Check for updates

# Physical limits in electromagnetism

Pengning Chao[1], Benjamin Strekha[1], Rodrick Kuate Defo[1], Sean Molesky[2]
and Alejandro W. Rodriguez[1]✉

Abstract | Photonic devices play an increasingly important role in advancing physics and engineering. Although improvements in nanofabrication and computational methods have driven progress in expanding the range of achievable optical characteristics, they have also greatly increased design complexity. These developments motivate the study of fundamental limits on optical response. Here we review recent progress in our understanding of these limits with special focus on an emerging theoretical framework that combines computational optimization with conservation laws to yield physical limits capturing all relevant wave effects. Results pertaining to canonical electromagnetic problems such as thermal emission, scattering cross-sections, Purcell enhancement and power routing are presented. Finally, we identify areas where further research is needed, including conceptual extensions and efficient numerical schemes for handling large-scale problems.

Photonics has become an indispensable tool in scientific discovery, enabling key advances in communications[1], sensing[2,3], photovoltaics[4], computing[5,6], quantum engineering[7,8] and many other fields. Central to the broad applicability of optical methods is a small, but powerful, set of design schemes for confining and transferring energy in time and space — notions such as index guiding, wave interference, polaritonics and effective-medium engineering — that provide physical intuition for extracting concrete functionality from the abstract mathematical richness of Maxwell's equations. Each of these schemes offers a mixture of distinctive capabilities and limitations, and determining the best approach (or combination thereof) for any particular application remains a challenge for photonic design[9].

As a concrete example, consider the problem of enhancing light–matter interactions via the photonic local density of states (LDOS) — the Purcell effect — reducing to the familiar Purcell factor $F_{\mathrm{P}} = \frac{6}{\pi^2}\left(\frac{\lambda}{2n}\right)^3\frac{Q}{V}$, with $\lambda$ denoting wavelength and $n$ refractive index, in the case where a single resonance dominates[10]. Integrated micro-resonators[11] based on index guiding can achieve extremely long lifetimes (high quality factors $Q$) at the expense of reduced spatial localization (large mode volumes $V$). Electronic plasmon- and phonon-polariton resonances allow for tight subwavelength confinement (small $V$) but suffer from high material absorption (small $Q$)[12]. Photonic crystals[10] and bandgap engineering provide a flexible low-loss platform for manipulating light at the wavelength scale but are limited in practice by the achievable bandwidths and a lack of forms exhibiting omnidirectional bandgaps. Metamaterials offer conceptual simplicity in engineering exotic dispersions and large LDOS[13], but are constrained by fabrication limitations, the breakdown of effective-medium approximations at large wavenumbers, and challenges related to light coupling[14,15].

Growing out of these general design principles, continued increases in computational power have enabled the development of inverse methods that, given a set of desired electromagnetic objectives and constraints, exploit global[16], gradient-based[17–19] and data-driven optimization algorithms[20,21] to search through potentially millions of structural degrees of freedom in pursuit of ideal response characteristics. This capacity has greatly expanded the accessible design space and has led to vast improvements in device performance (FIG. 1). However, for typical problems, the range of design possibilities and complicated interplay between standard objectives and constraints also makes it practically impossible to determine truly optimal structures, and one can at most hope for a well-performing local optimum. (Notwithstanding this, inverse methods may converge to structures that appear to reflect intuitive design principles, such as bowtie antenna and slot waveguide motifs for enhancing light–matter interaction[18,22,23]). Although the current arsenal of design techniques and algorithms provides enormous capability to tackle a wide range of engineering applications, it cannot rigorously answer a natural question of increasing relevance: what are the fundamental limits to optical control, and how close are existing devices to reaching them?

The notion of such fundamental limits, encoded in bedrock principles such as the finite speed of light and the second law of thermodynamics, is ubiquitous in physics. Beyond added theoretical understanding,

[1]Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ, USA.

[2]Department of Engineering Physics, Polytechnique Montréal, Montréal, Québec, Canada.

✉e-mail: arod@princeton.edu

https://doi.org/10.1038/s42254-022-00468-w

**Key points**

- Limits provide a rigorous theoretical basis to quantify the trade-offs and scaling laws associated with almost any common design parameter (device footprint, material choices, fixed separations and so on).
- Limits may guide practical design efforts and complement computational inverse design methods by confirming the achievement of near-optimality or clarifying possibilities for further improvement.
- Inverse design methods, such as topology optimization, involve a large number of degrees of freedom subject to non-convex constraints (Maxwell's wave equations); although gradient-based methods can efficiently find local optima, it is generally infeasible to determine the global optimum.
- Many limits are derived by simplifying the design problem in some fashion: for example, by working with geometric optics, maximizing per-mode contributions ignoring inter-mode constraints, only enforcing a subset of the relevant physics and so on.
- Mathematical optimization theory formalizes the notion of simplified ('relaxed') physics and offers tools for bounding the design performance in such settings. In turn, these values bound realistic design performance.

limits have and continue to play an important role as signposts for further technological improvement. Attempts to overcome the Abbe diffraction limit contributed greatly to the development of the field of super-resolution microscopy, with diverse techniques exploiting evanescent fields[24–26], nonlinear effects[25,27] and active temporal control[28]. Knowledge of the physical origins of the factors forming the Shockley–Queisser limit[29] for solar cell efficiency pointed the way to developments in concentrators[30], tandem[31] and intermediate band photovoltaics[32]. The breakdown of familiar blackbody limits to nanoscale separations sparked interest in super-Planckian thermal devices[33,34].

In this Review, we first present a historical overview of the development of electromagnetic limits with representative examples that showcase the broad range of problems for which limits are applicable. The derivations of these limits illustrate a general thematic evolution mirroring the history of optics itself: from simplifying and restrictive assumptions (homogeneity, quasistatics, ray optics and so on) pertinent to low-dimensional, deeply subwavelength and large-etalon systems, toward increasingly sophisticated wave arguments applicable to any length scale (FIG. 2).

These examples also provide context and motivation for the main focus of this Review: an emerging general methodology for evaluating photonic design bounds based on mathematical optimization theory and physical conservation principles. Originally developed as an instrument for investigating maximal scattering cross-sections[35–37], the framework is applicable to a broad range of design problems where the objective can be expressed as a quadratic function of the fields[38–40]. The constraints follow directly from Maxwell's equations and the identities of scattering theory, limiting both the amplitude of the polarization response — important to power transfer — and the extent to which the phase of the polarization response can be modified — with consequences for the engineering of resonances. To better handle problems involving near-field effects and rapidly varying length scales, spatially localized constraints can be introduced, with a denser distribution of local constraints giving tighter bounds at the expense of higher computational complexity[38,39]. In this sense, the framework emphasizes the complementary role of limits and structural optimization: structural optimization enforces Maxwell's equations exactly (up to computational discretization) and produces specific devices corresponding to local optima; limit calculations instead produces bounds that apply to all possible structures via conservation-law based constraints over spatial regions that can be viewed as a relaxation of Maxwell's equations.

