &\boldsymbol{\phi}^\dagger G_i \boldsymbol{\phi} + 2\,\mathrm{Re}\,\boldsymbol{c}_i^\dagger \boldsymbol{\phi} = 0 \quad \text{for } i = 1, \ldots, 2m\,, \end{aligned} \tag{7.8}$$

where $\boldsymbol{\phi}$ is a m by 1 vector from the discretizated $\phi(\boldsymbol{r})$. Those m by m Hermitian matrices A, G_i and m by 1 vectors $\boldsymbol{b}, \boldsymbol{c}_i$ are from the discretized operators and functions from the FOM f and Eq. (7.4). Here we have $2m$ constraints, two times the grid number, because the real and imaginary part of constraints from Eq. (7.4) are satisfied separately for computational convenience.

QCQPs in Eq. (7.8) are known to be NP-hard to solve, and we want to relax it into a convex problem for finding its upper bound. Semidefinite relaxation (SDR) often produces the tightest result among these relaxation methods [143], so we adopt this approach here. SDR is based on the following equivalence for a quadratic form with $A, \boldsymbol{b}$:

$$\begin{aligned} &\boldsymbol{\phi}^\dagger A \boldsymbol{\phi} + 2\,\mathrm{Re}\,\boldsymbol{b}^\dagger \boldsymbol{\phi} \leftrightarrow \begin{pmatrix} \boldsymbol{\phi} \\ 1 \end{pmatrix}^\dagger \begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{b}^\dagger & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi} \\ 1 \end{pmatrix} \\ \leftrightarrow \quad &\mathrm{Tr}\left\{ \begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{b}^\dagger & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi}\boldsymbol{\phi}^\dagger & \boldsymbol{\phi} \\ \boldsymbol{\phi}^\dagger & 1 \end{pmatrix} \right\} \\ \leftrightarrow \quad &\mathrm{Tr}\left\{ \begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{b}^\dagger & 0 \end{pmatrix} X \right\} \quad \text{with } X \geq 0,\ \mathrm{rank}(X) = 1,\ X_{m+1,m+1} = 1\,, \end{aligned} \tag{7.9}$$

if we replace the optimization variable $\boldsymbol{\phi}$ by an $m+1$ by $m+1$ semidefinite matrix X with rank one, and impose the right bottom element $X_{m+1,m+1}$ to be 1. We can apply this transform to the objective and all quadratic constraints, resulting in an equivalent problem in X to Eq. (7.8) in $\boldsymbol{\phi}$. We see that the fundamental difficult is revealed: the rank one constraint is not convex. SDR is done by simply ignoring the rank one constraint, so we

have a linear program over the cone of semidefinite matrices:

$$\begin{aligned} & \max_X \operatorname{Tr}\left\{\begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{b}^\dagger & 0 \end{pmatrix} X\right\} \\ \text{subject to} \quad & \operatorname{Tr}\left\{\begin{pmatrix} G_i & \boldsymbol{c}_i \\ \boldsymbol{c}_i^\dagger & 0 \end{pmatrix} X\right\} = 0 \quad \text{for } i = 1, \ldots, 2m\,, \\ & X_{m+1,m+1} = 1\,, X \geq 0\,, \end{aligned} \tag{7.10}$$

which is a convex problem and can be solved via interior point methods [144, 145]. The solution provide an upper bound to Eq. (7.8). As a result, we have obtained fundamental bounds to the original design problem Eq. (7.1). Any design $B(\boldsymbol{r})$ surpasses this upper-bound will necessarily violate one of the constraints in Eq. (7.4), a violation of local energy conservation.

7.4 Summary

This chapter introduced the general framework of nanophotonics design, including both the local and global optimization approach. The local optimization approach is known as "inverse design", where gradients of objectives with respect to a large number of design parameters can be computed efficiently. This approach often leads to high-performance and nonintuitive designs. Still, it usually requires repetitive initial restarts to get out of low quality local optima and provides little physical insights into final results. As a complementary approach, we introduced a global approach, where the initial nonconvex optimization problem is turned into a quadratically constraint quadratic program. Problems of this type can be relaxed into a convex problem via semidefinite relaxation, which can be solved for obtaining fundamental bounds to the original design problem. In the next chapter, we apply this bound framework to pulse design problems in quantum control.

Chapter 8

Fundamental Limits in Quantum Control

In this chapter, we develop a framework for computing fundamental limits to what is possible via the control of quantum systems. We show that quantum control problems can be transformed to quadratically constrained quadratic programs (QCQPs), with generalized probability conservation laws as the constraints, adapting a mathematical approach recently developed for light-matter interactions summarized in Sec. 7.3. Concepts and derivations in this chapter are conceptually similar to those in Chapter 7 and Table. 8.1 contains the comparison between concepts in these two fields. Very often these concepts are more intuitive in the setting of nanophotonics, so the reader can use Table. 8.1 to refer to Chapter 7.

The QCQP formulation enables global bounds via relaxations to semidefinite programs [146, 145]. We demonstrate the power and utility of our method on three prototype systems: (1) three-level system driving, where our bounds incorporate sophisticated information about the interference between levels and can account for constraints on undesirable transitions (as needed in transmons [147], for example), (2) upper bounds to the suppression of decoherence, and (3) the maximum fidelity of a control-based implementation of a single-qubit Hadamard gate. In each case we supplement our bounds with many local-optimization-based solutions, showing that they come quite close to (and in some cases achieve) our bounds, suggesting that our bounds are tight or nearly so. Our framework applies to

open and closed systems, can be extended to related domains in NMR [148, 149, 150] and quantum complexity [151, 152, 153], and should reveal the limits of what is possible with quantum control.

	Nanophotonics	Quantum control
Fields	$\psi(\boldsymbol{r}) = (\boldsymbol{E}(\boldsymbol{r}) \quad \boldsymbol{H}(\boldsymbol{r}))$	$\lvert\psi(t)\rangle$
Differential equation	Maxwell equations	Schrödinger equation
Integral form	Integral equation in Eq. (2.6)	Dyson equation in Eq. (8.1)
Design parameters	Material in space $B(\boldsymbol{r})$	Driving pulse in time $\varepsilon(t)$
"Polarization current"	$\phi = \Delta B(\boldsymbol{r})\psi(\boldsymbol{r})$	$\Phi = \varepsilon(t)H_c(t)U(t)$
Typical objectives	Max intensity $\left\lvert\psi_{\text{tar}}^{\dagger}\psi(\boldsymbol{r})\right\rvert^2$	Max probability $\lvert\langle\psi_{\text{tar}}\rvert\psi(t)\rangle\rvert^2$
Local optimization	Inverse design	GRAPE etc.
Optical theorem	Eq. (7.5)	Eq. (8.4)
D-matrix constraints	Eq. (7.4)	Eq. (8.3)

Table 8.1: Comparison between similar concepts in nanophotonics design and quantum control pulse engineering.

8.1 Quantum control landscape

Quantum control [154, 155, 156, 157, 158] refers to the design and synthesis of efficient control sequences that drive a quantum system to maximize the desired objective, such as maximizing overlap with a target state or minimizing error in the implementation of a gate operation. Recent experiments have demonstrated the power of optimal control for wide-ranging applications [159, 160, 161, 162, 163, 164].

Because the wave function $|\psi(t)\rangle$ that represents a quantum state is nonlinear in the control parameter $\varepsilon(t)$, it is generically difficult to identify globally optimal controls. One strategy is to use local numerical optimization over the control parameters (e.g. GRAPE [148, 165, 166, 167], the Krotov method [168, 169, 170, 171, 172, 173], and CRAB [174, 175]), optimizing over many initial conditions in the hopes of identifying high-performance local optima. Yet, except in the simplest systems, one is left uncertain about the best performance possible.

Alternatively, there are a variety of global bounds [176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 7, 8]; most famously, the

Mandelstam–Tamm (MT) bound. The MT bound is a prototype of "quantum speed limits," which more generally have varying levels of complexity but are essentially time-energy uncertainty relations [176, 177, 178, 179, 182, 186, 187, 194, 195, 7, 8, 196]. The energy measure is typically a matrix norm of the Hamiltonian, but more complex details of the system interactions are not captured. Another class of bounds is obtained by analytically solving Pontryagin's maximum principle [197], which is only possible in simple cases such as two-level systems [180, 181, 184, 185, 188, 189, 190]. Consequently, meaningful, accurate bounds cannot be computed for most quantum control systems of interest.

8.2 Formulation of bound framework

We consider a Hamiltonian of the form $H_0(t) + H_c'(t) = H_0(t) + \varepsilon(t)H_c(t)$, where H_0 is the non-controllable part of the Hamiltonian, H_c' is the controllable part, and ε is the control parameter to be optimized. We assume the control parameter is bounded between 0 and $\varepsilon_{\max}$ (any other minimum value can be shifted to 0 by replacing H_0 with $H_0 + \varepsilon_{\min}H_c$). Our method generalizes to any number of control parameters (cf. Appendix C.4), but for simplicity, we assume one throughout this chapter. Any smooth, continuous, bounded control can be approximated with arbitrary accuracy by a "bang–bang" binary control that only takes the values 0 and $\varepsilon_{\max}$ (cf. Appendix C.2), so we use bang-bang controls in our formulation. (Similar to the equivalence between binary and grayscale design in Sec. 7.1.1.) Instead of the differential Schrödinger equation for the time-evolution operator $U(t, t_0)$ (for an initial time t_0), we instead start with an integral form (equivalent to the Dyson equation [198, 199] in the interaction picture):

$$U(t, t_0) = U_0(t, t_0) - \frac{i}{\hbar}\int_{t_0}^{T} G_0^+(t, t')H_c'(t')U(t', t_0)\,\mathrm{d}t', \tag{8.1}$$

where U_0 and G_0^+ are the time-evolution operator and retarded Green's function in the absence of controls (i.e., for $H_0(t)$), and T is the final time. To derive conservation laws, we start by taking the product of Eq. (8.1) with $U^\dagger(t, t_0)H_c'(t)D_i(t)$ from the left and integrating

from an initial time t_0 to T:

$$\begin{aligned}
&\int_{t_0}^{T} U^\dagger(t,t_0) H_c'(t) D_i(t) U(t,t_0)\,\mathrm{d}t \\
&+ \frac{i}{\hbar}\int_{t_0}^{T}\int_{t_0}^{T} U^\dagger(t,t_0) H_c'(t) D_i(t) G_0^+(t,t') H_c'(t') U(t',t_0)\,\mathrm{d}t\,\mathrm{d}t' \\
&= \int_{t_0}^{T} U^\dagger(t,t_0) H_c'(t) D_i(t) U_0(t,t_0)\,\mathrm{d}t. \qquad (8.2)
\end{aligned}$$

The variable $D_i(t)$ can be any time-dependent operator and is an optimization hyperparameter below in Eq. (8.10); intuitively, allowing different possible choices of D_i enables the isolation of particular times and elements in Hilbert space for which Eq. (8.2) should be satisfied. The variable H_c is effectively a renormalization that simplifies the probabilistic interpretation below; equivalently, it can be omitted. The constraint of Eq. (8.2) depends on both the time-evolution degrees of freedom $U(t)$ and the control variable degrees of freedom $\varepsilon(t)$. However, if we define a new variable $\Phi(t) = \varepsilon(t) H_c(t) U(t,t_0)$, this variable (the analog of a polarization field in electrodynamics [200, 90]) can subsume both. Crucially, we can replace any instance of $\varepsilon(t)$ with $\varepsilon_{\max}$. This can be thought of as a two-step simplification: one could restrict the domains of the integrals to only times in which the control is on, in which case such a replacement is trivial. Next, $\varepsilon(t)$ only appears in a term of the form $\Phi^\dagger \varepsilon^{-1} \Phi$, which is zero even when $\varepsilon(t) = 0$, due to the quadratic dependence on Φ. Hence we can extend the domain of the integrals back to the entire time interval from t_0 to T. Such "domain-obliviousness" [90] arises from our inclusion of $\varepsilon(t)$ and $U^\dagger(t,t_0)$ in the product term. Finally, we have the constraints:

$$\int_{t_0}^{T}\int_{t_0}^{T} \Phi^\dagger(t) D_i(t) \left[\frac{H_c^{-1}(t)}{\varepsilon_{\max}}\delta(t-t') + \frac{i}{\hbar} G_0^+(t,t')\right] \Phi(t')\mathrm{d}t\,\mathrm{d}t' = \int_{t_0}^{T} \Phi^\dagger(t) D_i(t) U_0(t,t_0)\,\mathrm{d}t, \qquad (8.3)$$

where H_c^{-1} is taken to be the pseudo inverse if H_c is not invertible. For any $D_i(t)$, Eq. (8.3) is a quadratic equation in the variable $\Phi(t)$; the set of all possible $D_i(t)$ imply an infinite number of quadratic constraints.

8.2.1 Probabilistic interpretation of constraints

Equation (8.3) can be interpreted as a generalization of probability conservation. At any time t_1, conservation of probability implies unitarity of the time-evolution operator $U(t_1, t_0)$, such that $U^\dagger U = \mathcal{I}$, where $\mathcal{I}$ is the identity operator. From the integral equation for U, Eq. (8.1), the difference $U^\dagger U - \mathcal{I}$ can be written

$$\begin{aligned} &U^\dagger(t_1, t_0)U(t_1, t_0) - \mathcal{I} \\ &= \frac{1}{\hbar^2}\int_{t_0}^{t_1}\int_{t_0}^{t_1} \Phi^\dagger(t'', t_0)U_0(t'', t')\Phi(t', t_0)\,\mathrm{d}t'\,\mathrm{d}t'' + \frac{2}{\hbar}\,\mathrm{Im}\int_{t_0}^{t_1} U_0(t_0, t')\Phi(t', t_0)\,\mathrm{d}t'. \end{aligned} \tag{8.4}$$

If we take the imaginary part of Eq. (8.3) and choose $D_i(t)$ to be the identity operator from t_0 to t_1 (and zero otherwise), the resulting constraint is precisely the one that requires the right-hand side of Eq. (8.4) to be zero (cf. Appendix C.3). In other words, a subset of the constraints of Eq. (8.3) is those which enforce unitary evolution at all times. (In an open system described by a density matrix, unitarity is not preserved, and the corresponding constraints instead represent conservation of probability flow, cf. Appendix C.5.)

8.2.2 Equivalence between QCQP and original control problem

Although our derivation above implies only that the conservation laws of Eq. (8.3) are necessary conditions for describing quantum evolution, one can show that they are sufficient as well: any $\Phi(t)$ that satisfies all possible versions of Eq. (8.3) implies a corresponding time-evolution operator $U(t, t_0)$ that satisfies Eq. (8.1). In this section, we show that the conservation laws are also *sufficient* conditions: any solution of the ultimate quadratically constrained quadratic program (QCQP) must also be a solution of the original control problem. Hence the optimal solution of the QCQP must also be the optimal solution of the original problem. The two problems are equivalent in this sense. (A related observation was made in Ref. [201].)

The QCQP arises from optimizing a linear or quadratic objective $f(\Phi)$ subject to Eq. (8.3) being satisfied for all possible D_i. Let us consider a solution $\Phi(t)$ of the QCQP that satisfies Eq. (8.3) for all $D_i(t)$. In particular, we can assume that it satisfied Eq. (8.3)

for all $D_i(t)$ of the form

$$D^i_{jk}(t) = \delta_{j\ell}\delta_{km}\delta(t-t'), \tag{8.5}$$

where we have momentarily moved the i to a superscript to explicitly consider the jk element of $D^{(i)}$ in the Hilbert space, and t' can be any time in the interval of interest. Given this choice of $D^{(i)}$, we can rewrite Eq. (8.3) as

$$\Phi^*_{\ell i}(t')\left[\frac{H_c^{-1}(t')}{\varepsilon_{\max}}\Phi(t') + \frac{i}{\hbar}\int_{t_0}^{T} G_0^+(t',t'')\Phi(t'')\,\mathrm{d}t'' - U_0(t',t_0)\right]_{jm} = 0, \tag{8.6}$$

which must hold for all i, j, ℓ, m and t'. Since Eq. (8.6) must hold at all times, we can see that at any time t' there are two possibilities for $\Phi(t')$: either $\Phi(t') = 0$, or $\Phi(t')$ satisfies the expression in square brackets. This dichotomy will dictate how to find a corresponding $U(t,t_0)$ that satisfied the original (Schrödinger) constraint. At times where $\Phi(t') = 0$, we can take the control field $\varepsilon(t') = 0$. We can also note that the integrand will be zero at such times, and that the domain of the integral can then be restricted to all times for which Φ is nonzero. Equation (8.6) can equivalently be written:

$$\Phi^*_{\ell i}(t')\left[\frac{H_c^{-1}(t')}{\varepsilon_{\max}}\Phi(t') + \frac{i}{\hbar}\int_{\{t''\Phi_{\text{in}}[t_0,T]|\Phi(t'')\neq 0\}} G_0^+(t',t'')\Phi(t'')\,\mathrm{d}t'' - U_0(t',t_0)\right]_{jm} = 0, \tag{8.7}$$

Now we consider all t' for which $\Phi(t') \neq 0$. The term in square brackets must be zero at all such times. For such times, we can set $\varepsilon(t') = \varepsilon_{\max}$, and $U(t',t_0) = (H_c^{-1}(t')/\varepsilon_{\max})\Phi(t')$. These $U(t',t_0)$ must satisfy the term in square brackets, which can now be written:

$$U(t',t_0) + \frac{i}{\hbar}\int_{\{t''\Phi_{\text{in}}[t_0,T]|\varepsilon(t'')\neq 0\}} G_0^+(t',t'')\varepsilon_{\max}H_c(t'')U(t'',t_0)\,\mathrm{d}t'' - U_0(t',t_0) = 0. \tag{8.8}$$

Finally, since $\varepsilon(t)$ is either $\varepsilon_{\max}$ or 0 at all times, we can rewrite the integral domain to take place over all possible times, with $\varepsilon_{\max}$ replaced by $\varepsilon(t)$, implying that $U(t,t_0)$ must satisfy

$$U(t',t_0) = U_0(t',t_0) - \frac{i}{\hbar}\int_{t_0}^{T} G_0^+(t',t'')\varepsilon(t'')H_c(t'')U(t'',t_0)\,\mathrm{d}t''. \tag{8.9}$$

Equation (8.9) is exactly the integral form of Eq. (8.1), which is equivalent to the Dyson equation and the differential Schrodinger equation! Hence we have shown that any solution $\Phi(t)$ that satisfies Eq. (8.3) for all D_i implies a solution $U(t, t_0)$ of the original dynamical constraints. Therefore the optimal value of the QCQP will also be the optimal value of the original control problem.

8.2.3 Semidefinite relaxation

Because of the equivalence between the QCQP form to the original control problem, we can replace the differential or integral dynamical equations with the conservation-law constraints of Eq. (8.3). The optimal-control problem, for any objective f that is a linear or quadratic function of the time-evolution operator U, and therefore a linear or quadratic function of $\Phi = \varepsilon H_c U$, is then the QCQP:

$$\begin{aligned} &\max_{\Phi} \quad f(\Phi) \\ &\text{s.t.} \quad \text{Equation (8.3) satisfied for all } D_i(t). \end{aligned} \tag{8.10}$$

We assume the problem has been discretized in any standard basis [202]. If we denote $\boldsymbol{\Phi}$ to be a single column vector containing the full discretization of $\Phi(t)$, Eq. (8.10) is a maximization of an objective of the form $\boldsymbol{\Phi}^\dagger A \boldsymbol{\Phi} + \mathrm{Re}(\boldsymbol{b}^\dagger \boldsymbol{\Phi})$, where A is Hermitian, subject to constraints of the form $\boldsymbol{\Phi}^\dagger G_i \boldsymbol{\Phi} + \mathrm{Re}(\boldsymbol{c}_i^\dagger \boldsymbol{\Phi}) = 0$ for all i. QCQPs are generically NP-hard to solve, but bounds on their solutions can be computed efficiently after semidefinite relaxation (SDR) described in Sec. 7.3.3 to find the bound to Eq. (8.10). As the bounds are computed over all possible matrices D_i, we label them "D-matrix bounds." This framework applies broadly across quantum control; next, we demonstrate bounds for three prototypical systems.

8.3 Applications

8.3.1 Three-level quantum systems

First, we compute bounds on driving three-level quantum systems. We consider two three-level systems described by Hamiltonians $H = \hbar \sum_{i=1,2} \omega_j \, |i\rangle\langle i| - \varepsilon(t) \sum_{i,j=0,1,2} \mu_{ij} \, |i\rangle\langle j|$: one modeling an asymmetric double-well potential, with exact parameters from Sec. 2.8 of Ref. [155] and given in Appendix C.1, and a second modeling a weakly nonlinear harmonic oscillator with nearest-level couplings, as is typically used to model a transmon qubit [147, 203]. (We consider both systems as they have different features: the first, couplings between all levels, and the second, small anharmonicity with hard-to-avoid leakage.) In each case, we assume the system starts in the ground state, $|0\rangle$, and that we want to drive it to the first excited state, $|1\rangle$, as rapidly as possible. We denote the probability of occupying state i at time t by $P_i(t) = |\langle i|\psi(t)\rangle|^2$. There are two classes of bounds that we can compute: for a given amount of time T, the maximum probability in $|1\rangle$, $P_1(t)$; or, iteratively, the minimum amount of time to achieve near-unity probability in $|1\rangle$.

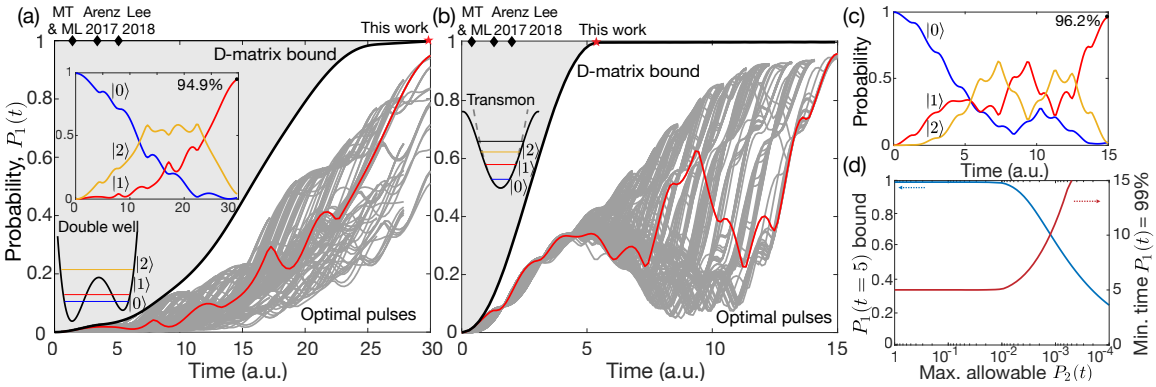

Figure 8.1: (a) Bounds on the maximum probability in state $|1\rangle$ as a function of time (solid black) for an asymmetric double-well potential, with shading above to indicate impossible values. Grey lines represent pulse evolutions optimized by gradient ascent, with the red line the very best evolution for final time 30. Inset: evolution of probabilities in states $|0\rangle, |1\rangle, |2\rangle$ for the optimal control, showing the complex dynamics captured by the bound. Black diamonds: evaluations of bounds of Mandelstam–Tamm, Margolus–Levitin, and Refs. [7, 8] for this problem. (b,c) Analogous to (a) but for a three-level model of a transmon qubit. (d) Incorporation of an additional constraint requiring small maximum allowable excitation probabilities of state 2. The bound on the maximum state-$|1\rangle$ probability (at time 5) decreases accordingly. The time to achieve 99% state-$|1\rangle$ probability increases substantially with smaller allowed leakage rates.

