Chapter 2
Fundamental Limits to Near-Field Optical Response

Owen D. Miller

Abstract Near-field optics is an exciting frontier of photonics and plasmonics. The tandem of strongly localized fields and enhanced emission rates offers significant opportunities for wide-ranging applications while also creating the following basic questions: How large can such enhancements be? To what extent do material losses inhibit optimal response? Over what bandwidths can these effects be sustained? This chapter surveys theoretical techniques for answering these questions. We start with physical intuition and mathematical definitions of the response functions of interest (LDOS, CDOS, SERS, NFRHT, etc.), after which we describe the general theoretical techniques for bounding such functions. Finally, we apply those techniques specifically to near-field optics, for which we describe known bounds, optimal designs, and open questions.

2.1 Introduction

Near-field optics is an exciting frontier of photonics and plasmonics. The near field is the region of space within much less than one electromagnetic wavelength of a source, and "near-field optics" refers to the phenomena that arise when optical-frequency sources interact with material structures in their near field. Free-space waves exhibit negligible variations over such small length scales, which might lead one to think this regime simply reduces to classical electrostatics and circuit theory. A new twist in the optical near field is the emergence of **polaritons**, modes that arise near the interfaces between negative- and positive-permittivity materials [1]. Polaritons emerge from an interplay of geometry and material susceptibility, instead

O. D. Miller (✉)
Department of Applied Physics, Yale University, New Haven, CT, USA
e-mail: owen.miller@yale.edu

of geometry and wave interference, to confine optical waves. Freedom from wave-interference requirements leads to a striking possibility: resonant fields whose size (spatial confinement) is *decoupled* from its wavelength. Highly confined polaritons enable two reciprocal effects: incoming free-space waves can be concentrated to spatial regions much smaller than the electromagnetic wavelength (well below the diffraction limit) and, conversely, that patterned materials close to a dipolar emitter can significantly amplify outgoing radiation.

The tandem of strongly localized fields and enhanced emission rates offers significant opportunities for applications including spectroscopy [2, 3], nanolasers [4], coherent plasmon generation [5], and broadband single-photon sources [6]. It also generates the following fundamental questions: How large can such enhancements be? Are there limits to field localization? All known polaritonic materials have significant or at least nontrivial amounts of material loss; to what extent does the loss affect these quantities? Over what bandwidths can these effects be sustained?

This chapter surveys theoretical techniques for answering these questions. The same features that make the near field appealing also make it theoretically challenging: there are not fixed photon flows, modal descriptions require exquisite care, and analytical descriptions are not possible except in the simplest high-symmetry scenarios. Over the past decade, thankfully, there has been a surge of interest in identifying what is possible in these systems. One key to the success of these approaches is to not attempt to develop models that apply to every possible instance of a given scattering scenario, but instead to develop techniques that identify *bounds* to the *extreme possibilities* of each scattering scenario. In this chapter, we describe these techniques in detail. We start with physical intuition and mathematical definitions of the response functions of interest (Sect. 2.2), after which we describe the general theoretical techniques for bounding such functions (Sect. 2.3). Finally, we apply those techniques specifically to near-field optics, for which we describe known bounds, optimal designs, and open questions (Sect. 2.4).

2.2 Near-Field Optical Response Functions

In this section, we summarize the background intuition and mathematical equations describing six key near-field optical response functions: local density of states (Sect. 2.2.1), which is proportional to the radiation of a single dipolar current; free-electron radiation (Sect. 2.2.2), which is the collective radiation of a line of current created by an electron beam; the cross density of states (Sect. 2.2.3), which measures modal or emission correlations across different spatial locations; surface-enhanced Raman scattering (Sect. 2.2.4), which is the simultaneous enhancement of incident radiation and outgoing luminescence, typically for imaging or sensing applications; near-field radiative heat transfer (Sect. 2.2.5), which is the transfer of radiative energy from a hot body to a cold one, at near-field separations; and mode volume (Sect. 2.2.6), which refers to the spatial confinement of a resonant mode. Many of these response functions are depicted in Fig. 2.1.
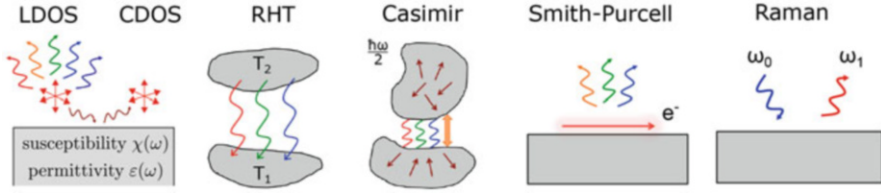
Fig. 2.1 An array of near-field optical response functions of broad interest (Adapted from Ref. [7])

2.2.1 LDOS

The first and arguably most important near-field response quantity is the *local density of states* (LDOS). The central role of LDOS is a result of the extent to which it underpins many connected ideas in near-field optics [8].

The first connection is to the power radiated by a dipole. In general, the work per time done by a field $\mathbf{E}$ on a current $\mathbf{J}$ in a volume V is given by $(1/2) \operatorname{Re} \int_V \mathbf{J}^* \cdot \mathbf{E}$. This is a generalized version of Watt's law in circuit theory, and it encodes the work done by the electric field mediating the electric force on the charges in the current, across a distance traveled by the charges given by the product of their speed and the time interval of interest. By Newton's second law, the work per time done by a current $\mathbf{J}$ on a field $\mathbf{E}$ is the negative of the expression above, $-(1/2) \operatorname{Re} \int_V \mathbf{J}^* \cdot \mathbf{E}$. We can convert the current density $\mathbf{J}$ to a dipole density $\mathbf{P}$ by the relation $\mathbf{J} = \partial \mathbf{P} / \partial t = -i\omega \mathbf{P}$ for harmonic frequency ω ($e^{-i\omega t}$ convention). Then the power radiated by a dipole at $\mathbf{x}_0$ with dipole moment $\mathbf{p}$ (and therefore dipole density $\mathbf{P} = \mathbf{p}\delta(\mathbf{x} - \mathbf{x}_0)$) is

$$
\begin{aligned}
P_{\text{rad}} &= -\frac{1}{2} \operatorname{Re} \int_V \mathbf{J}^* \cdot \mathbf{E} \, d\mathbf{x} \\
&= \frac{\omega}{2} \operatorname{Im} \int_V \mathbf{P}^* \cdot \mathbf{E} \, d\mathbf{x} \\
&= \frac{\omega}{2} \operatorname{Im} \left[\mathbf{p}^* \cdot \mathbf{E}(\mathbf{x}_0) \right].
\end{aligned}
$$

The electric field at $\mathbf{x}_0$, $\mathbf{E}(\mathbf{x}_0)$, is the field produced by a delta-function dipole source, which exactly coincides with the dyadic Green's function (GF) $\mathbb{G}$, evaluated at $\mathbf{x}_0$ from a source at $\mathbf{x}_0$, multiplied by the dipole moment $\mathbf{p}$, giving:

$$
P_{\text{rad}} = \frac{\omega}{2} \operatorname{Im} \left[\mathbf{p}^* \cdot \mathbb{G}(\mathbf{x}_0, \mathbf{x}_0) \mathbf{p} \right].
$$

The imaginary part of a complex number of the form $z^\dagger A z$ is $\operatorname{Im}(z^\dagger A z) = z^\dagger (\operatorname{Im} A) z$ by symmetry, where $\operatorname{Im} A$ refers to the anti-Hermitian part of A ($\operatorname{Im} A = (A - A^\dagger)/2i$). So we have

$$P_{\text{rad}} = \frac{\omega}{2} \mathbf{p}^\dagger \left[\text{Im} \, \mathbb{G} \left(\mathbf{x}_0, \mathbf{x}_0 \right) \right] \mathbf{p}. \tag{2.1}$$

This result gives us the first key near-field response function, the imaginary part of the Green's function evaluated at the source position,

$$\text{Im} \, \mathbb{G}(\mathbf{x}_0, \mathbf{x}_0), \tag{2.2}$$

which is proportional to the radiation rate of an electric dipole into any environment.

Spontaneous emission typically occurs via electric-dipole transitions in atomic or molecular systems, so the rate of spontaneous emission is governed by the imaginary part of the GF. It has been recognized for many decades that this rate is not an immutable constant, but a function of the environment. Just as specifying the amplitude of a current or voltage source in a circuit does not dictate the power delivered by the source, which depends on the impedance of the load, specifying the amplitude of a dipole moment does not dictate the power it delivers to its electromagnetic environment. This fact inspired the concept of a photonic bandgap [9] and photonic crystals [10, 11], with the goal of *inhibiting* spontaneous emission, originally to avoid laser power loss. It has conversely inspired significant effort toward *amplifying* spontaneous emission, for applications such as single-molecule imaging [2, 3]. An early recognition of this fact came from Purcell, who noted that an emitter radiating into a single-photonic-mode environment would have an altered spontaneous-emission rate [12]. Purcell recognized that for a single-mode resonator with quality factor Q and mode volume V, the density of states (per unit volume and per unit frequency) becomes $(Q/\omega)/V$. The relative change of the spontaneous-emission rate is the **Purcell factor**, which is proportional to $\lambda^3 Q/V$.

Purcell derived this expression in the context of enhancing magnetic-dipole transitions in spin systems, but exactly the same argument applies to electric-dipole transitions, where it is most used today. This expression drives many modern investigations of high-quality-factor and/or small-mode-volume cavity design [13–19], to reach the largest Purcell enhancement possible. It can be generalized to multimode, high-Q systems: if each mode has mode field $\mathbf{E}_i$, center frequency ω_i, and half-width-at-half-maximum γ_i (corresponding to a *mode lifetime* of $1/(2\gamma_i)$), the power radiated by a dipole with moment $\mathbf{p}$ located at position $\mathbf{x}_0$ is [20]

$$P_{\text{rad}} \approx \frac{\omega^2}{4} \sum_i \frac{\gamma_i |\mathbf{E}_i^\dagger(\mathbf{x}_0)\mathbf{p}|^2}{(\omega - \omega_i)^2 + \gamma_i^2} \tag{2.3}$$

In the limit of infinite Q, the Lorentzian lineshapes become delta functions, and the summation simplifies to delta functions multiplied by the overlap of modal fields with the dipole moment. The overlap of each mode with the dipole is a measure of the relative modal energy concentration at that particular point in space. Hence, the overall summation can be understood as a **local density of states**, or LDOS (with appropriate prefactors). The power radiated by a dipole into an electromagnetic environment, then, is directly proportional to the local density of electromagnetic

modes; inserting the correct prefactors leads to an LDOS expression in terms of $\operatorname{Im} \mathbb{G}$ [8, 21–23]:

$$\mathrm{LDOS}(\omega, \mathbf{x}) = \frac{1}{\pi \omega} \operatorname{Tr} \operatorname{Im} \mathbb{G}(\mathbf{x}_0, \mathbf{x}_0), \qquad (2.4)$$

where the trace encodes a summation over all independent polarizations. (Note that, e.g., Ref. [8] defines the Green's function with an extra $1/\omega^2$ factor, which leads to ω in the numerator of their analog to Eq. (2.4).) In free space, the LDOS coincides with the density of states (as there are no spatial variations) and is given by $\mathrm{LDOS}(\omega) = \omega^2/2\pi^2 c^3$. Technically, the expression of Eq. (2.4) is the *electric* LDOS; one can similarly define a magnetic LDOS through a summation over the relative magnetic-field strengths, or more generally by the power radiated by a magnetic dipole. For a magnetic Green's function $\mathbb{G}^{(HM)}$, denoting the magnetic field from a magnetic-dipole source, the magnetic LDOS is [8]

$$\mathrm{LDOS}^{(m)}(\omega, \mathbf{x}) = \frac{1}{\pi \omega} \operatorname{Tr} \operatorname{Im} \mathbb{G}^{(HM)}(\mathbf{x}_0, \mathbf{x}_0). \qquad (2.5)$$

The sum of Eqs. (2.4) and (2.5) is referred to as the *total* LDOS, representing the totality of electric- and magnetic-field energy localized to a point $\mathbf{x}_0$, at frequency ω, over all modes. (Significant alterations to the modal-decomposition expressions are needed, e.g., in plasmonic (and polaritonic) systems [24, 25].) Such descriptions are mathematically accurate only in the high-quality-factor, nonoverlapping-mode limit [26, 27], but the dipole-radiation interpretation generalizes to any linear scattering scenario.

To summarize, the imaginary part of the Green's function, $\operatorname{Im} \mathbb{G}(\mathbf{x}_0, \mathbf{x}_0)$, is a measure of the power radiated by electric and/or magnetic dipoles in an arbitrary environment, which is proportional to the spontaneous-emission rate of a dipolar emitter, and it encapsulates the Purcell factor, particularly the ratio Q/V, of high-quality-factor modes that concentrate energy at that point. We have extensively described LDOS due to its versatility and cross-cutting nature. The following quantities have more focused and niche applications and can be described more concisely.

2.2.2 Free-Electron Radiation

Radiation by a free-electron beam is closely related to LDOS, with the key distinction being that the current distribution is now a line source. An electron (charge $-e$) propagating through free space at constant velocity $v\hat{\mathbf{x}}$ comprises a free current density $\mathbf{J}(\mathbf{r}, t) = -\hat{\mathbf{x}} e v \delta(y)\delta(z)\delta(x - vt)$, which generates a frequency-dependent incident field [28]:

$$\mathbf{E}_{\text{inc}} = \frac{e\kappa_p e^{ik_v x}}{2\pi\omega\varepsilon_0}\left[\hat{\mathbf{x}}i\kappa_\rho K_0(\kappa_\rho\rho) - \hat{\boldsymbol{\rho}}k_v K_1(\kappa_\rho\rho)\right], \tag{2.6}$$

written in cylindrical coordinates (x, ρ, θ), where K_n is the modified Bessel function of the second kind, $k_v = \omega/v$, and $\kappa_\rho = \sqrt{k_v^2 - k^2} = k/\beta\gamma$ ($k = \omega/c$, free-space wavevector; $\gamma = 1/\sqrt{1 - \beta^2}$, Lorentz factor). When there is no back-action on the source from its interactions with a scatterer, then the free-electron photon emission and energy loss is a typical scattering problem, with Eq. (2.6) as the incident field.

An important feature of Eq. (2.6) is that the incident field is entirely evanescent (the asymptotic decay of the special function K_n is given by e^{-kr}/kr in the far field). This is expected on physical grounds, as an electron moving at constant velocity cannot radiate. Once a scattering body is brought close to the electron beam, however, the situation changes: the evanescent incident field can excite modes in the scatterer that couple to far-field radiation. (Physically, the electromagnetic-field-mediated interaction of the electron beam with the scatterer can lead to deceleration and therefore radiation.) The radiated power can be computed by an LDOS-like expression, $\frac{1}{2}\operatorname{Re}\int \mathbf{J}^* \cdot \mathbf{E}$, where $\mathbf{J}$ is the free-electron current density, but the bound techniques developed below for scattering bodies are most easily applied to the polarization fields $\mathbf{P}$ within the scatterer, so we prefer an equivalent expression in terms of $\mathbf{P}$. One option would be a linear combination of a direct-radiation term with a scatterer-interaction-radiation term, but the evanescent-only nature of the incident field implies that the direct-radiation term is zero. Instead, the only power lost by the electron beam is that which is extinguished by the scatterer, into absorption losses or far-field radiation. As we discuss more thoroughly in Sect. 2.3.1, the extinction of a scattering body V is given by

$$P_{\text{ext}} = \frac{\omega}{2}\operatorname{Im}\int_V \mathbf{E}_{\text{inc}}^*(\mathbf{x}) \cdot \mathbf{P}(\mathbf{x})\,\mathrm{d}\mathbf{x}, \tag{2.7}$$

which we will use to analyze the free-electron loss, as $P_{\text{loss}} = P_{\text{ext}}$.

When the beam passes by the scatterer without intersecting it, the resulting radiation is referred to as **Smith–Purcell radiation**. When the beam passes through the scatterer, causing radiation, it is referred to as **transition radiation**. And when the beam radiates while propagating *inside* a refractive medium (within which the modified speed of light can be smaller than the electron speed), it is referred to as **Cherenkov radiation**. The Smith–Purcell process resides squarely in the realm of near-field electromagnetism.

2.2.3 CDOS

In Sect. 2.2.1, we showed that the power radiated by a single dipole at position $\mathbf{x}$ is proportional to the LDOS at that point, which itself is proportional to $\operatorname{Im}\mathbb{G}(\mathbf{x}, \mathbf{x})$.

Consider now the power radiated by *two* dipoles, $\mathbf{p}_1$ and $\mathbf{p}_2$, at positions $\mathbf{x}_1$ and $\mathbf{x}_2$, for a total dipole density of $\mathbf{P}(\mathbf{x}) = \mathbf{p}_1 \delta(\mathbf{x} - \mathbf{x}_1) + \mathbf{p}_2 \delta(\mathbf{x} - \mathbf{x}_2)$. The power they jointly radiate is given by

$$
\begin{aligned}
P_{\text{rad}} &= \frac{\omega}{2} \int_V \int_{V'} \mathbf{P}(\mathbf{x}) \operatorname{Im} \mathbb{G}(\mathbf{x}, \mathbf{x}') \mathbf{P}(\mathbf{x}') \, d\mathbf{x} \, d\mathbf{x}' \\
&= \frac{\omega}{2} \left\{ \mathbf{p}_1^{\dagger} \left[\operatorname{Im} \mathbb{G}(\mathbf{x}_1, \mathbf{x}_1) \right] \mathbf{p}_1 + \mathbf{p}_2^{\dagger} \left[\operatorname{Im} \mathbb{G}(\mathbf{x}_2, \mathbf{x}_2) \right] \mathbf{p}_2 \right. \\
&\quad \left. + \mathbf{p}_1^{\dagger} \left[\operatorname{Im} \mathbb{G}(\mathbf{x}_1, \mathbf{x}_2) \right] \mathbf{p}_2 + \mathbf{p}_2^{\dagger} \left[\operatorname{Im} \mathbb{G}(\mathbf{x}_2, \mathbf{x}_1) \right] \mathbf{p}_1 \right\}.
\end{aligned}
\tag{2.8}
$$

The first two terms are the powers radiated by the two dipoles in isolation (or when incoherently excited); the second pair of terms is the positive or negative contribution that arises for constructive or destructive (coherent) interference between the two dipoles. For reciprocal media (of arbitrary patterning), the third and fourth terms are complex conjugates of each other, such that we can just consider one of them (say, the third term) in determining the two-dipole interference. By analogy with Eq. (2.4), we can define a **cross density of states (CDOS)** by the expression:

$$
\text{CDOS}_{ij}(\omega, \mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\pi \omega} \operatorname{Im} \mathbb{G}_{ij}(\mathbf{x}_1, \mathbf{x}_2),
\tag{2.9}
$$

which differs from Ref. [29] only by the absence of a 2 in the prefactor. The sign of the CDOS indicates the sign of the interference term, while its magnitude is a field-correlation strength between the two points of interest in a given electromagnetic environment. The amplification of emission that can occur when the sign is positive is an example of superradiance, while the reduction of emission when the sign is negative is an example of subradiance, in each case mediated by the local CDOS [30]. Because the CDOS is the off-diagonal part of a positive-definite matrix, it is straightforward to show that its magnitude is bounded above by the square root of the product of the diagonal terms in the matrix, i.e., the local densities of states of the two dipoles in isolation [31].

In systems that are closed, or approximately closed, there is another interesting interpretation of the CDOS [29, 31]. Just as the LDOS can be interpreted as a local modal density, the CDOS can be interpreted as a local modal *connectivity*—it is a measure of spatial coherence between two points. In Ref. [29], it was shown the one can compute local coherence lengths from spatial integrals of the CDOS. From these local coherence lengths, it was unambiguously demonstrated that "spatial squeezing" of eigenmodes occurs in systems of disordered plasmonic nanoparticles. This plausibly explains surprising experimental results when probing the local response of such disordered films [32], showing the value of CDOS as an independent concept from LDOS.

There are two other areas in which CDOS emerges as a key metric: Forster energy transfer [33–35] and quantum entanglement and super-radiative coupling between qubits [36–40]. The general idea in each case is a dipole $\mathbf{p}_1$ transferring energy to

a second dipole $\mathbf{p}_2$. In this scenario, $\mathbf{p}_1$ and $\mathbf{p}_2$ are considered *fixed*. By Poynting's theorem, the energy flux into a small bounding surface of $\mathbf{p}_2$, for a field $\mathbf{E}_1$ generated by $\mathbf{p}_1$, is

$$\frac{\omega}{2} \operatorname{Im} \left[\mathbf{p}_2^\dagger \mathbf{E}_1(\mathbf{x}_2) \right] = \frac{\omega}{2} \operatorname{Im} \left[\mathbf{p}_2^\dagger \mathbb{G}(\mathbf{x}_2, \mathbf{x}_1) \mathbf{p}_1 \right], \tag{2.10}$$

which is a form of the CDOS. The fixed nature of the second dipole, $\mathbf{p}_2$, is crucial for the CDOS metric to be the correct one. If the second dipole is *induced by the field emanating from the first dipole*, then $\mathbf{p}_2 = \alpha_2 \mathbf{E}_1(\mathbf{x}_2)$, and the correct energy-transfer expression would be the imaginary part of the polarizability multiplied by the squared absolute value of the Green's function.

2.2.4   Surface-Enhanced Raman Scattering (SERS)

Surface-enhanced Raman scattering is a technique whereby molecules are excited by a pump field, subsequently emitting Stokes- (or anti-Stokes-) shifted radiation (fluorescence) that can be used for imaging or identification [41–44]. The small cross sections of most chemical molecules result in very low pump and emission efficiencies in conventional Raman spectroscopy [45], but one can engineer the near-field environment to enhance both the concentration of the pump field and the emission rate. Efficiency improvements of up to 12 orders of magnitude have been demonstrated, enabling single-molecule detection and a variety of applications.

SERS is a nonlinear process, in which a single dipolar molecular sees both a pump enhancement and a spontaneous-emission enhancement. A key insight for understanding SERS is that the weakness of the nonlinearities of the individual molecules means that the nonlinear process can be treated as the *composition* of linear processes, in which the pump first enhances the excited-population densities (or, classically, the dipole amplitudes), and then the spontaneous-emission enhancements can be treated as a second step, essentially independent of the first.