Through instructive examples concerning thermal emission, scattering/absorption cross-sections, LDOS enhancement and power splitting, we describe in detail the physical implications behind various components of the framework and showcase its broad applicability. Well-known results such as the Chu limit for electrically small antennae and the geometric optics limit for cross-sections of large scatterers are shown to spring naturally from the framework. For readers interested in the mathematical details of the underlying optimization theory, we also recommend REF.[41].

Finally, we discuss the remaining challenges and opportunities, including the need for numerical methods that can handle larger systems, generalizations to other physical settings beyond photonics, and potential improvements to structural optimizations that may arise from knowledge of optimal fields.

## Historical overview

In this section, we highlight representative prior results on electromagnetic limits relevant to a wide range of topics in photonics.

***Speed of light limits.*** Since the first measurement of the speed of light in vacuum by Ole Rømer, and the subsequent postulates of special relativity demanding that information cannot travel faster than $c$, rigorous proofs of subluminal energy velocity $v_e \le c$ have been deduced from increasingly general assumptions: moving from homogeneous non-absorbing media, through the inclusion of dispersion, anisotropy and non-locality[42–45], to the simple unifying requirement of passivity: materials that do no net work on electromagnetic fields[45–47].

***Slow light.*** In complement, there has also been great interest in establishing limits on achieving minimal energy velocity or 'slow light', using engineered devices such as optical delay lines and buffers[48–52]. Under the approximation of zero bandwidth, the delay experienced by a light pulse can essentially be made indefinitely long, for example, near the band edge of a photonic crystal where the group velocity vanishes[53]. For finite bandwidths, however, the product of the delay time and operating bandwidth sets a fundamental lower bound (similar to the energy–time uncertainty principle in quantum mechanics). For a slow light waveguide with an idealized linear dispersion relation across the bandwidth of interest, the delay–bandwidth product limit takes the form $\Delta T \Delta f \le \frac{L}{\lambda_c}(n_{\text{avg}} - n_{\text{min}})$, where $\Delta T$ is the delay time, $\Delta f$ is the operating bandwidth, $L$ is the length of the device, $\lambda_c$ is the free-space bandwidth at the band centre, and $n_{\text{avg}}$, $n_{\text{min}}$ are the average and minimum effective

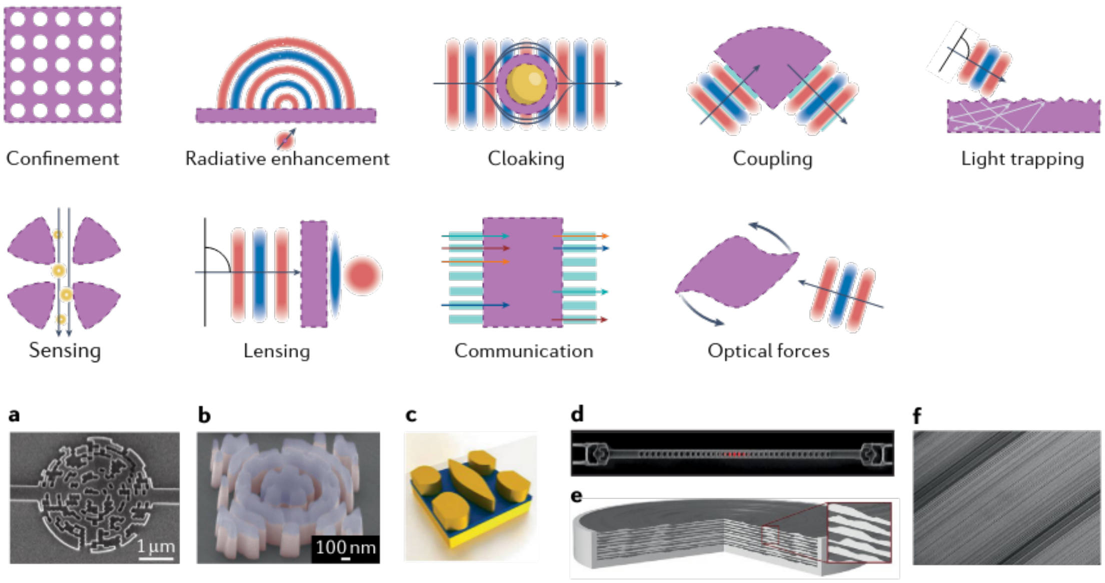

Fig. 1 | **Applications of photonics and computational design.** The top part of the figure illustrates a collection of representative photonic functionalities. In each instance, performance critically depends on the possibility of using (sub) wavelength-scale structuring to confine or transform optical fields and is intrinsically linked to physical limits on these phenomena. The bottom part of the figure depicts several examples of computationally synthesized devices for achieving improved performance in the operations shown above. **a** | A photon spin selector preferentially coupling left-handed incident photons to one output waveguide and right-handed incident photons to the other output waveguide, as part of a chip-scale alkali vapour magnetometer. **b** | A 'photon extractor' enhancing the emission collection rate of a nitrogen vacancy centre. **c** | A frequency-selective emitter for thermo-photovoltaic applications. **d** | A photon extractor (cavity and coupler) for diamond photonics. **e**,**f** | Large-area, high-efficiency metalenses. Panel **a** is reproduced from REF.[179], under a Creative Commons licence CC BY 4.0; panel **b** is reproduced with permission from REF.[8], Optica; panel **c** is reproduced with permission from REF.[180], AIP; panel **d** is reproduced from REF.[7], under a Creative Commons licence CC BY 4.0; panel **e** is reproduced with permission from REF.[146], Optica; panel **f** is adapted from REF.[145], under a Creative Commons licence CC BY 4.0.

index of refraction across the bandwidth, respectively[48]. For the simple case of explicitly 1D wave propagation and material structuring, more general bounds can be derived without the notion of an effective index of refraction and the assumption of idealized linear dispersion: $\Delta T\Delta f \leq \frac{1}{2\sqrt{3}}\frac{L}{\lambda_c}\frac{\max\|\varepsilon(z,f)-\varepsilon_{ro}\|}{\varepsilon_{ro}}$, where $\varepsilon(z,f)$ is the (possibly complex) relative dielectric constant as a function of the propagation coordinate $z$ and frequency $f$, $\varepsilon_{ro}$ is the relative dielectric constant of the background medium, and the maximum is taken across space and bandwidth[54,55]. Further work is needed to extend these results to accommodate rigorous full-wave descriptions of 3D structures.