The black curve of Fig. 8.1(a) is the computed bound on $P_1(t)$ for the asymmetric-double-well model, for a bounded control field with $|\varepsilon(t)| \leq 0.15$. The shaded region of

the figure is impossible to reach: our bounds indicate that any such evolution would necessarily violate at least one of the conservation laws. The grey lines are the results of local computational optimizations; we implemented a gradient-ascent optimization (similar to GRAPE) as described in Appendix C.6 for many different final times and initial pulse sequences. The gap between the local optimizations and the bounds arises from two sources—looseness in the bounds (from the SDR) or insufficient local exploration of the optimal pulses—though it is hard to pinpoint which source is more responsible. Also included in the figure are data points corresponding to evaluations of other bounds as applied to this problem: Mandelstam–Tamm (MT), Margolus–Levitin (ML), and Refs. [7, 8]. It takes some effort to map the various bounds to this problem, with varying degrees of looseness, which we discuss in detail in Appendix C.7. In particular, however, one can see that each of these bounds predicts minimal times an order of magnitude smaller than our approach. The inset provides a likely explanation: the optimal trajectory (highlighted in red) first populates the second excited state, then transitions to the first excited state through appropriate driving. Such complex dynamics cannot be captured by any previous bound approaches but can be captured by our approach.

Parts (b–d) of Fig. 8.1 show results for the transmon-qubit model, with $\omega_1 = 0.19$, $\omega_2 = 0.37$, $\mu_{10} = \mu_{01} = -1$, $\mu_{21} = \mu_{12} = -\sqrt{2}$ (all other $\mu_{ij} = 0$), and $|\varepsilon(t)| \leq 0.3$. Fig. 8.1(b,c) are the transmon analogs of Fig. 8.1(a). The key novelty that is possible in this case is the addition of a constraint on the excitation probability of the second excited state, $|2\rangle$. Such "leakage" can be highly detrimental to the practical control of such systems, as they can open up additional decoherence channels [204]. In our approach, we can simply add to Eq. (8.10) a (quadratic) constraint on the maximum allowed probability in $|2\rangle$. In Fig. 8.1(d), we show the bound for maximum $P_1(t)$ subject to varying constraints on the maximum allowed $P_2(t)$, at time $t = 5$, which shows the dramatic reduction that is required if state-$|2\rangle$ transitions are to be avoided. Conversely, also in Fig. 8.1(d), the minimum time for near-unity first-state probability increases dramatically with more stringent constraints (red). Such constraints could not be incorporated into previous bound approaches.

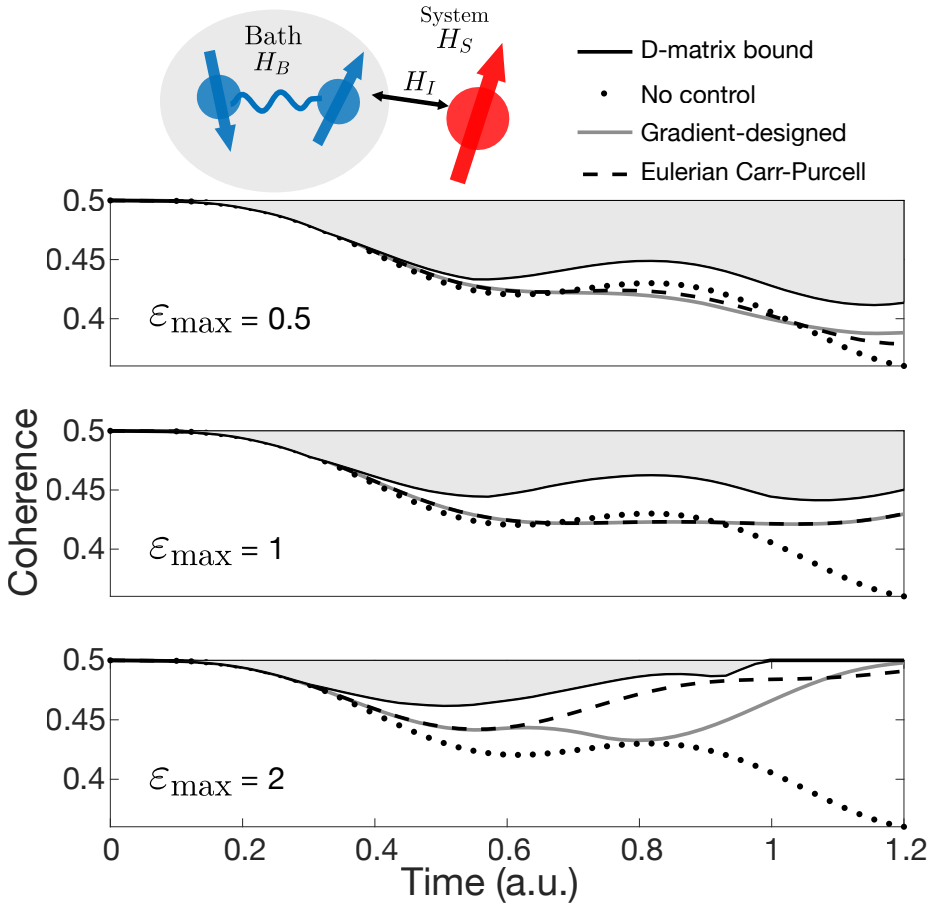

Figure 8.2: For a spin system interacting with a spin bath, the D-matrix approach enables bounds on maximum possible coherence as a function of time. The black solid line bounds the magnitude of the off-diagonal element of the system density matrix, $|\rho_{12}^S|$, for varying maximum control amplitudes $\varepsilon_{\max}$. The time evolutions of $|\rho_{12}^S|$ for pulses designed by gradient-ascent (solid grey) and finite Carr-Purcell (black dash lines) methods can closely approach the bounds.

8.3.2 Decoherence with a spin bath

A second example we consider is the extent to which one can prevent decoherence and dissipation due to interactions with the environment. The design of pulses to achieve such a goal has been studied extensively through semi-heuristic "dynamic decoupling" design schemes [205, 206, 207, 208], which may not be (and in many cases are not) globally optimal. A typical model of environmental effects is a spin system interacting with a spin bath. We consider a spin-bath system [209] with Hamiltonian $H_0 = H_S + H_E + H_{\text{int}}$, where H_S is the system Hamiltonian (two levels split by energy $\hbar\omega_0$), H_E is the Hamiltonian of the environmental bath ($H_E = -J\sum_{j=1}^{N}\left(\sigma_j^x\sigma_{j+1}^x + \lambda\sigma_j^z\right)$), and H_{int} is the interaction between the system and the bath, $H_{\text{int}} = -\nu\left|\downarrow\right\rangle\left\langle\downarrow\right| \otimes \sum_j \sigma_j^z$, with $\omega_0 = \pi, J = 1, \lambda = 0.5$, and

$\nu = 2$ here. The control Hamiltonian here is $H_c = \varepsilon(t)\sigma_x$ on the system only. Rather than use an approximation to the environmental coupling [210], we model the full dynamics of the wave function $|\psi(t)\rangle$. As a result, we only use a bath of size $N = 2$. Despite the bath being unrealistically small, it provide a qualitatively accurate description of the decoherence process [211] and serves as a proof of principle. The system initial state is $\frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{1}{\sqrt{2}}|\downarrow\rangle$, while the spin bath is in its ground state. The system density matrix ρ^S is found by tracing out the bath part of the full density matrix, $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$. The objective is to maximize $|\rho^S_{12}|$, the magnitude of the off diagonal elements of ρ^S, which represents the coherence of the system state. Instead of working with the absolute value (or its square, which is quartic in $|\psi\rangle$), we equivalently maximize $f = \mathrm{Re}\left(\rho^S_{12}e^{i\phi}\right)$ for a given ϕ, and then iterate over possible values of ϕ between 0 and 2π. Fig. 8.2 shows the bounds on maximal coherence as a function of time for three different bounded controls: $\varepsilon_{\max} = 0.5, 1$ and 2. Also included are actual evolutions for three cases: without control, with a pulse designed by gradient ascent, and pulses designed by a bounded-control version of dynamical decoupling termed "Eulerian Carr-Purcell" [212]. It is possible with strong controls to increase coherence at short times (as is particularly visible in Fig. 8.2(c)), but that would not be possible over longer time scales. We see that the bounds appear nearly tight, and provide information about what levels of coherence are possible as a function of time.

8.3.3 Hadamard gate

For the third application, we consider the implementation of a single-qubit Hadamard gate. For a two-level system with Hamiltonian $H = \hbar\omega_0\sigma_z - \mu\varepsilon(t)\sigma_x$ ($\omega_0 = 0.0784$, $\mu = 1$) [155], the target time-evolution operator is given by $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. The objective is to compute the maximal fidelity of a quantum gate at time T; for computational purposes, it is easier to work with the square of fidelity, $f^2 = \frac{1}{4}\left|\mathrm{Tr}\left\{U^\dagger_{\mathrm{tar}}U(T)\right\}\right|^2$. Identifying when the bound approaches 1 then indicates the minimum possible time to perform a gate operation. We consider a bounded control with $\varepsilon_{\max} = 1$. A crucial difference in the gate problem is that multiple inputs map to multiple outputs; the off-diagonal elements of the D matrices in Eq. (8.3) inherently enforce the corresponding orthogonal-evolution requirement. Fig. 8.3

shows the fidelity bound as a function of time (solid black), along with time evolutions for locally optimized pulse sequences in the colored lines (optimized for different end times). The bound is tight, or very nearly so, across all times.

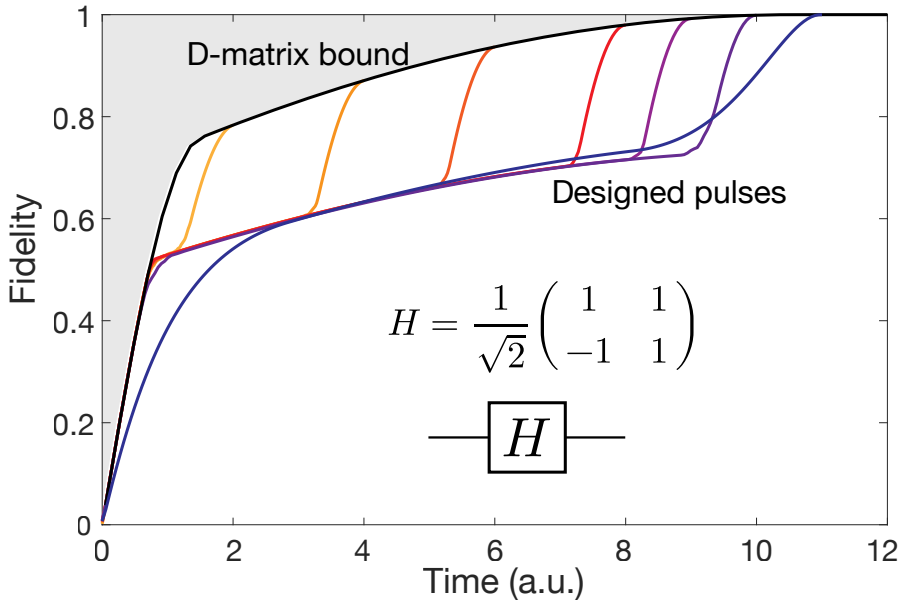

Figure 8.3: The black solid line bounds the maximum fidelity of a single-qubit Hadamard gate implemented in a two-level system with $H = \hbar\omega_0\sigma_z - \mu\varepsilon(t)\sigma_x$, and maximum control amplitude $\varepsilon_{\max} = 1$. Pulses optimized for different final times (colored lines) can achieve the upper bounds at all times.

8.4 Conclusion

Quadratic constraints representing generalized probability-conservation laws offer a framework for quantum control bounds. We have shown that this method can be significantly tighter than previous bounds and more widely applicable. There are further extensions that may be possible as well: in nanophotonic design problems, a hierarchy of bounds with varying analytical and semi-analytical complexity have been discovered as subsets of the D-matrix constraints [213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 142, 90, 140, 201]; the same may be possible in quantum control. In particular, environment-induced decoherence and dissipation are similar to material-absorption losses in electromagnetism, and may be amenable to general analytical bounds [213, 217]. From an algorithmic perspective, there are significant computational speed-ups that should make the bound computations competitive with local optimizations, as a function of the number of degrees of freedom

of the system, N (the product of time steps and Hilbert-space dimensionality). Global optimization is presumably NP-hard; local optimizations require $O(N)$ time for each iteration and a number of iterations that may be large but independent of N. To find good local optima, however, requires restarting the search a number of times proportional to the number of local optima, which should scale at least as $O(N)$, for a total scaling of at least $O(N^2)$ (which is likely optimistic). For the bound computations, the straightforward implementation used for this work, using all possible constraints and interior-point-methods oblivious to the structure of the problem, scales as $O(N^{4.5})$ [223]. Clever selection of the constraint matrices [90] can reduce the scaling to $O(N^{3.5})$. At the same time, exploitation of the integral operator structure (e.g. via fast-multipole-type methods [224, 225]) should further improve the scaling to $O(N^{2.5})$, making it highly competitive with local design methods. More broadly, our approach and extensions thereof can be applied to problems across the quantum-control landscape, ranging from speed limits and gate fidelity to areas like NMR [148, 149, 150] and quantum complexity [151, 152, 153].

Chapter 9

Outlook

In this thesis, we have developed a quasinormal coupled-mode theory (QCMT) framework as an exact generalization of conventional CMT in Chapter 4, extended the brightness theorem into wave scatterings in Chapter 6, and developed a general bound framework for discovering fundamental limits in quantum control in Chapter 8. In closing, we outline some directions for future work.

Having established the quasinormal mode (QNM) expansion of scattering matrices, a glaring opportunity is the development of a robust QNM solver for complex scatterers. The state-of-the-art QNM solver [226] used by specialists in this field is based on finding eigenvalues of the Maxwell operator with perfectly matched layers (PMLs) mimicking the radiation boundary condition. This eigenvalue problem is always non-Hermitian due to the boundary condition, and the operator can be far from a normal operator. (The presence of exceptional points is an extreme indication.) As a result, these eigenvalues are extremely sensitive to any perturbations, potentially from discretization errors and the approximated boundary condition by PMLs. Since the Maxwell differential operator does not admit high-order discretization schemes due to unavoidable numerical cancellations in finite difference methods, the current differential-equation-based non-Hermitian eigenvalue solver for QNMs will never produce accurate results. Instead, one should solve via the nonlinear eigenvalue problem of the volume integral equation. Although the nonlinear nature seems undesirable, we can simply sweep over part of the complex plane for locations of eigenvalues and then refine the accuracy with a couple of Newton steps. Combined with fast algorithms for

evaluating integral operators, this integral-equation-based approach will likely replace the current differential approach for computing QNMs with high accuracy.

The brightness concentration bound for wave scattering only works for the power transport of channels. It can be easily extended into a more general framework for arbitrary objectives that is quadratic in fields in terms of the eigenvalues of the objective matrix in the channel function basis and eigenvalues of the incoming density matrix. The final bound will take a particular form of the Wielandt-Hoffman inequality, where the current power concentration bound is subsumed as a special case. This generalized concentration bound will find many more applications whenever incoherence is present.

The fundamental bound framework in quantum control can be extended in multiple ways. All three examples in this thesis are essentially closed systems, but this framework is especially suited for dealing with open quantum systems with losses. Since current theoretical tools for quantum control of open systems are very limited, this framework can serve as a computation approach for identifying what is possible for the control of open systems. Ultimately, we need to incorporate the physical properties of quantum dynamics and devise specialized algorithms to accelerate the process of solving semidefinite programs. If this is achieved, the framework can be scaled up for more realistic quantum control problems in quantum computation and quantum circuit complexity theory.

Appendix A

Appendix of "Quasinormal coupled-mode theory"

A.1 Examples of S Matrix with Quasinormal Mode Expansion

In this section we provide details for the examples provided in the main text. The first example in Sec. A.1.1 is a Fabry-Perot slab, where the scattering channels are plane waves. The second example in Sec. A.1.2 is a Mie sphere, where the scattering channels are vector spherical waves. We list the explicit form of the channel functions and resonance mode expressions, through which the S matrix is constructed. In both examples, the material susceptibility is nonmagnetic, so that ΔB only has nonzero $\Delta\epsilon$, and we can work primarily with the electric field $\boldsymbol{E}$ only. The magnetic field $\boldsymbol{H}$, if not shown, can be found by Maxwell equations directly. We work in dimensionless unit and set $c = 1$, so $\widetilde{\omega}_m = \widetilde{k}_m$.

A.1.1 QNM expansion of Fabry-Perot slab

QNM basis

For a Fabry-Perot slab configuration in Fig. 2(a), the normalized QNMs inside the slab of refractive index n are given by (s and p polarizations are degenerate) [56]

$$\boldsymbol{E}_{\mathrm{qnm},m} = \begin{cases} \frac{1}{n\sqrt{L}} \cos\left(n\widetilde{k}_m x\right)\hat{\boldsymbol{z}}, & \text{for} - L/2 < x < L/2, m \text{ even}, \\ \frac{1}{n\sqrt{L}} \sin\left(n\widetilde{k}_m x\right)\hat{\boldsymbol{z}}, & \text{for} - L/2 < x < L/2, m \text{ odd}, \end{cases} \tag{A.1}$$

with resonant frequencies

$$\widetilde{k}_m = \frac{1}{nL}\left[m\pi - \mathrm{i}\ln\left(\frac{n+n_0}{n-n_0}\right)\right]. \tag{A.2}$$

For this one-dimensional example, the normalization integral is 0 outside of the slab, so a PML is not needed to normalize the QNMs.

Channel basis functions

The plane wave basis function has not been treated as systematically as the vector spherical waves and is normally used in a intuitive manner. As a result, first we need to formalize the plane wave basis to be consistent with the scattering framework developed in the main text. Here, we take the plane waves as incident basis Φ_{inc}, which is a traveling wave regular throughout all space. We use Heaviside step function $\eta(x)$ with plane waves to represent the incoming basis Φ_{in} and outgoing basis Φ_{out}, so that there are net power flow into/out the bounding surface Σ, which is normally two flat surfaces enclosing a unit cell of a periodic structure. The discontinuity due to $\eta(x)$ can be understood as the presence of additional sinks/sources, which is similar to the singularities in incoming and outgoing basis in vector spherical waves. Although here we only treat the plane wave basis without higher specular orders, the generalization is straightforward.

We choose $\boldsymbol{E}_{\mathrm{in},1} = \eta(-x)\sqrt{2}\mathrm{e}^{\mathrm{i}k(x+L/2)}\hat{\boldsymbol{z}}$, $\boldsymbol{E}_{\mathrm{in},2} = \eta(x)\sqrt{2}\mathrm{e}^{-\mathrm{i}k(x-L/2)}\hat{\boldsymbol{z}}$ as the incoming channel basis, and $\boldsymbol{E}_{\mathrm{out},1} = \eta(x)\sqrt{2}\mathrm{e}^{\mathrm{i}k(x-L/2)}\hat{\boldsymbol{z}}$, $\boldsymbol{E}_{\mathrm{out},2} = \eta(-x)\sqrt{2}\mathrm{e}^{-\mathrm{i}k(x+L/2)}\hat{\boldsymbol{z}}$, as the out-

going ones. One can easily check they are power orthonormal. Then we have

$$\begin{aligned} \boldsymbol{E}_{\mathrm{inc},1} &= \boldsymbol{E}_{\mathrm{in},1} + \mathrm{e}^{\mathrm{i}kL}\boldsymbol{E}_{\mathrm{out},1} = \sqrt{2}\mathrm{e}^{\mathrm{i}k(x+L/2)}\hat{\boldsymbol{z}}, \\ \boldsymbol{E}_{\mathrm{inc},2} &= \boldsymbol{E}_{\mathrm{in},2} + \mathrm{e}^{\mathrm{i}kL}\boldsymbol{E}_{\mathrm{out},2} = \sqrt{2}\mathrm{e}^{-\mathrm{i}k(x-L/2)}\hat{\boldsymbol{z}}\,, \end{aligned} \tag{A.3}$$

where $\eta(x)$ and $\eta(-x)$ adds up to unity and the discontinuity disappears. We can see that $\alpha = 1$ and $\beta = \mathrm{e}^{\mathrm{i}kL}$, due to the special phase choice.

S matrix construction

In the channel basis chosen above, the scattering matrix is of the form $S = \begin{pmatrix} t_1 & r_2 \\ r_1 & t_2 \end{pmatrix}$, and $S_{\mathrm{bg}} = \begin{pmatrix} \mathrm{e}^{\mathrm{i}kL} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}kL} \end{pmatrix}$ without the presence of the slab. Here the reflection coefficients $r_1 = r_2 = r$ and the transmission coefficients $t_1 = t_2 = t$, and the exact expressions are [227]

$$r = \frac{r_0(\mathrm{e}^{2\mathrm{i}nkL} - 1)}{1 - r_0^2\mathrm{e}^{2\mathrm{i}nkL}}, \qquad t = \frac{t_0^2\mathrm{e}^{\mathrm{i}nkL}}{n(1 - r_0^2\mathrm{e}^{2\mathrm{i}nkL})}, \tag{A.4}$$

where $r_0 = \frac{n-n_0}{n+n_0}$ and $t_0 = \frac{2n}{n+n_0}$.