We can write the key metric of SERS by considering these two steps in sequence, following a procedure outlined in Ref. [46]. First, an illumination field at frequency ω_0 impinges upon the molecule and its environment; in tandem, a total field of $\mathbf{E}_{\omega_0}(\mathbf{x}_0)$ is generated at the molecule. The Raman process generates a dipole moment at frequency ω_1 given by

$$\mathbf{p}_{\omega_1} = \boldsymbol{\alpha}_{\text{Raman}} \mathbf{E}_{\omega_0}(\mathbf{x}_0) \tag{2.11}$$

where $\boldsymbol{\alpha}_{\text{Raman}}$ is the molecular polarizability. Next, the power radiated at ω_1 by this dipole is given, per Eq. (2.1), by

$$P_{\text{rad},\omega_1} = \mathbf{p}_{\omega_1}^\dagger \left[\operatorname{Im} \mathbb{G}_{\omega_1}(\mathbf{x}_0, \mathbf{x}_0) \right] \mathbf{p}_{\omega_1}. \tag{2.12}$$

Hence, we see that there are two opportunities for amplification of SERS: concentrating the incoming field $\mathbf{E}_{\omega_0}$ that determines the dipole amplitude and enhancing the outgoing radiation by maximizing the LDOS, proportional to $\text{Im}\,\mathbb{G}_{\omega_1}(\mathbf{x}_0, \mathbf{x}_0)$, at the location of the dipole. To separate the two contributions, we can write the dipole moment as $\mathbf{p} = \|\boldsymbol{\alpha}\mathbf{E}\|\,(\boldsymbol{\alpha}\mathbf{E}/\|\boldsymbol{\alpha}\mathbf{E}\|)$, i.e., an amplitude multiplied by a unit vector. If we denote the unit vector as $\hat{\mathbf{p}}_{\omega_1}$, then we can write

$$P_{\text{rad},\omega_1} = \|\boldsymbol{\alpha}_{\text{Raman}}\mathbf{E}_{\omega_0}\|^2 \hat{\mathbf{p}}_{\omega_1}^{\dagger}\left[\text{Im}\,\mathbb{G}_{\omega_1}(\mathbf{x}_0, \mathbf{x}_0)\right]\hat{\mathbf{p}}_{\omega_1}, \tag{2.13}$$

where now the first term encapsulates ω_0-frequency concentration, and the second term encapsulates ω_1-frequency LDOS-enhancement. Straightforward arguments lead to a net SERS enhancement, relative to a base rate P_0 without any nearby surface, given by

$$\frac{P_{\text{rad},\omega_1}}{P_0} = \left(\frac{\|\boldsymbol{\alpha}_{\text{Raman}}\mathbf{E}_{\omega_0}\|^2}{\|\boldsymbol{\alpha}_{\text{Raman}}\|^2\|\mathbf{E}_{\text{inc},\omega_0}\|^2}\right)\left(\frac{\rho_{\hat{\mathbf{p}},\omega_1}}{\rho_{0,\omega_1}}\right), \tag{2.14}$$

where $\|\boldsymbol{\alpha}\|$ refers to the induced matrix norm of $\boldsymbol{\alpha}$, $\rho_{\hat{\mathbf{p}},\omega_1}$ is the ω_1-frequency LDOS for a $\hat{\mathbf{p}}$-polarized dipole, and ρ_{0,ω_1} in this expression is the background ω_1-frequency LDOS of a $\hat{\mathbf{p}}$-polarized dipole (not the typical summation over all polarizations). The two parenthetical terms in Eq. (2.14) must both be bounded to identify fundamental limits to SERS enhancements.

2.2.5 Near-Field Radiative Heat Transfer

The warming of the cold earth by the hot sun is mediated by radiative transfer, i.e., photons radiated from the sun to the earth. The maximum rate at which such a process could occur is of course given by the **blackbody** rate, which is determined only by the solid angle subtended by the earth from the sun (or vice versa). Determination of this blackbody rate requires no knowledge of multiple-scattering processes between the two bodies. In the far field, the only "channels" (carriers of power into and out of a scattering region) are propagating-wave channels; by Kirchhoff's law [47], one need only know the absorption or emission rates of the two bodies in isolation to know their maximum radiative-exchange rate. A more general viewpoint of far-field radiation, via the idea of communication channels, is discussed in Sect. 2.3.2.

It has been known for 75 years [48, 49] that two bodies separated by less than a thermal wavelength can exchange radiative heat at significantly larger rates than their far-field counterparts. Once in the near field, the bodies can exchange photons not only through radiative channels but also through evanescent channels; moreover, as the separation distance d is reduced, the number of evanescent channels that can be accessed increases dramatically, scaling as $1/d^2$. These channels can be accessed via any mechanism that produces strong near fields. Polaritonic surface

waves, via either plasmons or phonon–polariton materials, are a natural choice, and hyperbolic metamaterials (whose strongest effect is not surface waves but instead high-wavenumber bulk modes with nonzero evanescent tails) can provide similar performance [50, 51]. Photonic crystals can also support surface waves, but the confinement of those waves is typically related to the size of their bandgap [11], thereby scaling with frequency, yielding surface waves with significantly less confinement than their metallic counterparts.

The complexity of near-field radiative heat transfer (NFRHT) is daunting, both experimentally and theoretically. The first experimental demonstrations of enhancements in NFRHT via near-field coupling were not achieved until the 2000s [52–54], many decades after the original predictions [48, 49], and measurements in the extreme near field were not achieved until 2015 [55]. There are a number of technical hurdles to experimental measurements, especially maintaining consistent, nanometer-scale gap separations over large-scale device diameters while simultaneously measuring miniscule heat currents [55].

The theoretical challenge has been no less severe. NFRHT involves rapidly decaying near fields (requiring high resolution), typically over large-area surfaces (requiring a large simulation region), for spatially incoherent and broadband thermal sources (such that the equivalent of very many simulations is needed). The computational complexity of this endeavor has limited the analysis of NFRHT almost exclusively to high-symmetry structures (planar/spherical bodies, metamaterials, etc.) [56–61], small resonators [58, 62], two-dimensional systems [63], and the like. We review the planar-body interaction, which is informative, while emphasizing the need (and opportunity) for new theoretical tools to understand what is possible when exchanging radiative heat in the near field. (Casimir forces [64–66] are theoretical brethren of NFRHT, the only difference being that they arise from zero-point instead of thermal fluctuations. Though we do not review them here, there is at least one recent work developing theoretical bounds/upper limits using techniques similar to those described throughout this chapter [67].)

Consider two near-field bodies with temperatures T_1 and T_2, respectively. By the fluctuation–dissipation theorem, the incoherent currents in body 1, $\mathbf{J}_1$, have ensemble averages (denoted $\langle\rangle$) given by [58]

$$\langle \mathbf{J}_1(\mathbf{x}, \omega)\mathbf{J}_1^\dagger(\mathbf{x}', \omega)\rangle = \frac{4\varepsilon_0\omega}{\pi} \operatorname{Im}\left[\chi_1(\mathbf{x}, \omega)\right] \Theta(\omega, T_1)\delta(\mathbf{x} - \mathbf{x}')\mathcal{I}, \qquad (2.15)$$

where $\chi_1(\mathbf{x}, \omega)$ is the material susceptibility of body 1, $\mathcal{I}$ is the 3×3 identity matrix, and $\Theta(\omega, T)$ is the Planck distribution:

$$\Theta(\omega, T) = \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}. \qquad (2.16)$$

These currents radiate to body 2, at each frequency ω, at a rate that we denote $\Phi_{21}(\omega)$. The rate $\Phi_{21}(\omega)$ is given by the ensemble average of the flux into body 2, i.e., $\langle -\frac{1}{2}\operatorname{Re}\int_{S_2} \mathbf{E} \times \mathbf{H}^* \cdot \hat{\mathbf{n}}\rangle$, where S_2 is a bounding surface of V_2, $\hat{\mathbf{n}}$ is the

outward normal, and the field sources are given by Eq. (2.15), except without the Planck function. The Planck function is separated so that $\Phi_{21}(\omega)$ is independent of temperature and depends only on the electromagnetic environment. Then the radiative heat transfer rate into 2 from currents in 1, denoted H_{21}, is given by

$$H_{21} = \int \Phi_{21}(\omega)\Theta(\omega, T_1)\, d\omega. \tag{2.17}$$

Similarly, the rate of transfer from body 2 to body 1, H_{12}, is given by

$$H_{12} = \int \Phi_{12}(\omega)\Theta(\omega, T_2)\, d\omega, \tag{2.18}$$

and the net transfer rate is the difference between the two. For reciprocal bodies, the rates $\Phi_{12}(\omega)$ and $\Phi_{21}(\omega)$ are always equal (by exchanging the source and "measurement" locations), but this is also true more generally: for *two* bodies exchanging radiative heat in the near field, $\Phi_{12}(\omega)$ and $\Phi_{21}(\omega)$ must be equal, or else one could have net energy exchange with both bodies at equal temperatures, in violation of the second law of thermodynamics. Note that if three bodies are present, or either body radiates significant amounts of energy into the far field, this relation need not hold in nonreciprocal systems, and indeed "persistent currents" have been predicted in three-body systems in the near field [68]. Throughout this chapter, we will focus on the prototypical two-body case, so we can take

$$\Phi_{12}(\omega) = \Phi_{21}(\omega) = \Phi(\omega), \tag{2.19}$$

without assuming reciprocity. Hence, the *net* NFRHT rate between the two bodies is given by

$$H_{2\leftarrow 1} = \int \Phi(\omega)\left[\Theta(\omega, T_1) - \Theta(\omega, T_2)\right] d\omega. \tag{2.20}$$

Often, it is illuminating to reduce the problem to a single temperature T and study the differential heat transfer for a temperature differential ΔT. The net heat exchange divided by this temperature differential is the **heat transfer coefficient**, or HTC, which is given by Eq. (2.20), except that the temperature difference is replaced by a single derivative of $\Theta(\omega, T)$ with respect to temperature:

$$\text{HTC} = \int \Phi(\omega)\frac{\partial \Theta(\omega, T)}{\partial T}\, d\omega. \tag{2.21}$$

Hence, the quantity $\Phi(\omega)$ is the designable quantity in NFRHT and is the focus of the NFRHT bounds appearing across Sect. 2.4.

2.2.6 Mode Volume

Finally, we turn to a unique near-field quantity: mode volume. Intuitively, **mode volume** encapsulates an "amount of space" occupied by an electromagnetic mode. Obviously, defining the volume of a continuous density is necessarily subjective. But we can develop an intuitive approach to the common volume definition. The energy density of a mode m at any point $\mathbf{x}$ is proportional to $\varepsilon(\mathbf{x})|\mathbf{E}_m(\mathbf{x})|^2$. If the maximum energy density occurs at a point $\mathbf{x}_0$, we can define the volume of the mode as follows: let us redistribute the energy into a binary pattern in which at every point in space, it can only take the values 0 or $\varepsilon(\mathbf{x}_0)|\mathbf{E}_m(\mathbf{x}_0)|^2$. Let us also require that the total energy of the mode will not change in this binarization, i.e., $\int \varepsilon(\mathbf{x})|\mathbf{E}(\mathbf{x})|^2$ remains fixed. Then the corresponding redistributed field will occupy the volume:

$$V_m = \frac{\int \varepsilon(\mathbf{x})|\mathbf{E}_m(\mathbf{x})|^2}{\varepsilon(\mathbf{x}_0)|\mathbf{E}_m(\mathbf{x}_0)|^2}. \tag{2.22}$$

Typical modes of interest, which have strong field concentration and Gaussian- or Lorentzian-like energy decay, are well-suited to such an interpretation.

More rigorously, per Eq. (2.3), the modal field intensity is the quantity that determines the interaction of a dipole with a specific mode, and the contribution of that mode to the spontaneous emission of the dipole. Then an alternative interpretation of the quantity in Eq. (2.22) is that the numerator can be taken to be 1, for a normalized mode, and the denominator is the relevant coupling term in the Hamiltonian that is to be maximized. This alternative approach explains why a common mathematical objective is to minimize the expression in Eq. (2.22), without reference to any physical concept of volume.

A critical question around mode volume is whether such a concept is even valid. For closed (or periodic) systems with nondispersive, real-valued permittivities, the Maxwell operator is Hermitian, and there is an orthogonal basis of modal fields that can be orthonormalized. Dispersion in the material systems makes the eigenproblem nonlinear, but for Drude–Lorentz-like dispersions, one can introduce auxiliary variables, and in this higher-dimensional space, there is again a linear, Hermitian eigenproblem [69]. But once losses are introduced, either through open boundary conditions or material dissipation, the operator is no longer Hermitian, and the modes cannot be orthonormalized with an energy-related inner product [25]. Instead, one must work with *quasinormal modes* (QNMs), for which two issues arise.

If material losses are the dominant loss mechanism, as is typical in plasmonics, then the key new subtlety often is the modification of orthogonality: the modes are orthogonal in an *unconjugated* "inner product" (e.g., $\int \varepsilon \mathbf{E}_1 \cdot \mathbf{E}_2$ instead of $\int \varepsilon \mathbf{E}_1^* \cdot \mathbf{E}_2$), which then replaces the standard conjugated inner product in modal expansions such as Eq. (2.3). While this is mathematically convenient, it can stymie our typical intuition. A beautiful example is demonstrated in Ref. [24]. There, it is shown that the spontaneous emission near a two-resonator antenna can be dominated by two

QNMs, as expected. However, if one tries to attribute individual contributions from each QNM, one of the QNMs appears to contribute *negative* spontaneous emission. This is attributable to the modified inner product: modes that are orthogonal in the unconjugated inner product are not orthogonal in an energy inner product, and their contributions to a positive energy flow (such as spontaneous emission) are invariably linked; one can no longer separate a power quantity such as LDOS into individual contributions from constituent modes. Ultimately, one can define mode volume as a complex-valued quantity [70], in which case it no longer becomes an independent quantity of interest to minimize or maximize, but rather an ingredient for other scattering quantities of interest.

If radiation losses are the dominant loss mechanism, one faces a hurdle even before orthogonality: just *normalizing* the modal fields becomes tricky. If the modal fields eventually radiate in free space, they will asymptotically scale as $e^{ik_m r}/r$, where $k_m = \omega_m/c$ is the wavenumber of the mode and r is a distance from the scatterer. But the losses to radiation transform the resonant eigenvalues to poles in the lower-half of the complex-frequency plane, i.e., $\omega_m \to \omega_m^{(r)} - i\omega_m^{(i)}$, where $\omega_m^{(i)} > 0$. Hence, the modal fields grow exponentially, $\sim e^{\omega_m^{(i)} r/c}$, such that any integrals of the form $\int \mathbf{E}^2$ or $\int |\mathbf{E}|^2$ diverge. There are a few resolutions to this issue (cf. Sec. 4 of Ref. [71]). Perhaps the simplest is to use computational perfectly matched layers (PMLs) to confine the fields to a finite region. Then, for any accurate discretization of the Maxwell operator, one is simply left with a finite-sized, non-Hermitian matrix, whose eigenvectors will generically be orthonormalizable under the unconjugated inner product. (Exceptions to this occur at aptly named *exceptional points*, where modes coalesce, and one needs Jordan vectors to complete the basis [72, 73].) Orthogonalization of these modes requires the same modification of the inner product discussed above (even without material loss [27], as radiation is a loss mechanism). Moreover, there is one further difficulty: sometimes important contributions to energy expression can come from fields that *primarily reside in the PML region*. One can think of this as a condition for a complete basis. It is difficult to attribute physical intuition or meaning to such contributions.

In Sect. 2.4.4, where we develop bounds for mode volume, we will only deal with cases of lossless dielectric materials, we assume the quality factors are sufficiently high that the system is approximately closed, and we assume isolated resonances. This is the limit in which the mode volume as defined by Eq. (2.22) is exactly the quantity that enters the LDOS expression of Eq. (2.3), which is typically the underlying goal of minimizing mode volume in the first place. In scenarios where one must use quasinormal modes, it is probably better to eschew them altogether (if one wants a bound) and to instead work directly with the scattering quantity (e.g., LDOS) of interest.

2.3   Analytical and Computational Bound Approaches

Across many areas of science and technology, "fundamental limits" or "bounds" play an important role in technological selection, theoretical understanding, and optimal design. Examples abound:

- The **Shockley–Queisser limits** for solar-cell energy conversion efficiency. Originally developed for single-cell, all-angle solar absorption and energy conversion [74], the basic framework they developed identifies two required loss mechanisms in any solar cell: radiation back to the sun (at the open circuit condition [75]), and thermalization losses in the establishment of quasi-Fermi levels in each band. Almost any proposed solar-energy-conversion technique must be put through a Shockley–Queisser analysis to earn serious consideration as a technology.
- The **Yablonovitch** $4n^2$ **limit**, for the maximum broadband, all-angle absorption enhancement in any optically thick material [76]. The factor $4n^2$, for a refractive index n, arises from the density-of-states enhancement in a high-index material, a 2X enhancement from mirrors on the rear surface, and a 2X enhancement from the reorientation of mostly vertical rays into random angles.
- The **Wheeler–Chu limit** to antenna quality factor, Q [77, 78]. It is difficult for a subwavelength antenna (such as a cell phone antenna) to operate over a wide bandwidth, and the Wheeler–Chu (sometimes Harrington is also given credit [79]) limit imposes a bound on the maximum operational bandwidth. Most state-of-the-art antenna designs operate very close to the Wheeler–Chu limit [80].
- The **Bergman–Milton bounds** on the effective properties of a composite material [81–86].
- The **Abbe diffraction limit** on the maximum focusing of an optical beam. This limit can be circumvented in the near field [87, 88], or even in the far field if one is willing to tolerate side lobes [89–95].
- The **Shannon bounds** [96], a foundational idea in information theory [97].

Many of these examples involve electromagnetism, but typically only for noninteracting waves and simplified physical regimes. The Yablonovitch $4n^2$ limit applies in geometric (ray) optics, the Wheeler–Chu limit only arises in highly subwavelength structures, and the diffraction limit applies only to free-space (or homogeneous-medium) propagation. Is it possible to create an analogous theoretical framework for the full Maxwell equations, identifying fundamental spectral response bounds while accounting for the exceptional points [98, 99], speckle patterns [100], bound states in the continuum [101], and other exotic phenomena permitted by the wave equation? A flurry of work over the past decade suggests that in many scenarios, the answer should be "yes." In the following subsections, we outline the key new ideas that have been developed.

2.3.1 Global Conservation Laws

One approach particularly well-suited to formulating bounds is to replace the complexity of the full Maxwell-equation design constraints with a single constraint that encodes some type of conservation law. The Yablonovitch limit, discussed in the previous section, offers a powerful example: to identify maximum absorption enhancement in a geometric-optics setting, one can replace the complexity of ray-tracing dynamics with a single density-of-states constraint. Unfortunately, one cannot extend such density-of-states arguments to full-Maxwell and near-field settings, but other types of "conservation laws" can be identified. A **global conservation law** that has been particularly fruitful for nanophotonics is the **optical theorem**. The optical theorem [102–104] is a statement of global power conservation: the total power extinguished from an incident beam by a scattering body (or bodies) equals the sum of the powers scattered and absorbed by that body. Writing the extinguished, scattered, and absorbed powers as P_{ext}, P_{scat}, and P_{abs}, respectively, the optical theorem can be expressed as

$$P_{\text{ext}} = P_{\text{scat}} + P_{\text{abs}}. \tag{2.23}$$

Conventionally, the optical theorem is specified in terms of the far-field scattering amplitudes of a scattering body [102], in which case the extinction is shown to be directly proportional to the imaginary part of the forward-scattering amplitude. This expression can be interpreted as a mathematical statement of the physical intuition that the total power taken from an incident beam can be detected in the phase and amplitude of its shadow. The analysis does not have to be done in the far field; another common version is to relate the extinguished-, scattered-, and absorbed-power fluxes via surface integrals of the relevant Poynting fluxes [103]. Still one more version of the optical theorem, and the one that turns out to be most useful for wide-ranging bound applications, is to use the divergence theorem to relate the surface fluxes to the fields within the volume of the scatterer and write all powers in terms of the polarization currents and fields induced in those scatterers [104]. As we briefly alluded to in the discussion of free-electron radiation in Sect. 2.2.2, the work done by a field $\mathbf{E}$ on a polarization field $\mathbf{P}$ in a volume V is given by $\left(\frac{\omega}{2}\right) \operatorname{Im} \int_V \mathbf{E}^* \cdot \mathbf{P} = \left(\frac{\omega}{2}\right) \int_V \mathbf{P}^* \left[\operatorname{Im} \chi / |\chi|^2\right] \mathbf{P}$, where χ is the material susceptibility. (We assume throughout scalar, electric material susceptibilities χ. The generalizations to magnetic, anisotropic, and bianisotropic materials are straightforward in every case.) Extinction is the work done by the incident field on the induced polarization field, scattered power is the work done by that polarization field on the scattered fields $\mathbf{E}_{\text{scat}}$, and absorbed power is the work done by the total field on the polarization field. Hence, the optical theorem reads:

$$\mathrm{Im} \int_V \mathbf{E}_{\mathrm{inc}}^*(\mathbf{x}) \cdot \mathbf{P}(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \mathrm{Im} \int_V \int_V \mathbf{P}^*(\mathbf{x}) \cdot \mathbb{G}_0(\mathbf{x}, \mathbf{x}')\mathbf{P}(\mathbf{x}') \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{x}'$$

$$+ \int_V \mathbf{P}^*(\mathbf{x}) \cdot \frac{\mathrm{Im} \, \chi(\mathbf{x})}{|\chi(\mathbf{x})|^2} \mathbf{P}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \qquad (2.24)$$

where we have substituted $\mathbf{E}_{\mathrm{scat}}(\mathbf{x}) = \int_V \mathbb{G}_0(\mathbf{x}, \mathbf{x}')\mathbf{P}(\mathbf{x}') \, \mathrm{d}\mathbf{x}'$ for the scattered field and dropped the constant factor $(\omega/2)$ preceding every integral. Equation (2.24) relates extinction on the left-hand side to the sum of scattered and absorbed powers on the right-hand side. For intuition and compactness, it is helpful to rewrite equations like Eq. (2.24) in a matrix/vector form. We can assume any arbitrarily high-resolution discretization in which $\mathbf{P}(\mathbf{x})$ becomes a vector $\mathbf{p}$, the integral operator $\int_V \mathbb{G}(\mathbf{x}, \mathbf{x}') \, \mathrm{d}\mathbf{x}'$ becomes a matrix $\mathbb{G}_0$, and integrals of the conjugate of a field $\mathbf{a}(\mathbf{x})$ with another $\mathbf{b}(\mathbf{x})$ are replaced with vector inner products $\mathbf{a}^\dagger \mathbf{b}$. It is also helpful to define a material parameter $\xi(\mathbf{x}) = -1/\chi(\mathbf{x})$, and a corresponding (diagonal) matrix $\xi = -\chi^{-1}$. With these notational changes, Eq. (2.24) can be rewritten as

$$\mathrm{Im} \left(\mathbf{e}_{\mathrm{inc}}^\dagger \mathbf{p} \right) = \mathbf{p}^\dagger \left[\mathrm{Im} \, \mathbb{G}_0 + \mathrm{Im} \, \xi \right] \mathbf{p}. \qquad (2.25)$$

This is the vectorized version of the optical theorem, and it illuminates some of the mathematical structure embedded in this particular version of power conservation. The left-hand side is a linear function of the polarization field $\mathbf{p}$, while the right-hand side is a quadratic function. Moreover, in passive systems, the absorbed and scattered powers are nonnegative quantities. This nonnegativity is embedded in the matrices (operators) $\mathrm{Im} \, \mathbb{G}_0$ and $\mathrm{Im} \, \xi$, both of which are positive semidefinite (denoted by "≥ 0") in passive systems:

$$\mathrm{Im} \, \mathbb{G}_0 \geq 0, \qquad (2.26)$$

$$\mathrm{Im} \, \xi \geq 0. \qquad (2.27)$$

The positive semidefinite nature of these matrices implies that the right-hand side of Eq. (2.25) is a convex quadratic functional of $\mathbf{p}$. Hence, Eq. (2.25) can be interpreted as an ellipsoid (as opposed to a hyperboloid) in the high-dimensional space occupied by $\mathbf{p}$.

A key feature of Eq. (2.25), and the conservation laws to follow, is that it is **domain oblivious** [105]. Suppose we enforce that constraint on a high-symmetry domain, such as a sphere or half-space, where the operator $\mathbb{G}_0$ might be easy to construct. Of course, enforcing Eq. (2.25) will enforce power conservation on the sphere itself. But it *also enforces power conservation on all sub-domains of the sphere*. This is not obvious—the operator $\mathbb{G}_0$ is different for every choice of domain and range, and once we have chosen a sphere for both, it seems that we are stuck with only the sphere domain. The key, however, is the appearance of $\mathbf{p}$ in each term of Eq. (2.25), and twice on the right-hand side. To enforce Eq. (2.25) on a

smaller sub-domain, instead of changing the domain and range of the operator, we can instead enforce the polarization **p** to be zero at each point outside the sub-domain but inside the enclosing domain. On the right-hand side, this effectively changes both the domain and range of $\mathbb{G}_0$, while on the left-hand side, it nulls any extinction contribution from outside the sub-domain. Hence, the conservation law of Eq. (2.25), and all of the volume-integral-based conservation laws to follow, is domain oblivious.