***Optical cloaking.*** Related bandwidth arguments have also been used to derive performance bounds on optical cloaks — devices that eliminate or greatly reduce scattering of incident light[56]. Most commonly, metamaterial cloaks are designed via transformation optics to carry out prescribed phase and amplitude manipulations by mapping coordinate transformations, $\mathbf{x} \rightarrow \mathcal{J}\mathbf{x}'$, onto effective homogenized susceptibility parameters, $\{\varepsilon,\mu\} \rightarrow \{\varepsilon',\mu'\} = \frac{\mathcal{J}^{\mathrm{T}}\{\varepsilon,\mu\}\mathcal{J}}{\det \mathcal{J}}$, in a design volume surrounding the object. The phase velocity in a perfect cloak must be superluminal, since light must go around the cloaked object and maintain an unperturbed wavefront. This requires the existence of dispersion in order to obey causality, and thereby precludes perfect cloaking in vacuum over any finite bandwidth[57]. One way around this limitation is to consider 'ground-plane' cloaking, wherein the object to be concealed is positioned adjacent to a reflecting boundary. The original causality argument does not apply in this case, as the reflected waves from the cloak travel a shorter distance than reflected waves from the ground plane itself. However, because the wavefront within the cloak must now be delayed, the need to respect delay–bandwidth limits remains. In the simplifying case of 1D waves, this leads to a bound on the minimum thickness of the cloak $w \gtrsim \frac{h}{n(n-1)}\frac{\Delta\omega}{\omega}$ as a function of the object size $h$ and the bandwidth $\Delta\omega$. Generalizing to 3D (REF.[58]), any ground-plane cloak must obey the inequality $V_c \geq V'/B$, where $V_c$ is the volume of the cloak, $V'$ is combined volume of cloak and object, and $B$ represents the maximum achievable index contrast (eigenvalues of $\mathcal{J}^{\mathrm{T}}\mathcal{J}$). Although instructive, determining the attainable $\mathcal{J}$, and hence $B$, requires specific geometric analyses.

***Optical confinement.*** The intertwined concepts of optical delay and waveguide propagation naturally lead to questions concerning optical confinement[11]. Two main techniques are commonly used to confine light, without relying on coupling to material resonances[59]: index guiding, a generalization of total internal reflection to wavelength-scale cavities (such as whispering gallery or ring resonators), and bandgap confinement, a generalization of Bragg scattering with embodiments in

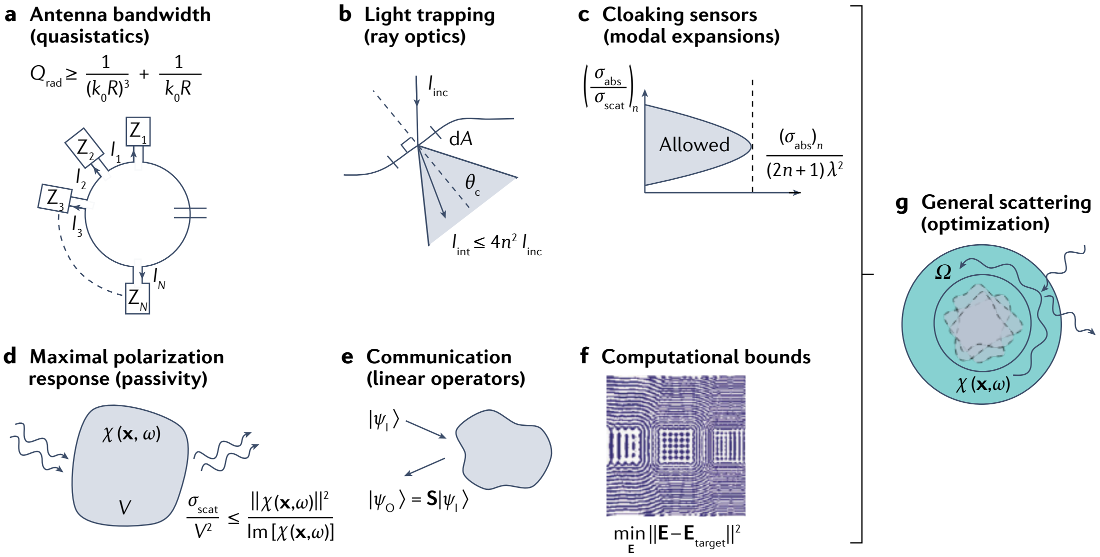
a Antenna bandwidth (quasistatics)
$Q_{rad} \geq \frac{1}{(k_0R)^3} + \frac{1}{k_0R}$
$Z_1$
$Z_2$
$Z_3$
$Z_N$
$I_1$
$I_2$
$I_3$
$I_N$
b Light trapping (ray optics)
$I_{inc}$
dA
$\theta_c$
$I_{int} \leq 4n^2 I_{inc}$
c Cloaking sensors (modal expansions)
$\left(\frac{\sigma_{abs}}{\sigma_{scat}}\right)_n$
Allowed
$\frac{(\sigma_{abs})_n}{(2n+1)\lambda^2}$
d Maximal polarization response (passivity)
$\chi(\mathbf{x}, \omega)$
V
$\frac{\sigma_{scat}}{V^2} \leq \frac{\|\chi(\mathbf{x},\omega)\|^2}{\mathrm{Im}[\chi(\mathbf{x},\omega)]}$
e Communication (linear operators)
$|\psi_I\rangle$
$|\psi_O\rangle = \mathbf{S}|\psi_I\rangle$
f Computational bounds
$\min_{\mathbf{E}} \|\mathbf{E} - \mathbf{E}_{target}\|^2$
g General scattering (optimization)
$\Omega$
$\chi(\mathbf{x},\omega)$