The task, then, is to test whether the QNM expressions for the scattering matrix produce results that are consistent with the exact expressions of Eq. (A.4). In the QCMT framework, the S matrix (as given in the main text), can be written

$$S = S_{\mathrm{bg}} + \frac{\mathrm{i}\omega}{4\alpha\beta^*}(\Phi_{\mathrm{inc}}^{\mathrm{TR}}, \Delta B\Phi_{\mathrm{inc}}) - \mathrm{i}K(\omega)\left[N(\omega)(\Omega - \omega)\right]^{-1} D^T(\omega)\,. \tag{A.5}$$

The reflection and transmission coefficients are the components of the S-matrix: $r_1 = S_{11}$, $r_2 = S_{22}$, $t_1 = S_{21}$, and $t_2 = S_{12}$. Here the material is non-dispersive so $N(\omega) = \mathbb{I}$, the identity matrix. From the definitions of the basis functions and coupling matrices, and since given our channel definitions we have $\frac{1}{\alpha\beta^*} = \mathrm{e}^{\mathrm{i}kL}$, $\frac{1}{\alpha\beta^*}\boldsymbol{E}_{\mathrm{inc},1}^{\mathrm{TR}} = \boldsymbol{E}_{\mathrm{inc},2}$ and $\frac{1}{\alpha\beta^*}\boldsymbol{E}_{\mathrm{inc},2}^{\mathrm{TR}} = \boldsymbol{E}_{\mathrm{inc},1}$,

the QCMT reflection and transmission coefficients are given by

$$r_1 = \frac{1}{4}\mathrm{i}\omega(\boldsymbol{E}_{\mathrm{inc},1}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},1}) + \frac{1}{4}(\mathrm{i}\omega)^2 \sum_m (\boldsymbol{E}_{\mathrm{inc},1}, \Delta\epsilon\boldsymbol{E}_{\mathrm{qnm},m}) \frac{1}{\mathrm{i}(\widetilde{\omega}_m - \omega)} (\boldsymbol{E}_{\mathrm{qnm},m}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},1}), \quad \text{(A.6)}$$

$$r_2 = \frac{1}{4}\mathrm{i}\omega(\boldsymbol{E}_{\mathrm{inc},2}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},2}) + \frac{1}{4}(\mathrm{i}\omega)^2 \sum_m (\boldsymbol{E}_{\mathrm{inc},2}, \Delta\epsilon\boldsymbol{E}_{\mathrm{qnm},m}) \frac{1}{\mathrm{i}(\widetilde{\omega}_m - \omega)} (\boldsymbol{E}_{\mathrm{qnm},m}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},2}), \quad \text{(A.7)}$$

$$t_1 = \mathrm{e}^{\mathrm{i}kL} + \frac{1}{4}\mathrm{i}\omega(\boldsymbol{E}_{\mathrm{inc},2}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},1}) + \frac{1}{4}(\mathrm{i}\omega)^2 \sum_m (\boldsymbol{E}_{\mathrm{inc},2}, \Delta\epsilon\boldsymbol{E}_{\mathrm{qnm},m}) \frac{1}{\mathrm{i}(\widetilde{\omega}_m - \omega)} (\boldsymbol{E}_{\mathrm{qnm},m}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},1}), \quad \text{(A.8)}$$

$$t_2 = \mathrm{e}^{\mathrm{i}kL} + \frac{1}{4}\mathrm{i}\omega(\boldsymbol{E}_{\mathrm{inc},1}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},2}) + \frac{1}{4}(\mathrm{i}\omega)^2 \sum_m (\boldsymbol{E}_{\mathrm{inc},1}, \Delta\epsilon\boldsymbol{E}_{\mathrm{qnm},m}) \frac{1}{\mathrm{i}(\widetilde{\omega}_m - \omega)} (\boldsymbol{E}_{\mathrm{qnm},m}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},2}), \quad \text{(A.9)}$$

where (A, B) here is $\int \mathrm{d}x A^T B$, a one-dimensional integral. It is obvious that $t_1 = t_2$ due to reciprocity. The symmetry $r_1 = r_2$ can be seen by a change of variable $x \to -x$ and noting $\boldsymbol{E}_{\mathrm{qnm},m}$ is either odd or even in x. The constructed quantities here agree precisely with the exact expressions of Eq. (A.4), as shown in Fig. 2(d) of the main text.

Pole expansion of S matrix

Both r and t are bounded as k goes to complex infinity, so one can apply Mittag-Leffler to r and t to obtain frequency-independent "background " and "resonant" terms. The QCMT S-matrix in such a case, as described in the main text, is given by the expression

$$S = S_{\mathrm{bg}}(\omega = 0) + \mathrm{i}\widetilde{K}\Omega^{-1}\widetilde{D}^T - \mathrm{i}\widetilde{K}(\Omega - \omega)^{-1}\widetilde{D}^T, \quad \text{(A.10)}$$

where the first two terms comprise the background, while the third term is the resonant term. The background and resonant transmission and reflection coefficients have been derived for the specific case of Fabry–Perot [72], giving:

$$r_{\mathrm{bg}} = \frac{2\mathrm{i}}{(1-n^2)L} \sum_m \frac{1}{\widetilde{k}_m}, \quad t_{\mathrm{bg}} = 0$$

and

$$r_{\mathrm{reso}} = \frac{2\mathrm{i}}{(1-n^2)L} \sum_m \frac{1}{k - \widetilde{k}_m}, \qquad t_{\mathrm{reso}} = \frac{2\mathrm{i}}{(1-n^2)L} \sum_m \frac{(-1)^{m+1}}{k - \widetilde{k}_m}.$$

As shown in Fig. 2(b,d) of the main text, the QCMT calculations from Eq. (A.43) are in exact agreement with these expressions, while generalizing to arbitrary scattering bodies.

A.1.2 QNM expansion of Mie sphere

For a spherical scattering body, the QNM fields can be expressed in terms of vector spherical waves (VSWs). Here we follow the convention of Ref. [80] for VSWs and write down QNM fields and channel functions. The resonant frequencies cannot be found analytically and must be numerically computed.

Vector spherical waves (VSWs)

VSWs in this convention have three indices, ℓ, m, σ, which we collectively denote by n. Besides the common angular momentum numbers ℓ and m, the σ index here takes value e, o, representing "even" and "odd" cases. This is because conventionally the spherical harmonics $Y_{\ell m}$, which are part of the VSW functions, have a factor $\mathrm{e}^{\mathrm{i}m\phi}$. However, here we separate it into $\cos(m\phi)$ and $\sin(m\phi)$, labeled by σ, taking value e, o respectively. This makes the angular part purely real, a convenient choice under the unconjugated inner product for QNMs. Due to this extra index σ, we denote the spherical harmonics by Y_n thereafter. They are given as

$$Y_n(\hat{\boldsymbol{r}}) = Y_{\sigma\ell m}(\hat{\boldsymbol{r}}) = \sqrt{\frac{\epsilon_m}{2\pi}}\sqrt{\frac{2\ell+1}{2}\frac{(\ell-m)!}{(\ell+m)!}}P_\ell^m(\cos\theta)\begin{Bmatrix}\cos\phi\\ \sin\phi\end{Bmatrix},$$

where $\hat{\boldsymbol{r}}$ is the unit position vector and $P_\ell^m(x)$ here are the associated Legendre polynomials, with

$$\sigma = e, o, \qquad m = 0, 1, 2, \ldots, \ell \qquad l = 0, 1, 2, \ldots, \qquad \epsilon_m = \begin{cases} 1, & m = 0 \\ 2, & m > 1 \end{cases}.$$

The angular part for a given n consists of

$$\boldsymbol{A}_{1n} = \frac{1}{\sqrt{\ell(\ell+1)}}\nabla \times (\boldsymbol{r}Y_n(\hat{\boldsymbol{r}})), \tag{A.11}$$

$$\boldsymbol{A}_{2n} = \frac{1}{\sqrt{\ell(\ell+1)}} r\nabla Y_n(\hat{\boldsymbol{r}}), \tag{A.12}$$

$$\boldsymbol{A}_{3n} = \hat{\boldsymbol{r}}Y_n(\hat{\boldsymbol{r}}), \tag{A.13}$$

which are an orthonormal set on the unit sphere since

$$\int_\Omega \boldsymbol{A}_{\tau n} \cdot \boldsymbol{A}_{\tau' n'} = \delta_{\tau\tau'}\delta_{nn'} \tag{A.14}$$

for $\tau, \tau' = 1, 2, 3$. We emphasize here again $\boldsymbol{A}_{\tau n}$ are real, so the above orthonormality relation is suitable for the unconjugated inner product.

The radial part consists of spherical Bessel functions $j_\ell(x)$, which is regular at $x = 0$, and spherical hankel functions of the first kind $h_\ell^{(1)}(x)$, which has an asymptotic form, for large x, proportional to outgoing spherical waves. As a result, this can be used to define outgoing channel functions, or to satisfy the radiation boundary condition of QNMs. Hence, we introduce the regular VSWs

$$\boldsymbol{v}_{1n}(k\boldsymbol{r}) = j_\ell(kr)\boldsymbol{A}_{1n}(\hat{\boldsymbol{r}}) \tag{A.15}$$

$$\boldsymbol{v}_{2n}(k\boldsymbol{r}) = \frac{(krj_\ell(kr))'}{kr}\boldsymbol{A}_{2n}(\hat{\boldsymbol{r}}) + \sqrt{\ell(\ell+1)}\frac{j_\ell(kr)}{kr}\boldsymbol{A}_{3n}(\hat{\boldsymbol{r}})\,, \tag{A.16}$$

and the outgoing VSWs

$$\boldsymbol{u}_{1n}(k\boldsymbol{r}) = h_\ell^{(1)}(kr)\boldsymbol{A}_{1n}(\hat{\boldsymbol{r}}) \tag{A.17}$$

$$\boldsymbol{u}_{2n}(k\boldsymbol{r}) = \frac{(krh_\ell^{(1)}(kr))'}{kr}\boldsymbol{A}_{2n}(\hat{\boldsymbol{r}}) + \sqrt{\ell(\ell+1)}\frac{h_\ell^{(1)}(kr)}{kr}\boldsymbol{A}_{3n}(\hat{\boldsymbol{r}})\,, \tag{A.18}$$

where $'$ represents the derivative of argument kr. With $h_\ell^{(1)}$ above replaced by $h_\ell^{(2)}(x) = (h_\ell^{(1)}(x))^*$, the spherical hankel functions of the second kind, the outgoing VSWs $\boldsymbol{u}_{\tau n}$ becomes the incoming ones $\boldsymbol{u}_{\tau n}^*$, which can be used to define incoming channel functions. Note that $j_\ell(x) = \frac{1}{2}h_\ell^{(1)}(x) + \frac{1}{2}h_\ell^{(2)}(x)$, so $\boldsymbol{v}_{1n}$ is a combination of incoming and outgoing

fields, $\boldsymbol{v}_{\tau n} = \frac{1}{2}\boldsymbol{u}_{\tau n} + \frac{1}{2}\boldsymbol{u}^*_{\tau n}$ with $\alpha = \beta = \frac{1}{2}$.

Power normalization

For $\boldsymbol{E} = \boldsymbol{u}_{1n}$, the power flow out of the unit sphere is $\frac{1}{2}\,\mathrm{Re}\int_\Omega \boldsymbol{E} \times \boldsymbol{H}^* \cdot \mathrm{d}\boldsymbol{S} = \frac{1}{2k^2}$, which is not power normalized. Hence, normalization is needed when the channel functions are chosen. Here we define

$$\boldsymbol{N}^{\mathrm{reg}}_{\sigma\ell m} = \sqrt{2}k\boldsymbol{v}_{1n}, \quad \boldsymbol{N}^{+}_{\sigma\ell m} = \sqrt{2}k\boldsymbol{u}_{1n}, \tag{A.19}$$

which is the notation used in Fig. 3(a) in the main text.

QNM and channel basis functions

With these functions define above, we are ready to write down QNMs and channel functions for the Mie spheres. We derive everything here for the e-polarization case, and the h-polarization counterpart will be similar.

We choose $\boldsymbol{E}_{\mathrm{out},n} = \sqrt{2}k\boldsymbol{u}_{1n}$, and $\boldsymbol{E}_{\mathrm{inc},n} = \sqrt{2}k\boldsymbol{v}_{1n}$, so that the outgoing channel basis is power normalized. If we denote $k_1 = nk$, the scattering matrix element can be found as [80]

$$S_{\ell m} = 1 + 2T_\ell, \quad T_\ell = -\frac{j_\ell(kR)(k_1 R j_\ell(k_1 R))' - (kR j_\ell(kR))' j_\ell(k_1 R)}{h^{(1)}_\ell(kR)(k_1 R j_\ell(k_1 R))' - (kR h^{(1)}_\ell(kR))' j_\ell(k_1 R)}, \tag{A.20}$$

where T_ℓ is the transition matrix elements [228], which are more commonly used for Mie scattering.

The resonance frequencies $\widetilde{\omega}_m$, for a particular n, can be found by searching for zeros of $\frac{1}{S_{\ell m}}$ (or $\frac{1}{T_\ell}$). Next, we start to solve for QNM field expressions. Since $\boldsymbol{v}_{1n}$ is regular at the origin, and $\boldsymbol{u}_{1n}$ satisfy the radiation boundary condition,

$$\boldsymbol{E}_{\mathrm{qnm},m} = \frac{1}{\sqrt{N_m}} \begin{cases} C\boldsymbol{v}_{1n}(n\widetilde{k}_m \boldsymbol{r}), & 0 \le r < R, \\ \boldsymbol{u}_{1n}(\widetilde{k}_m \boldsymbol{r}), & R \le r < \infty, \end{cases} \tag{A.21}$$

where N_m is the normalization constant and the constant C can be fixed by the continuity

of electric field across $r = R$, which gives $C = \frac{h_\ell^{(1)}(kR)}{j_\ell(nkR)}$. By the relation that $\nabla \times \boldsymbol{v}_{1n}(k\boldsymbol{r}) = k\boldsymbol{v}_{2n}(k\boldsymbol{r})$, and $\nabla \times \boldsymbol{u}_{1n}(k\boldsymbol{r}) = k\boldsymbol{u}_{2n}(k\boldsymbol{r})$, the magnetic fields are

$$\boldsymbol{H}_{\mathrm{qnm},m} = \frac{1}{\sqrt{N_m}} \begin{cases} -Cin\boldsymbol{v}_{2n}(n\widetilde{k}_m\boldsymbol{r}), & 0 \le r < R, \\ -\mathrm{i}\boldsymbol{u}_{2n}(\widetilde{k}_m\boldsymbol{r}), & R \le r < \infty. \end{cases} \tag{A.22}$$

To work out the normalization constant N_m, one way to use the PML method as in the main text. This can be done by a coordinate stretching of $r \to \bar{r} = r + \mathrm{i}\sigma(r)$, where $\sigma(r)$ turns on outside of the sphere [229]. Hence, by the orthonormality of $\boldsymbol{A}_{\tau n}$, the normalization integral $\int_V \epsilon \boldsymbol{E}^2_{\mathrm{qnm},m} - \mu \boldsymbol{H}^2_{\mathrm{qnm},m} = 2\int_V \epsilon \boldsymbol{E}^2_{\mathrm{qnm},m}$ becomes a radial integral of r only. The normalization condition gives

$$N_m = 2C^2 \int_0^R \mathrm{d}r r^2 n^2 j_\ell^2(n\widetilde{k}_m r) + 2\int_R^\infty \mathrm{d}\bar{r}\bar{r}^2 (h_\ell^{(1)}(\widetilde{k}_m\bar{r}))^2 \,. \tag{A.23}$$

For this particular example, this normalization method works, but it fails for purely imaginary $\widetilde{k}_m$. Besides, $\widetilde{k}_m\bar{r}$ has a large imaginary part for large r. The numerical evaluation of Bessel functions at large imaginary argument is unstable [230], and this could lead to potential numerical issues. Hence, an alternative normalization method can be found in Ref. [231], which is equivalent to the PML method in Ref. [232] and uses fields inside of the scatter only, can be adopted if numerical issues occur.

S matrix construction

Given the QNMs and channel functions as defined above, the QCMT S-matrix of Eq. (A.5) simplifies for the sphere to

$$S_{\ell m} = 1 + \mathrm{i}\omega(\boldsymbol{E}_{\mathrm{inc},n}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},n}) + (\mathrm{i}\omega)^2 \sum_m (\boldsymbol{E}_{\mathrm{inc},n}, \Delta\epsilon\boldsymbol{E}_{\mathrm{qnm},m}) \frac{1}{\mathrm{i}(\widetilde{\omega}_m - \omega)} (\boldsymbol{E}_{\mathrm{qnm},m}, \Delta\epsilon\boldsymbol{E}_{\mathrm{inc},n}) \,. \tag{A.24}$$

All inner product here, although 3-dimensional in nature, can be reduced into a radial integral by the orthonormality of $\boldsymbol{A}_{\tau n}$, as the normalization integral. This constructed $S_{\ell m}$ is plotted against the exact one of Eq. (A.20) in Fig. 3(e) in the main text.

A.2 Derivation of 2nd QCMT equation in special cases

In the main text, we derive the second QCMT equation,

$$\boldsymbol{c}_{\text{out}} = \left\{ S_{\text{bg}} + \frac{\mathrm{i}\omega}{4\alpha\beta^*} (\Phi_{\text{inc}}^{\text{TR}}, \Delta B \Phi_{\text{inc}}) \right\} \boldsymbol{c}_{\text{in}} + K(\omega)\boldsymbol{a}\,, \tag{A.25}$$

by applying the equivalence principle to simplify the time-reversed channel functions. In this section, we provide an alternative proof, which is less general but perhaps more conventional, in the special cases of Fabry–Perot slabs and Mie spheres. The basic idea is to decompose the background Green's function according to the special symmetry of the scatterer.

A.2.1 Fabry-Perot slabs

Consider a single slab of refractive index n (as in Fig. 4.2(a) of the main text). We isolate the x direction due to the y, z transnational symmetry, and denote (y, z) by $\boldsymbol{r}_\perp$, which is perpendicular to the x direction. The total field E satisfies the Helmholtz equation,

$$(\nabla^2 + k^2\epsilon)E = 0\,, \tag{A.26}$$

or

$$(\nabla^2 + k^2)E = -k^2\Delta\epsilon E\,. \tag{A.27}$$

Take the incoming field as $E_{\text{inc}} = \sqrt{2}\mathrm{e}^{\mathrm{i}k_x x}$ ($c_{\text{in},1} = 1$) and convert the above equation into the integral form and we have

$$E = E_{\text{inc}} + k^2 \int \mathrm{d}S'\mathrm{d}x' \, \frac{1}{4\pi} \frac{\mathrm{e}^{\mathrm{i}k|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|} \Delta\epsilon(x')E(x')\,, \tag{A.28}$$

with x integral isolated. As $\frac{1}{4\pi}\frac{\mathrm{e}^{\mathrm{i}k|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|}$ can be written as [233]

$$\frac{1}{4\pi} \frac{\mathrm{e}^{\mathrm{i}k|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|} = \int \frac{\mathrm{d}k'_z}{2\pi} \frac{\mathrm{d}k'_y}{2\pi} \, \mathrm{e}^{\mathrm{i}\boldsymbol{k}'_\perp \cdot (\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp)} \frac{\mathrm{i}}{2\sqrt{k^2 - k_\perp'^2}} \mathrm{e}^{\mathrm{i}\sqrt{k^2 - k_\perp'^2}|x-x'|}\,, \tag{A.29}$$

the S' integral produces a $\delta(\boldsymbol{k}'_\perp)$ and we have

$$E = \sqrt{2}\mathrm{e}^{\mathrm{i}kx} + \frac{\mathrm{i}k}{2}\int_{-L/2}^{L/2} \mathrm{d}x' \mathrm{e}^{\mathrm{i}k|x-x'|}\Delta\epsilon(x')E(x')\,. \tag{A.30}$$

At $x > L/2$, we can write $E = t\sqrt{2}\mathrm{e}^{\mathrm{i}k_x x}$. We then have

$$t = 1 + \frac{\mathrm{i}k}{4}\int \mathrm{d}x'\sqrt{2}\mathrm{e}^{-\mathrm{i}kx'}\Delta\epsilon(x')E(x')\,. \tag{A.31}$$

Similarly, at $x < -L/2$, we write $E(x) = \sqrt{2}\mathrm{e}^{-\mathrm{i}kx} + r\sqrt{2}\mathrm{e}^{\mathrm{i}kx}$. We then have

$$r = \frac{\mathrm{i}k}{4}\int \mathrm{d}x'\sqrt{2}\mathrm{e}^{\mathrm{i}kx'}\Delta\epsilon(x')E(x')\,. \tag{A.32}$$

If we express $E = \sqrt{2}\mathrm{e}^{\mathrm{i}kx} + a_m \sum_m E_{\mathrm{qnm},m}$, where a_m is the element of the expansion coefficients $\boldsymbol{a}$, we obtain the second QCMT equation, Eq. (A.25), for t and r, with outgoing channel functions $\eta(x)\sqrt{2}\mathrm{e}^{\mathrm{i}kx}$ and $\eta(-x)\sqrt{2}\mathrm{e}^{-\mathrm{i}kx}$. (See Section A.1.1 for the channel function definition.) If we solve for $\boldsymbol{a}$ from the first QCMT equation, t and r in Eq. (A.31) and Eq. (A.32) will be identical to Eq. (A.8) and Eq. (A.6), except for a phase factor of $\mathrm{e}^{\mathrm{i}kL}$.

A.2.2 Mie spheres

Consider a single spherical scattering body (as in Fig. 4.3(a) of the main text). The total field $\boldsymbol{E}$ satisfies

$$\nabla\times\nabla\times\boldsymbol{E} - \epsilon k^2\boldsymbol{E} = 0\,, \tag{A.33}$$

or

$$\nabla\times\nabla\times\boldsymbol{E} - k^2\boldsymbol{E} = \Delta\epsilon k^2\boldsymbol{E}\,. \tag{A.34}$$

For incident field $\boldsymbol{E}_{\mathrm{inc}}$, we convert the equation into an integral equation

$$\boldsymbol{E} = \boldsymbol{E}_{\mathrm{inc}} + k^2\int G_{EE}(\boldsymbol{r},\boldsymbol{r}')\Delta\epsilon\boldsymbol{E}(\boldsymbol{r}')\,, \tag{A.35}$$

where $G_{EE} = (\mathbf{1} + \frac{1}{k^2}\nabla\nabla)\frac{\mathrm{e}^{\mathrm{i}k|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|} = \mathrm{i}k\sum_{n,\tau=1,2}\boldsymbol{u}_{\tau n}(k\boldsymbol{r}_>)\boldsymbol{v}_{\tau n}(k\boldsymbol{r}_<)$ [80]. Here $\Delta\epsilon$ has spherical symmetry, thus angle independent, so orthonormality of $\boldsymbol{A}_{\tau n}$ enable us to isolate

each different τ and n in $\boldsymbol{E}_{\text{inc}}$. For $\boldsymbol{E}_{\text{inc}} = \sqrt{2}k\boldsymbol{v}_{1n}$ ($c_{\text{in},n} = c_{\text{out},n} = \frac{1}{2}$, $c_{\text{inc},n} = 1$) and $\boldsymbol{r}$ outside of the sphere, we have

$$\begin{aligned}
\boldsymbol{E}(\boldsymbol{r}) &= \sqrt{2}k \underbrace{\boldsymbol{v}_{1n}(\boldsymbol{r})}_{=\frac{1}{2}\boldsymbol{u}_{1n}+\frac{1}{2}\boldsymbol{u}_{1n}^*} + \boldsymbol{u}_{1n}(\boldsymbol{r})\mathrm{i}k^3 \int \boldsymbol{v}_{1n}(\boldsymbol{r}') \cdot \Delta\epsilon \cdot \boldsymbol{E}(\boldsymbol{r}') && \text{(A.36)} \\
&= \frac{1}{2}\sqrt{2}k\boldsymbol{u}_{1n}^*(\boldsymbol{r}) + \frac{1}{2}\sqrt{2}k\boldsymbol{u}_{1n}(\boldsymbol{r}) \underbrace{\left(1 + \mathrm{i}k \int \sqrt{2}k\boldsymbol{v}_{1n}(\boldsymbol{r}') \cdot \Delta\epsilon \cdot \boldsymbol{E}(\boldsymbol{r}')\right)}_{=S_{\ell m}} && \text{(A.37)}
\end{aligned}$$

If we express $\boldsymbol{E} = \sqrt{2}k\boldsymbol{v}_{1n} + a_m \sum_m \boldsymbol{E}_{\text{qnm},m}$, where a_m is the element of the expansion coefficients $\boldsymbol{a}$, we obtain the second QCMT equation, Eq. (A.25), for e-polarized $S_{\ell m}$ for Mie spheres , with power normalized outgoing channel $\sqrt{2}k\boldsymbol{u}_{1n}$. (See Section A.1.2 for the channel function definition.) Solving for $\boldsymbol{a}$ from the first QCMT equation, $S_{\ell m}$ in Eq. (A.37) will be identical to the one in Eq. (A.24).