Power conservation via the optical theorem has led to a surprisingly wide array of bounds and fundamental limits in electromagnetic systems. The key idea is to drop the full Maxwell-equation constraint that is implicit in any design problem and replace it with only the power-conservation expression of Eq. (2.25). Even with just this single constraint, surprisingly good bounds can be attained. As an example, consider systems where absorptive losses are more important than radiation/scattering losses. In such systems, we can drop the $\text{Im}\,\mathbb{G}_0$ term in the optical theorem of Eq. (2.25) and use its positivity to write a constraint that absorbed power, be less than or equal to extinction:

$$\mathbf{p}^{\dagger}\,(\text{Im}\,\xi)\,\mathbf{p} \leq \text{Im}\left(\mathbf{e}_{\text{inc}}^{\dagger}\mathbf{p}\right). \tag{2.28}$$

This constraint implies a bound on the strength of the polarization field, because the left-hand-side term is quadratic (and positive-definite) in **p**, while the right-hand side is linear in **p**. A few steps of variational calculus [106] can identify the largest polarization-field strength that can be induced in a scatterer:

$$\|\mathbf{p}\|^2 = \mathbf{p}^{\dagger}\mathbf{p} = \int_V |\mathbf{P}(\mathbf{x})|^2\,\mathrm{d}\mathbf{x} \leq \frac{\|\mathbf{e}_{\text{inc}}\|^2}{\text{Im}\,\xi} = \frac{|\chi|^2}{\text{Im}\,\chi}\int_V |\mathbf{E}_{\text{inc}}(\mathbf{x})|^2\,\mathrm{d}\mathbf{x}. \tag{2.29}$$

We have a first bound: in a lossy material, wherein $\text{Im}\,\chi > 0$, there is a bound on the largest polarization currents that can be induced in a scatterer, based only on the material properties and the energy of the *incident* wave in the scattering region. Polarization currents beyond this strength would have absorbed powers larger than their extinction, implying an unphysical negative scattered power.

Beyond the strength of the polarization field itself, one can use similar variational-calculus arguments to identify bounds on wide-ranging quantities: extinction, absorption, and scattering; in bulk materials [106], 2D materials [107], and lossy environments [108, 109]; high-radiative-efficiency scatterers [110]; and even near-field quantities such as local density of states [46, 106], near-field radiative heat transfer [107, 111], and Smith–Purcell radiation [112]. As a canonical example, let us consider the extinction, absorption, and scattering cross sections of a scattering body with volume V, susceptibility χ, and a plane-wave incident field. Cross sections $\sigma_{\text{ext,abs,scat}}$ are the relevant powers divided by the intensity of the incident wave; the corresponding bounds are

$$\frac{\sigma_{\text{abs,scat,ext}}}{V} \leq \frac{\beta\omega}{c} \frac{|\chi|^2}{\text{Im}\,\chi} \qquad \beta_{\text{abs,ext}} = 1, \beta_{\text{scat}} = \frac{1}{4}. \tag{2.30}$$

Per-volume cross sections are bounded above by the frequency of the incoming waves and the material susceptibilities. Plasmonic nanoparticles can approach these bounds [106, 107, 113].

One subtlety that arises in the near field (whose bounds are discussed in depth in Sect. 2.4) is which conservation laws to use. The absorption- and extinction-based constraint of Eq. (2.28) may not be ideal for local density of states, for example, as the power radiated by a dipole is not exactly the same as the power extinguished by a nearby scatterer. (There is a separate pathway for the dipole to radiate directly to the far field, and this radiation can destructively/constructively interfere with waves scattered by the scatterer.) The optical theorem of Eq. (2.25) arises from equating fluxes through a surface surrounding the scatterer. Instead, in the near field, one can draw a surface around the dipolar source itself. Then one can identify new conservation laws, which now relate the total power radiated by the dipole (the LDOS) to the sum of power absorbed in the scatterer and power radiated to the far field.

In some systems, radiation losses are the limiting factor rather than absorption losses. Prominent examples include metals at low frequencies and low-loss dielectrics. In these systems, the key component of the optical theorem of Eq. (2.25) is the radiation-loss term with $\text{Im}\,\mathbb{G}_0$, not the absorption-loss term. Of course, absorption must be positive, so we can drop it and replace the optical theorem with a second inequality version:

$$\mathbf{p}^\dagger \left(\text{Im}\,\mathbb{G}_0\right) \mathbf{p} \leq \text{Im}\left(\mathbf{e}_{\text{inc}}^\dagger \mathbf{p}\right). \tag{2.31}$$

Although the $\text{Im}\,\mathbb{G}_0$ matrix may appear daunting, we typically use high-symmetry volumes for our designable domains, and we can use analytical or semi-analytical forms of $\text{Im}\,\mathbb{G}_0$ in those domains. (Such usage does not restrict the validity of the bound to only the high-symmetry domain; as discussed above, this expression is domain oblivious.) One common high-symmetry domain is a sphere, in which case $\text{Im}\,\mathbb{G}_0$ can be written in a basis of vector spherical waves [114–116]. Application of this approach to the question of maximum cross sections yields different bounds from the ones of Eq. (2.30). One must limit the number of spherical waves that can contribute to the scattering process; allowing only the first N electric multipole leads to maximum cross sections proportional to the square of the wavelength, λ:

$$\sigma_{\text{abs,scat,ext}} \leq \frac{\beta\lambda^2}{\pi}\left(N^2 + 2N\right) \qquad \beta_{\text{scat,ext}} = 1, \beta_{\text{abs}} = \frac{1}{4}, \tag{2.32}$$

with double the value if the magnetic vector spherical waves can be equally excited. Note the different values of β for absorption and scattering in the absorption-limited case of Eq. (2.30) versus the radiation-limited case of Eq. (2.32). The different coef-

ficients arise because of the different conditions under which maximum extinction occurs. In an absorption-dominated system, arbitrarily small scattering is possible (in principle), such that the maximum for extinction and absorption coincide, while the scattered-power maximum requires a reduction in absorption relative to extinction and a 1/4 coefficient to account for the matching that must occur. The opposite occurs in scattering-limited systems, where absorption can be arbitrarily small (in principle), the maximum for extinction and scattering coincide, and an extra factor of 1/4 is introduced when absorption is to be maximized. The bound of Eq. (2.32) was originally derived for antenna applications or spherically symmetry scatterers via long and/or restrictive arguments [117–121]; the single conservation law of Eq. (2.31) is sufficient to derive Eq. (2.32) in quite general settings [122, 123]. (An interesting precursor to the global-conservation-law approach is Ref. [124], which identifies metrics that intrinsically have bounded optima over polarization currents, even *without* any constraints.)

Of course, in some settings, both absorption and radiation losses will be important to capture what is possible, and the bounds of Eqs. (2.30, 2.32) may not be sufficient. It is possible to capture both loss mechanisms in a single bound by using the entirety of the optical theorem, Eq. (2.25), without dropping either term. This was first recognized in Refs. [116, 125, 126]. Ref. [116] used this approach to derive bounds on the thinnest possible perfect absorber. (Or, conversely, the maximum absorption of an arbitrarily patterned thin film with a given maximum thickness.) Cross-sectional bounds given in Ref. [116, 125, 126] are generalizations of the two bounds listed above, Eqs. (2.30,2.32), containing each as separate asymptotic limits. At normal incidence, one can derive a simple transcendental equation for the minimum thickness, $h_{\min}$, of a perfect absorber with material parameter $\xi = -1/\chi$:

$$h_{\min} = \left(\frac{2\lambda}{\pi}\right) \frac{\operatorname{Im} \xi(\omega)}{1 - \operatorname{sinc}^2 (\omega h_{\min}/c)}. \tag{2.33}$$

This approach has been successfully applied to the identification of the minimum thickness of a metasurface reflector [127].

Finally, at the global-conservation level, one can go one step further, as first recognized in Refs. [125, 126]. The optical theorem of Eq. (2.25) represents the conservation of *real* power across the volume of a scatterer, which can be understood as the conservation of the real part of the Poynting vector through any bounding surface. Additionally, the imaginary part of the Poynting vector corresponds to what is known as *reactive* power [103]. The complex-valued version of the optical theorem is essentially the same as Eq. (2.25) but without the imaginary part in any of the terms; a careful analysis leads to the generalized optical theorem:

$$-\mathbf{p}^\dagger \mathbf{e}_{\mathrm{inc}} = \mathbf{p}^\dagger [\mathbb{G}_0 + \xi] \mathbf{p}. \tag{2.34}$$

The real and imaginary parts of Eq. (2.34) now offer *two* global conservation laws that must be satisfied in any scatterer. The real-power conservation law accounts for absorption- and radiation-loss pathways, while the reactive-power

conservation law accounts for resonance conditions in real materials. The latter has been shown to be beneficial for tightening bounds in plasmonic materials that are relatively large (wavelength-scale sizes are quite large for plasmonic resonators) or which have very large negative real susceptibilities and/or very small imaginary susceptibilities [126]. This approach has been applied to bounds in cloaks [128] and focusing efficiency [129]. Equation (2.34) can be derived in one step from the volume-integral equation [130] (or Lippmann–Schwinger equation), which in this notation reads $[\mathbb{G}_0 + \xi]\,\mathbf{p} = -\mathbf{e}_{\text{inc}}$, simply by taking the inner product of that equation with $\mathbf{p}$.

In this section, we have seen that the optical theorem, written over the volume polarization fields induced in a scatterer, offers a single (or two) global conservation laws that can be used to identify bounds in wide-ranging applications. In Sect. 2.3.3, we show that it is also a starting point for generating an infinite number of "local" conservation laws. First, however, we will explore an approach that is closely related to global conservation laws: so-called "channel" bounds.

2.3.2   Channel Bounds

In this section, we explore another technique for identifying bounds to what is possible: decomposing power transfer into a set of independent or orthogonal power-carrying "channels." Then the upper limits distill to the maximum power (or alternative objective) per channel multiplied by the number of possible channels.

A particularly elegant formulation of channels was proposed by D. A. B. Miller and colleagues in the early 2000s [131–134]. Consider a transmitter region that wants to communicate (i.e., send information/energy) to a receiver region, and a vacuum (or background) Green's function operator $\mathbb{G}_0$ comprising the fields in the receiver from sources in the transmitter. How many communication channels are possible? There is a simple, rigorous mathematical answer to this question: if one decomposes the $\mathbb{G}_0$ operator via a singular value decomposition (SVD) [135],

$$\mathbb{G}_0 = \mathbb{U}\mathbb{S}\mathbb{V}^{\dagger}, \qquad (2.35)$$

then each pair of singular vectors forms an independent channel. The singular value decomposition encodes orthogonality and normalization. For example, the first right singular value, which we can call $\mathbf{v}_1$, radiates *only* to the first left singular vector $\mathbf{u}_1$ in the receiver region, and the strength of this connection is given exactly by the first singular value, which we can call s_1. This triplet $(\mathbf{v}_1, \mathbf{u}_1, s_1)$ mathematically define a **communication channel**, as are all the pairs in the SVD. There cannot be an infinite number of such channels with arbitrarily large strengths, as the channel strengths obey a simple sum rule related to the integral of the Green's function over the transmitter and receiver volumes:

$$\sum_i |s_i|^2 = \mathrm{Tr}\left(\mathbb{S}^\dagger \mathbb{S}\right) = \mathrm{Tr}\left(\mathbb{G}_0^\dagger \mathbb{G}_0\right) = \int_{V_T} \int_{V_R} \|\mathbb{G}_0(\mathbf{x}_T, \mathbf{x}_S)\|^2 \, \mathrm{d}\mathbf{x}_T \, \mathrm{d}\mathbf{x}_R.$$

(2.36)

One can define more granular bounds as well: for any transmitter/receiver regions enclosed within high-symmetry bounding domains, one can identify upper limits for each individual singular value [136]. The singular values must decay exponentially in two-dimensional systems, whereas in three dimensions, their decay can be sub-exponential. This SVD-based decomposition of Eq. (2.35) implicitly uses a field-energy normalization; one can alternatively use power-transfer normalizations and arrive at related bounds for the communication strength between two volumes [137–139]. Each of these is a powerful approach for free-space communication systems such as MIMO [140, 141]. More generally, they capture a general truth about free-space propagation: it can always be decomposed into orthogonal, power-carrying channels.

In the near field, however, evanescent waves do *not* offer an equivalent set of power-carrying channels. Evanescent waves obey different mathematical orthonor-malization rules, which are consistent with the following fact: evanescent waves decaying (or growing) in one direction *cannot* carry power and power can be transmitted only in the presence of oppositely directed evanescent waves [142]. A prototypical example: a single interface can only exhibit total internal reflection alongside evanescent-wave excitation, whereas the introduction of second interface, and counter-propagating evanescent waves, can lead to the tunneling of power through a "barrier."

In lieu of the general SVD approach, in high-symmetry scenarios, it is often possible to decompose power transfer in a high-symmetry basis. For example, a spherically symmetric scatterer preserves the quantum numbers of incoming vector spherical waves and cannot scatter into waves of different quantum numbers, which implies that each vector spherical wave comprises a "channel" for incoming and outgoing radiation. Similarly, in planar systems, the in-plane (parallel) wavevector $\mathbf{k}$ is a conserved quantity, in which case one can isolate the scattering process into each $\mathbf{k}$-dependent propagating and evanescent plane wave. One cannot define free-space evanescent-wave channels, per the orthonormalization discussion above, but a more complete analysis can lead to $\mathbf{k}$-dependent transfer coefficients that are readily interpretable as a channel-based power decomposition. We discuss the successful application of these ideas to near-field radiative heat transfer in Sect. 2.4.1.4. A word of caution is important, however: the assumption of a high-symmetry structure dramatically limits the set of structures to which such bounds apply, and in many scenarios, it has been found that the symmetry-independent approaches of global conservation laws (previous section) and local conservation laws (next section) yield both tighter *and* more general bounds.

2.3.3 Local Conservation Laws

In the global-conservation-law section of Sect. 2.3.1, we discussed that one or two conservation-of-power constraints are already sufficient for bounds in many scenarios of interest. Of course, one or two constraints cannot capture every objective of interest: if, for example, one wanted to know the largest average response over multiple incident fields, certainly more constraints are needed. Thankfully, it turns out that there is a systematic way to generate a large number of conservation-law constraints for any nanophotonic design problem of interest.

The key is to identify **local** **conservation laws** that apply at every point within the scatterer [105, 115]. These conservation laws can be "built" from a volume-integral formulation of the underlying governing dynamics, but we will use a more intuitive approach to develop them. The "generalized optical theorem" is written in Eq. (2.34) in vector/matrix notation; the equivalent integral expression is

$$\int_V \int_V \mathbf{P}^*(\mathbf{x})\mathbb{G}_0(\mathbf{x}, \mathbf{x}')\mathbf{P}(\mathbf{x}') \, d\mathbf{x} \, d\mathbf{x}' + \int_V \mathbf{P}^*(\mathbf{x})\xi(\mathbf{x})\mathbf{P}(\mathbf{x}) \, d\mathbf{x} = -\int_V \mathbf{P}^*(\mathbf{x})\mathbf{E}_{\text{inc}}(\mathbf{x}) \, d\mathbf{x}.$$
(2.37)

To formulate local conservation laws, we simply recognize the following: for the first integral over the entire scatterer V that appears in every term, we can replace V with $V_{\mathbf{x}}$, where $V_{\mathbf{x}}$ is an infinitesimal volume centered around any point $\mathbf{x}$ within the scatterer. With this replacement, the dependence on $\mathbf{x}$ of each integrand becomes approximately constant (exactly constant in the zero-volume limit), and the integral simplifies to just multiplication by the volume $V_{\mathbf{x}}$, which appears in every term and can be cancelled, leaving

$$\int_V \mathbf{P}^*(\mathbf{x})\mathbb{G}_0(\mathbf{x}, \mathbf{x}')\mathbf{P}(\mathbf{x}') \, d\mathbf{x}' + \mathbf{P}^*(\mathbf{x})\xi(\mathbf{x})\mathbf{P}(\mathbf{x}) = -\mathbf{P}^*(\mathbf{x})\mathbf{E}_{\text{inc}}(\mathbf{x}).$$
(2.38)

More rigorous justifications are given in Refs. [105, 115] and can proceed either from the volume-integral formulation or, with equal validity, by converting the volume integrals around $V_{\mathbf{x}}$ into surface integrals (via the divergence theorem), in which case Eq. (2.38) is interpreted simply as flux conservation through the surface of $V_{\mathbf{x}}$. To convert Eq. (2.38) to the more compact vector notation, we denote new matrices $\mathbb{D}_i$ as diagonal matrices of all zeros except a single 1 at diagonal entry i, in which case Eq. (2.38) can be written as

$$\mathbf{p}^\dagger \mathbb{D}_i \left(\mathbb{G}_0 + \xi\right) \mathbf{p} = -\mathbf{e}_{\text{inc}}^\dagger \mathbb{D}_i \mathbf{x},$$
(2.39)

which must hold for all spatial locations' index by i. Equation (2.39) offers an infinite set of local conservation laws that must be satisfied for any (linear) scattering body. Moreover, just as for the global conservation laws, Eq. (2.39) is domain oblivious. Hence, if the constraints of Eq. (2.39) lead to a bound, then that bound will apply to all sub-domains (or "patterns") contained therein.

There is a systematic procedure that one can follow for identifying fundamental limits using the constraints of Eq. (2.39). If one discards the Maxwell differential (or integral) equations, and only imposes the constraints of Eq. (2.39), the resulting optimization problem has the form of a **quadratically constrained quadratic program**, or **QCQP**. QCQPs arise across many areas of science and engineering [143–148], and there are many mathematical approaches for solving them. One in particular is useful for identifying bounds: one can relax a QCQP to a **semidefinite program (SDP)** in a higher-dimensional space [145, 149], which can be solved for its global optimum by standard algorithms in polynomial time [150, 151]. The solution of the SDP is guaranteed to be a bound, or fundamental limit, on the solution of the problem of interest. (The semidefinite program can also be regarded as the "dual" [151] of the dual of the QCQP [152], which is another way to see that it leads to bounds.)

Thus, local conservation laws lead to a systematic procedure for identifying bounds, or fundamental limits, to electromagnetic quantities of interest. One replaces the governing Maxwell equations with the domain-oblivious conservation-law constraints of Eq. (2.39), forms a semidefinite program from the objective and constraints, and solves the SDP to find a bound. To avoid the computational complexity of using all of the constraints, one can iteratively select only the "maximally violated" constraints, for rapid convergence to the bound of interest [105]. A mathematically oriented review of bounds related to Eq. (2.39) is given in Ref. [153]. Extensions of various types are given in Ref. [154] (multi-functionality), Ref. [155] (quantum optimal control), Ref. [156] (efficiency metrics), and Ref. [157] (other physical equations).

2.3.4 Sum Rules

Whereas the three previous sections primarily emphasized fundamental limits across spatial degrees of freedom, at a single frequency, **sum rules** center around *spectral* degrees of freedom and constraints related to bandwidth. Sum rules are a prime example of applied complex analysis. Most often, they are taught and discussed in the context of material susceptibilities, so we will start there, before focusing on our key interest, scattering problems. In the Appendix Sect. 2.6, we provide a short review of key results from complex analysis, and the intuition behind their derivations, culminating in the Cauchy residue theorem that is used for all sum rules. Cauchy's residue theorem, for our purposes, can be distilled to the following statement. Consider a function $f(z)$ that is **analytic** (has no poles) in some domain D in the complex z plane. (Below, the analytic variable z will be the frequency ω.) Then the function $f(z)/(z-z_0)$ has a simple pole at z_0, for z_0 in D, and any integral of this function along a closed contour in D containing z_0 simplifies to the value of the function at the pole:

$$\oint_\gamma \frac{f(z)}{z - z_0} = 2\pi i f(z_0), \tag{2.40}$$

where $f(z_0)$ is the "residue" of the function $f(z)/(z - z_0)$. Now let us put Cauchy's residue theorem to use.

Consider a material susceptibility χ that relates an electric field $\mathbf{E}$ to an induced polarization field $\mathbf{P}$. Typically, we might directly consider the frequency-domain relationship of these variables:

$$\mathbf{P}(\omega) = \chi(\omega)\mathbf{E}(\omega), \tag{2.41}$$

where we are suppressing spatial dependencies in these expressions for simplicity. (All of the position dependencies are straightforward.) This multiplicative frequency-domain relation arises from a *convolutional* time-domain relationship: the polarization field at a given field is related to the electric field at all other times convolved with the susceptibility function (as a function of time):

$$\mathbf{P}(t) = \int \chi(t - t')\mathbf{E}(t')\,\mathrm{d}t'. \tag{2.42}$$

(We do not use different variables for the time- and frequency-domain definitions; the domain should be clear in each context.) **Causality** is the formal specification that *cause precedes effect*. Material susceptibilities are causal: the polarization field cannot arise before the electric field has arrived, which means that for some origin of time, the susceptibility function is identically zero at all preceding times:

$$\chi(t - t') = 0 \qquad \text{for } t < t'. \tag{2.43}$$

In the usual Fourier-transform relation between the time- and frequency-domain susceptibility functions, then, one can set the lower limit of the time-domain integral to be 0:

$$\chi(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(t)e^{i\omega t}\,\mathrm{d}t = \frac{1}{2\pi} \int_{0}^{\infty} \chi(t)e^{i\omega t}\,\mathrm{d}t. \tag{2.44}$$

Setting the lower limit of the integral to 0 has an important ramification. Let us assume the susceptibility takes a finite value for all real frequencies. (Metals are an exception, with divergent susceptibilities at zero frequencies, but known modifications to the rules below can be developed to account for this singularity [158, 159].) This implies that the integral of Eq. (2.44) converges to the correct finite value at each frequency. Now let us consider a **complex-valued** frequency $\omega = \omega_0 + i\,\Delta\omega$. If we insert this frequency into Eq. (2.44), we find

$$\chi(\omega_0 + i\,\Delta\omega) = \frac{1}{2\pi} \int_{0}^{\infty} \chi(t)e^{i\omega_0 t}e^{-\Delta\omega t}\,\mathrm{d}t, \tag{2.45}$$

which is equivalent to the integral of Eq. (2.44), except now that there is the additional exponential decay term $e^{-\Delta \omega t}$ in the integrand. This exponential decay term can only aid in convergence, and under appropriate technical assumptions (e.g., Titchmarsh's theorem [160]), one can prove the intuitive idea that Eq. (2.45) cannot diverge for any $\Delta \omega$. This implies that the material susceptibility $\chi(\omega)$ is analytic in the upper-half of the complex-frequency plane. (Conversely, frequencies in the lower half would have the exponentially diverging term $e^{\Delta \omega t}$ in their integrands, leading to divergences at certain frequencies, which is where the system *resonances* are located.) Hence, we can use Cauchy's integral theorem of Eq. (2.40) with $\chi(\omega)$ as the analytic function in the numerator of the integrand. The typical usage of the integral theorem is to select a pole on the real axis (or, technically, in the limit of approaching the real axis from above), and to use a contour C that follows the real line, includes a semicircular deformation around ω', and then closes along a semicircle approaching infinity in the upper-half plane. This contour actually does not enclose *any* poles, instead "side-stepping" the real-axis pole, at a frequency we denote by ω. Hence, we have

$$\oint_C \frac{\chi(\omega')}{\omega' - \omega} \, d\omega' = 0. \tag{2.46}$$

The integral over C can be broken into three components: the principal-valued integral along the real axis from negative infinity to infinity (skipping ω'), the semicircular arc going into the upper-half plane, and the semicircular arc rotating clockwise around ω. The second of these terms is zero (for sufficiently rapid decay of $\chi(\omega)$), while the third term is simply $-i\pi \chi(\omega)$ (half of the typical Cauchy residue term since it is half of a circle, with a negative sign for the clockwise rotation). Equating the negative of the third term to the first, we have

$$i\pi \chi(\omega) = \int_{-\infty}^{\infty} \frac{\chi(\omega')}{\omega' - \omega} \, d\omega'. \tag{2.47}$$

We can take the imaginary part of both sides, and use the symmetry of χ around the origin, $\chi(-\omega) = \chi^*(\omega)$, to arrive at one of the **Kramers–Kronig** (KK) relations for a material susceptibility:

$$\text{Re} \, \chi(\omega) = \frac{2}{\pi} \int_0^{\infty} \frac{\omega' \, \text{Im} \, \chi(\omega')}{(\omega')^2 - \omega^2} \, d\omega'. \tag{2.48}$$

The counterpart KK relation relates the imaginary part of $\chi(\omega)$ to an integral involving the real part. These KK relations are the foundations of sum rules. There are two special pole frequencies ω at which we may have additional information about the material response: infinity frequency and zero frequency (statics). In the limit of infinitely large frequencies, all materials become transparent, with a susceptibility that must scale as

$$\chi(\omega) \to -\frac{\omega_p^2}{\omega^2} \qquad \text{as } \omega \to \infty, \tag{2.49}$$

where ω_p is a constant proportional to the total electron density of the material [158, 159]. Inserting this asymptotic limit into the KK relation of Eq. (2.48), we find our first example of a sum rule:

$$\int_0^\infty \omega \, \text{Im} \, \chi(\omega) \, d\omega = \frac{\pi \omega_p^2}{2}. \tag{2.50}$$

Equation (2.50) is known as either the **TRK sum rule** or the f **sum rule** [158, 159]. It relates the weighted integral of the imaginary part of the susceptibility to simple constants multiplied by the electron density of the material of interest. The quantity $\omega \, \text{Im} \, \chi(\omega)$ is proportional to the *oscillator strengths* in single-electron susceptibility models [161]. Alternatively, in the low-frequency limit, one may know the static refractive index n_0 of a given material; inserting $\omega = 0$ in the KK relation of Eq. (2.48) gives the low-frequency sum rule:

$$\int_0^\infty \frac{\text{Im} \, \chi(\omega)}{\omega} \, d\omega = \frac{\pi}{2} \left(n_0^2 - 1 \right). \tag{2.51}$$

The two sum rules of Eqs. (2.50, 2.51) are well-known material sum rules that are useful for spectroscopy [158, 159] as well as for bounds on material properties [162–164]. We have repeated their well-known derivations to familiarize the reader with the machinery of KK relations and sum rules, which we apply next to scattering properties.