Fig. 2 | **Overview of representative electromagnetic limits.** The figure illustrates a collection of representative photonic bounds, arranged from **a** to **f** in roughly chronological order. **a** | Equivalent circuit of an omnidirectional antenna. For electrically small antennas, taking the quasistatic limit yields the celebrated Chu limit. **b** | Yablonovitch limits to light trapping at a rough interface, combining geometric (ray) optics with a statistical description of the scattering angle. **c**,**d** | Limits based on passivity: for a passive device both the absorbed and scattered power must be non-negative. Panel **c** shows the feasible region for cloaking as a function of absorption efficiency $\sigma_{abs}/\sigma_{scat}$ and net absorption $\sigma_{abs}$ (REF.[181]). Panel **d** illustrates upper bounds on scattering cross-sections dependent on the material susceptibility $\chi$. **e** | Schematic showing a scattering matrix description of an optical device connecting input waves to output waves. Tools of linear algebra such as singular value decompositions can be used to analyse limits on communication and power transfer using lightwaves[54,81,182]. **f** | Suggested initial design for a multimode 2D Helmholtz resonator, based on the solution of the Lagrange dual problem of the design optimization minimizing the norm-squared error between the electrical field **E** within the device and some target field $\mathbf{E}_{target}$. **g** | A Lagrange dual framework for evaluating general photonic bounds using only knowledge of the design region $\Omega$ and material susceptibility $\chi$, through constraints based on conservation laws derived from Maxwell's equations. Overall, we see a trend from early results that are problem- or regime-specific towards more recent limits with broader applicability. In panel **a**, $Q_{rad}$ is the radiative Q-factor of the antenna, $k_0$ the wavevector, $R$ the radius of a sphere constraining the size of the antenna, and $I_i$ and $Z_i$ are effective currents and impedances of the equivalent circuit model used by Chu[162]. In panel **b**, d$A$ is a small area element of the absorption surface under study, $I_{inc}$, $I_{int}$ are the intensities for the incident and transmitted light respectively, $\theta_c$ denotes the maximum angle from which light can escape from the high index material. In panel **c**, $(\sigma_{abs})_n$, $(\sigma_{scat})_n$ are the cross-sections for the scattering harmonic order $n$. In panel **d**, $\chi(\mathbf{x}, \omega)$ is the material susceptibility distribution over device volume $V$ as a function of position **x** and angular frequency $\omega$. In panel **e**, $|\psi\rangle_I$ and $|\psi\rangle_O$ are input and output waves connected by the scattering **S** matrix of an optical device. Panel **a** is adapted with permission from REF.[162], AIP; panel **b** is adapted with permission from REF.[87], Optica; panel **c** is adapted with permission from REF.[181], APS; panel **e** is adapted with permission from REF.[81], PNAS; panel **f** adapted with permission from REF.[160], ACS; panel **g** is adapted with permission from REF.[177], Optica.

photonic crystal waveguides and holey fibres[10]. Except for the simplest cases, determining the propagation characteristics of a specific design requires numerical computation. This makes sufficient conditions for the existence of guided modes (analogous to variational conditions for existence of bound states in quantum mechanics) conceptually and practically useful. For instance, the displacement field $\mathbf{D}_c$ of the fundamental mode of a waveguide with permittivity profile $\varepsilon$ and cladding profile $\varepsilon_c$ was shown to necessarily satisfy $\int \mathbf{D}_c^\star \cdot (\varepsilon^{-1} - \varepsilon_c^{-1})\mathbf{D}_c < 0$ within the cladding[60]. On a similar note, the degree of localization achievable by a photonic crystal defect mode is intuitively proportional to the bandgap size[10], leading to variational bounds on the minimum index contrast required to engineer 2D bandgaps[61]. Generalizations to incorporate conditions for dual-polarization and 3D localization, along with considerations of quasicrystalline[62] and disordered media[63] remain open problems.

***Optical density of states.*** Light confinement is also an integral tool for enhancing light–matter interactions in optical modulators, lasers and quantum devices. As detailed in the Introduction, a resonant mode enhances the power radiated by a nearby dipolar emitter in accordance with the Purcell factor, which grows proportionally with the quality factor $Q$ (longer lifetimes) and inversely with mode volume $V$ (higher field intensities). Beyond modal descriptions that do not readily generalize to multiresonant systems[64] or that refer to specific geometric designs, the electromagnetic LDOS stands as a fundamental figure of merit quantifying optical response in arbitrary settings. By enforcing passivity for the scattered power[65], quadratic optimization arguments have made it possible to constrain the magnitude of achievable polarization response independent of geometric or modal considerations. For a dipole emitter a small distance $d$ away from a device enclosed within a half space, such conservation arguments yield a material bound on the dominant contribution of evanescent fields to LDOS enhancement at a single wavelength, $\frac{\rho(\omega)}{\rho_0(\omega)} \leq \frac{1}{8\pi^3}(\frac{\lambda}{d})^3 \frac{\|\chi(\omega)\|^2}{\mathrm{Im}\chi(\omega)}$, where $\rho_0(\omega)$ is the free-space LDOS, $\rho(\omega)$ is the total LDOS and $\chi(\omega)$ is the material susceptibility, all as a function of the angular frequency $\omega$. As discussed later, the positivity of scattered power alone cannot fully account for all relevant wave effects and resonance conditions, and in fact is only one of several important constraints that limit optical response beyond quasistatic settings[35]. Passivity arguments based on the Kramers–Kronig conditions similarly place limits on the frequency-integrated material response of a medium, yielding sum rules of the form $\int_0^\infty \mathrm{d}\omega \ \frac{\rho(\mathbf{x},\omega) - \rho_0(\omega)}{\rho_0(\omega)} = 0$ (REFS[66,67]).

***Light focusing.*** Another important aspect of localization relates to the focusing of optical power from a source to a receiver. When restricted to systems described by geometric optics, the conservation of etendue[68] places a lower bound on how tight the light rays from a source can be focused down[69]. Accounting for wave effects, scattering concentration bounds over input–output modal channels of the form $\langle \|c_{\mathrm{out},\hat{\mathbf{u}}}\|^2 \rangle \leq \mathrm{maxeig}(\rho_{\mathrm{in}})$ yield the maximum concentration of power achievable for any linear combination of output channels represented by the unit vector $\hat{\mathbf{u}}$ in terms of the largest eigenvalue of $\rho_{\mathrm{in}}$, the density matrix describing the power flow and coherence across input channels[70]. Achromatic metalenses[71–76] that focus several incident beams onto the same focal spot are further restricted by causality and, thus, delay–bandwidth limitations[77].

***Optical communication.*** More broadly, limits on focusing are connected to the general theme of using light as a conduit for information and energy transfer. An early constraint related to energy transfer is Kirchhoff's law, equating the emissivity and absorptivity of an object, often associated with the second law of thermodynamics (detailed balance) and originally derived under assumptions of geometrical optics and reciprocity[78–80]. Generalizations of this concept via the formalism of a linear 'mode-converter' have been used as models of communication capacity, and are in principle capable of accounting for wave effects and non-reciprocal media[81]. In particular, information encoded in waves transferred between a source $V_{\mathrm{S}}$ and receiver region $V_{\mathrm{R}}$ in free space, can be quantified via a Frobenius norm $\int_{V_{\mathrm{S}}}\int_{V_{\mathrm{R}}} \|\mathbb{G}_0(\mathbf{x},\mathbf{x}')\|^2$ of the vacuum Green's function $\mathbb{G}_0$ connecting them. Adaptations to describe communication mediated by devices (such as lenses and multiplexers), which can strongly modify electromagnetic fields and thus 'channel capacity', remain an active area of investigation[40,82].