A.3 Applications of Mittag-Leffler expansion

In this section we show the details of applying Mittag-Leffler expansion to

$$\psi_{\text{scat}}(\boldsymbol{r}) = \mathrm{i}\omega \int \Gamma(\boldsymbol{r}, \boldsymbol{r}', \omega)\Delta B(\boldsymbol{r}', \omega)\psi_{\text{inc}}(\boldsymbol{r}')\,, \tag{A.38}$$

which leads to various QNM expansion formulae.

A.3.1 Equivalence between decomposition approaches

Green's function

If we use the full Green's function of the system and get ψ_{scat} from Eq. (A.38), rather than doing QNM expansions, we can identify the QNM expansion of Green's function

$$\Gamma(\boldsymbol{r}, \boldsymbol{r}') = \Phi_{\text{Rqnm}}(\boldsymbol{r}) \frac{1}{\mathrm{i}(\Omega - \omega)} \frac{1}{N(\omega)} \Phi^T_{\text{Lqnm}}(\boldsymbol{r}')\,. \tag{A.39}$$

This is different from the other more widely used form

$$\Gamma(\boldsymbol{r},\boldsymbol{r}') = \Phi_{\mathrm{Rqnm}}(\boldsymbol{r})\frac{1}{\mathrm{i}(\Omega-\omega)}\Phi^{T}_{\mathrm{Lqnm}}(\boldsymbol{r}')\,. \tag{A.40}$$

In deriving Eq. (A.39) we have used what is known as the orthogonality-decomposition approach, whereas the residue-decomposition approach is used to derive Eq. (A.40).

Note that at each resonant frequency $\widetilde{\omega}_i$, the ith row of the $N(\widetilde{\omega}_i)$ matrix is diagonal, and the non-zero element is responsible for the normalization of the ith mode. As a result, if we apply Mittag-Leffler to Eq. (A.39), we will obtain Eq. (A.40). This proves the equivalence between residue-docomposition approach to orthogonality-decomposition approach. This shows that the difference between Eq. (A.39) and Eq. (A.40) is because Mittag-Leffler expansion is applied to the same quality at different stages.

QNM expansion formulae

If we apply Mittag–Leffler to Γ in Eq. (A.40), we have $\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega) = \sum_i \frac{\psi_{R,i}(\boldsymbol{r})\psi^T_{L,i}(\boldsymbol{r}')}{\mathrm{i}(\widetilde{\omega}_i-\omega)}$ and put it back to Eq. (A.38) and we obtain one version of the residue-decomposition approach. Similarly, we apply ML to $\omega\Gamma\Delta B$, and by the sum rules of Γ and partial fractions, we have

$$\begin{aligned}
&\omega\Gamma(\boldsymbol{r},\boldsymbol{r}',\omega)\sum_n \frac{\sigma_n}{\omega-\omega_n}\\
=\;& \sum_{i,n}\frac{\psi_{R,i}(\boldsymbol{r})\psi^T_{L,i}(\boldsymbol{r}')}{\mathrm{i}(\widetilde{\omega}_i-\omega)}\frac{\widetilde{\omega}_i\sigma_n}{(\widetilde{\omega}_i-\omega_n)}\\
&\underbrace{-\sum_{i,n}\frac{1}{\mathrm{i}}\psi_{R,i}(\boldsymbol{r})\psi^T_{L,i}(\boldsymbol{r}')\frac{\sigma_n}{(\widetilde{\omega}_i-\omega_n)}}_{=\Gamma(\omega_n)\sigma_n=0}\\
&+\underbrace{\Gamma(\omega=\omega_n)}_{=0}\sum_n\frac{\omega_n\sigma_n}{(\omega-\omega_n)}+\underbrace{\omega\Gamma\sum_n\frac{\sigma_n}{\omega-\omega_n}|_{\omega=0}}_{=0}\\
=\;& \sum_{i,n}\frac{\psi_{R,i}(\boldsymbol{r})\psi^T_{L,i}(\boldsymbol{r}')}{\mathrm{i}(\widetilde{\omega}_i-\omega)}\frac{\widetilde{\omega}_i\sigma_n}{(\widetilde{\omega}_i-\omega_n)}\,.
\end{aligned} \tag{A.41}$$

This gives the QNM expansion formula

$$\begin{aligned} a_i &= \int \psi_{L,i}^T \left[(B_\infty - B_\mathrm{b}) \frac{\omega}{\widetilde{\omega}_i - \omega} + \frac{\widetilde{\omega}_i}{\widetilde{\omega}_i - \omega} \sum_n \frac{\sigma_n}{\widetilde{\omega}_i - \omega_n} \right] \psi_\mathrm{inc} \\ &= \int \psi_{L,i}^T \left[(B_\infty - B_\mathrm{b}) + \frac{\widetilde{\omega}_i}{\widetilde{\omega}_i - \omega} \Delta B(\widetilde{\omega}_i) \right] \psi_\mathrm{inc} \,, \end{aligned} \tag{A.42}$$

which is identical to the expression in Ref. [65] derived from an augmented Maxwell operator approach, and in Ref. [63] derived from residue expansion of Green's function together with more complex sum rules. Various other expansion expressions summaried in Ref. [56] can be proven to be equivalent, similar to pole methods above.

A.3.2 Resonant and background part

In this section, we derive the expression of $H(\omega)$, the frequency dependent part of the background part. When applying Mittag-Leffler expansion to Eq. (A.5) (dropping $N(\omega)$), since elements in the scattering matrix are normally not bounded at complex infinity, the $H(\omega)$ term is generally inevitable, rendering the pole expansion formula of S of no practical use. Although $H(\omega)$ cannot be obtained from Mittag-Leffler itself, by comparing

$$S = \underbrace{S_\mathrm{bg}(\omega = 0) + H(\omega) + \mathrm{i}\widetilde{K}\Omega^{-1}\widetilde{D}^T}_{=\text{“background part”}} - \underbrace{\mathrm{i}\widetilde{K}(\Omega - \omega)^{-1}\widetilde{D}^T}_{=\text{“resonant part”}} \,, \tag{A.43}$$

with Eq. (A.5), we conclude that

$$H(\omega) = S_\mathrm{bg}(\omega) - S_\mathrm{bg}(0) + \frac{\mathrm{i}\omega}{4\alpha\beta^*}(\Phi_\mathrm{inc}^\mathrm{TR}, \Delta B \Phi_\mathrm{inc}) + \widetilde{K}\frac{1}{\mathrm{i}\Omega}\widetilde{D}^T + K(\omega)\frac{1}{\mathrm{i}(\Omega - \omega)}D^T(\omega) - \widetilde{K}\frac{1}{\mathrm{i}(\Omega - \omega)}\widetilde{D}^T. \tag{A.44}$$

Due to cancellations of residues, the above expression is regular at each resonant frequency $\widetilde{\omega}_m$, hence an entire function. One can see that $H(\omega)$ contains the Born scattering term and effects of frequency-dependent coupling matrices.

Appendix B

Appendix of "Brightness Theorem for Nanophotonics"

B.1 Deriving the classical brightness theorem from its wave-scattering generalization

Here we show that the classical ray-optical brightness theorem follows from our wave-scattering generalization. Consider a ray-optical system with an entrance plane and an exit plane. Let us consider the power flow from within a differential area ΔA_1 on the entrance plane through a differential area ΔA_2 on the exit plane. In a wave-scattering framework, what are the power-normalized channels? And how many channels are there in a differential area ΔA?

In ray optics, the wavelength is taken to zero, such that even an infinitesimal area is arbitrarily large relative to the wavelength. Thus we can consider the differential area as the "box" (actually square) into which the states must fit with periodic boundary condition. If we take $\Delta A = \Delta x \Delta y$ (with z as the propagation direction), the states that satisfy periodic boundary conditions on ΔA are plane waves with k_x and k_y taking integral multiples of $2\pi/\Delta x$ and $2\pi/\Delta y$, and k_z fixed by the frequency and specific values of k_x and k_y. Thus

we can write non-normalized electric-field states as

$$\boldsymbol{E}_i = e^{i\boldsymbol{k}_i \cdot \boldsymbol{r}}. \tag{B.1}$$

where $\boldsymbol{k}_i$ is the corresponding wavevector for state i. In our manuscript, we choose for simplicity a channel definition such that the total power flowing through a given channel is 1. Since the intensity of a plane wave of amplitude $\boldsymbol{E}_0$ is $|\boldsymbol{E}_0|^2/2Z_0$, where Z_0 is the impedance of the medium, we can choose our properly normalized channel states as

$$\boldsymbol{E}_i = \sqrt{\frac{2Z_0}{\Delta A}} e^{i\boldsymbol{k}_i \cdot \boldsymbol{r}}. \tag{B.2}$$

Taking $\boldsymbol{H}_i$ to be the magnetic field of channel i, one can then verify that the real part of $(1/2)\int_{\Delta A} \boldsymbol{E}_i \times \boldsymbol{H}_j^*$ is indeed δ_{ij}, and we have a basis of power-orthonormal channels.

Now we can answer the question about the number of channels (states) within ΔA in the range from $\boldsymbol{k}_\perp$ to $\boldsymbol{k}_\perp + \Delta\boldsymbol{k}_\perp$ and k_z to $k_z + \Delta k_z$ (which determine the propagating direction), so that $k_\perp^2 + k_z^2 = k^2$. A state occupies a region $(2\pi/\Delta x)(2\pi/\Delta y) = 4\pi^2/\Delta A$ in $k_x k_y$-space, then the total number of states is (including an extra factor of two for polarization)

$$N = 2\frac{\Delta \boldsymbol{k}_\perp}{4\pi^2/\Delta A} = \frac{1}{2\pi^2}\Delta A\, k_\perp \Delta k_\perp \Delta\phi\,. \tag{B.3}$$

Since $k_\perp^2 = k^2 \sin\theta^2$, where θ is the angle between $\boldsymbol{k}$ and z axis, we have $k_\perp \Delta k_\perp = k^2 \cos\theta \sin\theta \Delta\theta$. Substitute this relation into Eq. (B.3), we obtain

$$N = \frac{1}{2\pi^2} k^2 \cos\theta \Delta A \underbrace{\sin\theta \Delta\theta \Delta\phi}_{=\Delta\Omega} = \frac{1}{2\pi^2}\frac{\omega^2}{c^2} \underbrace{n^2 \cos\theta \Delta A \Delta\Omega}_{=\text{ray optics étendue}}. \tag{B.4}$$

Since $\text{rank}\,(\boldsymbol{\rho}_{\text{out}}) = \text{rank}\,(\boldsymbol{\rho}_{\text{in}})$,the number of incoherent excitations will remain unchanged for an ideal system, we thus conclude that N is an invariant quantity in the course of wave propagation. Hence, we recover the law of étendue conservation.

Now reconsider power flowing from area $\mathrm{d}A_1$ and solid angle $\mathrm{d}\Omega_1$ at the entrance plane.

The number of channels that are (incoherently) excited is given by

$$N_1 = \frac{\omega^2}{2\pi^2 c^2} n_1^2 \cos\theta_1 \mathrm{d}A_1 \mathrm{d}\Omega_1. \tag{B.5}$$

How much power can flow into area ΔA_2 and solid angle $\Delta\Omega_2$ at the exit plane? From our modal-concentration inequality, we know that the power out on any given channel i on the exit plane is less than or equal to 1 divided by the number of excited incoming channels

$$P_{\mathrm{out},i} \leq \frac{1}{N_1} = \frac{1}{N_2}\,, \tag{B.6}$$

which is a ramification of the ray optics brightness theorem that it is impossible to concentrate rays to increase the brightness.

We could have equivalently derived the usual concentration-ratio bounds on optical concentrators, and by standard ray-optical arguments one can see that the étendue derived in Eq. (B.4) generalizes to non-ideal ray-optical systems as usual.

B.2 Alternative Proof for the brightness-concentration inequality

Here we provide a more compact—but with less physical intuition and non-constructive—proof of the fact that $\hat{\boldsymbol{u}}^\dagger \boldsymbol{S}\boldsymbol{S}^\dagger \hat{\boldsymbol{u}} \leq 1$. For any matrix $\boldsymbol{A}$, the matrices $\boldsymbol{A}^\dagger\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^\dagger$ have the same eigenvalues, as can be proven by inserting a singular valued decomposition of A into the expressions. By energy conservation, $\boldsymbol{S}$ is subunitary, and the eigenvalues of $\boldsymbol{S}^\dagger\boldsymbol{S}$, and therefore of $\boldsymbol{S}\boldsymbol{S}^\dagger$, must be smaller than 1. By the variational principle (i.e. Rayleigh quotient [108]), we therefore must have $\hat{\boldsymbol{u}}^\dagger \boldsymbol{S}\boldsymbol{S}^\dagger \hat{\boldsymbol{u}} \leq 1$ for any unit vector $\hat{\boldsymbol{u}}$.

B.3 Coupled-mode theory for étendue transmission

In this section we specify the step-by-step procedure to identify the rank of the transmission rank in coupled-mode theory, which plays the critical role in restricting étendue transmission through the system.

We start with standard coupled-mode theory: in a scattering system with N channels, we have $N \times 1$ incoming and outgoing coefficients $\boldsymbol{c}_{\text{in}}$ and $\boldsymbol{c}_{\text{out}}$, respectively, and an $M \times 1$ vector (for M resonances) of resonant-mode amplitudes $\boldsymbol{a}$. The three vectors are related by the coupled-mode equations:

$$\mathrm{i}(\boldsymbol{\Omega} - \omega)\boldsymbol{a} = \boldsymbol{K}^T \boldsymbol{c}_{\text{in}} \tag{B.7}$$

$$\boldsymbol{c}_{\text{out}} = \boldsymbol{C}\boldsymbol{c}_{\text{in}} + \boldsymbol{K}\boldsymbol{a}\,, \tag{B.8}$$

where $\boldsymbol{\Omega}$ is a matrix whose Hermitian part is diagonal, comprising the real parts of the resonant frequencies, and whose anti-Hermitian part encodes dissipation via external coupling. The matrix $\boldsymbol{K}$ denotes coupling between the modes and the incoming/outgoing channels, while $\boldsymbol{C}$ denotes direct-scattering processes, independent of the resonances. Typically one can take $\boldsymbol{C} = -\boldsymbol{I}$ (as the waveguide combiner cases in the main text), where $\boldsymbol{I}$ is the $N \times N$ identify matrix, essentially as a normalization stating that in the absence of the scatterer, all energy flows back into the channel it came in on, with a negative amplitude.

Next one can solve for $\boldsymbol{a}$ to get $\boldsymbol{c}_{\text{out}}$ in terms of $\boldsymbol{c}_{\text{in}}$:

$$\boldsymbol{c}_{\text{out}} = \big[\underbrace{\boldsymbol{C}}_{=\text{direct process}} - \underbrace{i\boldsymbol{K}\left(\boldsymbol{\Omega} - \omega\right)^{-1}\boldsymbol{K}^T}_{=\text{resonance assisted process}} \big]\boldsymbol{c}_{\text{in}}\,, \tag{B.9}$$

where the term in square brackets is the scattering matrix.

We are interested in the *resonance assisted* transmission from N_{inc} to an orthogonal subset of N_{trans} scattering channels. So we are invited to write Eq. (B.9) as

$$\underbrace{\boldsymbol{c}_{\text{out}} - \boldsymbol{C}\boldsymbol{c}_{\text{in}}}_{=\boldsymbol{c}_{\text{out}}^{\text{reson}}} = -i\boldsymbol{K}\left(\boldsymbol{\Omega} - \omega\right)^{-1}\boldsymbol{K}^T\boldsymbol{c}_{\text{in}}\,. \tag{B.10}$$

Because there are only N_{inc} non-zero excitations, it is convenient to work with a $N_{\text{inc}} \times 1$ vector $\boldsymbol{c}_{\text{inc}}$, a subset of $\boldsymbol{c}_{\text{in}}$. Similarly, as only transmissions from N_{trans} channels are collected, we can now work with a $N_{\text{trans}} \times 1$ vector $\boldsymbol{c}_{\text{trans}}$, a subset of $\boldsymbol{c}_{\text{out}}^{\text{reson}}$. Accordingly, only a submatrix of $\boldsymbol{K}$ of size $N_{\text{inc}} \times M$, denoted by $\boldsymbol{K}_{\text{inc}}$, contains the coupling information between $\boldsymbol{c}_{\text{inc}}$ to resonance modes; another submatrix of size $N_{\text{trans}} \times M$, denoted by $\boldsymbol{K}_{\text{trans}}$,

connects the resonance modes with $\boldsymbol{c}_{\text{trans}}$. Following the above decompositions, for the T-matrix (or transmission matrix) from $\boldsymbol{c}_{\text{inc}}$ to $\boldsymbol{c}_{\text{trans}}$, namely

$$\boldsymbol{c}_{\text{trans}} = \boldsymbol{T}\boldsymbol{c}_{\text{inc}}\,, \tag{B.11}$$

can be written as

$$\boldsymbol{T} = -i\boldsymbol{K}_{\text{trans}}\left(\boldsymbol{\Omega} - \omega\right)^{-1}\boldsymbol{K}_{\text{inc}}^{T}\,. \tag{B.12}$$

With the T-matrix definition of Eq. (B.11), we can analyze resonance assisted power flow into the transmission channels due to incoherent excitations of the incident channels with the same matrix-trace approach used for $\boldsymbol{c}_{\text{inc}}$, $\boldsymbol{c}_{\text{out}}$, and the scattering matrix in the main text. As a result, the average power in transmission channel i is given by

$$\langle|\boldsymbol{c}_{\text{trans},i}|^2\rangle = \hat{\boldsymbol{u}}^{\dagger}\boldsymbol{T}\boldsymbol{\rho}_{\text{inc}}\boldsymbol{T}^{\dagger}\hat{\boldsymbol{u}}\,. \tag{B.13}$$

Let us consider the total average power (equal to the average total power) transmitted onto all transmission channels, which is simply the sum of Eq. (B.13) over i:

$$\begin{aligned}
\sum_i \langle|\boldsymbol{c}_{\text{trans},i}|^2\rangle &= \sum_i \hat{\boldsymbol{u}}^{\dagger}\boldsymbol{T}\boldsymbol{\rho}_{\text{inc}}\boldsymbol{T}^{\dagger}\hat{\boldsymbol{u}} && \text{(B.14)}\\
&= \text{Tr}\left(\boldsymbol{T}\boldsymbol{\rho}_{\text{inc}}\boldsymbol{T}^{\dagger}\sum_i \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}^{\dagger}\right) && \text{(B.15)}\\
&= \text{Tr}\left(\boldsymbol{T}\boldsymbol{\rho}_{\text{inc}}\boldsymbol{T}^{\dagger}\right) = \text{Tr}\left(\boldsymbol{T}^{\dagger}\boldsymbol{T}\boldsymbol{\rho}_{\text{inc}}\right) && \text{(B.16)}\\
&= \text{Tr}\left(\boldsymbol{T}^{\dagger}\boldsymbol{T}\boldsymbol{\rho}_{\text{inc}}\sum_i \boldsymbol{v}_i\boldsymbol{v}_i^{\dagger}\right) && \text{(B.17)}\\
&= \sum_i \lambda_i \boldsymbol{v}_i^{\dagger}\boldsymbol{T}^{\dagger}\boldsymbol{T}\boldsymbol{v}_i\,, && \text{(B.18)}
\end{aligned}$$

where $\boldsymbol{v}_i$ is the eigenvector of $\boldsymbol{\rho}_{\text{inc}}$ with eigenvalue λ_i (we label those eigenvalues in decreasing order, i.e. $\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots$) . The summation over i includes all necessary channels so that $\sum_i \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}^{\dagger}$ and $\sum_i \boldsymbol{v}_i\boldsymbol{v}_i^{\dagger}$ are both identity. However, the rank of $\boldsymbol{T}$ restricts the meaningful inclusion of $\boldsymbol{v}_i$ – the number of $\boldsymbol{v}_i$ such that $\boldsymbol{T}\boldsymbol{v}_i \neq 0$ is at most rank $(\boldsymbol{T})$ as a consequence

of the rank-nullity theorem[234]. Besides, since there are at most $\mathrm{rank}(\boldsymbol{\rho}_{\mathrm{inc}})$ number of $\boldsymbol{v}_i$s are relevant, we can thus bound Eq. (B.18) by

$$\sum_i \langle |\boldsymbol{c}_{\mathrm{trans},i}|^2 \rangle \leq \sum_{i=1}^{U} \lambda_i \boldsymbol{v}_i^\dagger \boldsymbol{T}^\dagger \boldsymbol{T} \boldsymbol{v}_i \,, \tag{B.19}$$

where $U = \min\big(\,\mathrm{rank}(\boldsymbol{\rho}_{\mathrm{inc}}), \mathrm{rank}\,(\boldsymbol{T})\,\big)$.