Just as the origin for material sum rules was recognition of material susceptibility as a causal (linear) response function, for scattering sum rules, we want to start by recognizing that the electromagnetic field **E** generated by a source (presumably current) is also a causal linear response function: **E** cannot be nonzero before the current **J** is nonzero. Hence, the electric field at all times before an origin must be zero, which again leads to analyticity in the upper-half of the complex-frequency plane. Yet we do not want KK relations for the electric field at specific points in space; we want KK relations (and sum rules) for relevant power quantities. Typical expressions of interest might be the field intensity, $|\mathbf{E}(\mathbf{x}, \omega)|^2$, or the Poynting flux $(1/2) \, \text{Re} \left[\mathbf{E}^*(\mathbf{x}, \omega) \times \mathbf{H}(\mathbf{x}, \omega) \right]$, at a point **x**, but *neither* of these quantities is analytic in the upper-half plane. The problematic term in each case is $\mathbf{E}^*(\omega)$. Analyticity is not preserved under complex conjugation, and indeed by symmetry, we know that $\mathbf{E}^*(\omega) = \mathbf{E}(-\omega)$ on the real line; if we try to continue ω into the upper-half plane, the $-\omega$ argument moves into the lower-half plane, where the resonances reside. Hence, $\mathbf{E}^*(\omega)$ can have poles, and the corresponding power terms do not have simple KK relations or sum rules.

We are rescued, again, by the optical theorem. Whereas absorbed and scattered powers always involve conjugated total fields, *extinction*, by virtue of the optical theorem, takes a different form (Eq. (2.7)), which is proportional to the overlap

integral of the conjugate of the incident field with the induced polarization field, $\int_V \mathbf{E}_{\text{inc}}^* \cdot \mathbf{P}$. Many common incident fields, such as plane waves of the form $e^{i\omega x/c}$, are analytic *everywhere* in the complex plane, and their conjugates *can* be analytically continued. The polarization field is the product of the analytic material susceptibility with the analytic electric field, and thus is itself analytic. Hence, extinction expressions contain a term that will obey KK relations and sum rules, which we denote $s(\omega)$:

$$P_{\text{ext}}(\omega) = \frac{\omega}{2} \operatorname{Im} \underbrace{\int_V \mathbf{E}_{\text{inc}}^*(\mathbf{x}, \omega) \cdot \mathbf{P}(\mathbf{x}, \omega) \, d\mathbf{x}}_{s(\omega)}. \tag{2.52}$$

By the arguments laid out above, the quantity $s(\omega)$ is analytic in the upper-half plane. It satisfies the other required assumptions as well (e.g., sufficient decay at infinity) for incident fields such as plane waves; we can immediately write a KK relation for it:

$$\operatorname{Re} s(\omega) = \frac{2}{\pi} \int_0^\infty \frac{\omega' \operatorname{Im} s(\omega')}{(\omega')^2 - \omega^2} \, d\omega'. \tag{2.53}$$

Notice that the term in the numerator of the integrand is exactly proportional to extinction; hence, sum rules for the imaginary part of $s(\omega)$ (by analogy with the sum rules for $\operatorname{Im}\chi$) will necessarily be sum rules for extinction. Again paralleling the susceptibility analysis, we can take the limit as $\omega \to \infty$, in which case

$$s(\omega) = \int \mathbf{E}_{\text{inc}}^*(\mathbf{x}, \omega) \cdot \mathbf{P}(\mathbf{x}, \omega) \, d\mathbf{x}$$

$$\to -\frac{\omega_p^2}{\omega^2} \int_V |\mathbf{E}_{\text{inc}}(\mathbf{x}, \omega)|^2 \, d\mathbf{x}$$

$$= -\frac{\omega_p^2}{\omega^2} |\mathbf{E}_0|^2 \, V, \tag{2.54}$$

where $\mathbf{E}_0$ is the (constant) vector amplitude of the plane wave and V is the volume of the scatterer. Evaluating the KK relation for $s(\omega)$, Eq. (2.53), in the high-frequency limit gives a sum rule for the imaginary part of $s(\omega)$:

$$\int_0^\infty \omega \operatorname{Im} s(\omega) \, d\omega = \frac{\pi \omega_p^2}{2} |\mathbf{E}_0|^2 \, V, \tag{2.55}$$

which in turn implies a sum rule for extinction (via Eq. (2.52)):

$$\int_0^\infty P_{\text{ext}}(\omega) \, d\omega = \frac{\pi \omega_p^2}{4} |\mathbf{E}_0|^2 \, V. \tag{2.56}$$

Equation (2.56) dictates that the total integrated extinction of any scattering body is fixed by the amplitude of the incident plane wave and the total number of electrons in the scatterer (from the product of ω_p^2 with V) and is otherwise independent of the shape, resonance profile, and any other characteristics of the scattering body.

Just as for a material susceptibility, one can also derive a sum rule for P_{ext} by setting $\omega = 0$ in the KK relation for $s(\omega)$, Eq. (2.53). The key low-frequency information we can utilize is that the induced dipole moment of the scatterer is related to the incident field via a polarizability tensor $\boldsymbol{\alpha}$. Following a few algebraic steps [165] paralleling the low-frequency material sum rule, one similarly finds a sum rule for the integral of $P_{\text{ext}}(\omega)/\omega^2$. The term $(1/\omega^2)\,d\omega$ is exactly proportional to $d\lambda$, where $\lambda = 2\pi c/\omega$ is the wavelength, so this sum rule is often written as a sum rule over wavelength:

$$\int_0^\infty P_{\text{ext}}(\omega)\,d\lambda = \pi^2 \mathbf{E}_0 \cdot \boldsymbol{\alpha}\mathbf{E}_0. \tag{2.57}$$

There is an additional magnetic polarizability term in materials with a nonzero magnetostatic response [165]. Interestingly, Eq. (2.57) has different dependencies than Eq. (2.56): the polarizability has a weak dependence on material, but a strong dependence on shape. The low-frequency sum rule implies that scattering bodies with the same size and shape, but made of different materials, can have nearly identical wavelength-integrated extinctions. Moreover, electrostatic polarizabilities obey "domain monotonicity" bounds that dictate that the quantity $\mathbf{E}_0 \cdot \boldsymbol{\alpha}\mathbf{E}_0$ must increase as the scatterer domain increases in size, such that one can bound integrated extinction via high-symmetry enclosures for which the right-hand side of Eq. (2.57) often takes a simplified analytical form. Taken together, the high- and low-frequency sum rules of Eqs. (2.56, 2.57) comprise strong constraints on the possible scattering lineshapes of arbitrary scatterers.

Equations (2.56, 2.57) are classical sum rules with a long history. The high-frequency sum rule, Eq. (2.56), was known at least as early as 1963 [166], when the connection to material-susceptibility sum rules was first made. A specialized version of the low-frequency sum rule, Eq. (2.57), was first proposed by Purcell in 1969 [167], in order to bound the minimum volume occupied by interstellar dust. It was generalized to arbitrary scattering bodies in Ref. [165], where the monotonicity bounds (originally developed by Jones [168]) were connected to the low-frequency sum rules. For many years, it seemed that plane-wave extinction might be the *only* scattering quantity for which sum rules can be derived. In recent years, however, it has been recognized that near-field local density of states has a similar form—it is the real or imaginary part of an amplitude, instead of the squared magnitude of an amplitude—for which sum rules can also be derived. We describe this sum rule and its implications in Sect. 2.4.2.

2.4   Fundamental Limits in the Near Field

We have set the stage: we have introduced near-field optics, defined many of the response functions of interest, and described tools formulated for electromagnetic-response bounds. In this section, we describe how these ingredients come together for bounds and fundamental limits to near-field response. We identify different bounds–and the different techniques required to derive them–based on the frequency range of interest: a single frequency (Sect. 2.4.1), all frequencies (Sect. 2.4.2), and finite, nonzero bandwidths (Sect. 2.4.3). We leave bounds for mode volume, which seemingly requires very different techniques, to the final section of the chapter (Sect. 2.4.4).

2.4.1   Single-Frequency Bounds

In Sect. 2.3, we described two techniques that can be used to identify single-frequency bounds to any linear-electromagnetic response function of interest: conservation laws and channel decompositions. In this subsection, we summarize how one can adapt, specialize, and/or combine those approaches in the near field, for spontaneous-emission and CDOS engineering, Smith–Purcell radiation enhancements, and spectral NFRHT response.

2.4.1.1   Spontaneous Emission

The canonical near-field quantity is LDOS, which as discussed in Sect. 2.2.1 is proportional to the spontaneous-emission rate of an electric dipole at a given location. In a closed system, the LDOS is a sum of delta functions over the modes of the system, in which case the LDOS diverges at the modal frequencies. In an open system, however, the modal intuition no longer applies, leading to the more general Green's function expression of Eq. (2.4). This scattering quantity lends itself well to the conservation-law-based scattering-response bounds described in Sect. 2.3.3.

We can repeat here the Green's function expression for LDOS, which we will denote in this section by $\rho(\mathbf{x}, \omega)$:

$$\rho(\mathbf{x}, \omega) = \frac{1}{\pi \omega} \operatorname{Tr} \operatorname{Im} \mathbb{G}(\mathbf{x}, \mathbf{x}, \omega). \tag{2.58}$$

The trace of the Green's function can be computed with a summation over three orthogonal unit vectors $\mathbf{s}_j$, for $j = 1, 2, 3$, in which case the trace can be interpreted as the incoherent summation of the fields from three dipoles with amplitudes $\varepsilon_0 \mathbf{s}_j$. There is an initial impediment to applying the conservation-law framework to this expression: it is not written explicitly as a function of the polarization fields, whose constraints are critical to meaningful bounds. This impediment is easily hurdled: one

can decompose the Green's function into its incident and scattered components. The scattered fields are the convolutions of the free-space Green's function matrix $\mathbb{G}_0$ from the scattering domain to the dipole point; by reciprocity, the overlap of $\mathbf{s}_j$ with $\mathbb{G}_0$ is the field incident upon the scattering body V. By this line of reasoning, for a scalar isotropic medium (the general bianisotropic case is derived in Ref. [106]), one can rewrite LDOS as

$$\rho(\mathbf{x}, \omega) = \rho_0(\omega) + \frac{1}{\pi \omega} \operatorname{Im} \sum_j \int_V \mathbf{E}_{\mathrm{inc}, \mathbf{s}_j} \cdot \mathbf{P}_{s_j} \, dV, \tag{2.59}$$

where $\rho_0(\omega)$ is the free-space LDOS (which is position-independent and given below Eq. (2.4)), and the $\mathbf{s}_j$ subscript encodes the three dipole orientations. Using the same discretized vector/matrix notation as we initiated with Eq. (2.25), this expression can equivalently be written as

$$\rho(\mathbf{x}, \omega) = \rho_0(\omega) + \frac{1}{\pi \omega} \operatorname{Im} \sum_j \mathbf{e}_{\mathrm{inc}, \mathbf{s}_j}^T \mathbf{p}_{s_j}. \tag{2.60}$$

Now we see that LDOS is a linear function of the polarization fields induced in the scattering body. We want to know the largest possible value of LDOS, of Eq. (2.60), subject to the Maxwell equations, but of course the latter constraint contains all of the complexity of the design problem. Instead, we drop the Maxwell-equation constraint and impose only one of the conservation laws of Sect. 2.3. To start, we can impose the conservation law that absorbed power be smaller than extinguished power, of Eq. (2.28), which leads to the optimization problem:

$$\begin{aligned} \max_{\mathbf{p}_{\mathbf{s}_j}} \quad & \frac{1}{\pi \omega} \operatorname{Im} \sum_j \mathbf{e}_{\mathrm{inc}, \mathbf{s}_j}^T \mathbf{p}_{\mathbf{s}_j} \\ \text{s.t.} \quad & (\operatorname{Im} \xi) \, \mathbf{p}_{\mathbf{s}_j}^\dagger \mathbf{p}_{\mathbf{s}_j} \le \operatorname{Im} \left(\mathbf{e}_{\mathrm{inc}, \mathbf{s}_j}^\dagger \mathbf{p}_{\mathbf{s}_j} \right). \end{aligned} \tag{2.61}$$

Treating each dipole orientation $\mathbf{s}_j$ independently, one can find from a Lagrangian analysis that the optimal $\mathbf{p}_{\mathbf{s}_j}$ comprises a linear combination of $\mathbf{e}_{\mathrm{inc}, \mathbf{s}_j}$ and $\mathbf{e}_{\mathrm{inc}, \mathbf{s}_j}^*$; in the near field, where the incident field and its conjugate are nearly identical, and the LDOS is dominated by its scattered-field contribution, we ultimately find the following bound [106]:

$$\rho(\mathbf{x}, \omega) \le \frac{1}{\pi \omega} \frac{|\chi(\omega)|^2}{\operatorname{Im} \chi(\omega)} \sum_{\mathbf{s}_j} \left\| \mathbf{e}_{\mathrm{inc}, \mathbf{s}_j} \right\|^2 = \frac{1}{\pi \omega} \frac{|\chi(\omega)|^2}{\operatorname{Im} \chi(\omega)} \sum_{\mathbf{s}_j} \int_V \left| \mathbf{E}_{\mathrm{inc}, \mathbf{s}_j} \right|^2 \, d\mathbf{x}. \tag{2.62}$$

Normalizing by the free-space electric LDOS $\rho_0(\omega)$, and performing the integral over an enclosing half-space (and keeping only the term that decreases most rapidly

with separation distance d), one finds [106]

$$\frac{\rho(\mathbf{x}, \omega)}{\rho_0(\omega)} \leq \frac{1}{8(kd)^3} \frac{|\chi(\omega)|^2}{\mathrm{Im}\,\chi(\omega)}, \tag{2.63}$$

where $k = \omega/c$ is the free-space wavenumber. Equation (2.63) represents our first near-field bound. This bound only depends on two parameters of the system: the separation distance d, relative to the wavenumber, and the **material enhancement factor**:

$$\frac{|\chi(\omega)|^2}{\mathrm{Im}\,\chi}. \tag{2.64}$$

The material enhancement factor encodes a key trade-off: a large susceptibility magnitude implies large possible polarization currents, while a large imaginary part of the susceptibility implies losses that necessarily restrict resonant enhancement. In Drude metals with $\chi = -\omega_p^2/(\omega^2 + i\gamma\omega)$, the material enhancement factor is given by $\omega_p^2/\gamma\omega$, showing that the largest possible single-frequency response is achievable in materials with large electron densities and small losses. The material enhancement factor is described in further detail in Refs. [106, 169].

The second key parameter is the distance d; the factor $1/d^3$ encodes the dramatic enhancements that are possible in the near field. These enhancements are typically achieved with plasmonic modes, and the factor $1/d^3$ arises from the most rapidly decaying component of the free-space Green's function, $\sim 1/r^3$; squaring this term and integrating over a three-dimensional volume leads to the inverse-cubic dependence. The last point also suggests an important caveat: systems with a different dimensionality must have different scaling laws as a function separation distance. Designing for 2D materials, for example, leads to integrals over 2D (or very thin) domains, leading to a $1/d^4$ near-field enhancement factor. There are also more slowly increasing terms that arise from the mid-field and far-field contributions to the free-space Green's function.

Finally, it should be noted that certain constraints of interest can be seamlessly integrated into the optimization problem of Eq. (2.61). Of particular importance in plasmonics applications is **radiative efficiency**. When one finds a bound on extinction or LDOS, the bound may suggest very large enhancements, but all of that enhancement could be going into material absorption rather than far-field radiation or scattering. Suppose a given application requires a certain radiative efficiency, such as some fraction η of the total emission going into the far field. This can be written mathematically as the constraint that absorption be smaller than $(1 - \eta)$ multiplied by the extinction, or $P_{\mathrm{abs}} \leq (1 - \eta) P_{\mathrm{ext}}$. Absorption is quadratic in the polarization field, while extinction is linear in the polarization field, such that this expression represents an additional constraint that can be seamlessly incorporated into Eq. (2.61). Often the bound of interest, with this constraint, is analytically solvable. Ref. [110] identifies precisely such bounds on high-radiative-efficiency plasmonics, prescribing a trade-off between large response and radiative efficiency.

In Ref. [110], it is not only shown that high-radiative-efficiency bounds can be derived; it is also shown that hybrid dielectric-metal designs can approach the bounds *and* that they surpass the same fundamental limits evaluated for metal-only structures. This example showcases the power of using bounds to understand the broader landscape of a photonics application area of interest.

2.4.1.2 CDOS

Bounds to CDOS can be found along very similar lines to the LDOS bounds of above. We can define the trace of the CDOS via Eq. (2.9), taking

$$\rho(\mathbf{x}_1, \mathbf{x}_2, \omega) = \frac{1}{\pi \omega} \operatorname{Tr} \operatorname{Im} \mathbb{G}(\mathbf{x}_1, \mathbf{x}_2, \omega). \tag{2.65}$$

Then, we can separate out a scattered contribution coming from the polarization fields induced in the scatterer, just as for LDOS, and when this term dominates (i.e., the geometry primarily mediates the CDOS), we have

$$\rho(\mathbf{x}_1, \mathbf{x}_2, \omega) = \frac{1}{\pi \omega} \sum \mathbf{e}_{\text{inc}, \mathbf{s}_j, \mathbf{x}_1}^T \mathbf{p}_{\mathbf{s}_j, \mathbf{x}_2}, \tag{2.66}$$

where the position subscripts on $\mathbf{e}_{\text{inc}}$ and $\mathbf{p}$ denote the source positions of the $\mathbf{s}_j$-polarized dipoles. Hence, in CDOS, the field incident from one position is overlapped with the polarization field induced by a source from a second position. The bound for CDOS will be identical to that of Eq. (2.62), but with $\|\mathbf{e}_{\text{inc}, \mathbf{s}_j}\|^2$ replaced by $\|\mathbf{e}_{\text{inc}, \mathbf{s}_j, \mathbf{x}_1}\| \|\mathbf{e}_{\text{inc}, \mathbf{s}_j, \mathbf{x}_2}\|$. Finally, normalizing by free-space LDOS and dropping all except the most rapidly varying terms as a function of separation distances d_1, d_2, one arrives at the bound [7]:

$$\frac{\rho(\mathbf{x}_1, \mathbf{x}_2, \omega)}{\rho_0(\omega)} \leq \frac{1}{4k^3 \sqrt{d_1^3 d_2^3}} \frac{|\chi(\omega)|^2}{\operatorname{Im} \chi(\omega)}. \tag{2.67}$$

The discussion of the terms that appeared in the LDOS bound of Eq. (2.63) can be translated almost seamlessly here: the same material dependence shows up, corresponding to the same possibilities for plasmonic enhancement, and the same distance dependencies due to the same enhancements of the near fields of the two dipoles. There are likely two further enhancements that can be made to Eq. (2.67). First, Eq. (2.67) is a factor of 2 larger than Eq. (2.63), when the former is evaluated in the limit as $\mathbf{x}_1 \to \mathbf{x}_2$. This is almost certainly because the bound of Eq. (2.67) in Ref. [7] came from evaluating bounds for each diagonal element, simplifying, and *then* taking the trace. Taking the trace and then simplifying the bound would likely remove this factor of 2. Second, the bound of Eq. (2.67) does *not* depend on the distance between the two dipoles, d_{12}. This may be physical in certain limits, e.g., when a plasmon can maintain its amplitude in propagating from one dipole

to the other, but may not be physical when such propagation is not possible, and one would expect improved bounds to capture this. It is likely true that applying the many-conservation-law approach of Sect. 2.3.3 would incorporate such effects. Nevertheless, Eq. (2.67) is a good starting point to understand the upper limits to engineering CDOS in photonic environments.

2.4.1.3 Smith–Purcell Radiation

Another exciting application area for the single-frequency bound approach is to Smith–Purcell radiation, which is the radiation that occurs when a free electron passes near a structured material. A constant-velocity free electron produces only a near field, with no far-field component, but when the evanescent wave interacts with grating-like structures, the gratings can couple the near fields to propagating far fields, leading to a release of energy from the electron in the form of electromagnetic radiation. The natural question, then, is how large this energy release can be?