***Radiative heat transfer.*** A related perspective on communication can be gleaned by considering heat as a stochastic source of energy transfer. The blackbody limit as applied to radiative heat transfer constrains the flux emitted by a macroscopic body of area $A$ and temperature $T$ to be $H \leq \sigma T^4 A$, where $\sigma$ is the Stefan–Boltzmann constant[79]. However, this result is only applicable to objects where all characteristic lengths are substantially larger than the thermal wavelength $\lambda_{\mathrm{T}} = \frac{2\pi c\hbar}{k_{\mathrm{B}}T}$. It does not explain the power exchanged between two bodies held at different temperatures separated by a subwavelength vacuum gap $d$; nor does it account for the material constraints subsumed in the assumption of perfect absorption, such as the difficulty of engineering absorption over a wide bandwidth in a device of a limited size[83–85]. Just as with LDOS, the positivity of the scattered power sets material constraints on the achievable polarization response that waves originating in one body may excite in another[65], yielding an upper bound on the mutual absorption of light $\propto \frac{\|\chi_{\mathrm{S}}\|^2}{\mathrm{Im}\chi_{\mathrm{S}}} \frac{\|\chi_{\mathrm{R}}\|^2}{\mathrm{Im}\chi_{\mathrm{R}}} \int_{V_{\mathrm{S}}}\int_{V_{\mathrm{R}}} \|\mathbb{G}_0(\mathbf{x},\mathbf{x}')\|^2$ (generalizing the aforementioned communication bounds to incorporate material considerations in the source/receivers) and a corresponding upper bound on the spectrally integrated heat transfer of $H \leq \sigma T^4 A \frac{2}{7(kd)^2} \frac{\|\chi\|^3}{\mathrm{Im}\chi}$. Going further, accounting for radiative losses due to mutual scattering between bodies reveals the much tighter bound of $H \leq \sigma T^4 A \frac{8}{7(kd)^2} \frac{\mathrm{Im}\chi}{\|\chi\|} \ln[\frac{\|\chi\|^4}{4(\mathrm{Im}\chi)^2}]$. The origin of this reduced material scaling lies in the infeasibility of achieving resonant optical response for all waves[86].

***Light trapping.*** The difficulty in engineering a blackbody response is directly related to limits on the absorption of incident radiation, also known as light trapping in the context of photovoltaic applications. The Yablonovitch limit[87], originally derived via a statistical description of rays scattering off rough surfaces, posits a maximum absorption enhancement factor $F \leq 4n^2$, compared with the expected single-pass absorption $\alpha w$ of a weakly absorbing bulk film of thickness $w$ and absorption coefficient $\alpha$. The dependence on the refractive index $n$ enters via the total internal reflection angle $\theta_{\mathrm{c}} = \arcsin(1/n)$, which sets the emission cone from which light can escape. Analyses of maximum absorptivities for films of thickness $w \lesssim \lambda$ have been carried out through modal decomposition techniques[88], allowing the associated limit to be expressed as a $\frac{1}{\Delta\omega}\sum_m \sigma_{m,\max}$, with the maximum spectral absorption cross-section for each mode $\sigma_{m,\max}$ determined by specific material and geometric considerations. In the simplifying regime of a thin film supporting a single guided band for each polarization, this approach gives a limit absorption enhancement of $F \leq (\frac{\lambda}{w}\frac{\alpha_{wg}}{\alpha}\frac{1}{2n_{wg}})4n_{wg}^2$, where $n_{wg}$ is the group index of the mode(s) and $\frac{\alpha_{wg}}{\alpha}$ characterizes the spatial overlap between the mode profile and absorption layer. As examined in later sections, aside from their practical utility in predicting performance for specific geometries, such modal summations can be used to gain a qualitative understanding of achievable absorption characteristics. In contrast, limits based on maximal material response of the form $\frac{\sigma_{\mathrm{abs}}}{V} \leq \frac{\|\chi\|^2}{\mathrm{Im}\chi}$ (REF.[65]) remedy the need of geometric specificity (beyond a linear volumetric $V$ dependence), but can be shown to be loose beyond quasistatic settings, or in cases where it is not possible to achieve resonant response.

***Optical forces.*** Besides transferring energy, light can also impart force[89,90]: the elastic scattering of impinging photons on a body of much larger size than $\lambda$ transfers a momentum of $\Delta p = 2h/\lambda$. For bodies with wavelength-scale features, the impact of wave effects and nanostructuring on the scattering cross-section becomes pronounced. For quantum and thermal waves originating within bodies — often associated with van der Waals and Casimir forces — the situation is even more complicated, owing to the broadband and incoherent nature of thermodynamic fluctuations. Despite these challenges, a no-go theorem establishing the impossibility of repulsive interactions between mirror-symmetric bodies separated through vacuum[91] exists, as do bounds on Casimir–Polder forces on nanoparticles[92].

***Cross-sections.*** Finally, as can be seen from the preceding subsections, the concept of a scattering cross-section is central to a great range of electromagnetic problems, and consequently it will occupy much of our ensuing discussion. For bodies of dimensions $a$ much greater than $\lambda$, the scattering cross-section $\sigma_{\text{scat}} \propto a^2$ is known to scale like the geometric area. For electrically small dielectric particles with $a \ll \lambda$, the well-known Rayleigh scattering result is $\sigma_{\text{scat}} \propto a^6$. Resonant metallic particles in the quasistatic regime provide larger relative optical response $\propto V$, captured in the aforementioned maximal material response bounds. As discussed in later sections, more recently developed limit techniques make it now possible to interpolate between these asymptotic regimes.

***General scattering bounds.*** A common feature across the panoply of electromagnetic limits mentioned so far is the search for simplifying assumptions that 'relax' physical constraints and thereby make the analysis feasible: working in the geometric optics regime; maximizing modal contributions without regards for geometric constraints; maximizing material response by application of optical theorems based on passivity. Over the past few years, numerous works have formalized the notion of physical relaxations through the mathematical language of optimization theory[35–37,65], and it is this perspective that will dominate the subsequent discussion.