In the main text, we used a coherent-scattering example to show that for each individual i, $\hat{\boldsymbol{u}}^\dagger \boldsymbol{S}\boldsymbol{S}^\dagger \hat{\boldsymbol{u}} \leq 1$. Exactly the same proof, but with the inc/trans channels replacing the in/out channels, leads to the same inequality for $\boldsymbol{T}$, i.e. $\hat{\boldsymbol{u}}^\dagger \boldsymbol{T}\boldsymbol{T}^\dagger \hat{\boldsymbol{u}} \leq 1$ and $\boldsymbol{v}_i^\dagger \boldsymbol{T}^\dagger \boldsymbol{T} \boldsymbol{v}_i \leq 1$. Thus all of the eigenvalues of $\boldsymbol{T}\boldsymbol{T}^\dagger$ and $\boldsymbol{T}^\dagger\boldsymbol{T}$ thus both less than or equal to one (and they are greater than or equal to zero because $\boldsymbol{T}\boldsymbol{T}^\dagger$ and $\boldsymbol{T}^\dagger\boldsymbol{T}$ are positive semidefinite). This reduce Eq. (B.18) into

$$\sum_i \langle |\boldsymbol{c}_{\mathrm{trans},i}|^2 \rangle \leq \sum_{i=1}^{U} \lambda_i \,. \tag{B.20}$$

We can still do better by bounding the rank of $\boldsymbol{T}$ using the fact that $\mathrm{rank}(AB) \leq \min\big(\,\mathrm{rank}(A), \mathrm{rank}(B)\big)$. We write again the coupled-mode expression for $\boldsymbol{T}$, Eq. (B.12), now enumerating the number of rows and columns of each matrix:

$$\boldsymbol{T} = -i \underbrace{\boldsymbol{K}_{\mathrm{trans}}}_{N_{\mathrm{trans}}\times M} \underbrace{(\boldsymbol{\Omega} - \omega)^{-1}}_{M\times M} \underbrace{\boldsymbol{K}_{\mathrm{inc}}^T}_{M\times N_{\mathrm{inc}}} \,. \tag{B.21}$$

By recursive application of the rank inequality for matrix products, we have

$$\mathrm{rank}(\boldsymbol{T}) \leq \min\left(N, M, N_{\mathrm{trans}}\right). \tag{B.22}$$

Finally, combining Eq. (B.20), Eq. (B.22) and the fact $\mathrm{rank}(\boldsymbol{\rho}_{\mathrm{in}}) \leq N_{\mathrm{inc}}$ gives

$$\sum_i \langle |\boldsymbol{c}_{\mathrm{trans},i}|^2 \rangle \leq \sum_{i=1}^{\min(\mathrm{rank}(\boldsymbol{\rho}_{\mathrm{inc}}), M, N_{\mathrm{trans}})} \lambda_i, \tag{B.23}$$

where eigenvalues are placed in decreasing order with $\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots$. In the case of

equal, incoherent excitations (for which all of the eigenvalues of $\boldsymbol{\rho}_{\mathrm{inc}}$ are $1/N_{\mathrm{inc}}$):

$$\sum_i \langle |\boldsymbol{c}_{\mathrm{trans},i}|^2 \rangle \leq \frac{\min\left(N_{\mathrm{inc}}, M, N_{\mathrm{trans}}\right)}{N_{\mathrm{inc}}} \,. \tag{B.24}$$

Appendix C

Appendix of "Quantum Control Bounds"

C.1 Three-level Hamiltonians

The asymmetric double-well case is taken from Sec. 2.8 of [155],

$$
\begin{aligned}
H &= \begin{pmatrix} \omega_0 & 0 & 0 \\ 0 & \omega_1 & 0 \\ 0 & 0 & \omega_2 \end{pmatrix} - \epsilon(t) \begin{pmatrix} \mu_{00} & \mu_{01} & \mu_{02} \\ \mu_{10} & \mu_{11} & \mu_{12} \\ \mu_{20} & \mu_{21} & \mu_{22} \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.1568 & 0 \\ 0 & 0 & 0.7022 \end{pmatrix} - \epsilon(t) \begin{pmatrix} -2.5676 & 0.3921 & 0.6382 \\ 0.3921 & 2.3242 & -0.7037 \\ 0.6382 & -0.7037 & -0.5988 \end{pmatrix} . \qquad \text{(C.1)}
\end{aligned}
$$

We consider a maximum control amplitude of $|\epsilon| \leq 0.15$ in this example.

For the transmon example, the exact Hamiltonian is

$$\begin{aligned} H &= \begin{pmatrix} \omega_0 & 0 & 0 \\ 0 & \omega_1 & 0 \\ 0 & 0 & \omega_2 \end{pmatrix} - \epsilon(t) \begin{pmatrix} 0 & \mu_{01} & 0 \\ \mu_{10} & 0 & \mu_{12} \\ 0 & \mu_{21} & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1.9 & 0 \\ 0 & 0 & 3.7 \end{pmatrix} + \epsilon(t) \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & \sqrt{2} \\ 0 & \sqrt{2} & 0 \end{pmatrix}, \end{aligned} \tag{C.2}$$

with a maximum control amplitude $|\epsilon| \leq 0.3$ in this example.

C.2 Finding a binary pulse equivalent to a continuous one by local averaging

Given a Hamiltonian $H(t) = H_0(t) + \varepsilon(t)H_c(t)$ with some control $\varepsilon(t)$, designed controls might be of "bang–bang" type, taking only two discrete values, or they might be smooth, continuous controls, possibly bounded in magnitude. From a theoretical bound perspective, the latter case is subsumed by the former: any smooth, continuous control can be approximated to arbitrarily high accuracy with a particular bang–bang control. Intuitively, one can see that this might be true: if you oscillate a bang–bang control at high enough frequency, oscillating over smaller time scales than any transitions (real or virtual) induced by the Hamiltonian, then the wave function will not respond to the particular high-frequency details of the control but rather to a homogenized average (which will lie between the two extremes). Mathematically, there is a rich literature on the field of "homogenization" theory [235], for applications from material science to optics. Here, we prove that any smooth, continuous control can be approximated by a bang–bang control, which then allows us to use bang–bang controls in the formulation of our bounds. (Note the key point that such bang–bang controls would not ever need to be implemented; they simply inform us of the generality of the bounds.)

Suppose that we are interested in the time evolution of the system from t_0 to t_f, with

continuous control $\varepsilon_c(t)$ and propagator $U_c(t_f, t_0)$. We can divide $[t_0, t_f]$ into N intervals of equal length $\tau = (t_f - t_0)/N$: $[t_0, t_1], [t_1, t_2], \ldots, [t_{N-1}, t_N]$ ($t_f = t_N$), such that each interval is shorter than any timescale of the Hamiltonian. We can then write $U_c(t_f, t_0) = U_c(t_N, t_{N-1}) \cdots U_c(t_2, t_1) U_c(t_1, t_0)$. Since each interval is smaller than any transition time scale, we can apply first-order perturbation theory to each $U(t_{i+1}, t_i)$ to analyze the accuracy of a bang–bang-control approximation. First, we analyze the accuracy of using a first-order approximation to the time-evolution operator over each interval:

$$
\begin{aligned}
U_c(t_{i+1}, t_i) &= U_0(t_{i+1}, t_i) - \frac{\mathrm{i}}{\hbar} \int_{t_i}^{t_{i+1}} \mathrm{d}t \, U_0(t_{i+1}, t) H_I U_0(t, t_i) \epsilon_c(t) + e_i' && \text{(C.3)} \\
&\quad \text{(Taylor expand} \quad U_0(t_{i+1}, t) H_I U_0(t, t_i) \quad \text{around} \quad t = t_i) \\
&= \underbrace{U_0(t_{i+1}, t_i) - \frac{\mathrm{i}}{\hbar} U_0(t_{i+1}, t_i) H_I \int_{t_i}^{t_{i+1}} \mathrm{d}t \, \epsilon_c(t)}_{= \tilde{U}_c(t_{i+1}, t_i)} + e_i && \text{(C.4)}
\end{aligned}
$$

where U_0 is the propagator of H_0 and e_i' is the error term of the perturbation series of $\mathrm{O}(\tau^2)$. Together with the error due to Taylor expansion of $O(\tau^2)$, the overall local error e_i is of $O(\tau^2)$. If we stitch these local approximations back to the global propagator, i.e. $U_c(t_f, t_0) = U_c(t_N, t_{N-1}) \cdots U_c(t_2, t_1) U_c(t_1, t_0)$, we see that the global error is $\sum_{i=0}^{N-1} U_0(t_N, t_{i+1}) e_i U_0(t_i, t_0)$, which is at most N ($\sim 1/\tau$) copies of $O(\tau^2)$, which is of $O(\tau)$. Hence the global error goes to zero as $\tau \to 0$, and we can approximate U_c with the local apprxomations $\tilde{U}_c$:

$$
U_c(t_f, t_0) \sim \tilde{U}_c(t_N, t_{N-1}) \cdots \tilde{U}_c(t_2, t_1) \tilde{U}_c(t_1, t_0) \quad \text{as} \quad \tau \to 0. \tag{C.5}
$$

The above perturbation analysis shows that if we can find a bang–bang control $\epsilon_b(t)$ with propagator $U_b(t_f, t_0)$ whose local approximation $\tilde{U}_b(t_{i+1}, t_i)$ agrees with $\tilde{U}_c(t_{i+1}, t_i)$ up to the first order in τ, then we will have

$$
U_c(t_f, t_0) \sim \tilde{U}_c(t_N, t_{N-1}) \cdots \tilde{U}_c(t_1, t_0) \sim \tilde{U}_b(t_N, t_{N-1}) \cdots \tilde{U}_b(t_1, t_0) \sim U_b(t_f, t_0) \tag{C.6}
$$

as $\tau \to 0$. This will show the global equivalence between $U_c(t_f, t_0)$ and $U_b(t_f, t_0)$. Such an $\epsilon_b(t)$ is easy to find: by the form of the perturbation in Eq. (C.4), any choice of ϵ_b is valid

as long as $\int_{t_i}^{t_{i+1}} \mathrm{d}t\, \epsilon_b(t) = \int_{t_i}^{t_{i+1}} \mathrm{d}t\, \epsilon_c(t)$ for all i, which can be achieve simply by requiring the average of the bang–bang control equaling the average of the continuous control. (This choice of $\epsilon_b(t)$ over $[t_i, t_{i+1}]$ has $\epsilon_b = \varepsilon_{\min}$ for a duration t' and then switch to $\epsilon_b = \varepsilon_{\max}$ for the rest of the time to t_{i+1}, for $t' = \frac{\varepsilon_{\max}\tau - M_i}{\varepsilon_{\max} - \varepsilon_{\min}}$, where $M_i = \int_{t_i}^{t_{i+1}} \mathrm{d}t\, \epsilon_c(t)$. Certainly, there are other choices as well.)

We illustrate the above proof numerically through the transmon example in the main text. We apply a continuous control $\varepsilon_c(t) = E\cos(\omega t)$ with $\omega = \omega_1$ and $E = 0.5$ from $t_0 = 0$ to $T = 12$. We want to find a binary pulse with $\varepsilon_{\max} = -\varepsilon_{\min} = E$, such that the dynamics $\psi_b(t)$ approaches $\psi_c(t)$ under the continuous pulse when τ goes to zeros. We construct such $\varepsilon_b(t)$ according to the method described in the previous paragraph. In Fig. C.1, we compare ψ_b with ψ_c for $\tau = 1.2$ for (a), 0.3 for (c) and 0.15 for (e) and show their respective pulses in (b), (d) and (f). We see from In Fig. C.1 the the evolution under each $\varepsilon_b(t)$ (red) gets closer to the evolution under $\varepsilon_c(t)$ (black dash) as τ decreases. More explicitly, we plot in Fig. C.2 the relative difference between evolution of ε_c and ε_b, measured by $\frac{(\int_{t_0}^{T} |\psi_b(t) - \psi_c(t)|^2)^{\frac{1}{2}}}{(\int_{t_0}^{T} |\psi_c(t)|^2)^{\frac{1}{2}}}$, against $\frac{\tau}{T}$. We can see that the convergence demonstrates the $O(\tau)$ behavior proved above. We emphasize again that the purpose of numerical results here is to show the equivalence between continuous and binary pulses when τ goes to zero. In the bound computation, we do not need to find such $\varepsilon_b(t)$ and the bound converges automatically and faster than $O(\tau)$ in the proof.

To conclude, we have proven that for any continuously valued control $\varepsilon_c(t)$ there is a corresponding binary control $\varepsilon_b(t)$ producing the same time evolution as $\epsilon_c(t)$, when the local averaging time τ approaches zero. Hence, a bound derived for bang–bang controls will include all possible bounded continuous controls as well.

C.3 Local probability conservation laws as a subset of D-matrix constraints

In this section, we show explicitly that Eq. 4 in the main text, which represents local probability conservation, is a subset of constraints of Eq. 3 in the main text with $D_i(t) = \eta(t_1 - t)\mathcal{I}$ with $t \geq t_0$, i.e. identity operator $\mathcal{I}$ supported from t_0 to t_1.

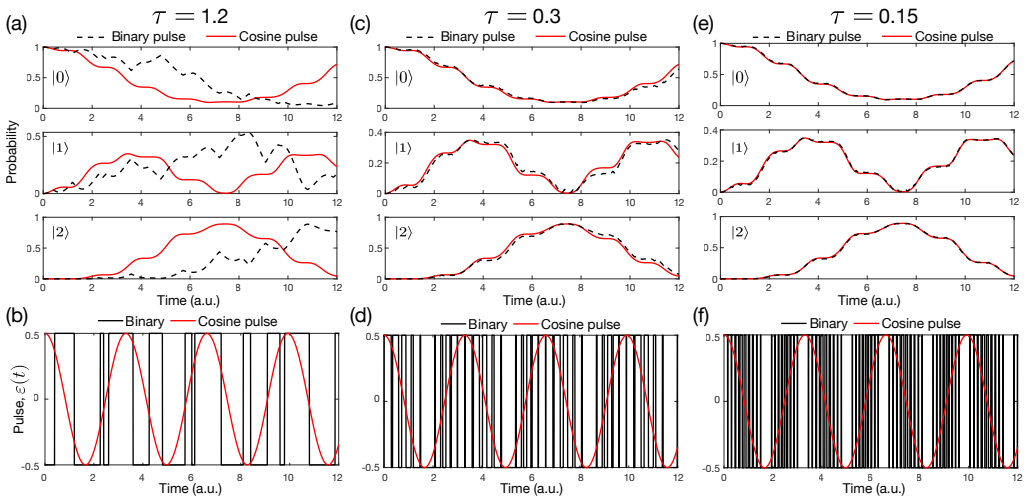

Figure C.1: (a) Comparisons between probability time evolution under $\varepsilon_c(t)$(red) and $\varepsilon_b(t)$(black dash) for all three levels, from $t_0 = 0$ to $T = 12$. (b) $\varepsilon_c(t) = E\cos(\omega t)$ is plotted in red and the $\varepsilon_b(t)$ in black, obtained through averaging for $\tau = 1.2$. (c) and (d), (e) and (f) are similar to (a) and (b), but the former pair is for $\tau = 0.3$ and the latter is for $\tau = 0.15$. One can see that as τ gets smaller, the $\varepsilon_b(t)$ oscillates more rapidly and the evolution gets closer to the one produce by $\varepsilon_c(t)$. For $\tau = 0.15$, the effect of $\varepsilon_b(t)$ and $\varepsilon_c(t)$ are almost identical.

Now consider Eq. 2 in the main text ,

$$\int_{t=t_0}^{T}\int_{t'=t_0}^{T}\Phi^\dagger(t)D_i(t)\left[\frac{H_c^{-1}}{\varepsilon_{\max}}\delta(t-t')+\frac{i}{\hbar}G_0^+(t,t')\right]\Phi(t')\mathrm{d}t\,\mathrm{d}t' = \int_{t_0}^{T}\Phi^\dagger(t)D_i(t)U_0(t,t_0)\mathrm{d}t, \quad \text{(C.7)}$$

where $G_0^+(t,t') = U_0(t,t')\eta(t-t')$. Now with $D_i(t) = \eta(t_1 - t)\mathcal{I}$, it becomes

$$\int_{t_0}^{t_1}\Phi^\dagger(t)\frac{H_c^{-1}}{\varepsilon_{\max}}\Phi(t)\mathrm{d}t+\frac{i}{\hbar}\int_{t=t_0}^{t_1}\int_{t'=t_0}^{t}\Phi^\dagger(t)U_0(t,t')\Phi(t')\mathrm{d}t\,\mathrm{d}t' = \int_{t_0}^{t_1}\Phi^\dagger(t)U_0(t,t_0)\mathrm{d}t\,. \quad \text{(C.8)}$$

Taking Hermitian conjugation both side of Eq. (C.8), and using $U_0^\dagger(t,t') = U_0(t',t)$ and $(H_c^{-1})^\dagger = H_c^{-1}$, produces

$$\int_{t_0}^{t_1}\Phi^\dagger(t)\frac{H_c^{-1}}{\varepsilon_{\max}}\Phi(t)\mathrm{d}t-\frac{i}{\hbar}\int_{t=t_0}^{t_1}\int_{t'=t_0}^{t}\Phi^\dagger(t')U_0(t',t)\Phi(t)\mathrm{d}t\,\mathrm{d}t' = \int_{t_0}^{t_1}U_0^\dagger(t,t_0)\Phi(t)\mathrm{d}t \quad \text{(C.9)}$$

Note we can exchange the t, t' integration in the the following integral

$$\int_{t=t_0}^{t_1}\int_{t'=t_0}^{t}\ldots\mathrm{d}t'\mathrm{d}t \quad = \quad \int_{t'=t_0}^{t_1}\int_{t=t'}^{t_1}\ldots\mathrm{d}t\mathrm{d}t'\,.$$

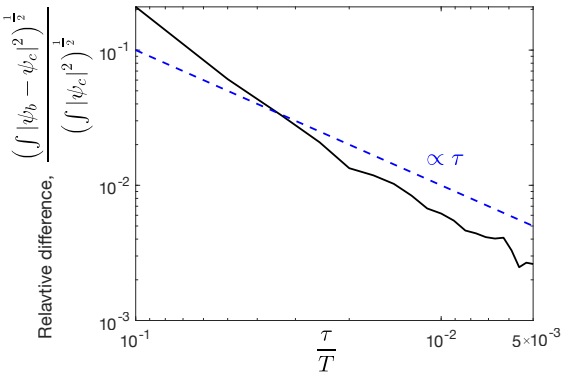

Figure C.2: The relative difference between evolution of ε_c and ε_b, measured by $\frac{(\int_{t_0}^{T} |\psi_b(t)-\psi_c(t)|^2)^{\frac{1}{2}}}{(\int_{t_0}^{T} |\psi_c(t)|^2)^{\frac{1}{2}}}$ is plotted against $\frac{\tau}{T}$. It shows the $O(\tau)$ convergence for small τ.

As a result, the double integral in Eq. (C.9) becomes

$$\begin{aligned} & \int_{t=t_0}^{t_1}\int_{t'=t_0}^{t} \Phi^\dagger(t')U_0(t',t)\Phi(t)\mathrm{d}t\,\mathrm{d}t' \quad \text{(exchange } t,t' \text{ integrations)} \\ = & \int_{t'=t_0}^{t_1}\int_{t=t'}^{t_1} \Phi^\dagger(t')U_0(t',t)\Phi(t)\mathrm{d}t\,\mathrm{d}t' \quad \text{(relabeling by } t\to t', t'\to t) \\ = & \int_{t=t_0}^{t_1}\int_{t'=t}^{t_1} \Phi^\dagger(t)U_0(t,t')\Phi(t')\mathrm{d}t\,\mathrm{d}t' \,. \end{aligned}$$

Now we subtract Eq. (C.8) by Eq. (C.9) to obtain the imaginary part of Eq. (C.8), and we get

$$\frac{1}{\hbar}\int_{t_0}^{t_1}\int_{t_0}^{t_1} \Phi^\dagger(t)U_0(t,t')\Phi(t')\mathrm{d}t\,\mathrm{d}t' = 2\,\mathrm{Im}\int_{t_0}^{t_1} \Phi^\dagger(t)U_0(t,t_0)\mathrm{d}t\,, \tag{C.10}$$

which is exactly Eq. 4 in the main text, since $U^\dagger U = \mathcal{I}$ by probability conservation. $\int_{t_0}^{t_1} \Phi^\dagger(t)\frac{H_c^{-1}}{\varepsilon_{\max}}\Phi(t)\mathrm{d}t$ disappears since its imaginary part is zero. This makes sense since H_c is a Hermitian matrix and no dissipation is present.

Often $D_i(t) = \delta(t-t')D$ is localized at time t' (D is some constant matrix). To connect such $D_i(t)$ with the above result, we notice that $\int_{t_0}^{t_1}\delta(t-t')\mathrm{d}t' = \eta(t_1-t)$ for $t \geq t_0$. As a result, we can interpret the above conservation law as the imaginary part of a linear combination of constraints with $D_i(t) = \delta(t-t')\mathcal{I}$ for $t_0 \leq t' \leq t_1$, a subset of all D-matrix constraints. More generally, D is not necessarily $\mathcal{I}$ and the real part is also a valid D-matrix constraint. We thus can interpret the entire set of D-matrix constraints as generalized local probability conservation laws.

C.4 Bound formulation for problems with multiple controllable terms

In the main text, we only consider problems with a single control. We show here that our framework extends naturally to problems with multiple controls. Consider a Hamiltonian $H(t) = H_0(t) + \sum_{i=1}^{N} \varepsilon_i(t) H_c^i(t)$, where $H_0(t)$ is the non-controllable part and $\sum_{i=1}^{N} \varepsilon_i(t) H_c^i$ is the controllable part. To avoid overloading the notation, we here formulate the bound problem for a two-term problem with $H(t) = H_0(t) + \varepsilon_1(t) H_c^1 + \varepsilon_2(t) H_c^2$, which can be extended to more terms in a straightforward way.