Mathematically, this question is identical to the question of the work done by a dipole (i.e., LDOS), except that the incident field is different in this case, and is given by Eq. (2.6). Maximizing the overlap of *this* incident field with the induced polarization field, subject to the same constraint of Eq. (2.61), leads to a bound on the Smith–Purcell emission spectral probability given by [112]:

$$\Gamma(\omega) \leq \frac{\alpha}{2\pi c} \frac{|\chi|^2}{\mathrm{Im}\,\chi} \frac{L\theta}{\beta} \left[(\kappa_\rho d) K_0(\kappa_\rho d) K_1(\kappa_\rho d) \right], \tag{2.68}$$

where $\Gamma = P/\hbar\omega$ for emission power P, α is the fine-structure constant, $\beta = v/c$ is the normalized electron velocity, L and θ are the height and opening azimuthal angle of the cylindrical sector containing the patterned material, $\kappa_\rho = k/\beta\gamma$ is the wavenumber divided by β and the Lorentz factor γ, d is the distance of the beam from the surface, and K_n is the modified Bessel function of the second kind. Although the exact expression is somewhat complex, we see that Smith–Purcell radiation also directly benefits from the material enhancement factor $|\chi|^2/\mathrm{Im}\,\chi$. A seemingly surprising conclusion also emerged from Eq. (2.68): *slow* electrons, at small enough separations, can lead to *greater* radiation enhancements than *fast* (i.e., high-energy) electrons. All constant-velocity electrons do not radiate when their speed is smaller than the speed of light in the background medium, and emit only near fields. But high-speed electrons are closer to surpassing the Cherenkov threshold, and hence the fields they generate decay more slowly, out to larger distances. By contrast, low-speed electrons have very strong but very tightly confined near fields. But if one brings a patterned surface close enough, the strong very near fields of slow electrons have greater potential for radiation enhancements than the more moderate near fields of fast electrons.

Some of the general trends, and absolute numerical values, of the bound of Eq. (2.68) were validated theoretically and experimentally in Ref. [112]. In particular, Fig. 2.2 shows an experimental setup for measuring the Smith–Purcell
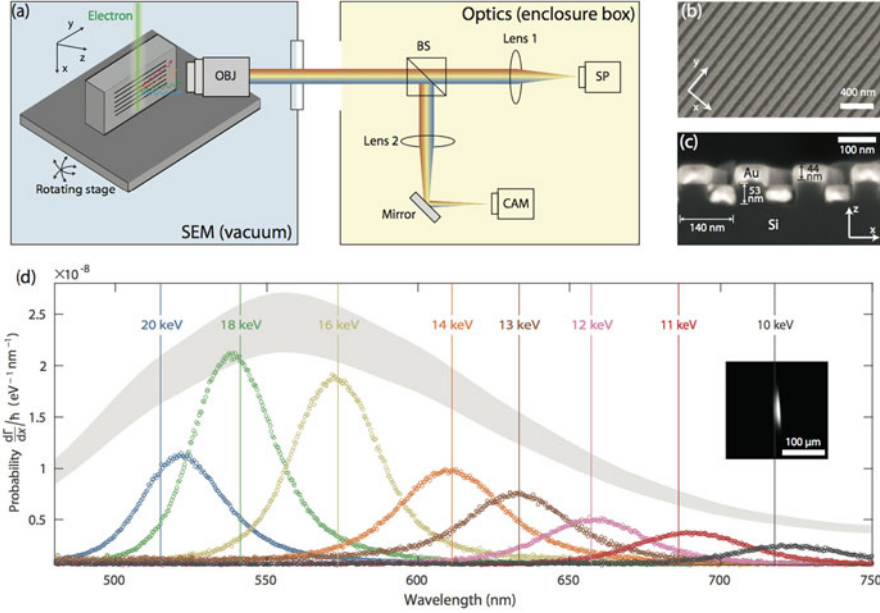
Fig. 2.2 The bounds of Eq. (2.68) dictate upper limits to Smith–Purcell emission rates. (**a–d**) The experiments of Ref. [112] quantitatively confirm that designed metallic gratings can approach the fundamental performance limits (Adapted from Ref. [112])

radiation for electron beams with varying energies, as well as designed gold-on-silicon gratings whose parameters were optimized for maximum response. The key result is shown in panel (d), where the gray region indicates the fundamental bounds, as a function of photon wavelength, with some width to account for experimental uncertainties. The colored data points are quantitatively measured probabilities (with no fitting parameters), showing that both the quantitative values of the bounds are nearly approachable and that the complex wavelength dependence (emerging from an interplay between the material enhancement factor and the optical near fields) correctly captures the response of high-performance designs.

2.4.1.4 Spectral NFRHT

Near-field radiative heat transfer, NFRHT, introduced in Sect. 2.2.5, offers an extraordinary challenge for fundamental limits. It comprises rapidly decaying, large-area, broadband thermal sources for which little has been understood about upper bounds for quite some time. While we tackle the question of broadband enhancements in Sect. 2.4.3, in this section, we describe the recent progress in understanding maximum NFRHT at a single frequency. There are three key results

that we can highlight: channel bounds for planar bodies [57, 60], material-loss bounds [111], and an amalgamation of the two [170, 171].

Channel bounds to NFRHT are described as "Landauer bounds," due to their similarities with Landauer transport. For planar bodies with in-plane translational (and therefor rotational) symmetries, the in-plane wavenumber is a conserved quantity, and the energy flux from one body to another can be decomposed into propagating and evanescent plane-wave channels with no cross-channel scattering. One can decompose the fields emanating from the emitting body into normalized plane-wave modes, insert them into the fluctuation-averaged flux, i.e., the average of the integral $\frac{1}{2} \int_A \mathbf{E} \times \mathbf{H}^* \cdot \hat{\mathbf{n}}$, for separating plane A and normal vector $\hat{\mathbf{n}}$. This results in an expression for the flux rate $\Phi(\omega)$, of Eq. (2.20) and Eq. (2.21), given by

$$\Phi(\omega) = \sum_{j=s,p} \frac{1}{2\pi} \int \frac{d^2\kappa}{4\pi^2} T_j^{12}(\omega, \kappa, d), \qquad (2.69)$$

where κ is the in-plane wave propagation constant (and κ its magnitude), j is a polarization index, k_0 is the free-space wavenumber, and the T_i are "transmission coefficients," which depend on the specific Fresnel reflection coefficients of the two interfaces [60]. This expression has an elegant interpretation: NFRHT is the composition of plane-wave fluxes, each contributing with a weight T_i. Moreover, the coefficients T_i are bounded above by 1, for both the propagating and evanescent waves [57, 60, 172]. Then, if there is a limit to the largest wavenumber across which a nonzero transmission can be achieved, one will have a bound on the maximum spectral RHT.

Hence, it is possible to identify a maximal rate of NFRHT which is given by power transferred with "Landauer" transmission unity over all possible plane waves [57, 172]. While intuitive, however, this bound has two serious drawbacks. The first is that if one literally computes the integral of Eq. (2.69) over all possible waves, the result is infinite, as there are an infinite number of plane-wave channels. Of course one cannot reasonably expect to achieve unity transmission over channels with infinitely large in-plane wavenumbers (as they decay exponentially fast), implying that there must be a maximal channel at which the sum should be terminated. But how to choose this value? One proposal, from Ref. [57], was that the maximal accessible channel should be proportional to $1/a$, where a is the lattice spacing of the material, the reasoning being that beyond this limit, the use of a continuum model of the materials would not be valid. Another proposal, from Ref. [172], is that the maximal accessible channel wavenumber is given by $k_{\max} = 1/d$, where d is the separation between the two bodies, the reasoning being that the exponential decay of the evanescent waves makes it difficult to achieve large transmission beyond $1/d$. Each of the resulting bounds (one from $k_{\max} = 1/a$ and the other from $k_{\max} = 1/d$) has shortcomings: the lattice-spacing-defined bound is extraordinarily high for any reasonable lattice constant, well beyond all other bounds discussed below. And the separation-defined bound is in fact not a true bound: it can

be superseded with reasonable material parameters [111], which in fact do show nontrivial transmission beyond $1/d$. Hence, the two known versions of the bound are either far too large or surpassable.

The second serious drawback of using Eq. (2.69) is that it only applies to planar bodies with translational symmetry in all directions. The use of conservation laws for bounds, discussed next, leads to bounds that can be applied to planar bodies with any patterning while also being tighter than the channel bounds resulting from Eq. (2.69).

The first use of conservation laws for spectral NFRHT bounds appeared in Ref. [111]. The mathematical procedure is sufficiently complex that we will not go through it in detail here, but the intuition can be explained. The idea is to use the global conservation law requiring $P_{\text{abs}} \leq P_{\text{ext}}$ in the spectral NFRHT problem. The difficulty is that the sources are embedded *within* one of the scattering bodies, which leads to divergences if one blindly applies the constraint $P_{\text{abs}} \leq P_{\text{ext}}$. However, the radiative exchange of heat can be decomposed into two subsequent scattering problems, both of which have sources separated from scatterers. In the first step, the incident field is given by the field emanating from body 1 *in the presence of body 1*, with only the second body serving as the scatterer. The absorption in this second body is bounded by the extinction by this second body, which leaves a bound in terms of the second material and the "incident field" emanating from body 1. Of course, we do not know exactly what this field is for any pattern. At this point, however, we can use reciprocity to rewrite the field emanating from body 1 in terms of fields emanating from the free space of body 2's domain, being absorbed by body 1. The constraint $P_{\text{abs}} \leq P_{\text{ext}}$ can be applied to this scattering process again, ultimately yielding a single-frequency, flux-per-area A bound given by [111]

$$\frac{\Phi(\omega)}{A} \leq \frac{1}{16\pi^2 d^2} \frac{|\chi_1|^2}{\text{Im } \chi_1} \frac{|\chi_2|^2}{\text{Im } \chi_2}, \tag{2.70}$$

where d is the separation distance between the two bodies and χ_1 and χ_2 are their optical susceptibilities, respectively. This bound includes two key dependencies: the material enhancement factor $|\chi|^2/\text{Im } \chi$ and a $1/d^2$ dependence arising from the rapidly decaying near fields in the electromagnetic Green's function. The bound of Eq. (2.70) is promising, as it suggests significant possible enhancements of spectral NFRHT, and it is plausible: the actual NFRHT of two planar bodies with equal susceptibilities, on resonance, is given by $\Phi(\omega)/A = 1/(4\pi^2 d^2) \ln\left[|\chi|^4/(4(\text{Im } \chi)^2)\right]$, with nearly identical dependencies as Eq. (2.70), except for the logarithmic dependence on the material enhancement. Can this be overcome, with instead linear enhancements in $|\chi|^2/\text{Im } \chi$? For some materials, the answer is "yes," as shown with computational inverse design in Ref. [173]. More generally, however, such linear enhancements are not generic, and one can further tighten the bound of Eq. (2.70).

Refs. [170, 171] showed that one can tighten the bound of Eq. (2.70) by combining the use of a global conservation law with that of a channel decomposition. If one decomposes the *general* (not specific to translation-symmetric) scattering response into plane waves, and further imposes conservation laws for absorption and

extinction (of the bodies in tandem as well as in isolation), then a long mathematical process leads to a tighter bound. If we define $\mathbb{G}_{0,AB}$ to be the free-space Green's function matrix for sources in body A to measurement points in body B, and g_i the singular values of $\mathbb{G}_{0,AB}$, then the resulting bound is given by Molesky et al. [170]

$$
\Phi(\omega) \leq \sum_i \left[\frac{1}{2\pi} \Theta(\zeta_A \zeta_B g_i^2 - 1) + \frac{2}{\pi} \frac{\zeta_A \zeta_B g_i^2}{(1 + \zeta_A \zeta_B g_i^2)^2} \Theta(1 - \zeta_A \zeta_B g_i^2) \right],
$$
(2.71)

where $\zeta_{A,B} = |\chi_{A,B}|^2 / \operatorname{Im} \chi_{A,B}$. One can see that the expression of Eq. (2.71) has components of both material response (in $\zeta_{A,B}$) and channels (in the g_i factors) in it. Strikingly, in the near-field limit, expression Eq. (2.71) is given by Venkataram et al. [171]
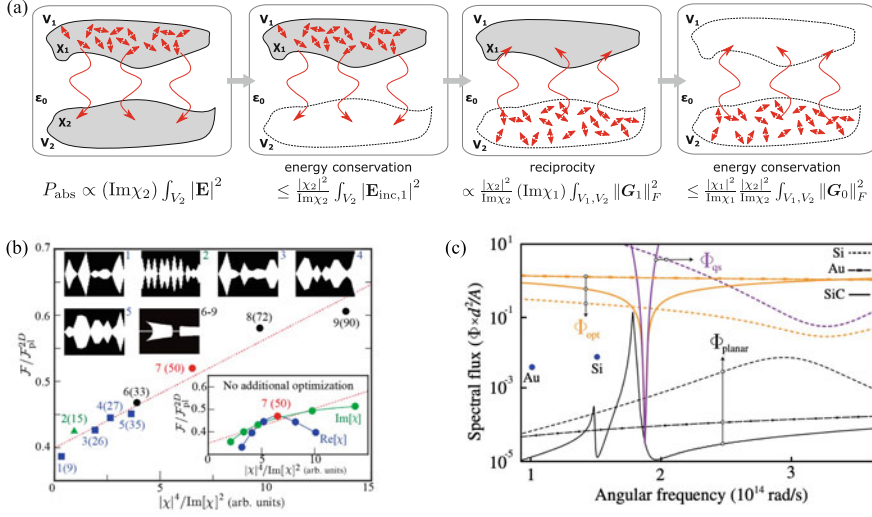
$$
\Phi(\omega) \frac{d^2}{A} \leq \frac{1}{4\pi^2} \ln\left(1 + \frac{\zeta_A \zeta_B}{4}\right)
$$
$$
+ \frac{\Theta(\zeta_A \zeta_B - 4)}{8\pi^2} \left\{ \ln(\zeta_A \zeta_B) + \frac{1}{4} \left[\ln\left(\frac{\zeta_A \zeta_B}{4}\right) \right]^2 - 2\ln\left(1 + \frac{\zeta_A \zeta_B}{4}\right) \right\},
$$
(2.72)

which correctly captures the *logarithmic* material dependence that is seen in planar bodies. This significantly tightens the bound of Eq. (2.70) for plasmonic materials such as silver or gold which have large material enhancement factors $|\chi|^2 / \operatorname{Im} \chi$. The genesis and utility of the bounds of Eqs. (2.70)–(2.72) are illustrated in Fig. 2.3, which contains the derivation of the conservation-law bounds of Eq. (2.70) in Fig. 2.3a, the design of structures showing the material dependence of Eq. (2.70) in Fig. 2.3b, and the more general combination of conservation law and channel-decomposition approach of Eq. (2.72) in Fig. 2.3c.

Generically, it is not possible to find "tighter" single-frequency dependencies than those that arise in Eq. (2.72), as both the distance and material enhancement dependencies are achievable in realistic-material planar designs. The only possible improvements are the coefficient prefactors, as well as the correct material dependence away from the surface-plasmon frequency, suggesting that Eq. (2.72) indeed captures the key trade-offs in single-frequency NFRHT. A key remaining question, then, is what is possible over a broad bandwidth? This question is resolved in Sect. 2.4.3.

2.4.2 All-Frequency Sum Rules

In Sect. 2.3.4, we developed the key elements needed for sum rules: a causal linear response function, an objective that does not involve the conjugate of that function,

$$P_{\text{abs}} \propto (\text{Im}\chi_2) \int_{V_2} |\mathbf{E}|^2 \qquad \leq \frac{|\chi_2|^2}{\text{Im}\chi_2} \int_{V_2} |\mathbf{E}_{\text{inc},1}|^2 \qquad \propto \frac{|\chi_2|^2}{\text{Im}\chi_2} (\text{Im}\chi_1) \int_{V_1,V_2} \|\boldsymbol{G}_1\|_F^2 \qquad \leq \frac{|\chi_1|^2}{\text{Im}\chi_1} \frac{|\chi_2|^2}{\text{Im}\chi_2} \int_{V_1,V_2} \|\boldsymbol{G}_0\|_F^2$$

Fig. 2.3 A collection of bounds on single-frequency near-field radiative heat transfer. (**a**) The approach of Ref. [111] using material loss as the only constraint, exploiting reciprocity to bound the response given that the sources are embedded *within* one of the arbitrarily patterned scattering bodies. (Adapted from Ref. [111].) (**b**) Bounds and designs from Ref. [173] showing the feasibility, in specific regimes, of achieving enhancements proportional to the square of the material enhancement factor $|\chi|^2/\text{Im}\,\chi$. (Adapted from Ref. [173].) (**c**) Tightened bounds from Refs. [170, 171], precluding the possibility of extraordinary response at frequencies away from the surface-polariton frequency of a material of interest (Adapted from Ref. [171])

and certain technical conditions (e.g., sufficient decay). Optical extinction is the prototype example, as the optical theorem prescribes that extinction should be proportional to the imaginary part of the overlap of the incident field with the induced polarization field, a quantity that is analytic (for plane-wave incident fields) in the upper-half plane. Within the past few years [7, 174], it has been realized that there is a near-field analog of extinction: the local density of states, or LDOS. As derived in Sect. 2.2.1, (electric) LDOS is given by the trace of the imaginary part of the (electric) Green's function, evaluated at the source location:

$$\text{LDOS}(\mathbf{x}, \omega) = \text{Im}\,\text{Tr}\left[\frac{1}{\pi\omega}\mathbb{G}(\mathbf{x}, \mathbf{x}, \omega)\right]. \qquad (2.73)$$

The key similarity with extinction is that LDOS is the imaginary part of an amplitude, rather than a squared norm (which depends on the complex conjugate of that amplitude). At first blush, then, it would appear that one can port exactly the derivation used for extinction to derive sum rules for LDOS. However, there are three obstacles that must be overcome.

First, LDOS diverges at high frequencies. Ignoring the effects of a scatterer (which are effectively infinitely far away at infinitely large frequencies), and as seen

in Eq. (2.4), the free-space photon density of states scales as ω^2 as frequency goes to infinity. A diverging LDOS violates the asymptotic-decay requirement of KK relations, prohibiting a sum rule. The resolution, however, is straightforward: one should subtract the free-space LDOS $\rho_0(\omega)$ from the total LDOS, leaving only the scatterer-based contribution $\rho_s(\omega)$:

$$\rho_s(\mathbf{x}, \omega) = \rho(\mathbf{x}, \omega) - \rho_0(\omega) = \operatorname{Im} \operatorname{Tr} \left[\frac{1}{\pi \omega} \left(\mathbb{G}(\mathbf{x}, \mathbf{x}, \omega) - \mathbb{G}_0(\mathbf{x}, \mathbf{x}, \omega) \right) \right]$$

$$= \operatorname{Im} \operatorname{Tr} \left[\frac{1}{\pi \omega} \mathbb{G}_s(\mathbf{x}, \mathbf{x}, \omega) \right], \tag{2.74}$$

where we define $\mathbb{G}_s$ as the scattered-field part of the Green's function. After isolating the scatterer's contribution to the LDOS, one can verify that the "scattered LDOS" indeed decays sufficiently quickly at high frequencies [7]. Hence, this approach of subtracting the free-space LDOS, an approach generalized in "dispersion relations with one subtraction" [160], resolves the first issue of diverging LDOS.

The second issue is that one is not free to arbitrarily choose the pole frequency for a KK relation involving the scattered LDOS. The Green's function itself is finite and generically nonzero at every real frequency, but by definition, the LDOS includes a factor of $1/\omega$, as in Eq. (2.73). (This does not correspond to a divergent LDOS at zero frequency, as the *imaginary* part of the Green's function goes to zero at frequency, but the real part does not generically go to 0.) This function, then, already has a pole at the origin. One could try to move the pole to infinite frequency, for example, by multiplying by $\omega/(\omega - \omega_0)$ and taking the limit as $\omega_0 \to \infty$, but the high-frequency asymptotic behavior of LDOS is quite complicated. Hence, there is likely only a single meaningful sum rule for near-field LDOS, which arises from the intrinsic pole at zero frequency.

The third issue is that the real part of the Green's function diverges, since the source and measurement locations coincide; sum rules relate the integral of the imaginary part to the real part (or vice versa), which leads to the impermissible evaluation of an infinite quantity. (Such an integral *should* diverge; the free-space LDOS increases with frequency, meaning that any integral over all frequencies will of course diverge.) One resolution to this issue was proposed in Ref. [175]: to remove the longitudinal contribution to the Green's function, which removes the singularity and suggests that over all frequencies there can be no net change in spontaneous-emission enhancements. But this removal thereby precludes the possibility for near-to-far-field coupling that is crucial for spontaneous-emission engineering, which is why a conventional refractive-index sum rule is recovered. Instead, it was recognized in Refs. [7, 174] that there is an alternative mechanism for overcoming this obstacle: to subtract out the free-space LDOS term from the total term. The free-space term is the one responsible for the diverging real part, yet the free-space LDOS is exactly known and hence there is no need for a KK relation for that part anyhow. Hence, this obstacle is resolved by the same procedure as the first one, and we can proceed to deriving a scattered-LDOS sum rule.
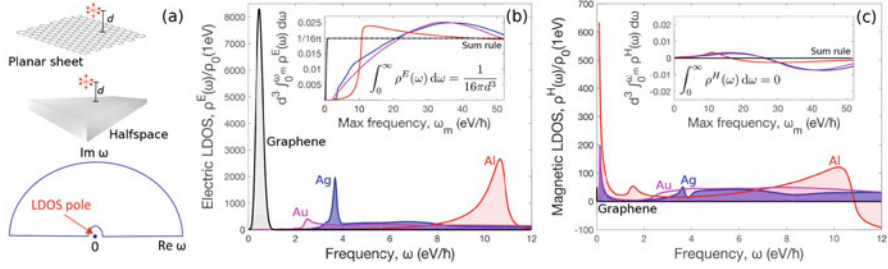
Fig. 2.4 (**a**) Sum rules, derived using the techniques of Sect. 2.3.4 and the contour on the lower left, impose strong constraints on LDOS lineshapes. (**b**) Electric LDOS of various material half-spaces and 2D sheets, with different resonance peaks and bandwidths. The inset, however, shows that the integral converges to identical values for each scenario. (**c**) Similarly with magnetic LDOS, whose sum rule is now zero. The sum rules are for the scattered-field contributions to the LDOS, which can be negative at frequencies where spontaneous emission is suppressed by the presence of a scatterer (Adapted from Ref. [7])

The hemispherical contour (with hemispherical bump at the origin), in tandem with the same Cauchy-residue arguments for far-field sum rules in Sect. 2.3.4, leads to a sum rule for $\rho - \rho_0$ analogous to the far-field case [7]:

$$\int_0^\infty \rho_s(\omega, \mathbf{x}) \, d\omega = \frac{1}{2} \operatorname{Re} \operatorname{Tr} \mathbb{G}_s(\mathbf{x}, \mathbf{x})\big|_{\omega=0} = \alpha_{\text{LDOS}}. \tag{2.75}$$

Now we have connected the all-frequency scattered-field component of electric LDOS to its electrostatic Green's function. Is that informative? It turns out to be quite informative, because there are near-field "domain monotonicity" theorems [7] that ensure that this shape-dependent Green's function term is bounded above by its form in any enclosure, and we can choose high-symmetry enclosures where it has a simple analytical form. For example, for a planar half-space, the near-field electrostatic constant is simply

$$\alpha_{\text{LDOS,plane}} = \frac{1}{16\pi d^3} \left[\frac{\varepsilon(0) - 1}{\varepsilon(0) + 1} \right], \tag{2.76}$$

where $\varepsilon(0)$ is the zero-frequency (electrostatic) permittivity. For conductive materials whose permittivity diverges at zero frequency, the corresponding fraction in Eq. (2.76) is simply 1, which can also be used as a general bound for any material. Notably, for the *magnetic* LDOS above an electric material, the right-hand side of the counterpart to Eq. (2.76) is *zero*: the scattering contribution to the magnetic LDOS must average out to zero (i.e., it provides suppression and enhancement of the free-space LDOS in equal amounts).