Before moving to this topic, it is important to recognize that there are other closely related lines of investigation. With regards to antenna design, considerable progress has been made in deriving flexible limits to various aspects of antenna performance by formulating the limits as the solutions of convex optimizations over possible current distributions of particular antenna geometries[93–96]. For problems in which an ideal target field distribution $\hat{\mathbf{E}}(\mathbf{x})$ is known, the design may be specified as minimizing the norm-squared deviation $\|\mathbf{E} - \hat{\mathbf{E}}\|^2$ subject to the constraint of Maxwell's equations $\mathbb{M}\mathbf{E} = \mathrm{i}\omega\mu_0\mathbf{J}$, with the Maxwell operator $\mathbb{M} = \nabla\times\nabla\times - \varepsilon_0\varepsilon(\mathbf{x})k_0^2$ (REF.[10]) and $\mathbf{J}$ being fixed sources of the problem (note that SI units are used throughout the text with $\varepsilon_0$, $\mu_0$ being the permittivity and permeability of vacuum respectively, $\omega$ is the angular frequency and $k_0$ is the vacuum wavenumber). Both the field distribution $\mathbf{E}(\mathbf{x})$ and material distribution $\varepsilon(\mathbf{x})$ are then treated as optimization degrees of freedom, resulting in a non-convex optimization problem where finding the minimum possible deviation is computationally difficult[41]. Nevertheless, the minimum deviation can be bounded by the global optimum of the convex Lagrangian dual relaxation[97], giving a limit on how closely $\hat{\mathbf{E}}$ can be realized in practice; an analogous procedure was used to obtain bounds on minimum achievable mode volumes of dielectric resonators, given constraints on device size and material[98]. More broadly, the method can be extended to any separable functions of the field at different positions $f(\mathbf{E}) = \Sigma_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{E}(\mathbf{x}))$, which can be of great relevance to design problems concerning the actualizing of specific field transformations, for example those in optical analogue computing[99]. For other types of objectives, especially if there is no rigorous way to assert that some particular field solution is optimal, it may be difficult to evaluate the form of the dual function and thereby obtain limits[41].

## Technical description

Continuing the perspective outlined at the beginning of the previous subsection, we now present a tutorial overview of one of the general (unifying) approaches by which limits on attainable electromagnetic properties may be determined via optimization theory.

***Scattering preliminaries.*** The perspective offered by scattering theory is helpful for formulating a widely applicable framework for obtaining limits on attainable response. First, by working from the basic definitions of scattering theory, the relationship between a given spatial permittivity and permeability profile (defined by a particular choice of material and geometry) and the polarization current generated in response to some incident electromagnetic field becomes explicit. Second, as scattering descriptions innately lead to integral equations, the constraints of any scattering theory are naturally organized into a hierarchy that meshes well with both physical intuition and formalization. These two aspects, taken together, help establish the crucial link between the features that may be imparted to waves through material structuring and the standard forms and techniques of optimization theory explored below. Throughout the following text, capital letters in blackboard font (such as $\mathbb{V}(\mathbf{x},\omega)$) are used for linear operators, and the constitutive relations $\mathbf{D}(\mathbf{x},\omega) = \varepsilon_0\mathbb{V}(\mathbf{x},\omega)\cdot\mathbf{E}(\mathbf{x},\omega)$ and $\mathbf{B}(\mathbf{x},\omega) = \mu_0\mathbf{H}(\mathbf{x},\omega)$ are assumed for simplicity. However, much of the subsequent development can be carried out in greater generality, for example in magnetic media, non-reciprocal media and so on[39]. It is also assumed that the boundary conditions on the domain under study, generally denoted as $\Omega$, are specified so as to enforce outgoing scattered fields[100,101]. In reference to scattering quantities, i superscripts are used for 'incident' or 'initial' fields and s superscripts for 'scattered' fields[102]. 'Total' fields are superscript free.

At the finest level of detail that the design of a photonic device may be described ('true physics'), Maxwell's wave equation associates each inhomogeneous permittivity profile $\mathbb{V}(\mathbf{x},\omega)$ with a unique differential equation,

$$\begin{aligned}&\nabla\times\nabla\times\mathbf{E}(\mathbf{x},\omega) - k_0^2\,\mathbb{V}(\mathbf{x},\omega)\cdot\mathbf{E}(\mathbf{x},\omega)\\&= \mathrm{i}\omega\mu_0\,\mathbf{J}^{\mathrm{i}}(\mathbf{x},\omega),\end{aligned} \tag{1}$$

where $k_0 = \omega/c = 2\pi/\lambda$, $\lambda$ is the free-space wavelength, and $\mathbf{J}^{\mathrm{i}}(\mathbf{x},\omega)$ is the 'free' source polarization current density. Unless $\mathbb{V}(\mathbf{x},\omega)$ exhibits special symmetries, equations such as equation (1) do not typically have closed-form solutions, meaning that it is usually not possible to fully analyse how changes to the potential $\mathbb{V}$ alter the total field $\mathbf{E}$ beyond local ('Taylor-like') expansions[103]. Yet this does not mean that the influence of $\mathbb{V}$ on $\mathbf{E}$ is completely opaque. Suppose that $\mathbf{J}^{\mathrm{i}}(\mathbf{x},\omega)=\mathbf{0}$, and that the boundary conditions on $\Omega$ are set to describe an incoming (incident) electromagnetic field with electric component $\mathbf{E}^{\mathrm{i}}(\mathbf{x},\omega)$. Introducing $\mathbb{I}_{\Omega}(\mathbf{x},\omega)$ as the dyadic identity operator, $\mathbb{I}_{\Omega}\cdot\mathbf{A}=\mathbf{A}$, equation (1) may be rewritten as

$$\begin{aligned}&\nabla\times\nabla\times\mathbf{E}(\mathbf{x},\omega)-k_0^2\,\mathbb{I}_{\Omega}(\mathbf{x},\omega)\cdot\mathbf{E}(\mathbf{x},\omega)\\&=k_0^2\,(\mathbb{V}(\mathbf{x},\omega)-\mathbb{I}_{\Omega}(\mathbf{x},\omega))\cdot\mathbf{E}(\mathbf{x},\omega).\end{aligned}\tag{2}$$