The first step is to write the integral-equation form (analog of Eq. (1) of the main text) for multiple controls. To do this, we will assume the two controls cannot be turned on at the same times. (Note that two controls that can be turned on at the same time is equivalent to three controls that cannot be turned on at the same time, so there is no loss of generality here.) With two independent controls, the integral equation for $U(t, t_0)$ is:

$$U(t,t_0) = U_0(t,t_0) - \frac{i}{\hbar}\int_{t_0}^{T} G_0^+(t,t') H_c^1(t) \varepsilon_1(t') U(t',t_0) - \frac{i}{\hbar}\int_{t_0}^{T} G_0^+(t,t') H_c^2(t) \varepsilon_2(t') U(t',t_0) \tag{C.11}$$

For times where $\varepsilon_2(t)$ is zero, we have

$$U(t,t_0) = U_0(t,t_0) - \frac{i}{\hbar}\int_{\{\varepsilon_1(t)\neq 0\}} G_0^+(t,t') H_c^1 \varepsilon_1(t') U(t',t_0) \mathrm{d}t' \,, \tag{C.12}$$

and for times where $\varepsilon_1(t)$ is zero,

$$U(t,t_0) = U_0(t,t_0) - \frac{i}{\hbar}\int_{\{\varepsilon_2(t)\neq 0\}} G_0^+(t,t') H_c^2 \varepsilon_2(t') U(t',t_0) \mathrm{d}t' \,. \tag{C.13}$$

We then left multiply $U^\dagger(t,t_0)\varepsilon_1(t) H_c^1 D_i(t)$ to Eq. (C.12) and $U^\dagger(t,t_0)\varepsilon_2(t) H_c^2 D_j(t)$ to Eq. (C.13), and integrate over t from t_0 to T. Then Eq. (C.12) and Eq. (C.13) become

$$\begin{aligned}
&\int_{t_0}^{T} \mathrm{d}t\, U^\dagger(t,t_0)\varepsilon_1(t) H_c^1 D_i(t) U(t,t_0) = \int_{t_0}^{T} \mathrm{d}t\, U^\dagger(t,t_0)\varepsilon_1(t) H_c^1 D_i(t) U_0(t,t_0) \\
&-\frac{i}{\hbar}\int_{t_0}^{T} \mathrm{d}t \int_{\{\varepsilon_1(t)\neq 0\}} \mathrm{d}t'\, U^\dagger(t,t_0)\varepsilon_1(t) H_c^1 D_i(t) G_0^+(t,t') H_c^1 \varepsilon_1(t') U(t',t_0) \,,
\end{aligned} \tag{C.14}$$

and

$$\int_{t_0}^{T} \mathrm{d}t\, U^{\dagger}(t,t_0)\varepsilon_2(t)H_c^2 D_j(t)U(t,t_0) = \int_{t_0}^{T} \mathrm{d}t\, U^{\dagger}(t,t_0)\varepsilon_2(t)H_c^2 D_j(t)U_0(t,t_0)$$
$$-\frac{i}{\hbar}\int_{t_0}^{T}\mathrm{d}t \int_{\{\varepsilon_2(t)\neq 0\}} \mathrm{d}t'\, U^{\dagger}(t,t_0)\varepsilon_2(t)H_c^2 D_j(t)G_0^{+}(t,t')H_c^2\varepsilon_2(t')U(t',t_0)\,, \qquad \text{(C.15)}$$

We define $\Phi_1(t) = \varepsilon_1(t)H_c^1 U(t,t_0)$ and $\Phi_2(t) = \varepsilon_2(t)H_c^2 U(t,t_0)$. The above equations are quadratic in Φ_1 and Φ_2, thus they hold for t' from t_0 and T. Consequently, we can replace the integral range of t' by t_0 and T and obtain constraints

$$\int_{t_0}^{T}\mathrm{d}t\int_{t_0}^{T}\mathrm{d}t'\,\Phi_1^{\dagger}(t)D_i(t)\left(\delta(t-t')\frac{(H_c^1)^{-1}}{\varepsilon_{1\max}}+\frac{i}{\hbar}G_0^{+}(t,t')\right)\Phi_1(t') = \int_{t_0}^{T}\mathrm{d}t\,\Phi_1^{\dagger}(t)D_i(t)U_0(t,t_0)\,, \qquad \text{(C.16)}$$

and

$$\int_{t_0}^{T}\mathrm{d}t\int_{t_0}^{T}\mathrm{d}t'\,\Phi_2^{\dagger}(t)D_j(t)\left(\delta(t-t')\frac{(H_c^2)^{-1}}{\varepsilon_{2\max}}+\frac{i}{\hbar}G_0^{+}(t,t')\right)\Phi_2(t') = \int_{t_0}^{T}\mathrm{d}t\,\Phi_2^{\dagger}(t)D_j(t)U_0(t,t_0)\,, \qquad \text{(C.17)}$$

similar to Eq. (3) in the main text, but now with two different components. In addition to these two sets of constraints, we also need to impose that only one of $\varepsilon_1(t)$ and $\varepsilon_2(t)$ can be nonzero at any given time, which is represented in the constraint

$$\Phi_1(t)\Phi_2^{\dagger}(t) = 0 \quad \text{for all components,} \qquad \text{(C.18)}$$

for all times t. Hence bounds for two controls can be formulated as

$$\begin{aligned}
\max_{\Phi_1,\Phi_2}\quad & f(\Phi_1,\Phi_2)\\
\text{s.t.}\quad & \text{Equation (C.16) satisfied for all } D_i(t)\,,\\
& \text{Equation (C.17) satisfied for all } D_j(t)\,,\\
& \Phi_1(t)\Phi_2^{\dagger}(t) = 0 \quad \text{for all } t\,, \text{for all components}\,.
\end{aligned}$$

One can see immediately how to generalize to multiterm pulse designs. For $H(t) = H_0(t) + \sum_{i=1}^{N}\varepsilon_i(t)H_c^i$, we will have N different Φ_i with its corresponding sets of quadratic constraints. We also need to impose that $\Phi_i(t)\Phi_j^{\dagger}(t) = 0$ for all $i \neq j$.

Finally, note that if both $\varepsilon_1(t)$ and $\varepsilon_2(t)$ can be nonzero at the same time, it is equivalent to considering a problem with three controllable terms: $\varepsilon_1(t)H_c^1$, $\varepsilon_2(t)H_c^2$, and $\varepsilon_1(t)H_c^1 + \varepsilon_2(t)H_c^2$, only one of which can be nonzero at any given time.

C.5 Bound formulation for open systems governed by master equations

In the main text, we only consider problems whose dynamics is governed by the Schrödinger equation. Now we show how to apply our bound framework to open systems described by differential equations for the density matrix ρ. In principle, this framework applies to any equation that is linear in ρ. Here we specifically consider the Lindblad master equation [236]:

$$\frac{\partial}{\partial t}\rho = -\frac{i}{\hbar}[H,\rho] + \sum_i \gamma_i \left(A_i \rho A_i^\dagger - \frac{1}{2}\left\{ A_i^\dagger A_i, \rho \right\} \right) , \tag{C.19}$$

where we also assume the Hamiltonian has the form $H = H_0 + \varepsilon(t)H_c$, and the second term is the dissipative part. A common trick is to vectorize the equation to form "super-operators" [236], through flattening the n^2 by n^2 density matrix ρ into a column vector $\boldsymbol{r}$ by ordering the columns of ρ in a single vector [237]. Then all operators are transformed according to this new ordering (basis):

$$\begin{aligned}
[H,\rho] &\longrightarrow \left(I \otimes H - H^T \otimes I\right) \boldsymbol{r} \\
&= \left(I \otimes H_0 - H_0^T \otimes I\right) \boldsymbol{r} + \left(I \otimes H_c - H_c^T \otimes I\right) \boldsymbol{r} && \text{(C.20)} \\
A_i \rho A_i^\dagger &\longrightarrow \left((A_i^\dagger)^T \otimes A_i \right) \boldsymbol{r} && \text{(C.21)} \\
A_i A_i^\dagger \rho &\longrightarrow \left(I \otimes A_i^\dagger A_i \right) \boldsymbol{r} && \text{(C.22)} \\
\rho A_i^\dagger A_i &\longrightarrow \left((A_i^\dagger A_i)^T \otimes I \right) \boldsymbol{r} . && \text{(C.23)}
\end{aligned}$$

If we define the vectorized $-\frac{i}{\hbar}[H_0, \rho]$ by L_0, the vectorized dissipative term by L_D, and vectorized $-\frac{i}{\hbar}[H_c, \rho]$ by L_c, we turn the original equation into

$$\frac{\partial}{\partial t}\boldsymbol{r} = (L_0 + L_D)\boldsymbol{r} + L_c\boldsymbol{r}\,. \tag{C.24}$$

This equation again can be turned into the integral form

$$\boldsymbol{r}(t) = \boldsymbol{r}_0(t) + \int_{t_0}^{T} \mathrm{d}t' G^+(t,t')L_c(t')\boldsymbol{r}(t'), \tag{C.25}$$

where $\boldsymbol{r_0}(t)$ is the evolution of $\boldsymbol{r}$ without the presence of L_c, and $G^+(t,t')$ is the retarded Green's function of $\frac{\partial}{\partial t}\boldsymbol{r} = (L_0 + L_D)\boldsymbol{r}$. The dissipative nature of the problem is encoded in $G^+(t,t')$.

Eq. (C.25) is the starting point of the bound formulation, just as Eq. (1) in the main text for the Schrödinger equation. Carrying out the same manipulations as in the main text gives the constraints

$$\begin{aligned} &\int_{t_0}^{T} \mathrm{d}t\, \boldsymbol{r}^\dagger(t)L_c(t)D_i(t)\boldsymbol{r}(t) + \int_{t_0}^{T}\mathrm{d}t \int_{t_0}^{T} \mathrm{d}t'\, \boldsymbol{r}^\dagger(t)L_c(t)D_i(t)G^+(t,t')L_c(t')\boldsymbol{r}(t') \\ = &\int_{t_0}^{T} \mathrm{d}t \boldsymbol{r}^\dagger(t)L_c(t)D_i(t)\boldsymbol{r}(t)\,, \end{aligned} \tag{C.26}$$

which are analogous to Eq. (2) in the main text. The quadratic constraints can then be formed analogously.

C.6 Numerical details

In this section, we provide details for the numerical scheme used in the main text. Unlike conventional time-stepping methods to solve initial-value differential equations, we directly use the integral equation (Eq. (1) of the main text) to solve for the dynamics. We do this for three reasons: (1) In our bound method, constraints come from the integral equation, so the discretized form of the integral equation is directly available after solving the bound problem. (2) Pulse design is straightforward to perform with methods based on the integral equation. (3) Methods based on the integral equation are equally viable as those based on

the differential equation.

C.6.1 Nyström discretization

We consider the integral equation in Eq. (1). We consider systems with L levels, modeling either discrete levels or a subset of continuous states, over times $[t_0, T]$. Here we isolate a single column of the time-evolution operator, essentially selecting an initial state $|\psi_0\rangle$ and the time-dependent state $|\psi(t)\rangle = U(t, t_0)\,|\psi_0\rangle$. Then the integral equation for $|\psi(t)\rangle$ is

$$|\psi(t)\rangle = |\psi_0(t)\rangle - \frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} U_0(t,t')H_c\varepsilon(t')\,\big|\psi(t')\big\rangle\,\mathrm{d}t'\,, \tag{C.27}$$

where we have replaced the retarded Green's function with the time-evolution operator multiplied by a step function (changing the upper limit of the integral to t). This is in the form of a Volterra integral equation of the second kind.

The Nyström method [238] for discretizing this equation consistent of defining quadrature nodes $\{t_i\}_{i=0}^{N}\Phi_{\mathrm{in}}[t_0, T]$ and corresponding weights , and enforcing Eq. (C.27) to hold at $\{t_i\}_{i=0}^{N}$

$$|\psi(t_i)\rangle = |\psi_0(t_i)\rangle - \frac{\mathrm{i}}{\hbar}\int_{t_0}^{t_i} U_0(t_i,t')H_c\varepsilon(t')\,\big|\psi(t')\big\rangle\,\mathrm{d}t'\,, \quad i = 0, 1, 2, \ldots, N\,, \tag{C.28}$$

and then approximating the integral operator by

$$\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t_i} U_0(t_i,t')H_c\varepsilon(t')\,\big|\psi(t')\big\rangle\,\mathrm{d}t' \approx \sum_{j=1}^{t_i} w_{ij}\,\underbrace{\frac{\mathrm{i}}{\hbar}U_0(t_i,t_j)H_c\varepsilon(t_j)}_{=K_{ij}}\,|\psi(t_j)\rangle\ . \tag{C.29}$$

for some weights w_{ij}. Then Eq. (C.27) becomes a linear system

$$|\psi(t_i)\rangle = \psi_0(t_i) - \sum_{j=0}^{i} w_{ij}K_{ij}\,|\psi(t_j)\rangle\ , \quad i = 0, 1, 2, \ldots, N\,, \tag{C.30}$$

or more compactly as

$$\boldsymbol{\psi} = \boldsymbol{\psi}_0 - \boldsymbol{A}\boldsymbol{\psi}\,, \tag{C.31}$$

where $\boldsymbol{\psi}$ and $\boldsymbol{\psi}_0$ are column vector of length $L(N+1)$ with the ith column block $\boldsymbol{\psi}_i = |\psi(t_i)\rangle$

and $(\boldsymbol{\psi}_0)_i = |\psi_0(t_i)\rangle$ respectively. And $\boldsymbol{A}$ is a matrix of size $L(N+1) \times L(N+1)$ whose explicit form we will show later. Then $\boldsymbol{\psi}$ can be solved as $(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{\psi}_0$.

We note that the convergence rate of $\boldsymbol{\psi}$ to the exact value ψ is the same as the order of quadrature used in Eq. (C.29). The matrix norm of $\boldsymbol{A}$ is provably small, so $\boldsymbol{I}+\boldsymbol{A}$ is always well-conditioned. To conclude, the scheme we described above is stable and convergences quickly for good choice of $\{t_i\}_i^N$ and weights.

In this work, we use the trapezoidal rule

$$\int_a^b f(x)\mathrm{d}x \approx \frac{h}{2}\left(f(a) + 2f(a+h) + \ldots + 2f(b-h) + f(b)\right) + O(h^2) \tag{C.32}$$

where h is the spacing between the equally spaced nodes from a to b. As a result, the method we are using has convergence rate equivalent to Runge Kutta of order 2. Without special techniques for the matrix inversion process, this method is not as efficient as Runge Kutta of order 2. Despite of this drawback and the fact that higher order methods are available, the method here is sufficient for the examples we consider. If we apply the trapezoidal rule to Eq. (C.30), then the nodes $\{t_i\}_{i=0}^N \Phi_{\text{in}}[t_0, T]$ are $t_0, t_1, \ldots, t_N = T$ with $t_i = t_0 + ih$ and $h = \frac{T-t_0}{N}$. Eq. (C.30) then becomes

$$\begin{aligned}
|\psi(t_0)\rangle &= |\psi_0(t_0)\rangle \\
|\psi(t_1)\rangle &= |\psi_0(t_1)\rangle - \frac{h}{2}(K_{10}\,|\psi(t_0)\rangle + K_{11}\,|\psi(t_1)\rangle) \\
|\psi(t_2)\rangle &= |\psi_0(t_2)\rangle - \frac{h}{2}(K_{20}\,|\psi(t_0)\rangle + 2K_{21}\,|\psi(t_1)\rangle + K_{22}\,|\psi(t_2)\rangle) \\
&\vdots \\
|\psi(t_N)\rangle &= |\psi_0(t_N)\rangle - \frac{h}{2}(K_{N0}\,|\psi(t_0)\rangle + 2\sum_{j=1}^{N-1} K_{Nj}\,|\psi(t_j)\rangle + K_{NN}\,|\psi(t_N)\rangle)\,,
\end{aligned} \tag{C.33}$$

which is Eq. (C.31) in component form with the trapezoidal rule.

To solve for U, we simply replace the vector $\boldsymbol{\psi}$ by the matrix $\boldsymbol{U} = \begin{pmatrix} \boldsymbol{\psi}^1 & \boldsymbol{\psi}^2 & \ldots \boldsymbol{\psi}^L \end{pmatrix}$ and $\boldsymbol{\psi}_0$ by the matrix $\boldsymbol{U}_0 = \begin{pmatrix} \boldsymbol{\psi}_0^1 & \boldsymbol{\psi}_0^2 & \ldots \boldsymbol{\psi}_0^L \end{pmatrix}$, both of size $L(N+1) \times L$, where $\boldsymbol{\psi}^i$ represents the discretized version of different states wave function $\left|\psi^i(t)\right\rangle = U(t,t_0)\left|\psi_0^i(t_0)\right\rangle$.

By comparing Eq. (C.33) and Eq. (C.30), we can see that the weight $w_{ij} = h$ for $i \neq j$ and $w_{ij} = h/2$ for $i = j$ and $w_{ij} = 0$ for $i < j$. Then the discretized version of Eq. (1) in the main text is given by

$$\boldsymbol{U} = \boldsymbol{U}_0 - \boldsymbol{A}\boldsymbol{U}\,, \tag{C.34}$$

and $\boldsymbol{U} = (\boldsymbol{I} + \boldsymbol{A})^{-1}\boldsymbol{U}_0\,.$

C.6.2 Discretization for quadratic constraints

With the understanding of how the integral equation is discretized, we are ready to give the discretization form of Eq. (3), which constitutes the constraints for computing the bound. Constraints in Eq. (3) are given by

$$\begin{aligned}&\int_{t_0}^{T} \mathrm{d}t \int_{t_0}^{T} \mathrm{d}t'\, \Phi^\dagger(t) D_i(t) \left(\delta(t-t')\frac{H_c^{-1}}{\varepsilon_{\max}} + \frac{i}{\hbar} G_0^+(t,t')\right) \Phi(t') \\ &= \int_{t_0}^{T} \mathrm{d}t\, \Phi^\dagger(t) D_i(t) U_0(t,t_0)\,,\end{aligned} \tag{C.35}$$

or

$$\begin{aligned}&\int_{t_0}^{T} \mathrm{d}t \int_{t_0}^{t} \mathrm{d}t'\, \Phi^\dagger(t) D_i(t) \frac{i}{\hbar} U_0(t,t') \Phi(t') + \int_{t_0}^{T} \mathrm{d}t\, \Phi^\dagger(t) D_i(t) \frac{H_c^{-1}}{\varepsilon_{\max}} \Phi(t) \\ &= \int_{t_0}^{T} \mathrm{d}t\, \Phi^\dagger(t) D_i(t) U_0(t,t_0)\,,\end{aligned} \tag{C.36}$$

We have seen in the previous section on how to discretize the t' integral. The t integral is in fact much easier with fixed integration limit, and we can apply the trapezoidal rule directly to the integral $\int_{t_0}^{T} \mathrm{d}t f(t) \approx \sum_{i=0}^{N} w_i f(t_i)$ with the same equally spaced nodes $\{t_i\}_{i=0}^{N}$ and weights w_i in Eq. (C.32). Then Eq. (C.36) becomes

$$\boldsymbol{\Phi}^\dagger \boldsymbol{D}_i \boldsymbol{W} \boldsymbol{G} \boldsymbol{\Phi} + \boldsymbol{\Phi}^\dagger \boldsymbol{D}_i \boldsymbol{W} \frac{\boldsymbol{H}_c^{-1}}{\varepsilon_{\max}} \boldsymbol{\Phi} = \boldsymbol{\Phi}^\dagger \boldsymbol{D}_i \boldsymbol{W} \boldsymbol{U}_0\,, \tag{C.37}$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{U}_0$ are matrix of size $L(N+1) \times L$. $\boldsymbol{G}$ is the discretized version of the integral operator $\int_{t_0}^{t} \mathrm{d}t' U(t,t')$. Similar to the way we discretized Eq. (C.29), we have the ij block

of $\boldsymbol{G}$

$$\boldsymbol{G}_{ij} = w_{ij}\frac{\mathrm{i}}{\hbar}U_0(t_i, t_j) \tag{C.38}$$

with w_{ij} the weight in Eq. (C.33). $\boldsymbol{W}$ is the matrix contains the effect of weights w_i from the t integral, and it is of the from

$$\boldsymbol{W} = \begin{pmatrix} w_0 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_N \end{pmatrix} \otimes \boldsymbol{I}_{L\times L}\,, \tag{C.39}$$

where $\otimes$ is the kroncker product for matrices and $\boldsymbol{I}_{L\times L}$ represents an identity matrix of dimension L. Here $\boldsymbol{H}_c^{-1}$ is

$$\boldsymbol{H}_c^{-1} = \boldsymbol{I}_{(N+1)\times(N+1)} \otimes H_c^{-1} \tag{C.40}$$

Also, since $D_i = \delta(t - t_i)D$ and D is some constant matrix of dimension L by L, we have

$$\boldsymbol{D}_i = \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix} \otimes D\,, \tag{C.41}$$

where the first diagonal matrix is the discretized $\delta(t - t_i)$ and is nonzero only in the ith diagonal entry. D of dimension L by L consists of L^2 unit basis matrices. For a two level system, all possible independent D consists of $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Higher dimension systems can be generalized accordingly.

By varying D_i over all possible t_i and D, we obtain all possible constraints in our problem in the form of a L by L matrix equality, which must hold componentwise. $\boldsymbol{\Phi}$ can be flattened column by column into a vector, and the equality constraints for each components can then be transformed into a quadratic constraint of the flattened $\boldsymbol{\Phi}$. Then semidefinite relaxation can be carried out to form the semidefinite program, which can be

solved by standard convex optimization package.

C.6.3 Pulse designs via local optimization

In this section, we describe how to efficiently compute the gradient $\frac{\partial f(U)}{\partial \varepsilon(t)}$ of the objective $f(U)$ for all t from $[t_0, T]$ by solving for the dynamics twice only. Now, we return to Eq. (C.34), the discretized version of Eq. (1) in the main text. To make the variables $\varepsilon(t_i)$ here more prominent, we separate $\varepsilon(t_i)$ from $\boldsymbol{A}$ in Eq. (C.34) and write it as

$$\boldsymbol{U} = \boldsymbol{U}_0 - \underbrace{\boldsymbol{G}\boldsymbol{H}_c\boldsymbol{\varepsilon}}_{=\boldsymbol{A}}\boldsymbol{U}\,, \tag{C.42}$$

where

$$\boldsymbol{H}_c = \boldsymbol{I}_{(N+1)\times(N+1)} \otimes H_c\,, \tag{C.43}$$

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon(t_0) & & & \\ & \varepsilon(t_1) & & \\ & & \ddots & \\ & & & \varepsilon(t_N) \end{pmatrix} \otimes \boldsymbol{I}_{L\times L}\,, \tag{C.44}$$

and $\boldsymbol{G}$ is given in Eq. (C.38). We also denote the column vector $\begin{pmatrix} \varepsilon(t_0) & \varepsilon(t_1) & \dots & \varepsilon(t_N) \end{pmatrix}^T$ by $\boldsymbol{p}$.

The goal here is to compute the discretized gradient $\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{p}}$ in a way that the work required is independent of the number of variable $\varepsilon(t_i)$. To do so, we first use the CR calculus to formally treat $\boldsymbol{U}$ and $\boldsymbol{U}^*$ as independent variables [239]; the chain rule yields

$$\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{p}} = \frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{U}}\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{p}} + \frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{U}^*}\frac{\partial \boldsymbol{U}^*}{\partial \boldsymbol{p}} = 2\,\mathrm{Re}\,\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{U}}\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{p}}\,. \tag{C.45}$$

Then we differentiate Eq. (C.42) with respect to $\boldsymbol{p}$ and obtain

$$\boldsymbol{G}\boldsymbol{H}_c\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{p}}\boldsymbol{U} = -(\boldsymbol{I} + \boldsymbol{A})\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{p}}\,, \tag{C.46}$$

and we can solve for $\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{p}} = -(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{G}\boldsymbol{H}_c\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{p}}\boldsymbol{U}$. Substituting this into Eq. (C.45) gives

$$\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{p}} = -2\,\mathrm{Re}\,\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{U}}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{G}\boldsymbol{H}_c\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{p}}\boldsymbol{U}. \tag{C.47}$$

Since $\boldsymbol{G}$, $\boldsymbol{H}_c$ are known $\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{p}}$ can be calculated analytically, then if we define the so-called "adjoint" solution [240, 23]

$$(\boldsymbol{I}+\boldsymbol{A})^T\boldsymbol{U}_{\mathrm{adj}} = \left(\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{U}}\right)^T, \tag{C.48}$$

we can obtain the gradient as

$$\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{p}} = -2\,\mathrm{Re}\,\boldsymbol{U}_{\mathrm{adj}}^T\boldsymbol{G}\boldsymbol{H}_c\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{p}}\boldsymbol{U}\,, \tag{C.49}$$

by solving for $\boldsymbol{U}_{\mathrm{adj}} = (\boldsymbol{I}+\boldsymbol{A}^T)^{-1}\left(\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{U}}\right)^T$ from Eq. (C.48) and $\boldsymbol{U} = (\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{U}_0$ from Eq. (C.42) only. This allows rapid computation of $\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{p}}$. Then standard optimization methods, such as gradient ascent, can be applied to maximize or minimize $f(\boldsymbol{U})$. All designed pulses in the main text are obtained by this approach described here.