An example of the utility of the LDOS sum rule is given in Fig. 2.4. The electric LDOS is shown for three typical metals: gold (Au), silver (Ag), and aluminum (Al), as well as for a single graphene sheet (with Fermi level 0.6 eV). These four

systems show LDOS peaks at quite different frequencies, from below 1 eV to beyond 10 eV, with very different quality factors leading to quite different "spreads" in their spectral response. Yet as is made clear by the inset of Fig. 2.4, the integrated response is exactly equal for each of these systems, as must be true from Eq. (2.76) (the material constant α for each system is exactly 1). Sum rules illuminate unifying principles that must apply across seemingly disparate systems.

2.4.3 Finite, Nonzero Bandwidth

The techniques of the previous two sections apply to single-frequency and all-frequency scenarios. In this section, we probe an intermediate regime: finite, nonzero bandwidth. Techniques that work for any arbitrary bandwidth would be tantalizingly powerful, as they would incorporate the single- and all-frequency results as asymptotic limits of a more general theory. Yet the techniques of the previous section would seem incapable of extension to nonzero, finite bandwidths: there is no single scattering problem for which power-conservation laws can be imposed, nor can the contour integrals of the sum-rule approaches be easily modified to a finite bandwidth. In this section, we describe two recently developed approaches to tackle finite-bandwidth bounds: first, transforming bandwidth-averaged response to a complex frequency (largely following Ref. [7]), and second, identifying an oscillator-based representation of any scattering matrix (largely following Ref. [176]).

2.4.3.1 Complex-Frequency Bounds

Ref. [7] recognized an intermediate route that utilized both techniques in one fell swoop. The idea can be summarized succinctly: finite-bandwidth average response can be transformed to a scattering problem at a single, *complex*-valued frequency, where quadratic constraints analogous to power conservation can be imposed. The complex frequency accounts for bandwidth, while the power-conservation analog imposes a finite bound. We now develop this intuition mathematically.

To compute the bandwidth average of a response function such as LDOS, one must define a "window function" that encodes the center frequency, the bandwidth, and the nature of the averaging. A common choice is a linear combination of step functions, but this choice turns out to be mathematically treacherous. A simple (and mathematically serendipitous) choice is a Lorentzian function. Uses of tailored window functions for bandwidth averaging were first proposed in Refs. [17, 177]; in the first, bandwidth-averaged extinction was analyzed for scaling laws for optical cloaking, while in the second, they were used to regularize the computational inverse design of maximum LDOS. Our quantity of interest, the frequency-averaged LDOS, $\langle \rho \rangle$, can be written as [7]

$$\langle \rho \rangle = \int_{-\infty}^{\infty} \rho(\omega) H_{\omega_0, \Delta\omega}(\omega) \, d\omega, \tag{2.77}$$

where $H_{\omega_0, \Delta\omega}(\omega)$ is the Lorentzian window function:

$$H_{\omega_0, \Delta\omega}(\omega) = \frac{\Delta\omega/\pi}{(\omega - \omega_0)^2 + (\Delta\omega)^2}, \tag{2.78}$$

where ω_0 is the center frequency and $\Delta\omega$ is the bandwidth of interest. In Eq. (2.77), we define the frequency integral from $-\infty$ instead of 0 for smoothness; typically, the window function will be narrow enough to render this difference negligible; conversely, in the all-frequency limit, the symmetry of the LDOS around zero frequency ensures we are working with the correct quantity. We are interested only in the near-field enhancements of ρ, so we will drop the free-space LDOS, as was useful in the sum-rule section to avoid spatial and spectral divergences. Then, consider the integral of Eq. (2.77): it already covers the entire real line, and we can imagine adding to it a hemispherical contour in the UHP that will contribute infinitesimally. Then the integral is a closed contour, and we can use complex-analytic techniques based on the analyticity of the integrand and the locations of the poles of the integrand. The integrand is not analytic, but the LDOS can be written as $\rho(\omega) = \mathrm{Im}\, s(\omega)$, where $s(\omega)$, proportional to the trace of the imaginary part of the scattered component of the Green's function, *is* analytic. Taking the imaginary part outside the integral, the remainder of the integrand of Eq. (2.77) has two poles away from the lower-half plane: one at zero, thanks to the $1/\omega$ term in the LDOS, and a second at $\omega_0 + i\,\Delta\omega$. Then, a few lines of algebra give the frequency average of $\rho(\omega)$ as [7]

$$\langle \rho \rangle = \mathrm{Im}\, s(\omega_0 + i\,\Delta\omega) + 2H_{\omega_0, \Delta\omega}(0)\alpha_{\mathrm{LDOS}}. \tag{2.79}$$

The second term comes from the contribution of the sum rule at a given frequency and ensures that the ultimate expression will give the sum rule in the asymptotic limit $\Delta\omega \to \infty$. Here, for simplicity and pedagogy, we will assume a sufficiently narrow bandwidth that the second term can be ignored. (It can always be reintroduced in the final expression.) The first term is the imaginary part of the LDOS scattering amplitude, evaluated at the *complex* frequency $\tilde{\omega} = \omega_0 + i\,\Delta\omega$. What is the largest this term can be?

To bound the complex-frequency term, we can develop a generalization of the real-frequency conservation-law approach. In Ref. [7], we developed such a generalization via a somewhat complicated line of differential-equation reasoning; here, we develop a simpler (but no less general) integral-equation form. The starting point is the complex-valued integral equation:

$$\left[\mathbb{G}_0(\tilde{\omega}) + \xi(\tilde{\omega})\right] \mathbf{p}(\tilde{\omega}) = -\mathbf{e}_{\mathrm{inc}}(\tilde{\omega}), \tag{2.80}$$

where we have momentarily included all frequency arguments to emphasize that Eq. (2.80) is evaluated at the complex frequency $\tilde{\omega}$. Next, we will multiply on the left by $\mathbf{p}^\dagger/\tilde{\omega}$, and take the imaginary part of the entire equation, to arrive at

$$\mathbf{p}^\dagger \left\{ \mathrm{Im}\left[\frac{\mathbb{G}_0}{\tilde{\omega}} + \frac{\xi}{\tilde{\omega}} \right] \right\} \mathbf{p} = \mathrm{Im}\left[\left(\frac{\mathbf{e}_{\mathrm{inc}}}{\tilde{\omega}} \right)^\dagger \mathbf{p} \right], \tag{2.81}$$

This equation can be regarded as a complex-valued extension of the real-valued, global conservation law of Eq. (2.25). In particular, the two terms on the left are both positive semidefinite, as can be proven by causality (cf. Sec. IX of the SM of Ref. [105]). To remove the shape dependence and focus on the material dependence, then we can drop the first term on the left-hand side of Eq. (2.81) and rewrite this equation as an inequality:

$$\mathbf{p}^\dagger \left[\mathrm{Im}\left(\frac{\xi}{\tilde{\omega}} \right) \right] \mathbf{p} \leq \mathrm{Im}\left[\left(\frac{\mathbf{e}_{\mathrm{inc}}}{\tilde{\omega}} \right)^\dagger \mathbf{p} \right], \tag{2.82}$$

Equation (2.82) imposes a constraint on the strength of the complex-frequency polarization field that enters the near-field scattering amplitude $s(\tilde{\omega})$. The exact expression for the scattering amplitude is $s(\tilde{\omega}) = \frac{1}{\pi\tilde{\omega}} \mathrm{Tr}\, \mathbb{G}_0(\mathbf{x}, \mathbf{x}, \tilde{\omega})$. One can maximize the imaginary part of this amplitude subject to the constraint of Eq. (2.82) by exactly the procedure outlined in Sec. IX of the SM of Ref. [7]; doing so, one arrives at a simple result (remembering that we have dropped the sum-rule term):

$$\langle \rho \rangle \leq \frac{1}{\pi} \frac{|\chi(\tilde{\omega})|^2}{\mathrm{Im}[\tilde{\omega}\chi(\tilde{\omega})]} \mathbf{e}_{\mathrm{inc}}^\dagger \mathbf{e}_{\mathrm{inc}}. \tag{2.83}$$

As a reminder, the inner product of the incident field with itself is a volume integral of the square of the incident fields. The deep near field is dominated by the most rapidly decaying term in the incident fields; integrating only this contribution at the complex frequency gives $\mathbf{e}_{\mathrm{inc}}^\dagger \mathbf{e}_{\mathrm{inc}} = \frac{1}{16\pi d^3}$, where we have taken the arbitrary scattering body to fit in a half-space enclosure separated from the source by a distance d. Inserting this expression into the inequality, and normalizing by the free-space LDOS evaluated at $|\tilde{\omega}|$, we finally have a bandwidth-averaged bound [7]:

$$\frac{\langle \rho \rangle}{\rho_0(|\tilde{\omega}|)} \leq \frac{1}{8|k|^3 d^3} f(\omega), \tag{2.84}$$

where $f(\omega)$ is the bandwidth-averaged generalization of the material enhancement factor (discussed at real frequencies in Sect. 2.4.1.1):

$$f(\omega) = \frac{|\tilde{\omega}\chi|^2}{|\tilde{\omega}| \, \mathrm{Im}\,(\tilde{\omega}\chi)}. \tag{2.85}$$

The material enhancement factor of Eq. (2.85) is slightly simpler than that of Ref. [7], thanks to our use of the simpler integral-equation constraint of Eq. (2.81).

The bound of Eq. (2.84) is the key result: the bandwidth-averaged LDOS has an upper bound that is similar to that of the single-frequency LDOS, but reduced by the presence of a complex frequency. This reduction is significant for low-loss materials, for which $\text{Im}\,\chi$ might be quite small, in which case $\text{Im}(\tilde{\omega}\chi) \approx (\Delta\omega)\chi$, wherein the bandwidth effectively provides the relevant loss. There is also an additional broadening due to dispersion, as χ is evaluated at the complex frequency $\tilde{\omega}$, at which $\text{Im}\,\chi$ will generally be larger. (There is *another* additional term in the more general version of the bound of Eq. (2.84) that exponentially decays with bandwidth, which we excluded for simplicity.) Hence, the bound of Eq. (2.84) has three properties that are quite theoretically pleasing. First, in the single-frequency limit, it asymptotically approaches the previously derived single-frequency bound. Second, in the all-frequency limit, it asymptotically approaches the previously derived sum rule. And, finally, in the nonzero- and finite-bandwidth regime, it intermediates between the two, with a smaller average response than the single-frequency bound, and a smaller total integrated response than the sum rule. This approach was extended to CDOS and NFRHT in Ref. [7], with similar features emerging. One interesting comparison point is to Ref. [178], which examined optimal materials for planar NFRHT designs. Unlike the power–bandwidth bounds, which increase with electron density and decrease with material loss, Ref. [178] found that the key material parameters in planar systems are simply the (ideally small) frequency at which surface polaritons are strongest, and the bandwidth over which they are strong. This finding has been experimentally corroborated [179], and it emerges theoretically in the more general NFRHT bounds of the next subsection.

Ref. [7] probed the feasibility of approaching the upper bounds in certain prototypical systems. Four key results were identified. First, for center frequencies close to the surface-plasmon frequencies of metals, planar systems supporting such plasmons are able to closely approach the bounds across a wide range of bandwidths. Second, double-cone (bowtie-antenna-like) antennas show a performance that can closely approach (nearly within 2X) their bounds across a wide range of bandwidths, for center frequencies coincident with their resonant frequencies. Third, these bounds were the first to enable systematic comparison of dielectric- and metal-based systems. Unlike the single-frequency case, the complex-frequency material enhancement factor does not diverge for lossless dielectrics (at nonzero bandwidth), which enables predictions of the center frequencies and bandwidths at which metals can be categorically superior to dielectrics, and vice versa. Finally, these bounds also enabled predictions of when 2D materials can be superior to bulk materials, and vice versa. The results highlight the power of fundamental limits more generally: they enable a high-level understanding of the landscape of a given physical design problem, identifying the material and architectural properties that really matter.

The "power–bandwidth" approach of Ref. [7] was recently generalized in Ref. [180]. Notice that the constraint of Eq. (2.81) is a global conservation law for real power; at the time that Ref. [7] was published, the reactive global conservation

law, as well as the local-conservation-law approach, had not yet local-conservation-law approach had not yet been invented. Ref. [180] remedies this gap and shows that for dielectric scatterers, the use of additional conservation laws can significantly improve the resulting bounds. There is an interesting interplay between the quality factor of the sources and the bandwidth of interest, and there are useful semi-analytical bounds that can be derived from the global conservation laws applied to large-scale devices. Moreover, inverse-design structures are shown to come quite close to the improved complex-frequency, bounds.

2.4.3.2   Oscillator-Representation Bounds

An alternative to the complex-frequency approach to bandwidth averaging was very recently proposed in Ref. [176]. We will briefly summarize the (detailed) mathematical apparatus developed and highlight the key result for our purposes: a new, nearly tight bound for bandwidth-averaged NFRHT.

Before delving into scattering bodies, consider the bulk optical susceptibility of a material. It is known that the response of an isotropic passive material can be written as a linear combination of Drude–Lorentz oscillators:

$$\chi(\omega) = \sum_i \frac{\omega_p^2}{\omega_i^2 - \omega^2 - i\gamma\omega} c_i, \tag{2.86}$$

where ω_p is the "plasma frequency" of the material (related to its electron density [161, 181]), ω_i are the oscillator frequencies, γ are infinitesimal oscillator loss rates, and c_i are "oscillator strengths" that sum to unity, thanks to the sum rule of Eq. (2.50) discussed in Sect. 2.3.4. Often this representation is derived in single-electron quantum-material frameworks [161], but it applies more generally as a consequence of causality and passivity. (The technically rigorous mathematical statement uses the theory of Herglotz functions [182].) Any linear material's susceptibility must conform to the Drude–Lorentz linear combination of Eq. (2.86), perhaps not with a small number of oscillators (it is well known that effects such as inhomogeneous broadening lead to other lineshapes, such as the "Voigt" line-shape [183]), but with sufficiently many oscillators. It may seem counterintuitive to work with a representation that may need 1,000, or even 100,000 oscillators, instead of a different model with fewer parameters. From an optimization perspective, however, this is not correct. In the Drude–Lorentz representation of Eq. (2.86), the only degrees of freedom are the c_i coefficients, and the susceptibility is *linear* in these degrees of freedom. In many scenarios, large linear optimization problems are significantly easier to solve (sometimes even analytically) than large, nonlinear (and nonconvex) optimization problems.

Causality and passivity create three key ingredients that together lead to the Drude–Lorentz representation of Eq. (2.86): a Kramers–Kronig relation, a sum rule, and positivity of the imaginary part of the susceptibility. The exact sequence of

transforming those ingredients to the Drude–Lorentz representation is detailed in Ref. [164]. One intuitive description is that the imaginary part of the susceptibility is a positive quantity and can be discretized into coefficients at many discrete frequencies along the real axis. Passivity implies that these coefficients are real, while the sum rule implies that their sum is constrained. Finally, the Kramers–Kronig relation guarantees that the imaginary parts of the susceptibilities are the *only* degrees of freedom; the real parts are entirely determined by the imaginary parts. Compiling the mathematical details of these steps leads to Eq. (2.86), which is a relation that many find intuitive, thanks largely to the fact that it can be derived in single-electron quantum mechanics.

The key idea of Ref. [176] is that there is a *wave-scattering operator* that exhibits nearly identical mathematical properties to material susceptibilities. This operator is the "$\mathbb{T}$" matrix. The $\mathbb{T}$ matrix is a scattering matrix that relates the polarization field induced in any scattering body to the incident fields impinging upon it [184]:

$$\mathbf{P}(\mathbf{x}, \omega) = \int_V \mathbb{T}(\mathbf{x}, \mathbf{x}', \omega) \mathbf{E}_{\text{inc}}(\mathbf{x}', \omega) \, d\mathbf{x}', \qquad (2.87)$$

or, in vector notation:

$$\mathbf{p} = \mathbb{T}\mathbf{e}_{\text{inc}}. \qquad (2.88)$$

The $\mathbb{T}$ matrix is a *causal linear response function*, as the polarization field at $\mathbf{x}$ cannot be excited before the incident field exciting it reaches $\mathbf{x}'$. Just as causality implies a Kramers–Kronig relation for material susceptibilities, it was recognized in Ref. [176] that causality implies a Kramers–Kronig relation for $\mathbb{T}$ matrices. Sum rules come from the low- and high-frequency asymptotic behavior of Kramers–Kronig relations, and the $\mathbb{T}$ matrix satisfies a matrix-valued analog of the f-sum rule for material oscillator strengths. Finally, just as passivity implies that the imaginary parts of susceptibilities are positive, it similarly implies that the anti-Hermitian part of the $\mathbb{T}$ matrix is positive semidefinite. Together, these three ingredients imply a matrix-valued analog of Eq. (2.86) for any $\mathbb{T}$ matrix:

$$\mathbb{T}(\omega) = \sum_i \frac{\omega_p^2}{\omega_i^2 - \omega^2 - i\gamma\omega} \mathbb{T}_i, \qquad (2.89)$$

where the Drude–Lorentz parameters are exactly the same as in Eq. (2.86) and the $\mathbb{T}_i$ are now matrix-valued coefficient degrees of freedom. The exact expression of Eq. (2.89) is for the case of reciprocal materials; in nonreciprocal terms, there is an extra term that makes the calculations more tedious but has no effect on most applications of interest. Analogous to the constraints on material oscillator strengths, passivity and the $\mathbb{T}$-matrix sum rule lead to constraints on the $\mathbb{T}_i$:

$$\sum_i \mathbb{T}_i = \mathbb{I}, \qquad \mathbb{T}_i \geq 0, \tag{2.90}$$

where $\mathbb{I}$ is the identity matrix. Equation (2.89), and its nonreciprocal analog, must hold for *any* linear electromagnetic scattering process. Even in scattering processes with complex interference phenomena, Fano resonances, etc., $\mathbb{T}(\omega)$ must exhibit lineshapes consistent with Eq. (2.89), which is shown in Ref. [176] to reveal surprising structure even in typical scattering problems.

Our interest in this chapter, however, is in fundamental limits, so we will focus on the utility of Eq. (2.89) to identify upper bounds in the application considered in Ref. [176], which is NFRHT. The approach in the paper requires a dozen or so mathematical steps explained in Sec. IX of the SM of Ref. [176]; the key is to transform the problem from one of thermal sources inside the hot body radiating power to the cold one to one of incoherent sources *between* the bodies radiating back to the emitter body. There are various other key steps, such as an appropriate renormalization of the point sources between the bodies. Ultimately, the culmination is the following: NFRHT is rewritten in terms of the total $\mathbb{T}$ matrix of the collective bodies, at which point the representation of Eq. (2.89) is inserted. Then, the entire frequency dependence of the problem is given by the collective products of the Drude–Lorentz oscillators and the Planck function, whose integrals can be determined analytically. Then one is left with a linear summation of given coefficients multiplying the unknown $\mathbb{T}_i$ degrees of freedom. The optimization over all possible $\mathbb{T}_i$, subject to the constraints of Eq. (2.90), has many unknowns, but can be done *analytically*, leading to a simple yet completely general bound on thermal HTC:

$$\mathrm{HTC} \leq \beta \frac{T}{d^2}, \tag{2.91}$$

where T is the temperature, d is the separation, and $\beta \approx 0.11 k_B^2/\hbar$ is a numerical constant. Equation (2.91) is an unsurpassable limit that captures the key constraints imposed on every scattering $\mathbb{T}$ matrix. Strikingly, despite the relative simplicity of the approach, it offers the tightest bounds on NFRHT to date, only a factor of 5 larger than the best theoretical designs [178]. Previous approaches suggested strong material dependencies, with bounds that increased with electron density, whereas planar designs show the reverse trend. In this bound, use of a low-frequency sum rule in the $\mathbb{T}$-matrix representation leads to an electron-density-independent bound. Moreover, the optimization over $\mathbb{T}_i$ predicts precisely the same optimal peak transfer frequency as the best designs [176].

There are two sets of relaxations used to arrive at the bound of Eq. (2.91): first, beyond the representation theorem, no other Maxwell-equation constraints are imposed. Hence, the optimal $\mathbb{T}_i$ may not actually be physically realizable. Potentially, one could impose such constraints exactly by the local-conservation-law approach discussed above. Second, the heat transfer process is relaxed to the emission of the sources between the bodies into both the source and emitter,

whereas the exact expression is the *difference* between the radiation into the emitter and receiver bodies. The latter relaxation leads to a linear dependence on $\mathbb{T}(\omega)$, as opposed to the quadratic dependence in the exact expression. It may be possible to optimize over the exact quadratic expression using manifold optimization techniques [176, 185, 186]. Tightening these relaxations may lead to a further tightening of the bound. Conversely, they may lead to the same bound, and improved design techniques [187] may identify structures that can achieve them.

2.4.4   Mode Volume

In this final section, we turn to the question of bounds on mode volume. Mode volume is a very different response function than any of those previously considered, as it is a property of an eigenfunction rather than a scattering quantity. There is no incident field in the definition of a mode volume, and hence the power-conservation and causality-based approaches of the previous sections are not immediately useful. In this section, we describe a method for bounding minimum mode volumes based on the optimization-theoretic notion of duality.

In optimization theory, the **dual** of an optimization problem is a second optimization problem, related to but distinct from the original, "primal" optimization problem [151]. The dual problem is formed by incorporating all constraints into the Lagrangian of the original optimization problem, introducing Lagrange multipliers as coefficients of the constraints, and optimizing out the primal variables, leaving only the Lagrange multipliers as degrees of freedom. An equivalent interpretation is that if one interprets a generic minimization optimization problem as the *minimax* of a Lagrangian, the dual problem is the *maximin* of the same Lagrangian. The dual program has two properties that can be quite useful for optimization and bounds: it is always a concave maximization problem (equivalent to a convex minimization problem, and therefore efficiently solvable by standard convex-optimization techniques), and its maximum is guaranteed to be a lower bound for the original, primal, minimization problem.

For many optimization problems, the dual cannot be expressed in a simple form; even among those problems for which it has a simple expression, it often has the trivial solution $-\infty$ as its maximum, giving a trivial lower bound. Ref. [188] showed that a very special class of electromagnetic design problems have a nontrivial, semi-analytical dual problem. In particular, for design problems in which the objective function to be minimized is the norm of a difference between the electric field $\mathbf{E}$ and some target field $\mathbf{E}_{\text{target}}$,

$$\mathcal{F} = \|\mathbf{E} - \mathbf{E}_{\text{target}}\|^2, \tag{2.92}$$

then one can impose the full Maxwell-equation constraints and identify a nontrivial, semi-analytical dual problem. One might suspect that objectives of the form of

Eq. (2.92) might be quite common: after all, a focusing metalens could have a target field that matches an Airy beam along a focal plane, a surface-pattern design intended to maximize spontaneous-emission enhancements could target the field at the location of the dipole, and so forth. But these cases do not work for the expression of Eq. (2.92): for a nontrivial dual problem, the field $\mathbf{E}_{\text{target}}$ must be specified *at every spatial point of the entire domain*. This includes, for example, the points within the scatterer, the points within any PML regions, etc. Knowing a target field at a single point, or on a focal plane, is not sufficient. And it is hard to think of any application in which we know the target field across the entire domain.