By defining $\mathbb{M}_0(\mathbf{x},\omega)=\nabla\times\nabla\times-k_0^2\,\mathbb{I}_{\Omega}(\mathbf{x},\omega)$, setting $\mathbb{R}(\mathbf{x},\omega)=\mathbb{V}(\mathbf{x},\omega)-\mathbb{I}(\mathbf{x},\omega)$, decomposing $\mathbf{E}(\mathbf{x},\omega)$ as $\mathbf{E}(\mathbf{x},\omega)=\mathbf{E}^{i}(\mathbf{x},\omega)+\mathbf{E}^{s}(\mathbf{x},\omega)$, and noting that $\mathbb{M}_0(\mathbf{x},\omega)\cdot\mathbf{E}^{\mathrm{i}}(\mathbf{x},\omega)=0$ — since the incident field is presumed to satisfy Maxwell's wave equation in free space — equation (2) becomes $\mathbb{M}_0(\mathbf{x},\omega)\cdot\mathbf{E}^{\mathrm{s}}(\mathbf{x},\omega)=k_0^2\,\mathbb{R}(\mathbf{x},\omega)\cdot\mathbf{E}(\mathbf{x},\omega)$. Hence, the scheme of equation (2) for $\mathbf{E}^{\mathrm{s}}(\mathbf{x},\omega)$ is no different from that of equation (1) as applied to vacuum and the polarization 'source' $(-\mathrm{i}k_0/Z)\mathbb{R}(\mathbf{x},\omega)\cdot\mathbf{E}(\mathbf{x},\omega)$, with $Z$ the impedance of free space $Z=\sqrt{\mu_0/\varepsilon_0}$. Accordingly, equation (2) implies an implicit integral relation on $\mathbf{E}$ mediated by the vacuum Green's function (a volume integral formulation[104,105]) that, while offering a distinct conceptual perspective[106] and possible computational advantages[107–109], is functionally equivalent to Maxwell's equations. That is, regardless of what the shape specified by $\mathbb{R}(\mathbf{x},\omega)$ actually is, the Green's function relation between the total field and 'source' stated above shows that $\mathbf{E}(\mathbf{x},\omega)$ must obey the self-referential (alternatively referred to as a Lippmann–Schwinger[110] or Liouville–Neumann series relation[102,111]):

$$\begin{aligned}\mathbf{E}^{\mathrm{s}}(\mathbf{x},\omega)&=\int_{\Omega}\mathrm{d}\mathbf{x}'\,\mathbb{G}_0(\mathbf{x},\mathbf{x}',\omega)\cdot\mathbb{R}(\mathbf{x}',\omega)\\&\quad\cdot\mathbf{E}(\mathbf{x}',\omega)\\&=\int_{\Omega}\mathbb{G}_0\cdot\mathbb{R}\cdot\mathbf{E}^{\mathrm{i}}+\iint_{\Omega}\mathbb{G}_0\cdot\mathbb{R}\cdot\mathbb{G}_0\cdot\mathbb{R}\\&\quad\cdot\mathbf{E}^{\mathrm{i}}+\ldots,\end{aligned}\tag{3}$$

where, taking $\mathbf{r}$ to be $k_0(\mathbf{x}-\mathbf{x}')$, $r=||\mathbf{r}||$, $\hat{\mathbf{r}}\otimes\hat{\mathbf{r}}$ to be the vector outer product of $\hat{\mathbf{r}}=\mathbf{r}/||\mathbf{r}||$ with itself, and $\overline{\overline{\mathbf{id}}}$ the $3\times3$ vector identity matrix,

$$\begin{aligned}\mathbb{G}_0(\mathbf{x}-\mathbf{x}',\omega)=\frac{k_0^3}{4\pi}\frac{\mathrm{e}^{\mathrm{i}r}}{r}\Bigg[&\left(1+\frac{\mathrm{i}r-1}{r^2}\right)\overline{\overline{\mathbf{id}}}\\&-\left(1+3\frac{\mathrm{i}r-1}{r^2}\right)\hat{\mathbf{r}}\otimes\hat{\mathbf{r}}\Bigg],\end{aligned}\tag{4}$$

is the vacuum Green's function for the left-hand side of equation (2) — $(1/k_0^2)\nabla\times\nabla\times\mathbb{G}_0(\mathbf{x},\mathbf{x}',\omega)-\mathbb{G}_0(\mathbf{x},\mathbf{x}',\omega)=\boldsymbol{\delta}(\mathbf{x}-\mathbf{x}')$ with the additional factor of $k_0^2$ included in equation (4) compared with the standard definition[112] so that all length quantities that appear in the volume integral relation between the fields and the Green's function may be defined relative to the wavelength. Recalling that a scattered field may be defined as the difference between incident and total fields, consistency demands that any scattered field must also be defined as the field resulting from the total polarization current density (here $\mathbf{J}(\mathbf{x},\omega)$) generated in response to the incident field (here $\mathbf{E}^{\mathrm{i}}(\mathbf{x},\omega)$). As such, equation (3) shows that $\mathbf{J}(\mathbf{x},\omega)=-\frac{ik_0}{Z}\mathbb{R}(\mathbf{x},\omega)\cdot\mathbf{E}(\mathbf{x},\omega)$. Using this result, acting on equation (3) with $-\frac{ik_0}{Z}\mathbb{R}(\mathbf{x},\omega)$ from the left gives (the Fredholm integral form[105,113])

$$\begin{aligned}&\mathbf{J}(\mathbf{x},\omega)-\int_{\Omega}\mathrm{d}\mathbf{x}'\,\mathbb{R}(\mathbf{x},\omega)\cdot\mathbb{G}_0(\mathbf{x},\mathbf{x}',\omega)\cdot\mathbf{J}(\mathbf{x}',\omega)\\&=-\frac{\mathrm{i}k_0}{Z}\,\mathbb{R}(\mathbf{x},\omega)\cdot\mathbf{E}^{\mathrm{i}}(\mathbf{x},\omega)=\mathbf{J}^{\mathrm{i}}(\mathbf{x},\omega),\end{aligned}\tag{5}$$

where $\mathbf{J}^{\mathrm{i}}$ is the 'initial' polarization current density setup in response to the initial electric field; see FIG. 3. (Note that it is also possible to derive results completely analogous to eqns. (5)–(9) using scattered electromagnetic fields[41], as opposed to the polarization current density perspective used here).