C.7 Guide to computing prior-literature bounds

In this section, we describe how we computed the bounds from prior literature in Fig. 1 of the main text: Mandelstam–Tamm (MT), Margolus–Levitin (ML), and the bounds from Refs. [7, 8]. We start with the MT bound: the minimum time, which we denote τ_{MT}, for a given evolution is given by

$$\tau_{\mathrm{MT}} = \frac{\pi}{2\Delta H}, \tag{C.50}$$

where the standard deviation ΔH is given by $\Delta H = \sqrt{\langle H^2\rangle - \langle H\rangle^2}$. We do not know the mean of the Hamiltonian over the optimal trajectory, but can bound its square below by zero, which gives $\Delta H \leq \sqrt{\langle H^2\rangle}$. We don't know the value of $\langle H^2\rangle$ over the optimal trajectory either, but it can be bounded above by the square of the largest possible

eigenvalue of the Hamiltonian, which we denote $\lambda_{\max}$. This eigenvalue can be found by a grid-search-based optimization over all possible instantaneous Hamiltonians (assuming bounded controls), and we are left with the bound:

$$\tau_{\mathrm{MT}} \geq \frac{\pi}{2\lambda_{\max}}. \tag{C.51}$$

The ML bound, which we denote τ_{ML}, follows similar reasoning. Now given by $\tau_{\mathrm{ML}} = \pi/2\langle H\rangle$, we again do not know the mean energy, but can now bound it above by $\lambda_{\max}$, which implies that

$$\tau_{\mathrm{ML}} \geq \frac{\pi}{2\lambda_{\max}}. \tag{C.52}$$

By this reasoning, the MT and ML bounds coincides when presented with uncertainty about the optimal trajectory and bounded in this way.

Next, we consider the bound of Ref. [7]. In this case, one defines two functions incorporating information about the controllable and non-controllable Hamiltonians, H_c and H_0, respectively:

$$C(U_g, H_c) = \frac{\sqrt{2\left(d - \sum_{j=1}^{d}\left|\langle\phi_j^{(c)}|U_g|\phi_j^{(c)}\rangle\right|\right)}}{2}, \tag{C.53}$$

$$C(U_g, H_0) = \frac{\sqrt{2\left(d - \sum_{j=1}^{d}\left|\langle\phi_j^{(0)}|U_g|\phi_j^{(0)}\rangle\right|\right)}}{2}, \tag{C.54}$$

where U_g is a target evolution operator, d is the dimensionality, and $\phi_j^{(c)}$ and $\phi_j^{(0)}$ are the eigenfunctions of the controllable and non-controllable Hamiltonians, respectively. From these functions, the time bounds of Ref. [7] can be written:

$$T \geq \max\left\{\frac{2C(U_g, H_c)}{\|H_0\|}, \frac{2C(U_g, H_0)}{|f_{\max}|\|H_c\|}\right\}, \tag{C.55}$$

where $|f_{\max}|$ is the maximum amplitude of the bounded control parameter. Equation (C.55) is straightforward to compute, though hard to apply to the maximum transition probability in a multilevel system, as there is no specific target unitary evolution. In this case, to

compute a meaningful bound at all (for the sake of comparison), we instead consider the best numerically-optimized control sequence, and use its evolution operator evaluated at the final time as the target unitary. In this way, we take a known achievable propagator and can generate a meaningful bound to compare it to. It should be noted that if one were to want to derive a strict bound in this way, it would be required to find the smallest bound over all possible unitaries that correspond to unity probability in the desired state; hence, our computation of a bound is actually an over-estimate of the "true" bound using this approach, which is necessarily smaller (and thereby looser).

The bound of Ref. [8] is similarly for a target unitary, but now uses a different expression. Now, the minimum time is given by:

$$T \geq \max_{V \Phi_{\text{in}} \text{Stab}(iH_c)} \frac{\|[U_g, V]\|}{\|[H_0, V]\|}. \tag{C.56}$$

This bound arises from maximizing the ratio of norms on the right-hand side of Eq. (C.56) over all matrices that commute with H_c. Generically, there is no simple way to optimize over all such matrices. To compute a reasonable bound, then, we choose H_c for V, as it is guaranteed to commute with itself. Then we evaluate the ratio in Eq. (C.56), again using a single unitary as described above (instead of minimizing over all possible unitaries). In both Eq. (C.55) and Eq. (C.56), the Frobenius norm is the norm that is used.

Bibliography

[1] C. Sauvan, J.-P. Hugonin, I. Maksymov, and P. Lalanne. Theory of the spontaneous optical emission of nanosize photonic and plasmon resonators. *Physical Review Letters*, 110(23):237401, 2013.

[2] G. Smestad, H. Ries, R. Winston, and E. Yablonovitch. The thermodynamic limits of light concentrators. *Solar Energy Materials*, 21(2):99 – 111, 1990.

[3] B. Kress and M. Shin. Diffractive and holographic optics as optical combiners in head mounted displays. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 1479–1482, 2013.

[4] J. S. Jensen and O. Sigmund. Topology optimization for nano-photonics. *Laser & Photonics Reviews*, 5(2):308–321, 2011.

[5] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659–670, 2018.

[6] R. Pestourie, C. Pérez-Arancibia, Z. Lin, W. Shin, F. Capasso, and S. G. Johnson. Inverse design of large-area metasurfaces. *Optics Express*, 26(26):33732–33747, 2018.

[7] C. Arenz, B. Russell, D. Burgarth, and H. Rabitz. The roles of drift and control field constraints upon quantum control speed limits. *New J. Phys.*, 19:103015, 2017.

[8] J. Lee, C. Arenz, H. Rabitz, and B. Russell. Dependence of the quantum speed limit on system size and control complexity. *New J. Phys.*, 20:063002, 2018.

[9] J. W. Strutt. On the reflection of light from a regularly stratified medium. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 93(655):565–577, 1917.

[10] D. Polder and M. Van Hove. Theory of radiative heat transfer between closely spaced bodies. *Physical Review B*, 4(10):3303, 1971.

[11] A. W. Rodriguez, M. H. Reid, and S. G. Johnson. Fluctuating-surface-current formulation of radiative heat transfer for arbitrary geometries. *Physical Review B*, 86(22):220302, 2012.

[12] C. R. Otey, L. Zhu, S. Sandhu, and S. Fan. Fluctuational electrodynamics calculations of near-field heat transfer in non-planar geometries: A brief overview. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 132:3–11, 2014.

[13] S. Noda, M. Fujita, and T. Asano. Spontaneous-emission control by photonic crystals and nanocavities. *Nature photonics*, 1(8):449–458, 2007.

[14] K. J. Russell, T.-L. Liu, S. Cui, and E. L. Hu. Large spontaneous emission enhancement in plasmonic nanocavities. *Nature Photonics*, 6(7):459–462, 2012.

[15] G. M. Akselrod, C. Argyropoulos, T. B. Hoang, C. Ciracì, C. Fang, J. Huang, D. R. Smith, and M. H. Mikkelsen. Probing the mechanisms of large purcell enhancement in plasmonic nanoantennas. *Nature Photonics*, 8(11):835–840, 2014.

[16] M. Moskovits. Surface-enhanced spectroscopy. *Reviews of Modern Physics*, 57(3):783, 1985.

[17] S. Nie and S. R. Emory. Probing single molecules and single nanoparticles by surface-enhanced raman scattering. *Science*, 275(5303):1102–1106, 1997.

[18] V. V. Thacker, L. O. Herrmann, D. O. Sigle, T. Zhang, T. Liedl, J. J. Baumberg, and U. F. Keyser. Dna origami based assembly of gold nanoparticle dimers for surface-enhanced raman scattering. *Nature Communications*, 5(1):1–7, 2014.

[19] F. Aieta, M. A. Kats, P. Genevet, and F. Capasso. Multiwavelength achromatic metasurfaces by dispersive phase compensation. *Science*, 347(6228):1342–1345, 2015.

[20] Z. Lin, B. Groever, F. Capasso, A. W. Rodriguez, and M. Lončar. Topology-optimized multilayered metaoptics. *Physical Review Applied*, 9(4):044030, 2018.

[21] A. Arbabi, Y. Horie, M. Bagheri, and A. Faraon. Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission. *Nature nanotechnology*, 10(11):937, 2015.

[22] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch. Adjoint shape optimization applied to electromagnetic design. *Optics Express*, 21(18):21693–21701, September 2013.

[23] O. D. Miller. Photonic design: From fundamental solar cell physics to computational inverse design. *arXiv preprint arXiv:1308.0212*, 2013.

[24] H. Chung and O. D. Miller. High-na achromatic metalenses by inverse design. *Optics Express*, 28(5):6945–6965, 2020.

[25] H. Zhang and O. D. Miller. Quasinormal coupled mode theory. *arXiv preprint arXiv:2010.08650*, 2020.

[26] H. Zhang, C. W. Hsu, and O. D. Miller. Scattering concentration bounds: brightness theorems for waves. *Optica*, 6(10):1321–1327, 2019.

[27] H. Zhang, Z. Kuang, S. Puri, and O. D. Miller. Conservation-law-based global bounds to quantum optimal control. *Physical Review Letters*, 127(11):110506, 2021.

[28] D. A. Miller. Waves, modes, communications, and optics: a tutorial. *Advances in Optics and Photonics*, 11(3):679–825, 2019.

[29] C. Mahaux and H. A. Weidenmüller. *Shell-model approach to nuclear reactions.* 1969.

[30] C. W. Beenakker. Random-matrix theory of quantum transport. *Reviews of Modern Physics*, 69(3):731, 1997.

[31] C. Viviescas and G. Hackenbroich. Field quantization for open optical cavities. *Physical Review A*, 67(1):013805, 2003.

[32] D. Lentrodt and J. Evers. Ab initio few-mode theory for quantum potential scattering problems. *Physical Review X*, 10(1):011008, 2020.

[33] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade. *Photonic crystals: molding the flow of light*. Princeton University Press, 2011.

[34] S. Fan and J. D. Joannopoulos. Analysis of guided resonances in photonic crystal slabs. *Physical Review B*, 65(23):235112, 2002.

[35] S. Fan, W. Suh, and J. D. Joannopoulos. Temporal coupled-mode theory for the fano resonance in optical resonators. *JOSA A*, 20(3):569–572, 2003.

[36] M. F. Yanik, S. Fan, M. Soljačić, and J. D. Joannopoulos. All-optical transistor action with bistable switching in a photonic crystal cross-waveguide geometry. *Optics Letters*, 28(24):2506–2508, 2003.

[37] R. E. Hamam, A. Karalis, J. Joannopoulos, and M. Soljačić. Coupled-mode theory for general free-space resonant scattering of waves. *Physical Review A*, 75(5):053801, 2007.

[38] L. Verslegers, Z. Yu, P. B. Catrysse, and S. Fan. Temporal coupled-mode theory for resonant apertures. *JOSA B*, 27(10):1947–1956, 2010.

[39] L. Verslegers, Z. Yu, Z. Ruan, P. B. Catrysse, and S. Fan. From electromagnetically induced transparency to superscattering with a single structure: a coupled-mode theory for doubly resonant structures. *Physical Review Letters*, 108(8):083902, 2012.

[40] Z. Ruan and S. Fan. Superscattering of light from subwavelength nanostructures. *Physical Review Letters*, 105(1):013901, 2010.

[41] Z. Ruan and S. Fan. Temporal coupled-mode theory for light scattering by an arbitrarily shaped object supporting a single resonance. *Physical Review A*, 85(4):043828, 2012.

[42] Z. Yu, A. Raman, and S. Fan. Fundamental limit of nanophotonic light trapping in solar cells. *Proceedings of the National Academy of Sciences*, 107(41):17491–17496, 2010.

[43] C. W. Hsu, B. G. DeLacy, S. G. Johnson, J. D. Joannopoulos, and M. Soljacic. Theoretical criteria for scattering dark states in nanostructured particles. *Nano Letters*, 14(5):2783–2788, 2014.

[44] L. Huang, Y. Yu, and L. Cao. General modal properties of optical resonances in subwavelength nonspherical dielectric structures. *Nano Letters*, 13(8):3559–3565, 2013.

[45] W. Suh, Z. Wang, and S. Fan. Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities. *IEEE Journal of Quantum Electronics*, 40(10):1511–1518, 2004.

[46] C. Manolatou, M. Khan, S. Fan, P. R. Villeneuve, H. Haus, and J. Joannopoulos. Coupling of modes analysis of resonant channel add-drop filters. *IEEE Journal of Quantum Electronics*, 35(9):1322–1331, 1999.

[47] C. W. Hsu, B. Zhen, W. Qiu, O. Shapira, B. G. DeLacy, J. D. Joannopoulos, and M. Soljačić. Transparent displays enabled by resonant nanoparticle scattering. *Nature Communications*, 5(1):1–6, 2014.

[48] J. B. Khurgin. How to deal with the loss in plasmonics and metamaterials. *Nature Nanotechnology*, 10(1):2–6, 2015.

[49] S. Ishii, R. P. Sugavaneshwar, and T. Nagao. Titanium nitride nanoparticles as plasmonic solar heat transducers. *The Journal of Physical Chemistry C*, 120(4):2343–2348, 2016.

[50] P. Lalanne and P. Chavel. Metalenses at visible wavelengths: past, present, perspectives. *Laser & Photonics Reviews*, 11(3):1600295, 2017.

[51] M. Khorasaninejad and F. Capasso. Metalenses: Versatile multifunctional photonic components. *Science*, 358(6367), 2017.

[52] S. Banerji, M. Meem, A. Majumder, F. G. Vasquez, B. Sensale-Rodriguez, and R. Menon. Imaging with flat optics: metalenses or diffractive lenses? *Optica*, 6(6):805–810, 2019.

[53] C. W. Hsu, S. F. Liew, A. Goetschy, H. Cao, and A. D. Stone. Correlation-enhanced control of wave focusing in disordered media. *Nature Physics*, 13(5):497–502, 2017.

[54] W. R. Sweeney, C. W. Hsu, and A. D. Stone. Theory of reflectionless scattering modes. *arXiv preprint arXiv:1909.04017*, 2019.

[55] S. Rotter and S. Gigan. Light fields in complex media: Mesoscopic scattering meets wave control. *Reviews of Modern Physics*, 89(1):015005, 2017.

[56] P. Lalanne, W. Yan, K. Vynck, C. Sauvan, and J.-P. Hugonin. Light interaction with photonic and plasmonic resonances. *Laser & Photonics Reviews*, 12(5):1700113, 2018.

[57] M. Liertzer, L. Ge, A. Cerjan, A. Stone, H. E. Türeci, and S. Rotter. Pump-induced exceptional points in lasers. *Physical Review Letters*, 108(17):173901, 2012.

[58] H. Hodaei, A. U. Hassan, S. Wittek, H. Garcia-Gracia, R. El-Ganainy, D. N. Christodoulides, and M. Khajavikhan. Enhanced sensitivity at higher-order exceptional points. *Nature*, 548(7666):187–191, 2017.

[59] A. Pick, B. Zhen, O. D. Miller, C. W. Hsu, F. Hernandez, A. W. Rodriguez, M. Soljačić, and S. G. Johnson. General theory of spontaneous emission near exceptional points. *Optics Express*, 25(11):12325–12348, 2017.

[60] M.-A. Miri and A. Alu. Exceptional points in optics and photonics. *Science*, 363(6422), 2019.

[61] P. Leung, S. Liu, and K. Young. Completeness and orthogonality of quasinormal modes in leaky optical cavities. *Physical Review A*, 49(4):3057, 1994.

[62] B. J. Hoenders and M. Bertolotti. The (quasi) normal natural mode description of the scattering process by dispersive photonic crystals. In *Photonic Crystal Materials and Devices III (ie V)*, volume 6182, page 61821F. International Society for Optics and Photonics, 2006.

[63] E. Muljarov and W. Langbein. Resonant-state expansion of dispersive open optical systems: Creating gold from sand. *Physical Review B*, 93(7):075417, 2016.

[64] J.-M. Jin. *Theory and computation of electromagnetic fields*. John Wiley & Sons, 2011.

[65] W. Yan, R. Faggiani, and P. Lalanne. Rigorous modal analysis of plasmonic nanoresonators. *Physical Review B*, 97(20):205422, 2018.

[66] E. A. Muljarov and W. Langbein. Exact mode volume and purcell factor of open optical systems. *Physical Review B*, 94(23):235438, 2016.

[67] P. T. Kristensen, K. Herrmann, F. Intravaia, and K. Busch. Modeling electromagnetic resonators using quasinormal modes. *Advances in Optics and Photonics*, 12(3):612–708, 2020.

[68] P. T. Kristensen and S. Hughes. Modes and mode volumes of leaky optical cavities and plasmonic nanoresonators. *ACS Photonics*, 1(1):2–10, 2014.

[69] R.-C. Ge, P. T. Kristensen, J. F. Young, and S. Hughes. Quasinormal mode approach to modelling light-emission and propagation in nanoplasmonics. *New Journal of Physics*, 16(11):113048, 2014.

[70] A. Gras, W. Yan, and P. Lalanne. Quasinormal-mode analysis of grating spectra at fixed incidence angles. *Optics Letters*, 44(14):3494–3497, 2019.

[71] F. Alpeggiani, N. Parappurath, E. Verhagen, and L. Kuipers. Quasinormal-mode expansion of the scattering matrix. *Physical Review X*, 7(2):021035, 2017.

[72] T. Weiss and E. A. Muljarov. How to calculate the pole expansion of the optical scattering matrix from the resonant states. *Physical Review B*, 98(8):085433, 2018.

[73] J. A. Kong. *Theory of Electromagnetic Waves*, volume 1. Wiley-Interscience, New York, NY, 1975.

[74] G. Lecamp, J.-P. Hugonin, and P. Lalanne. Theoretical and computational concepts for periodic optical waveguides. *Optics Express*, 15(18):11042–11060, 2007.

[75] E. T. Whittaker and G. N. Watson. *A course of modern analysis*. Dover Publications, 2020.

[76] M. J. Ablowitz, A. S. Fokas, and A. S. Fokas. *Complex variables: introduction and applications*. Cambridge University Press, 2003.

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

[78] Y. Luo, A. Fernandez-Dominguez, A. Wiener, S. A. Maier, and J. Pendry. Surface plasmons and nonlocality: a simple model. *Physical Review Letters*, 111(9):093901, 2013.

[79] A. Raman, W. Shin, and S. Fan. Upper bound on the modal material loss rate in plasmonic and metamaterial systems. *Physical Review Letters*, 110(18):183901, 2013.

[80] G. Kristensson. *Scattering of Electromagnetic Waves by Obstacles*. The Institution of Engineering and Technology, 2016.

[81] H. A. Haus. *Waves and fields in optoelectronics*. Prentice-Hall,, 1984.

[82] N. Yu and F. Capasso. Flat optics with designer metasurfaces. *Nature Materials*, 13(2):139–150, 2014.

[83] H. Chung and O. D. Miller. Tunable metasurface inverse design for 80% switching efficiencies and 144 ° angular deflection. *ACS Photonics*, 7(8):2236–2243, 2020.

[84] M. De Zoysa, T. Asano, K. Mochizuki, A. Oskooi, T. Inoue, and S. Noda. Conversion of broadband to narrowband thermal emission through energy recycling. *Nature Photonics*, 6(8):535–539, 2012.

[85] A. Lenert, D. M. Bierman, Y. Nam, W. R. Chan, I. Celanović, M. Soljačić, and E. N. Wang. A nanophotonic solar thermophotovoltaic device. *Nature Nanotechnology*, 9(2):126–130, 2014.

[86] F. Presutti and F. Monticone. Focusing on bandwidth: achromatic metalens limits. *Optica*, 7(6):624–631, 2020.

[87] S. M. Barnett and R. Loudon. Sum rule for modified spontaneous emission rates. *Physical Review Letters*, 77(12):2444, 1996.

[88] H. Shim, L. Fan, S. G. Johnson, and O. D. Miller. Fundamental limits to near-field optical response over any bandwidth. *Physical Review X*, 9(1):011043, 2019.

[89] S. Molesky, P. Chao, W. Jin, and A. W. Rodriguez. Global t operator bounds on electromagnetic scattering: Upper bounds on far-field cross sections. *Physical Review Research*, 2(3):033172, 2020.

[90] Z. Kuang and O. D. Miller. Computational bounds to light–matter interactions via local conservation laws. *Phys. Rev. Lett.*, 125:263607, December 2020.

[91] M. Gustafsson, K. Schab, L. Jelinek, and M. Capek. Upper bounds on absorption and scattering. *New Journal of Physics*, 2020.

[92] R. F. Horton. Optical design for a high-etendue imaging fourier-transform spectrometer. In *Imaging Spectrometry II*, volume 2819, pages 300–315. SPIE, 1996.

[93] R. Winston, J. C. Miñano, P. G. Benitez, et al. *Nonimaging Optics*. Elsevier, 2005.

[94] L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, New York, NY, 1995.

[95] A. Walther. Radiometry and Coherence. *J. Opt. Soc. Am.*, 58(9):1256, 1968.

[96] A. T. Friberg. On the existence of a radiance function for finite planar sources of arbitrary states of coherence. *J. Opt. Soc. Am.*, 69(1):192–198, 1979.

[97] R. G. Littlejohn and R. Winston. Corrections to Classical Radiometry. *J. Opt. Soc. Am. A-Optics Image Sci. Vis.*, 10(9):2024–2037, 1993.

[98] R. G. Littlejohn and R. Winston. Generalized radiance and measurement. *J. Opt. Soc. Am. A*, 12(12):2736–2743, 1995.

[99] M. A. Alonso. Radiometry and wide-angle wave fields. I. Coherent fields in two dimensions. *J. Opt. Soc. Am. A.*, 18(4):902–909, 2001.

[100] M. A. Alonso. Radiometry and wide-angle wave fields III: partial coherence. *J. Opt. Soc. Am. A*, 18(10):2502–2511, 2001.

[101] M. E. Testorf, B. M. Hennelly, and J. Ojeda-Castañeda. *Phase-space Otics: Fundamentals and Applications*. McGraw-Hill, 2010.

[102] M. A. Alonso. Wigner functions in optics: describing beams as ray bundles and pulses as particle ensembles. *Adv. Opt. Photonics*, 3(4):272, 2011.

[103] L. Waller, G. Situ, and J. W. Fleischer. Phase-space measurement and coherence synthesis of optical beams. *Nat. Photonics*, 6(7):474–479, 2012.

[104] A. Yariv. Phase Conjugate Optics and Realtime Holography. *IEEE J. Quantum Electron.*, 14(9):650–660, 1978.

[105] A. Yariv. *Quantum Electronics*. John Wiley & Sons, 3rd edition, 1989.

[106] Y. Chong, L. Ge, H. Cao, and A. D. Stone. Coherent perfect absorbers: time-reversed lasers. *Physical review letters*, 105(5):053901, 2010.

[107] S. Withington and J. A. Murphy. Modal analysis of partially coherent submillimeter-wave quasi-optical systems. *IEEE Transactions on Antennas and Propagation*, 46(11):1651–1659, 1998.

[108] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, 2nd edition, 2013.

[109] L. D. Landau and E. Lifshitz. *Course of Theoretical Physics Vol 3 Quantum Mechanics*. Pergamon Press, 1958.

[110] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, 2nd edition, 2013.

[111] O. Cakmakci and J. Rolland. Head-worn Displays: A Review. *J. Disp. Technol.*, 2(3):199–216, 2006.