It turns out, however, that mode-volume minimization can be reformulated to target an objective specified over the entire domain. Mode volume, as specified in Eq. (2.22), is given by the integral of the field energy over all space divided by the field energy at a single point. Typically, the integral is treated as a normalization constant (taken to equal 1), and maximization of the field energy at a single point is the key objective. In Ref. [189], it was recognized that this convention could be reversed: the field energy at the point of interest can be fixed as a normalization constant, equal to 1, while minimizing the integral of the field energy can be the objective. Such an objective is exactly of the form of Eq. (2.92), with a target field of 0 everywhere! Physically, this makes intuitive sense: a minimum mode volume tries to minimize the field energy at every point, except for the "origin" of interest; everywhere else, it wants to drive the field as close to a target of 0 as possible.

Given this transformation, and a few others described in Ref. [189], one can use the formulation of Ref. [188] to specify a dual program for the mode-volume minimization problem. The solutions of this dual program can be formulated with the modeling language CVX [190] and solved with Gurobi [191], and those solutions represent fundamental lower bounds on the mode volume, given only a designable region and a refractive index of the material to be patterned.

First, the 2D TE case encapsulates scalar-wave physics: without vector fields, there are no field discontinuities across boundaries that can be responsible for large field amplitudes in "slot-mode" configurations [16, 18, 19]. There is also no near field for scalar waves, in the sense of large nonpropagating fields that culminate in a singularity at the location of a point source. In this case, the argument for a trivially small mode volume near a perfectly sharp tip *fails*: the lack of a singularity means that one cannot drive the field at the location of the source arbitrarily high. If there is no sharp-tip enhancement (as we will see), then dimensional arguments would require mode volume to scale with the square of the wavelength (in 2D), restoring the notion of a "diffraction-limited" mode volume. The only question, then, is the value of the coefficient of the squared wavelength. The duality-computed bounds confirm indeed that below some separation distance d, the mode-volume bounds asymptotically flatten out, to a small fraction of the square wavelength. This bound depends only on the available refractive index of the designable region and has been closely approached by inverse-designed structures [17, 189].

The 2D TM case is fundamentally different: sharp field discontinuities occur across material boundaries, and singularities in the near field of point sources imply the possibility for zero mode volume unless fabrication constraints, or similarly a

nonzero source–scatterer separation distance, are enforced. In this case, the duality-based approach finds quite different scaling: the 2D TM mode-volume bounds scale as d^2, where d is the relevant source–scatterer distance (or sharp-tip radius of curvature), with no dependence on the wavelength. Intriguingly, this scaling is faster than the typical structure used for mode-volume minimization: a "bowtie antenna" [18, 19], whose optimal mode volume appears to scale only linearly with d (and hence linearly with wavelength, λ, as well). In Ref. [189], it is shown that inverse-designed structures appear to exhibit mode volumes that scale roughly as $d^{1.4}$, faster than the linear scaling of bowtie antennas but not quite as fast as the duality-based bound. At smaller length scales, these differences can be dramatic. For minimum feature sizes $d \approx 0.01\lambda$, the inverse-design curve falls about 5X below the bowtie-antenna curve, which itself is 40X above the mode-volume bound. Resolving this gap, either through identifying better designs or by identifying tighter bounds, could lead to significant reductions in mode volume through near-field engineering.

2.5 Summary and Looking Forward

Near-field optical response can require significant mathematical machinery, and the techniques to bound them even more so. We were careful above to give correct and sometimes nearly complete mathematical descriptions. Here, we can give a high-level summary of three of the prototypical response functions and application areas covered:

- LDOS, arguably the most important near-field response function, has single-frequency bounds that scale as $1/d^2$ and $|\chi(\omega)|^2/\operatorname{Im}\chi(\omega)$ [106]. This bound can be achieved at the surface-plasmon frequency of a given material; away from that frequency, inverse designs have shown good performance that can be relatively close to the bound, but generally it is also true that tighter bounds can be computed by using additional constraints. A sum rule is known for all-frequency LDOS [7, 174], which depends on the separation but *not* on the material; over finite bandwidth, bounds similar to the single-frequency expression can be found, albeit evaluated at the complex frequency. Again, these bounds are nearly achievable when the frequency range is centered around the surface-plasmon frequency of a material, but can be tightened in other scenarios (e.g., dielectric materials) [115]. The key open questions around LDOS are twofold: first, is there an analytical or semi-analytical bound that can be derived that is nearly achievable across all frequencies? And can one identify achievable bounds for only the *radiative* part of the LDOS, i.e., that fraction of power that is emitted to the far field?
- Near-field radiative heat transfer is one of the most technically challenging areas of near-field optics, both experimentally and theoretically, but an abundance of work makes it perhaps the area where we have the best understanding of what is possible. For planar bodies, there are simple and powerful transmission expres-

sions for NFRHT [57, 60], as well as an understanding of the optimal materials that lead to the largest response [178, 192, 193]. At a single frequency, semi-analytical bounds have been derived [171] that scale as $1/d^2$ with separation distance and *logarithmically* with $|\chi(\omega)|^2/\operatorname{Im}\chi$, both dependencies of which are exhibited by planar structures. Finally, when averaging against the Planck function to account for the thermal nature of the radiation, the recently developed oscillator theory of $\mathbb{T}$ matrices [176] enables a bound proportional only to $1/d^2$ and $k_B^2 T/\hbar$, with no material dependence. This bound can be approached within a factor of 5 by the best theoretical designs, showing a comprehensive understanding of what is possible in NFRHT, and the materials and structures needed to achieve that performance. One interesting open question is how this bound changes when one of the bodies must have a bandgap, as is required, for example, in thermophotovoltaics.

- Finally, mode volume is quite different from the other response functions considered above. It is a property of an eigenmode, instead of a scattered field, and hence some of the techniques based on power conservation do not lead to useful bounds in this case. The only approach we know of that leads to useful bounds relies on the *duality* technique of optimization theory. The most important question surrounding mode volume is how it scales with minimum feature size d. Ideally, it would scale as d^n, where n is the dimensionality of the system (either 2D or 3D), with no dependence on wavelength; this scaling would lead to the largest enhancements at highly subwavelength feature sizes. Certainly, such scaling is possible with plasmonic structures, but plasmonic structures are too lossy, and the concept of mode volume itself must be modified for plasmonic mode volume [25]. The question, then, is the optimal scaling for dielectric materials. Interestingly, the duality-based bounds of Ref. [189] suggest exactly d^n scaling. However, bowtie-antenna structures show d^{n-1} scaling, while inverse designs appear to show a scaling between these two. Hence, progress has been made on this crucial question, but it is still not fully resolved: what is the best possible scaling of mode volume with minimum feature size?

The theory of fundamental limits to near-field optical response is now sufficiently rich to be summarized in a book chapter, as we have done here. But the story is not complete: as we have seen in numerous examples, including the three above, there are still many response functions, material regimes, and frequency ranges at which there are gaps between the best known device structures and the best known bounds. Many of the bound techniques described herein have only been discovered in the past few years, and there are likely still significant strides to be made. The optical near field continues to offer a fertile playground for theoretical discovery, experimental demonstration, and new devices and technological applications.

2.6   Appendix: Complex Analysis for Sum Rules

Here we provide a brief summary of the basic rules of complex analysis, and how they are derived, emphasizing the key results relevant to sum rules. More expansive discussions of these ideas can be found in any good complex-analysis textbook.

First, we start with the definition of **complex differentiable**: a function f is complex differentiable if the limit

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h} \tag{2.93}$$

exists for h along *any path* in the complex plane. The equality along any path is a very strong constraint and leads to the Cauchy–Riemann conditions on the derivatives of the real and imaginary parts of f. A function that is complex differentiable at every point on some domain Ω is **holomorphic** on Ω. A major theorem of complex analysis is that all such functions are also **complex analytic** (which means they have a convergent power series in a neighborhood of every point in Ω). From complex differentiability, it is a straight path to **Cauchy's integral theorem**: for f holomorphic on Ω, and a closed contour γ in Ω,

$$\oint_\gamma f(z)\, \mathrm{d}z = 0, \tag{2.94}$$

which can be proven by setting $f = u + iv$, $\mathrm{d}z = \mathrm{d}x + i\mathrm{d}y$, applying Green's/Stokes theorem, and using the Cauchy–Riemann conditions.

An important technique for integrals over open contours is **contour shifting**: if γ and $\tilde{\gamma}$ are contours with the same endpoints, then

$$\int_\gamma f(z)\, \mathrm{d}z = \int_{\tilde{\gamma}} f(z)\, \mathrm{d}z. \tag{2.95}$$

This follows directly from reversing the second contour, combining it with the first to make a closed contour, and applying Cauchy's integral theorem. Contour shifting is common in Casimir physics, for example, where the standard transformation is a "Wick rotation" from the positive real axis to the positive imaginary axis [194].

One can use contour shifting to prove an important integral formula. Consider the closed-contour integral $\oint_\gamma \frac{f(z)}{z - z_0}\, \mathrm{d}z$, where f is holomorphic on γ, but there is now a singularity in the integrand. For any arbitrary closed contour γ, one can follow the prescription of Fig. 2.5: first, make a tiny perforation in the contour and then use that perforation to shift to a modified contour that comprises two straight lines (whose integrals cancel by directionality) and a tiny circle at the origin. On the tiny circle, we can write $f(z) \approx f(z_0)$. On the circle, $z = z_0 + \varepsilon e^{i2\pi t}$, for t from 0 to 1, where ε is the radius of the circle on $\tilde{\gamma}$, such that
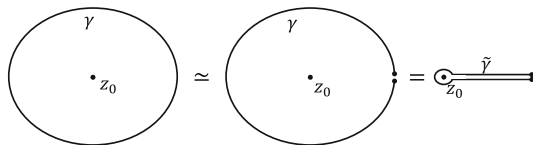
Fig. 2.5 Equivalent contours –the latter two by contour shifting– simplify the integration of *any* closed contour around a singularity (left) to that of a circle arbitrarily close to the singularity (right)

$$\oint_{\tilde{\gamma}} \frac{f(z)}{z - z_0} \approx f(z_0) \oint_{\tilde{\gamma}} \frac{1}{z - z_0} \, dz$$

$$= f(z_0) \frac{1}{\varepsilon} \oint e^{-i2\pi t} \, d\left(\varepsilon e^{i2\pi t}\right)$$

$$= 2\pi i f(z_0). \tag{2.96}$$

Equation (2.96) is **Cauchy's integral formula**.

One can take derivatives of Eq. (2.96) with respect to z_0 to yield an expression for the first derivative:

$$f'(z_0) = \frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{(z - z_0)^2} \, dz, \tag{2.97}$$

and more generally **Cauchy's differentiation formula**:

$$f^{(n-1)}(z_0) = \frac{(n-1)!}{2\pi i} \oint_{\gamma} \frac{f(z)}{(z - z_0)^n} \, dz, \tag{2.98}$$

It is then one final step to get from Cauchy's differentiation formula to the residue theorem. Set the integrand in Eq. (2.98) to a function $g(z)$, which has a pole of order n at z_0. By a Laurent expansion, we can write *any* function with a pole of order n at z_0 in this form. Then we have the **residue theorem**:

$$\int_{\gamma} g(z) \, dz = 2\pi i \sum_{\rho} \text{Res}(f; z_0), \tag{2.99}$$

where the **residue** of f at z_0 is defined as

$$\text{Res}(f; z_0) = \frac{1}{(n-1)!} \lim_{z \to z_0} \frac{d^{n-1}}{dz^{n-1}} \left[(z - z_0)^n f(z)\right]. \tag{2.100}$$

For $n = 1$, a simple pole, the residue is given by

$$\lim_{z \to z_0} \left[(z - z_0) f(z)\right]. \tag{2.101}$$

References

1. S.A. Maier, H.A. Atwater, Plasmonics: localization and guiding of electromagnetic energy in metal/dielectric structures. J. Appl. Phys. **98**(1), 10 (2005)
2. E. Betzig, R.J. Chichester, Single molecules observed by near-field scanning optical microscopy. Science **262**(5138), 1422–1425 (1993)
3. T. Taminiau, F. Stefani, F.B. Segerink, N. Van Hulst, Optical antennas direct single-molecule emission. Nat. Photonics **2**(4), 234–237 (2008)
4. M. Khajavikhan, A. Simic, M. Katz, J. Lee, B. Slutsky, A. Mizrahi, V. Lomakin, Y. Fainman, Thresholdless nanoscale coaxial lasers. Nature **482**(7384), 204–207 (2012)
5. R.F. Oulton, V.J. Sorger, T. Zentgraf, R.-M. Ma, C. Gladden, L. Dai, G. Bartal, X. Zhang, Plasmon lasers at deep subwavelength scale. Nature **461**(7264), 629–632 (2009)
6. I. Maksymov, M. Besbes, J.-P. Hugonin, J. Yang, A. Beveratos, I. Sagnes, I. Robert-Philip, P. Lalanne, Metal-coated nanocylinder cavity for broadband nonclassical light emission. Phys. Rev. Lett. **105**(18), 180502 (2010)
7. H. Shim, L. Fan, S.G. Johnson, O.D. Miller, Fundamental limits to near-field optical response over any bandwidth. Phys. Rev. X **9**, 011043 (2019)
8. K. Joulain, R. Carminati, J.-P. Mulet, J.-J. Greffet, Definition and measurement of the local density of electromagnetic states close to an interface. Phys. Rev. B: Condens. Matter Mater. Phys. **68**, 245405 (2003)
9. E. Yablonovitch, Inhibited spontaneous emission in solid-state physics and electronics. Phys. Rev. Lett. **58**(20), 2059 (1987)
10. E. Yablonovitch, Photonic crystals. J. Mod. Opt. **41**(2), 173–194 (1994)
11. J.D. Joannopoulos, S.G. Johnson, J.N. Winn, R.D. Meade, *Photonic Crystals: Molding the Flow of Light* (Princeton University Press, Princeton, 2011)
12. E.M. Purcell, Spontaneous emission probabilities at radio frequencies. Phys. Rev. **69**, 681 (1946)
13. M. Pelton, J. Vukovic, G.S. Solomon, A. Scherer, Y. Yamamoto, Three-dimensionally confined modes in micropost microcavities: quality factors and purcell factors. IEEE J. Quantum Electron. **38**(2), 170–177 (2002)
14. T. Kippenberg, S. Spillane, K. Vahala, Demonstration of ultra-high-q small mode volume toroid microcavities on a chip. Appl. Phys. Lett. **85**(25), 6113–6115 (2004)
15. N. Stoltz, M. Rakher, S. Strauf, A. Badolato, D. Lofgreen, P. Petroff, L. Coldren, D. Bouwmeester, High-quality factor optical microcavities using oxide apertured micropillars. Appl. Phys. Lett. **87**(3), 031105 (2005)
16. J.T. Robinson, C. Manolatou, L. Chen, M. Lipson, Ultrasmall mode volumes in dielectric optical microcavities. Phys. Rev. Lett. **95**, 143901 (2005)
17. X. Liang, S.G. Johnson, Formulation for scalable optimization of microcavities via the frequency-averaged local density of states. Opt. Express **21**, 30812–30841 (2013)
18. H. Choi, M. Heuck, D. Englund, Self-similar nanocavity design with ultrasmall mode volume for single-photon nonlinearities. Phys. Rev. Lett. **118**, 223605 (2017)
19. S. Hu, M. Khater, R. Salas-Montiel, E. Kratschmer, S. Engelmann, W.M. Green, S.M. Weiss, Experimental realization of deep-subwavelength confinement in dielectric optical resonators. Sci. Adv. **4**(8) (2018)
20. A. Taflove, A. Oskooi, S.G. Johnson, *Advances in FDTD Computational Electrodynamics: Photonics and Nanotechnology* (Artech house, Norwood, 2013)
21. E.N. Economou, *Green's Functions in Quantum Physics* (Springer Science & Business Media, Berlin, 2006)
22. F. Wijnands, J.B. Pendry, F.J. Garcia-Vidal, P.M. Bell, P.J. Roberts, L.M. Moreno, Green's functions for Maxwell's equations: application to spontaneous emission. Opt. Quantum Electron. **29**(2), 199–216 (1997)
23. Y. Xu, R.K. Lee, A. Yariv, Quantum analysis and the classical analysis of spontaneous emission in a microcavity. Phys. Rev. A **61**(3), 033807 (2000)

24. C. Sauvan, J.P. Hugonin, I.S. Maksymov, P. Lalanne, Theory of the spontaneous optical emission of nanosize photonic and plasmon resonators. Phys. Rev. Lett. **110**, 237401 (2013)
25. P. Lalanne, W. Yan, K. Vynck, C. Sauvan, J.-P. Hugonin, Light interaction with photonic and plasmonic resonances. Laser Photon. Rev. **12**, 1700113 (2018)
26. H. Zhang, O.D. Miller, Quasinormal coupled mode theory. arXiv:2010.08650 (2020)
27. D. Pellegrino, D. Balestri, N. Granchi, M. Ciardi, F. Intonti, F. Pagliano, A.Y. Silov, F.W. Otten, T. Wu, K. Vynck, et al., Non-lorentzian local density of states in coupled photonic crystal cavities probed by near-and far-field emission. Phys. Rev. Lett. **124**(12), 123902 (2020)
28. F.G. De Abajo, Optical excitations in electron microscopy. Rev. Mod. Phys. **82**(1), 209 (2010)
29. A. Cazé, R. Pierrat, R. Carminati, Spatial coherence in complex photonic and plasmonic systems. Phys. Rev. Lett. **110**, 063903 (2013)
30. R. Carminati, M. Gurioli, Purcell effect with extended sources: the role of the cross density of states. Opt. Express **30**, 16174–16183 (2022)
31. A. Canaguier-Durand, R. Pierrat, R. Carminati, Cross density of states and mode connectivity: probing wave localization in complex media. Phys. Rev. A **99**, 013835 (2019)
32. V. Krachmalnicoff, E. Castanié, Y. De Wilde, R. Carminati, Fluctuations of the local density of states probe localized surface plasmons on disordered metal films. Phys. Rev. Lett. **105**, 183901 (2010)
33. H.T. Dung, L. Knöll, D.-G. Welsch, Intermolecular energy transfer in the presence of dispersing and absorbing media. Phys. Rev. A **65**, 043813 (2002)
34. D. Martín-Cano, L. Martín-Moreno, F.J. García-Vidal, E. Moreno, Resonance energy transfer and superradiance mediated by plasmonic nanowaveguides. Nano Lett. **10**, 3129–3134 (2010)
35. J.A. Gonzaga-Galeana, J.R. Zurita-Sánchez, A revisitation of the Förster energy transfer near a metallic spherical nanoparticle: (1) efficiency enhancement or reduction? (2) the control of the Förster radius of the unbounded medium. (3) the impact of the local density of states. J. Chem. Phys. **139**(24), 244302 (2013)
36. J. Kästel, M. Fleischhauer, Suppression of spontaneous emission and superradiance over macroscopic distances in media with negative refraction. Phys. Rev. A **71**(1), 011804 (2005)
37. J. Kästel, M. Fleischhauer, Quantum electrodynamics in media with negative refraction. Laser Phys. **15**(1), 135–145 (2005)
38. D. Dzsotjan, A.S. Sørensen, M. Fleischhauer, Quantum emitters coupled to surface plasmons of a nanowire: a Green's function approach. Phys. Rev. B **82**(7), 075427 (2010)
39. D. Martín-Cano, A. González-Tudela, L. Martín-Moreno, F.J. García-Vidal, C. Tejedor, E. Moreno, Dissipation-driven generation of two-qubit entanglement mediated by plasmonic waveguides. Phys. Rev. B **84**, 235306 (2011)
40. A. Gonzalez-Tudela, D. Martin-Cano, E. Moreno, L. Martin-Moreno, C. Tejedor, F.J. Garcia-Vidal, Entanglement of two qubits mediated by one-dimensional plasmonic waveguides. Phys. Rev. Lett. **106**, 020501 (2011)
41. A. Otto, I. Mrozek, H. Grabhorn, W. Akemann, Surface-enhanced raman scattering. J. Phys. Condens. Matt. **4**(5), 1143 (1992)
42. S. Nie, S.R. Emory, Probing single molecules and single nanoparticles by surface-enhanced raman scattering. Science **275**(5303), 1102–1106 (1997)
43. K. Kneipp, Y. Wang, H. Kneipp, L.T. Perelman, I. Itzkan, R.R. Dasari, M.S. Feld, Single molecule detection using surface-enhanced raman scattering (sers). Phys. Rev. Lett. **78**(9), 1667 (1997)
44. K. Kneipp, H. Kneipp, I. Itzkan, R.R. Dasari, M.S. Feld, Surface-enhanced raman scattering and biophysics. J. Phys. Condens. Matt. **14**(18), R597 (2002)
45. D.A. Long, *Raman Spectroscopy* (McGraw-Hill, New York, 1977)
46. J. Michon, M. Benzaouia, W. Yao, O.D. Miller, S.G. Johnson, Limits to surface-enhanced raman scattering near arbitrary-shape scatterers. Opt. Express **27**, 35189–35202 (2019)
47. J.H. Lienhard IV, J.H. Lienhard V, *A Heat Transfer Textbook*, 4th ed. (Dover, Downers Grove, 2011)

48. D. Polder, M. Van Hove, Theory of radiative heat transfer between closely spaced bodies. Phys. Rev. B: Condens. Matter Mater. Phys. **4**(10), 3303–3314 (1971)

49. S.M. Rytov, Y.A. Kravtsov, V.I. Tatarskii, *Principles of Statistical Radiophysics* (Springer, New York, 1988)

50. S.-A. Biehs, M. Tschikin, R. Messina, P. Ben-Abdallah, Super-Planckian near-field thermal emission with phonon-polaritonic hyperbolic metamaterials. Appl. Phys. Lett. **102**(13), 131106 (2013)

51. O.D. Miller, S.G. Johnson, A.W. Rodriguez, Effectiveness of thin films in lieu of hyperbolic metamaterials in the near field. Phys. Rev. Lett. **112**, 157402 (2014)

52. S. Shen, A. Narayanaswamy, G. Chen, Surface phonon polaritons mediated energy transfer between nanoscale gaps. Nano Lett. **9**, 2909–2913 (2009)

53. E. Rousseau, A. Siria, G. Jourdan, S. Volz, F. Comin, J. Chevrier, J.-J. Greffet, Radiative heat transfer at the nanoscale. Nat. Photonics **3**(9), 514–517 (2009)

54. B. Song, Y. Ganjeh, S. Sadat, D. Thompson, A. Fiorino, V. Fernández-Hurtado, J. Feist, F.J. Garcia-Vidal, J.C. Cuevas, P. Reddy, et al., Enhancement of near-field radiative heat transfer using polar dielectric thin films. Nat. Nanotechnol. **10**(3), 253–258 (2015)

55. K. Kim, B. Song, V. Fernández-Hurtado, W. Lee, W. Jeong, L. Cui, D. Thompson, J. Feist, M.T.H. Reid, F.J. García-Vidal, J.C. Cuevas, E. Meyhofer, P. Reddy, Radiative heat transfer in the extreme near field. Nature **528**, 387–391 (2015)

56. J.J. Loomis, H.J. Maris, Theory of heat transfer by evanescent electromagnetic waves. Phys. Rev. B: Condens. Matter Mater. Phys. **50**(24), 18517–18524 (1994)

57. J.B. Pendry, Radiative exchange of heat between nanostructures. J. Phys. Condens. Matter **11**(35), 6621–6633 (1999)

58. K. Joulain, J.-P. Mulet, F. Marquier, R. Carminati, J.-J. Greffet, Surface electromagnetic waves thermally excited: radiative heat transfer, coherence properties and casimir forces revisited in the near field. Surf. Sci. Rep. **57**, 59–112 (2005)

59. P. Ben-Abdallah, K. Joulain, J. Drevillon, G. Domingues, Near-field heat transfer mediated by surface wave hybridization between two films. J. Appl. Phys. **106**(4), 44306 (2009)