[112] T. Levola. Diffractive optics for virtual reality displays. *J. Soc. Inf. Disp.*, 14:467, 2006.

[113] J. S. Price, X. Sheng, B. M. Meulblok, J. A. Rogers, and N. C. Giebink. Wide-angle planar microtracking for quasi-static microcell concentrating photovoltaics. *Nat. Commun.*, 6:6223, 2015.

[114] M. A. Shameli and L. Yousefi. Absorption enhancement in thin-film solar cells using an integrated metasurface lens. *JOSA B*, 35(2):223–230, 2018.

[115] N. Yu and F. Capasso. Flat optics with designer metasurfaces. *Nat. Mater.*, 13(2):139–150, 2014.

[116] D. Lin, P. Fan, E. Hasman, and M. L. Brongersma. Dielectric gradient metasurface optical elements. *Science*, 345(6194):298–302, 2014.

[117] A. Jameson, L. Martinelli, and N. A. Pierce. Optimum Aerodynamic Design Using the Navier-Stokes Equations. *Theor. Comput. Fluid Dyn.*, 10(1-4):213–237, 1998.

[118] O. Sigmund and J. Søndergaard Jensen. Systematic design of phononic band–gap materials and structures by topology optimization. *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Eng. Sci.*, 361(1806):1001, 2003.

[119] J. Lu, S. Boyd, and J. Vucković. Inverse design of a three-dimensional nanophotonic resonator. *Opt. Express*, 19(11):10563–10570, 2011.

[120] J. S. Jensen and O. Sigmund. Topology optimization for nano-photonics. *Laser & Photonics Rev.*, 5(2):308–321, mar 2011.

[121] O. D. Miller. *Photonic Design: From Fundamental Solar Cell Physics to Computational Inverse Design*. PhD thesis, University of California, Berkeley, 2012.

[122] V. Ganapati, O. D. Miller, and E. Yablonovitch. Light trapping textures designed by electromagnetic optimization for subwavelength thick solar cells. *IEEE Journal of Photovoltaics*, 4(1):175–182, January 2014.

[123] N. Aage, E. Andreassen, B. S. Lazarov, and O. Sigmund. Giga-voxel computational morphogenesis for structural design. *Nature*, 550(7674):84, 2017.

[124] M. Moharam and T. Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *JOSA*, 71(7):811–818, 1981.

[125] V. Liu and S. Fan. S4 : A free electromagnetic solver for layered periodic structures. *Computer Physics Communications*, 183(10):2233 – 2244, 2012.

[126] J. Chaves. *Introduction to Nonimaging Optics*. CRC Press, Madrid, Spain, 2016.

[127] Y. Ding, J. Xu, F. Da Ros, B. Huang, H. Ou, and C. Peucheret. On-chip two-mode division multiplexing using tapered directional coupler-based mode multiplexer and demultiplexer. *Optics Express*, 21(8):10376–10382, 2013.

[128] R. Ji, L. Yang, L. Zhang, Y. Tian, J. Ding, H. Chen, Y. Lu, P. Zhou, and W. Zhu. Five-port optical router for photonic networks-on-chip. *Optics Express*, 19(21):20258–20268, 2011.

[129] A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vucković. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nat. Photonics*, 9(6):374–377, 2015.

[130] B. Shen, P. Wang, R. Polson, and R. Menon. An integrated-nanophotonics polarization beamsplitter with 2.4 x 2.4 μm^2 footprint. *Nat. Photonics*, 9(6):378–382, 2015.

[131] S. Fan, P. R. Villeneuve, J. D. Joannopoulos, and H. Haus. Channel drop tunneling through localized states. *Physical Review Letters*, 80(5):960, 1998.

[132] S. A. Lerner and B. Dahlgrenn. Etendue and Optical System Design. *Proc. SPIE*, 6338:633801, 2006.

[133] F. Fournier and J. Rolland. Design methodology for high brightness projectors. *Journal of Display Technology*, 4(1):86–91, 2008.

[134] K. Fang, Z. Yu, and S. Fan. Realizing effective magnetic field for photons by controlling the phase of dynamic modulation. *Nat. Photonics*, 6(11):782–787, 2012.

[135] L. D. Tzuang, K. Fang, P. Nussenzveig, S. Fan, and M. Lipson. Non-reciprocal phase shift induced by an effective magnetic flux for light. *Nat. Photonics*, 8(9):701–705, 2014.

[136] D. L. Sounas and A. Alù. Non-reciprocal photonics based on time modulation. *Nature Photonics*, 11(12):774, 2017.

[137] S. A. Cummer, J. Christensen, and A. Alù. Controlling sound with acoustic metamaterials. *Nature Reviews Materials*, 1(3):16001, 2016.

[138] A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink. Controlling waves in space and time for imaging and focusing in complex media. *Nature Photonics*, 6(5):283, 2012.

[139] C. W. Hsu, A. Goetschy, Y. Bromberg, A. D. Stone, and H. Cao. Broadband coherent enhancement of transmission and absorption in disordered media. *Physical Review Letters*, 115(22):223901, 2015.

[140] S. Molesky, P. Chao, and A. W. Rodriguez. Hierarchical mean-field T operator bounds on electromagnetic scattering: Upper nounds on near-field radiative Purcell enhancement. *Phys. Rev. Res.*, 2(4):043398, 2020.

[141] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

[142] Z. Kuang, L. Zhang, and O. D. Miller. Maximal single-frequency electromagnetic response. *Optica*, 7(12):1746–1757, November 2020.

[143] J. Park and S. Boyd. A semidefinite programming method for integer convex quadratic minimization. *Optimization Letters*, 12(3):499–518, 2018.

[144] L. Vandenberghe and S. Boyd. Semidefinite Programming. *SIAM Rev.*, 38(1):49–95, mar 1996.

[145] Z. Q. Luo, W. K. Ma, A. So, Y. Ye, and S. Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Process. Mag.*, 27(3):20–34, 2010.

[146] M. Laurent and F. Rendl. Semidefinite Programming and Integer Programming. *Handbooks Oper. Res. Manag. Sci.*, 12(C):393–514, 2005.

[147] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm. Simple pulses for elimination of leakage in weakly nonlinear qubits. *Physical Review Letters*, 103(11):110501, 2009.

[148] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser. Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms. *J. Magn. Reson.*, 172(2):296–305, 2005.

[149] N. C. Nielsen, C. Kehlet, S. J. Glaser, and N. Khaneja. Optimal Control Methods in NMR Spectroscopy. *Encycl. Magn. Reson.*, mar 2010.

[150] Z. Tošner, R. Sarkar, J. Becker-Baldus, C. Glaubitz, S. Wegner, F. Engelke, S. J. Glaser, and B. Reif. Overcoming Volume Selectivity of Dipolar Recoupling in Biological Solid-State NMR Spectroscopy. *Angew. Chemie Int. Ed.*, 57(44):14514–14518, 2018.

[151] M. A. Nielsen, M. R. Dowling, M. Gu, and A. C. Doherty. Optimal control, geometry, and quantum computing. *Phys. Rev. A*, 73(6):1–7, 2006.

[152] R. A. Jefferson and R. C. Myers. Circuit complexity in quantum field theory. *J. High Energy Physics*, 2017:107, oct 2017.

[153] S. Chapman, M. P. Heller, H. Marrochio, and F. Pastawski. Toward a Definition of Complexity for Quantum Field Theory States. *Phys. Rev. Lett.*, 120(12):121602, 2018.

[154] A. P. Peirce, M. A. Dahleh, and H. Rabitz. Optimal control of quantum-mechanical systems: Existence, numerical approximation, and applications. *Phys. Rev. A*, 37(12):4950, 1988.

[155] J. Werschnik and E. K. Gross. Quantum optimal control theory. *J. Phys. B At. Mol. Opt. Phys.*, 40(18), 2007.

[156] D. D'Alessandro. *Introduction to quantum control and dynamics*. CRC press, Boca Raton, FL, 2007.

[157] C. Brif, R. Chakrabarti, and H. Rabitz. Control of quantum phenomena: Past, present and future. *New J. Phys.*, 12, 2010.

[158] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm. Training Schrödinger's cat: Quantum optimal control: Strategic report on current status, visions and goals for research in Europe. *Eur. Phys. J. D*, 69(12), 2015.

[159] M. G. Bason, M. Viteau, N. Malossi, P. Huillery, E. Arimondo, D. Ciampini, R. Fazio, V. Giovannetti, R. Mannella, and O. Morsch. High-fidelity quantum driving. *Nat. Phys.*, 8(2):147–152, 2012.

[160] S. Van Frank, M. Bonneau, J. Schmiedmayer, S. Hild, C. Gross, M. Cheneau, I. Bloch, T. Pichler, A. Negretti, T. Calarco, and S. Montangero. Optimal control of complex atomic quantum systems. *Sci. Rep.*, 6:34187, 2016.

[161] R. W. Heeres, P. Reinhold, N. Ofek, L. Frunzio, L. Jiang, M. H. Devoret, and R. J. Schoelkopf. Implementing a universal gate set on a logical qubit encoded in an oscillator. *Nat. Commun.*, 8(1):1–7, 2017.

[162] R. E. Goetz, C. P. Koch, and L. Greenman. Quantum Control of Photoelectron Circular Dichroism. *Phys. Rev. Lett.*, 122(1):13204, 2019.

[163] A. Vepsäläinen, S. Danilin, and G. S. Paraoanu. Superadiabatic population transfer in a three-level superconducting circuit. *Sci. Adv.*, 5:eaau5999, 2019.

[164] M. R. Lam, N. Peter, T. Groh, W. Alt, C. Robens, D. Meschede, A. Negretti, S. Montangero, T. Calarco, and A. Alberti. Demonstration of Quantum Brachistochrones between Distant States of an Atom. *Phys. Rev. X*, 11(1):11035, 2021.

[165] P. De Fouquieres, S. G. Schirmer, S. J. Glaser, and I. Kuprov. Second order gradient ascent pulse engineering. *J. Magn. Reson.*, 212(2):412–417, oct 2011.

[166] F. Dolde, V. Bergholm, Y. Wang, I. Jakobi, B. Naydenov, S. Pezzagna, J. Meijer, F. Jelezko, P. Neumann, T. Schulte-Herbrüggen, J. Biamonte, and J. Wrachtrup. High-fidelity spin entanglement using optimal control. *Nat. Commun.*, 5(1):1–9, feb 2014.

[167] B. E. Anderson, H. Sosa-Martinez, C. A. Riofrío, I. H. Deutsch, and P. S. Jessen. Accurate and Robust Unitary Transformations of a High-Dimensional Quantum System. *Phys. Rev. Lett.*, 114(24):240401, jun 2015.

[168] V. F. Krotov. Global Methods in Optimal Control Theory. In *Adv. Nonlinear Dyn. Control A Rep. from Russ.*, pages 74–121. Birkhäuser Boston, 1993.

[169] J. Somlói, V. A. Kazakov, and D. J. Tannor. Controlled dissociation of I2 via optical transitions between the X and B electronic states. *Chem. Phys.*, 172(1):85–98, may 1993.

[170] W. Zhu and H. Rabitz. A rapid monotonically convergent iteration algorithm for quantum optimal control over the expectation value of a positive definite operator. *J. Chem. Phys.*, 109(2):385–391, jul 1998.

[171] Y. Ohtsuki, G. Turinici, and H. Rabitz. Generalized monotonically convergent algorithms for solving quantum optimal control problems. *J. Chem. Phys.*, 120(12):5509–5517, mar 2004.

[172] S. G. Schirmer and P. De Fouquieres. Efficient algorithms for optimal control of quantum dynamics: The Krotov method unencumbered. *New J. Phys.*, 13(35pp):73029, jul 2011.

[173] M. H. Goerz, D. Basilewitsch, F. Gago-Encinas, M. G. Krauss, K. P. Horn, D. M. Reich, and C. P. Koch. Krotov: A Python implementation of Krotov's method for quantum optimal control. *SciPost Phys.*, 7(6):080, 2019.

[174] T. Caneva, T. Calarco, and S. Montangero. Chopped random-basis quantum optimization. *Phys. Rev. A*, 84:022326, 2011.

[175] P. Doria, T. Calarco, and S. Montangero. Optimal control technique for many-body quantum dynamics. *Phys. Rev. Lett.*, 106(19):1–4, 2011.

[176] L. Mandelstam and I. Tamm. The Uncertainty Relation Between Energy and Time in Non-relativistic Quantum Mechanics. *J. Phys. USSR*, 9:249–254, 1945.

[177] G. N. Fleming. A unitarity bound on the evolution of nonstationary states. *Nuovo Cim. A*, 16(2):232–240, 1973.

[178] L. Vaidman. Minimum time for the evolution to an orthogonal quantum state. *Am. J. Phys.*, 60(2):182–183, feb 1992.

[179] N. Margolus and L. B. Levitin. The maximum speed of dynamical evolution. *Phys. D Nonlinear Phenom.*, 120(1-2):188–195, 1998.

[180] N. Khaneja, R. Brockett, and S. J. Glaser. Time optimal control in spin systems. *Phys. Rev. A - At. Mol. Opt. Phys.*, 63(3):032308, 2001.

[181] N. Khaneja, S. J. Glaser, and R. Brockett. Sub-Riemannian geometry and time optimal control of three spin systems: Quantum gates and coherence transfer. *Phys. Rev. A*, 65(3):032301, jan 2002.

[182] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum limits to dynamical evolution. *Phys. Rev. A - At. Mol. Opt. Phys.*, 67(5):8, 2003.

[183] V. Giovannetti, S. Lloyd, and L. Maccone. The role of entanglement in dynamical evolution. *Europhys. Lett.*, 62(5):615–621, jun 2003.

[184] A. Carlini, A. Hosoya, T. Koike, and Y. Okudaira. Time-optimal quantum evolution. *Phys. Rev. Lett.*, 96(6):1–4, 2006.

[185] A. Carlini, A. Hosoya, T. Koike, and Y. Okudaira. Time-optimal unitary operations. *Phys. Rev. A - At. Mol. Opt. Phys.*, 75(4):1–8, 2007.

[186] S. Deffner and E. Lutz. Quantum speed limit for non-Markovian dynamics. *Phys. Rev. Lett.*, 111(1):1–5, 2013.

[187] A. Del Campo, I. L. Egusquiza, M. B. Plenio, and S. F. Huelga. Quantum speed limits in open system dynamics. *Phys. Rev. Lett.*, 110(5), 2013.

[188] P. M. Poggi, F. C. Lombardo, and D. A. Wisniacki. Quantum speed limit and optimal evolution time in a two-level system. *EPL (Europhysics Lett.*, 104(4):40005, 2013.

[189] G. C. Hegerfeldt. Driving at the quantum speed limit: Optimal control of a two-level system. *Phys. Rev. Lett.*, 111(26):1–5, 2013.

[190] G. C. Hegerfeldt. High-speed driving of a two-level system. *Phys. Rev. A - At. Mol. Opt. Phys.*, 90(3):1–10, 2014.

[191] B. Russell and S. Stepney. Zermelo navigation and a speed limit to quantum information processing. *Phys. Rev. A - At. Mol. Opt. Phys.*, 90(1):1–6, 2014.

[192] D. C. Brody and D. M. Meier. Solution to the quantum Zermelo navigation problem. *Phys. Rev. Lett.*, 114(10):1–5, 2015.

[193] B. Russell and S. Stepney. Zermelo navigation in the quantum brachistochrone. *J. Phys. A Math. Theor.*, 48(11):115303, 2015.

[194] M. R. Frey. Quantum speed limits—primer, perspectives, and potential future directions. *Quantum Inf. Process.*, 15(10):3919–3950, 2016.

[195] S. Deffner and S. Campbell. Quantum speed limits: From Heisenberg's uncertainty principle to optimal quantum control. *J. Phys. A Math. Theor.*, 50:453001, 2017.

[196] D. Burgarth, J. Borggaard, and Z. Zimborás. Quantum distance to uncontrollability and quantum speed limits. *arXiv preprint arXiv:2010.16156*, 2020.

[197] G. Knowles. *An Introduction to Applied Optimal Control*. Academic Press, Inc., New York, NY, 1981.

[198] B.-G. Englert. *Lectures On Quantum Mechanics-Volume 3: Perturbed Evolution*. World Scientific Publishing Company, 2006.

[199] J. J. Sakurai. *Modern Quantum Mechanics*. Addison-Wesley Publishing Co., Reading, MA, 1994.

[200] J. D. Jackson. *Classical Electrodynamics, 3rd Ed.* John Wiley & Sons, 1999.

[201] G. Angeris, J. Vučković, and S. Boyd. Heuristic methods and performance bounds for photonic design. *Opt. Express*, 29(2):2827–2854, jan 2021.

[202] J. Thijssen. *Computational Physics*. Cambridge University Press, 2012.

[203] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver. A quantum engineer's guide to superconducting qubits. *Applied Physics Reviews*, 6(2):021318, 2019.

[204] C. J. Wood and J. M. Gambetta. Quantification and characterization of leakage errors. *Phys. Rev. A*, 97(3):032306, 2018.

[205] L. Viola, E. Knill, and S. Lloyd. Dynamical decoupling of open quantum systems. *Physical Review Letters*, 82(12):2417, 1999.

[206] G. S. Uhrig. Concatenated control sequences based on optimized dynamic decoupling. *Physical Review Letters*, 102(12):120502, 2009.

[207] J. R. West, D. A. Lidar, B. H. Fong, and M. F. Gyure. High fidelity quantum gates via dynamical decoupling. *Physical Review Letters*, 105(23):230503, 2010.

[208] A. M. Souza, G. A. Alvarez, and D. Suter. Robust dynamical decoupling for quantum computing and quantum memory. *Physical Review Letters*, 106(24):240501, 2011.

[209] D. Rossini, P. Facchi, R. Fazio, G. Florio, D. A. Lidar, S. Pascazio, F. Plastina, and P. Zanardi. Bang-bang control of a qubit coupled to a quantum critical spin bath. *Physical Review A*, 77(5):052112, 2008.

[210] R. Kosloff. Quantum thermodynamics and open-systems modeling. *The Journal of Chemical Physics*, 150(20):204105, 2019.

[211] K. Khodjasteh and D. A. Lidar. Fault-tolerant quantum dynamical decoupling. *Physical Review Letters*, 95(18):180501, 2005.

[212] L. Viola and E. Knill. Robust dynamical decoupling of quantum systems with bounded controls. *Physical Review Letters*, 90(3):037901, 2003.

[213] O. D. Miller, A. G. Polimeridis, M. T. H. Reid, C. W. Hsu, B. G. Delacy, J. D. Joannopoulos, M. Soljačič, and S. G. Johnson. Fundamental limits to optical response in absorptive systems. *Optics Express*, 24(4):3329, Sep 2016.

[214] Y. Yang, O. D. Miller, T. Christensen, J. D. Joannopoulos, and M. Soljačič. Low-loss plasmonic dielectric nanoresonators. *Nano Letters*, 17(5):3238–3245, 2017.

[215] H. Shim, L. Fan, S. G. Johnson, and O. D. Miller. Fundamental Limits to Near-Field Optical Response over Any Bandwidth. *Phys. Rev. X*, 9:11043, mar 2019.

[216] S. Molesky, W. Jin, P. S. Venkataram, and A. W. Rodriguez. T Operator Bounds on Angle-Integrated Absorption and Thermal Radiation for Arbitrary Objects. *Phys. Rev. Lett.*, 123(25):257401, 2019.

[217] Y. Ivanenko, M. Gustafsson, and S. Nordebo. Optical theorems and physical bounds on absorption in lossy media. *Opt. Express*, 27(23):34323–34342, 2019.

[218] H. Shim, H. Chung, and O. D. Miller. Maximal free-space concentration of electromagnetic waves. *Physical Review Applied*, 14:014007, July 2020.

[219] H. Shim, Z. Kuang, and O. D. Miller. Optical materials for maximal nanophotonic response (Invited). *Optical Materials Express*, 10:1561–1585, July 2020.

[220] S. Molesky, P. Chao, W. Jin, and A. W. Rodriguez. Global T operator bounds on electromagnetic scattering: Upper bounds on far-field cross sections. *Phys. Rev. Res.*, 2(3):033172, 2020.

[221] M. Gustafsson, K. Schab, L. Jelinek, and M. Capek. Upper bounds on absorption and scattering. *New J. Phys.*, 22(7):073013, sep 2020.

[222] S. Molesky, P. S. Venkataram, W. Jin, and A. W. Rodriguez. Fundamental limits to radiative heat transfer: Theory. *Phys. Rev. B*, 101(3):35408, 2020.

[223] W.-K. Ma, C.-C. Su, J. Jaldén, and C.-Y. Chi. Some results on 16-qam mimo detection using semidefinite relaxation. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2673–2676. IEEE, 2008.

[224] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73(2):325–348, 1987.

[225] R. Coifman, V. Rokhlin, and S. Wandzura. The fast multipole method for the wave equation: A pedestrian prescription. *IEEE Antennas and Propagation magazine*, 35(3):7–12, 1993.

[226] P. Lalanne, W. Yan, A. Gras, C. Sauvan, J.-P. Hugonin, M. Besbes, G. Demésy, M. Truong, B. Gralak, F. Zolla, et al. Quasinormal mode solvers for resonators with dispersive materials. *JOSA A*, 36(4):686–704, 2019.

[227] P. Yeh and M. Hendry. Optical waves in layered media. *PhT*, 43(1):77, 1990.

[228] P. Waterman. Matrix formulation of electromagnetic scattering. *Proceedings of the IEEE*, 53(8):805–812, 1965.

[229] W. Chew, J. Jin, and E. Michielssen. Complex coordinate stretching as a generalized absorbing boundary condition. *Microwave and Optical Technology Letters*, 15(6):363–369, 1997.

[230] W. J. Lentz. Generating bessel functions in mie scattering calculations using continued fractions. *Applied Optics*, 15(3):668–671, 1976.

[231] E. A. Muljarov, W. Langbein, and R. Zimmermann. Brillouin-wigner perturbation theory in open electromagnetic systems. *EPL (Europhysics Letters)*, 92(5):50010, 2011.

[232] P. T. Kristensen, R.-C. Ge, and S. Hughes. Normalization of quasinormal modes in leaky optical cavities and plasmonic resonators. *Physical Review A*, 92(5):053810, 2015.

[233] J. Schwinger, L. L. DeRaad Jr, K. Milton, and W.-y. Tsai. *Classical electrodynamics*. Westview Press, 1998.

[234] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, 2nd edition, 2013.

[235] G. W. Milton. *The Theory of Composites*. Cambridge University Press, 2002.

[236] H.-P. Breuer, F. Petruccione, et al. *The theory of open quantum systems*. Oxford University Press on Demand, 2002.

[237] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[238] K. E. Atkinson. The numerical solution of fredholm integral equations of the second kind. *SIAM Journal on Numerical Analysis*, 4(3):337–348, 1967.

[239] K. Kreutz-Delgado. The complex gradient operator and the cr-calculus. *arXiv preprint arXiv:0906.4835*, 2009.

[240] G. Strang. *Computational science and engineering*. Wellesley-Cambridge Press, 2007.