60. S.-A. Biehs, E. Rousseau, J.-J. Greffet, Mesoscopic description of radiative heat transfer at the nanoscale. Phys. Rev. Lett. **105**, 234301 (2010)

61. M. Krüger, T. Emig, M. Kardar, Nonequilibrium electromagnetic fluctuations: heat transfer and interactions.' Phys. Rev. Lett. **106**, 210404 (2011)

62. J.P. Mulet, K. Joulain, R. Carminati, J.J. Greffet, Nanoscale radiative heat transfer between a small particle and a plane surface. Appl. Phys. Lett. **78**(19), 2931–2933 (2001)

63. A.W. Rodriguez, O. Ilic, P. Bermel, I. Celanovic, J.D. Joannopoulos, M. Soljačić, S.G. Johnson, Frequency-Selective Near-Field radiative heat transfer between photonic crystal slabs: a computational approach for arbitrary geometries and materials. Phys. Rev. Lett. **107**, 114302 (2011)

64. D. Dalvit, P. Milonni, D. Roberts, F. Da Rosa, *Casimir Physics*, vol. 834 (Springer, Berlin, 2011)

65. S.Y. Buhmann, *Dispersion Forces I: Macroscopic Quantum Electrodynamics and Ground-State Casimir, Casimir–Polder and van der Waals forces*, vol. 247 (Springer, Berlin, 2013)

66. A.W. Rodriguez, F. Capasso, S.G. Johnson, The casimir effect in microstructured geometries. Nat. Photonics **5**(4), 211–221 (2011)

67. P.S. Venkataram, S. Molesky, P. Chao, A.W. Rodriguez, Fundamental limits to attractive and repulsive casimir-polder forces. Phys. Rev. A **101**(5), 052115 (2020)

68. L. Zhu, S. Fan, Persistent directional current at equilibrium in nonreciprocal many-body near field electromagnetic heat transfer. Phys. Rev. Lett. **117**(13), 134303 (2016)

69. A. Raman, S. Fan, Photonic band structure of dispersive metamaterials formulated as a hermitian eigenvalue problem. Phys. Rev. Lett. **104**, 087401 (2010)

70. T. Wu, M. Gurioli, P. Lalanne, Nanoscale light confinement: the q's and v's. ACS Photonics **8**(6), 1522–1538 (2021)

71. C. Sauvan, T. Wu, R. Zarouf, E.A. Muljarov, P. Lalanne, Normalization, orthogonality, and completeness of quasinormal modes of open systems: the case of electromagnetism. Opt. Express **30**(5), 6846–6885 (2022)
72. T. Kato, *Perturbation Theory for Linear Operators*, vol. 132 (Springer Science & Business Media, Berlin, 2013)
73. D.C. Brody, E.-M. Graefe, Information geometry of complex hamiltonians and exceptional points. Entropy **15**(9), 3361–3378 (2013)
74. W. Shockley, H.J. Queisser, Detailed balance limit of efficiency of p-n junction solar cells. J. Appl. Phys. **32**, 510–519 (1961)
75. O.D. Miller, E. Yablonovitch, S.R. Kurtz, Strong internal and external luminescence as solar cells approach the Shockley-Queisser limit. IEEE J. Photovoltaics **2**, 303–311 (2012)
76. E. Yablonovitch, Statistical ray optics. J. Opt. Soc. Am. **72**(7), 899–907 (1982)
77. H.A. Wheeler, Fundamental limitations of small antennas. Proc. IRE **35**, 1479–1484 (1947)
78. L.J. Chu, Physical limitations of Omni-Directional antennas. J. Appl. Phys. **19**(12), 1163–1175 (1948)
79. R.F. Harrington, Effect of antenna size on gain, bandwidth, and efficiency. J. Res. Natl. Bur. Stand. Sect. D Radio Propag. **64D**, 1 (1960)
80. D.F. Sievenpiper, D.C. Dawson, M.M. Jacob, T. Kanar, S. Kim, J. Long, R.G. Quarfoth, Experimental validation of performance limits and design guidelines for small antennas. IEEE Trans. Antennas Propag. **60**, 8–19 (2012)
81. D.J. Bergman, Exactly solvable microscopic geometries and rigorous bounds for the complex dielectric constant of a two-component composite material. Phys. Rev. Lett. **44**(19), 1285–1287 (1980)
82. G.W. Milton, Bounds on the complex dielectric constant of a composite material. Appl. Phys. Lett. **37**(3), 300 (1980)
83. D.J. Bergman, Bounds for the complex dielectric constant of a two-component composite material. Phys. Rev. B **23**(6), 3058–3065 (1981)
84. G.W. Milton, Bounds on the complex permittivity of a two-component composite material. J. Appl. Phys. **52**(1981), 5286–5293 (1981)
85. G.W. Milton, Bounds on the transport and optical properties of a two-component composite material. J. Appl. Phys. **52**(8), 5294–5304 (1981)
86. C. Kern, O.D. Miller, G.W. Milton, Tight bounds on the effective complex permittivity of isotropic composites and related problems. Phys. Rev. Appl. **14**, 054068 (2020)
87. N. Fang, H. Lee, C. Sun, X. Zhang, Sub-diffraction-limited optical imaging with a silver superlens. Science **308**, 534–537 (2005)
88. R. Merlin, Radiationless electromagnetic interference: evanescent-field lenses and perfect focusing. Science **317**, 927–929 (2007)
89. C.W. McCutchen, Superresolution in microscopy and the Abbe resolution limit. J. Opt. Soc. Am. **57**(10), 1190–1192 (1967)
90. E.H.K. Stelzer, Beyond the diffraction limit? Nature **417**, 806–807 (2002)
91. N.I. Zheludev, What diffraction limit? Nat. Mat. **7**(6), 420–422 (2008)
92. D. Slepian, Prolate spheroidal wave functions, fourier analysis and uncertainty— I. Bell Syst. Tech. J. **40**(1), 43–63 (1961)
93. H.J. Landau, H.O. Pollak, Prolate spheroidal wave functions, fourier analysis and uncertainty—II. Bell Syst. Tech. J. **40**(1), 65–84 (1961)
94. P.J. Ferreira, A. Kempf, Superoscillations: faster than the nyquist rate. IEEE Trans. Sig. Proces. **54**(10), 3732–3740 (2006)
95. H. Shim, H. Chung, O.D. Miller, Maximal free-space concentration of electromagnetic waves. Phys. Rev. Appl. **14**(1), 014007 (2020)
96. C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949)
97. T.M. Cover, *Elements of Information Theory* (John Wiley & Sons, Hoboken, 1999)
98. W.D. Heiss, The physics of exceptional points. J. Phys. A: Math. Theor. **45**, 444016 (2012)
99. M.-A. Miri, A. Alù, Exceptional points in optics and photonics. Science **363**, eaar7709 (2019)

100. N. Bender, H. Yılmaz, Y. Bromberg, H. Cao, Creating and controlling complex light. APL Photonics **4**, 110806 (2019)
101. C.W. Hsu, B. Zhen, A.D. Stone, J.D. Joannopoulos, M. Soljačić, Bound states in the continuum. Nat. Rev. Mat. **1**, 1–13 (2016)
102. R.G. Newton, Optical theorem and beyond. Am. J. Phys. **44**(7), 639–642 (1976)
103. J.D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley & Sons, Hoboken, 1999)
104. D.R. Lytle, P.S. Carney, J.C. Schotland, E. Wolf, Generalized optical theorem for reflection, transmission, and extinction of power for electromagnetic fields. Phys. Rev. E **71**, 056610 (2005)
105. Z. Kuang, O.D. Miller, Computational bounds to light–matter interactions via local conservation laws. Phys. Rev. Lett. **125**, 263607 (2020)
106. O.D. Miller, A.G. Polimeridis, M.T. Homer Reid, C.W. Hsu, B.G. DeLacy, J.D. Joannopoulos, M. Soljacic, S.G. Johnson, Fundamental limits to optical response in absorptive systems. Opt. Express **24**, 3329–3364 (2016)
107. O.D. Miller, O. Ilic, T. Christensen, M.T.H. Reid, H.A. Atwater, J.D. Joannopoulos, M. Soljacic, S.G. Johnson, Limits to the optical response of graphene and two-dimensional materials. Nano Lett. **17**, 5408–5415 (2017)
108. Y. Ivanenko, M. Gustafsson, S. Nordebo, Optical theorems and physical bounds on absorption in lossy media. Opt. Express **27**, 34323–34342 (2019)
109. S. Nordebo, G. Kristensson, M. Mirmoosa, S. Tretyakov, Optimal plasmonic multipole resonances of a sphere in lossy media. Phys. Rev. B Condens. Matt. **99**, 054301 (2019)
110. Y. Yang, O.D. Miller, T. Christensen, J.D. Joannopoulos, M. Soljacic, Low-loss plasmonic dielectric nanoresonators. Nano Lett. **17**, 3238–3245 (2017)
111. O.D. Miller, S.G. Johnson, A.W. Rodriguez, Shape-independent limits to near-field radiative heat transfer. Phys. Rev. Lett. **115**, 204302 (2015)
112. Y. Yang, A. Massuda, C. Roques-Carmes, S.E. Kooi, T. Christensen, S.G. Johnson, J.D. Joannopoulos, O.D. Miller, I. Kaminer, M. Soljacic, Maximal spontaneous photon emission and energy loss from free electrons. Nat. Phys. **14**, 894–899 (2018)
113. O.D. Miller, K. Park, R.A. Vaia, Towards maximum optical efficiency of ensembles of colloidal nanorods. Opt. Express **30**, 25061–25077 (2022)
114. L. Tsang, J.A. Kong, K.-H. Ding, *Scattering of Electromagnetic Waves: Theories and Applications* (John Wiley & Sons, Inc., New York, 2000)
115. S. Molesky, P. Chao, A.W. Rodriguez, Hierarchical mean-field T operator bounds on electromagnetic scattering: upper bounds on near-field radiative purcell enhancement. Phys. Rev. Res. **2**, 043398 (2020)
116. Z. Kuang, L. Zhang, O.D. Miller, Maximal single-frequency electromagnetic response. Optica **7**, 1746–1757 (2020)
117. R.E. Hamam, A. Karalis, J.D. Joannopoulos, M. Soljačić, Coupled-mode theory for general free-space resonant scattering of waves. Phys. Rev. A **75**, 053801 (2007)
118. D.-H. Kwon, D.M. Pozar, Optimal characteristics of an arbitrary receive antenna. IEEE Trans. Antennas Propag. **57**, 3720–3727 (2009)
119. Z. Ruan, S. Fan, Design of subwavelength superscattering nanospheres. Appl. Phys. Lett. **98**(4), 043101 (2011)
120. I. Liberal, I. Ederra, R. Gonzalo, R. Ziolkowski, Upper bounds on scattering processes and Metamaterial-Inspired structures that reach them. IEEE Trans. Antennas Propag. **62**, 6344–6353 (2014)
121. I. Liberal, Y. Ra'di, R. Gonzalo, I. Ederra, S.A. Tretyakov, R.W. Ziolkowski, Least upper bounds of the powers extracted and scattered by bi-anisotropic particles. IEEE Trans. Antennas Propag. **62**, 4726–4735 (2014)
122. J.-P. Hugonin, M. Besbes, P. Ben-Abdallah, Fundamental limits for light absorption and scattering induced by cooperative electromagnetic interactions. Phys. Rev. B: Condens. Matter Mater. Phys. **91**, 180202 (2015)
123. A.E. Miroshnichenko, M.I. Tribelsky, Ultimate absorption in light scattering by a finite obstacle. Phys. Rev. Lett. **120**(3), 033902 (2018)

124. M. Gustafsson, M. Cismasu, Physical bounds and optimal currents on antennas. IEEE Trans. Antennas Propag. **60**(6), 2672–2681 (2012)
125. M. Gustafsson, K. Schab, L. Jelinek, M. Capek, Upper bounds on absorption and scattering. New J. Phys. **22**, 073013 (2020)
126. S. Molesky, P. Chao, W. Jin, A.W. Rodriguez, Global T operator bounds on electromagnetic scattering: upper bounds on far-field cross sections. Phys. Rev. Res. **2**, 033172 (2020)
127. M.I. Abdelrahman, F. Monticone, How thin and efficient can a metasurface reflector be? Universal bounds on reflection for any direction and polarization. arXiv:2208.05533 (2022)
128. L. Jelinek, M. Gustafsson, M. Capek, K. Schab, Fundamental bounds on the performance of monochromatic passive cloaks. Opt. Express **29**(15), 24068–24082 (2021)
129. K. Schab, L. Jelinek, M. Capek, M. Gustafsson, Upper bounds on focusing efficiency. Opt. Express **30**(25), 45705–45723 (2022)
130. W.C. Chew, M.S. Tong, B. Hu, Integral equation methods for electromagnetic and elastic waves. Synth. Lect. Comput. Electromagn. **3**, 1–241 (2008)
131. D.A. Miller, Spatial channels for communicating with waves between volumes. Opt. Lett. **23**(21), 1645–1647 (1998)
132. D.A. Miller, Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths. Appl. Opt. **39**(11), 1681–1699 (2000)
133. R. Piestun, D.A. Miller, Electromagnetic degrees of freedom of an optical system. JOSA A **17**(5), 892–902 (2000)
134. D.A. Miller, Waves, modes, communications, and optics: a tutorial. Adv. Opt. Photonics **11**(3), 679–825 (2019)
135. L.N. Trefethen, D. Bau, *Numerical Linear Algebra* (Society for Industrial and Applied Mathematics, Philadelphia, 1997)
136. Z. Kuang, D.A.B. Miller, O.D. Miller, Bounds on the coupling strengths of communication channels and their information capacities. arXiv:2205.05150 (2022)
137. C. Ehrenborg, M. Gustafsson, Fundamental bounds on mimo antennas. IEEE Antennas Wireless Propag. Lett. **17**(1), 21–24 (2017)
138. C. Ehrenborg, M. Gustafsson, Physical bounds and radiation modes for mimo antennas. IEEE Trans. Antennas Propag. **68**(6), 4302–4311 (2020)
139. C. Ehrenborg, M. Gustafsson, M. Capek, Capacity bounds and degrees of freedom for mimo antennas constrained by q-factor. IEEE Trans. Antennas Propag. (2021)
140. Y. Gao, H. Vinck, and T. Kaiser, Massive mimo antenna selection: switching architectures, capacity bounds, and optimal antenna selection algorithms. IEEE Trans. Sig. Proces. **66**(5), 1346–1360 (2017)
141. S. Asaad, A.M. Rabiei, R.R. Müller, Massive mimo with antenna selection: fundamental limits and applications. IEEE Trans. Wireless Commun. **17**(12), 8502–8516 (2018)
142. S.G. Johnson, P. Bienstman, M.A. Skorobogatiy, M. Ibanescu, E. Lidorikis, J.D. Joannopoulos, Adiabatic theorem and continuous coupled-mode theory for efficient taper transitions in photonic crystals. Phys. Rev. E **66**, 66608 (2002)
143. S. Boyd, L. El Ghaoui, E. Feron, V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory* (SIAM, Philadelphia, 1994)
144. M.X. Goemans, D.P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J. ACM **42**(6), 1115–1145 (1995)
145. Z.Q. Luo, W.K. Ma, A. So, Y. Ye, S. Zhang, Semidefinite relaxation of quadratic optimization problems. IEEE Signal Process. Mag. **27**(3), 20–34 (2010)
146. S. Sojoudi, J. Lavaei, Physics of power networks makes hard optimization problems easy to solve, in *2012 IEEE Power and Energy Society General Meeting* (IEEE, Piscataway, 2012), pp. 1–8
147. E.J. Candes, T. Strohmer, V. Voroninski, Phaselift: exact and stable signal recovery from magnitude measurements via convex programming. Commun. Pure Appl. Math. **66**(8), 1241–1274 (2013)
148. R. Horstmeyer, R.Y. Chen, X. Ou, B. Ames, J.A. Tropp, C. Yang, Solving ptychography with a convex relaxation. New J. Phys. **17**(5), 053044 (2015)

149. M. Laurent, F. Rendl, Semidefinite programming and integer programming. Handbooks Oper. Res. Manag. Sci. **12**(C), 393–514 (2005)
150. L. Vandenberghe, S. Boyd, Semidefinite programming. SIAM Rev. **38**, 49–95 (1996)
151. S. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004)
152. J. Park, S. Boyd, General Heuristics for Nonconvex Quadratically Constrained Quadratic Programming. arXiv:1703.07870 [math] (2017)
153. G. Angeris, J. Vučković, S. Boyd, Heuristic methods and performance bounds for photonic design. Opt. Express **29**(2), 2827–2854 (2021)
154. H. Shim, Z. Kuang, Z. Lin, O.D. Miller, Fundamental limits to multi-functional and tunable nanophotonic response. arXiv: 2112.10816 (2021)
155. H. Zhang, Z. Kuang, S. Puri, O.D. Miller, Conservation-law-based global bounds to quantum optimal control. Phys. Rev. Lett. **127**, 110506 (2021)
156. G. Angeris, T. Diamandis, J. Vučković, S. Boyd, Bounds on efficiency metrics in photonics. Preprint. arXiv:2204.05243 (2022)
157. G. Angeris, A note on generalizing power bounds for physical design. arXiv:2208.04411 (2022)
158. F.W. King, Sum rules for the optical constants. J. Math. Phys. **17**(8), 1509–1514 (1976)
159. V. Lucarini, J.J. Saarinen, K.-E. Peiponen, E. M. Vartiainen, *Kramers-Kronig Relations in Optical Materials Research* (Springer Science & Business Media, Berlin, 2005)
160. H.M. Nussenzveig, *Causality and Dispersion Relations* (Academic Press, New York, 1972)
161. E. Kaxiras, J.D. Joannopoulos, *Quantum Theory of Materials* (Cambridge University Press, Cambridge, 2019)
162. J. Skaar, K. Seip, Bounds for the refractive indices of metamaterials. J. Phys. D Appl. Phys. **39**, 1226–1229 (2006)
163. M. Gustafsson, D. Sjöberg, Sum rules and physical bounds on passive metamaterials. New J. Phys. **12**, 043046 (2010)
164. H. Shim, F. Monticone, O.D. Miller, Fundamental limits to the refractive index of transparent optical materials. Adv. Mat. **33**, 2103946 (2021)
165. C. Sohl, M. Gustafsson, G. Kristensson, Physical limitations on broadband scattering by heterogeneous obstacles. J. Phys. A: Math. Theor. **40**, 11165–11182 (2007)
166. R.G. Gordon, Three sum rules for total optical absorption cross sections. J. Chem. Phys. **38**(7), 1724 (1963)
167. E.M. Purcell, On the absorption and emission of light by interstellar grains. Astrophys. J. **158**, 433–440 (1969)
168. D.S. Jones, Scattering by inhomogeneous dielectric particles. Q. J. Mech. Appl. Mech. **38**(1), 135–155 (1985)
169. H. Shim, Z. Kuang, O.D. Miller, Optical materials for maximal nanophotonic response (Invited). Opt. Mat. Express **10**, 1561–1585 (2020)
170. S. Molesky, P.S. Venkataram, W. Jin, A.W. Rodriguez, Fundamental limits to radiative heat transfer: theory. Phys. Rev. B **101**(3), 035408 (2020)
171. P.S. Venkataram, S. Molesky, W. Jin, A.W. Rodriguez, Fundamental limits to radiative heat transfer: the limited role of nanostructuring in the near-field. Phys. Rev. Lett. **124**(1), 013904 (2020)
172. P. Ben-Abdallah, K. Joulain, Fundamental limits for noncontact transfers between two bodies. Phys. Rev. B: Condens. Matter Mater. Phys. **82**, 121419 (2010)
173. W. Jin, S. Molesky, Z. Lin, A.W. Rodriguez, Material scaling and frequency-selective enhancement of near-field radiative heat transfer for lossy metals in two dimensions via inverse design. Phys. Rev. B **99**(4), 041403 (2019)
174. S. Sanders, A. Manjavacas, Analysis of the limits of the local density of photonic states near nanostructures. ACS Photonics **5**, 2437–2445 (2018)
175. S.M. Barnett, R. Loudon, Sum rule for modified spontaneous emission rates. Phys. Rev. Lett. **77**(12), 2444–2446 (1996)

176. L. Zhang, F. Monticone, O.D. Miller, All electromagnetic scattering bodies are matrix-valued oscillators. arXiv: 2211.04457 (2022)
177. H. Hashemi, C.-W. Qiu, A.P. Mccauley, J.D. Joannopoulos, S.G. Johnson, Diameter-bandwidth product limitation of isolated-object cloaking. Phys. Rev. A **86**, 013804 (2012)
178. L. Zhang, O.D. Miller, Optimal materials for maximum large-area near-field radiative heat transfer. ACS Photonics **7**, 3116–3129 (2020)
179. R. Mittapally, J.W. Lim, L. Zhang, O.D. Miller, P. Reddy, E. Meyhofer, Probing the limits to near-field heat transfer enhancements in phonon-polaritonic materials. Nano Lett. **23**, 2187–2194 (2023)
180. P. Chao, R.K. Defo, S. Molesky, A. Rodriguez, Maximum electromagnetic local density of states via material structuring. Nanophotonics **12**, 549–557 (2022)
181. Z.-J. Yang, T.J. Antosiewicz, R. Verre, F.J. Garcia de Abajo, S.P. Apell, M. Kall, Ultimate limit of light extinction by nanophotonic structures. Nano Lett. **15**(11), 7633–7638 (2015)
182. A. Bernland, A. Luger, M. Gustafsson, Sum rules and constraints on passive systems. J. Phys. A Math. Theoret. **44**(14), 145205 (2011)
183. J.-M. Hartmann, C. Boulet, D. Robert, *Collisional Effects on Molecular Spectra: Laboratory Experiments and Models, Consequences for Applications* (Elsevier, Amsterdam, 2021)
184. R. Carminati, J.C. Schotland, *Principles of Scattering and Transport of Light* (Cambridge University Press, Cambridge, 2021)
185. P.-A. Absil, R. Mahony, R. Sepulchre, *Optimization Algorithms on Matrix Manifolds* (Princeton University Press, Princeton, 2009)
186. N. Boumal, B. Mishra, P.-A. Absil, R. Sepulchre, Manopt, a Matlab toolbox for optimization on manifolds. J. Mach. Learn. Res. **15**(42), 1455–1459 (2014)
187. S. Gertler, Z. Kuang, C. Christie, O.D. Miller, Many physical design problems are sparse QCQPs. arXiv:2303.17691 (2023)
188. G. Angeris, J. Vuckovic, S.P. Boyd, Computational bounds for photonic design. ACS Photonics **6**(5), 1232–1239 (2019)
189. Q. Zhao, L. Zhang, O.D. Miller, Minimum dielectric-resonator mode volumes. arXiv: 2008.13241 (2020)
190. M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx (2014)
191. Gurobi Optimization, LLC, Gurobi optimizer reference manual (2019)
192. E. Rousseau, M. Laroche, J.-J. Greffet, Asymptotic expressions describing radiative heat transfer between polar materials from the far-field regime to the nanoscale regime. J. Appl. Phys. **111**(1), 014311 (2012)
193. M. Pascale, G.T. Papadakis, Tight bounds and the role of optical loss in polariton-mediated near-field heat transfer. Phys. Rev. Appl. **19**(3), 034013 (2023)
194. S.G. Johnson, Numerical methods for computing casimir interactions. In: Dalvit, D., Milonni, P., Roberts, D., da Rosa, F. (eds.) Casimir Physics. Lecture Notes in Physics, Springer, Berlin, Heidelberg. **834**, 175–218 (2011). https://doi.org/10.1007/978-3-642-20288-9_6