Equation (3) (equation (5), respectively) rests as the foundation of scattering theory and definition of the $\mathbb{T}$-operator[102,114] (or $\mathbb{S}$-matrix[115–117] (3), every incident electric field is related to a specific polarization current density (the polarization that it generates) by

$$\begin{aligned}\mathbb{I}_{\mathrm{R}}\mathbf{E}^{\mathrm{i}}(\mathbf{x},\omega)&=\mathbb{I}_{\mathrm{R}}(\mathbf{E}(\mathbf{x},\omega)-\mathbf{E}^{\mathrm{s}}(\mathbf{x},\omega))\\&=\int_{\mathrm{R}}\mathrm{d}\mathbf{x}'(\boldsymbol{\delta}(\mathbf{x}-\mathbf{x}')-\mathbb{G}_0(\mathbf{x},\mathbf{x}',\omega)\cdot\mathbb{R}(\mathbf{x}',\omega))\\&\quad\cdot\mathbf{E}(\mathbf{x}',\omega)\\&=\frac{\mathrm{i}Z}{k_0}\int_{\mathrm{R}}\mathrm{d}\mathbf{x}'(\mathbb{R}^{-1}(\mathbf{x}',\omega)-\mathbb{G}_0(\mathbf{x},\mathbf{x}',\omega))\cdot\mathbf{J}(\mathbf{x}',\omega),\end{aligned}\tag{6}$$

where R is the subdomain of $\Omega$ where $\mathbb{R}(\mathbf{x},\omega)\neq\mathbf{0}$ (the 'material' extent of the scattering potential) and $\mathbb{R}^{-1}(\mathbf{x},\omega)$ is the pseudo inverse of $\mathbb{R}$ (the inverse over the subdomain where $\mathbb{R}$ is non-zero). Because $\mathbb{R}$ and $\mathbb{G}$ describe causal, passive, linear system responses, the total linear operator relating $\mathbf{J}$ to $\mathbf{E}^{\mathrm{i}}$ within R must be

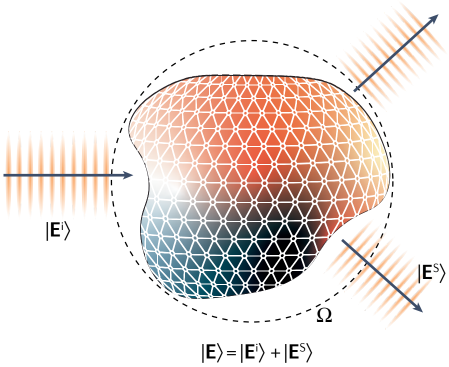
$|\mathbf{E}^{\mathrm{i}}\rangle$
$|\mathbf{E}^{\mathrm{S}}\rangle=\frac{\mathrm{i}Z}{k_0}\,\mathrm{G}_0|\mathbf{J}\rangle$
$\Omega$
$|\mathbf{E}\rangle=|\mathbf{E}^{\mathrm{i}}\rangle+|\mathbf{E}^{\mathrm{S}}\rangle$

Fig. 3 | **Schematic of scattering theory.** The basis of scattering theory broadly rests on the existence, for each particular object (scattering potential $\mathbb{V}$), of an exact relation (the $\mathbb{T}$-operator) between incident (initial or free) electromagnetic fields and total (generated or net) polarization currents. Once this relation is known, the total electromagnetic field $|\mathrm{E}\rangle$ (the solution of Maxwell's equations) is given by applying the Green's function of the free (empty) domain $\mathbb{G}_0$ to the total current associated with the incident field $|\mathrm{J}\rangle$ to obtain the scattered field $|\mathrm{E}^{\mathrm{s}}\rangle$ and summing with the incident field $|\mathrm{E}^{\mathrm{i}}\rangle$.

'invertible'[118,119]. Accordingly, equation (3) delineates the existence of the inverse relation

$$\mathbf{J}(\mathbf{x},\omega) = -\frac{\mathrm{i}k_0}{Z}\int_{\mathrm{R}} \mathrm{d}\mathbf{x}'\ \mathbb{T}(\mathbf{x},\mathbf{x}',\omega)\cdot\mathbf{E}^{\mathrm{i}}(\mathbf{x}',\omega) \tag{7}$$

defining for each unique scattering potential $\mathbb{V}$ a unique linear response function $\mathbb{T}$ that relates any incident field $\mathbf{E}^{\mathrm{i}}$ to the polarization current density $\mathbf{J}$ that self-consistently solves Maxwell's equations through equation (3). The operator relation governing the $\mathbb{T}$-operator, the Green's function, and $\mathbb{V}$ resulting from equation (7),

$$\begin{aligned}\mathbb{I}_{\mathrm{R}} = \int_{\mathrm{R}} &\mathrm{d}\mathbf{x}'(\mathbb{R}^{-1}(\mathbf{x}',\omega) - \mathbb{G}_0(\mathbf{x},\mathbf{x}',\omega))\\ &\cdot\mathbb{T}(\mathbf{x}',\mathbf{x}'',\omega),\end{aligned} \tag{8}$$

like equations (3) and (6), is fully equivalent to Maxwell's equations and serves as an advantageous starting point for deriving conserved quantities.

From equation (6), relatively little must be done to reframe the determination of an optimal scattering object $\mathbb{V}$ in terms of the polarization current density. Integrating over $\int_{\mathrm{P}} \mathrm{d}\mathbf{x}\ \mathbf{J}^{\star}(\mathbf{x},\omega)\cdot\mathbb{P}(\mathbf{x},\omega)\cdot$, where P is any known subdomain of $\Omega$, and $\mathbb{P}(\mathbf{x},\omega)$ is a local 'polarization' projection matrix — a linear response that does not mix distinct spatial points — equation (6) is transformed into

$$\begin{aligned}&\int_{\mathrm{P}} \mathrm{d}\mathbf{x}\,\mathbf{J}^{\star}(\mathbf{x},\omega)\cdot\mathbb{P}(\mathbf{x},\omega)\cdot(\underline{\mathbb{R}}^{-1}(\mathbf{x},\omega)\cdot\mathbf{J}(\mathbf{x},\omega)\\ &-\int_{\Omega}\mathrm{d}\mathbf{x}'\ \mathbb{G}_0(\mathbf{x}-\mathbf{x}',\omega)\cdot\mathbf{J}(\mathbf{x}',\omega))\\ &= -\frac{\mathrm{i}k_0}{Z}\int_{\mathrm{P}}\mathrm{d}\mathbf{x}\ \mathbf{J}^{\star}(\mathbf{x},\omega)\cdot\mathbb{P}(\mathbf{x},\omega)\cdot\mathbf{E}^{\mathrm{i}}(\mathbf{x});\end{aligned} \tag{9}$$

where, crucially, the dependence of the domains of integration on the spatial structure of $\mathbb{R}(\mathbf{x},\omega)$ has been removed. More fully, because of the inclusion of $\mathbf{J}^{\star}(\mathbf{x},\omega)$ in the integrand, and the fact that $\mathbf{J}(\mathbf{x},\omega) = \mathbf{0}$ implies that $\mathbb{R}(\mathbf{x},\omega) = \mathbf{0}$, it is completely equivalent to integrate $\mathbf{x}'$ over either R (which requires knowledge of the geometry of $\mathbb{V}(\mathbf{x},\omega)$) or $\Omega$ (which does not). By the same token, the content of equation (9) is also unchanged for any choice of response function $\underline{\mathbb{R}}^{-1}$ on $\Omega$ so long as $\underline{\mathbb{R}}^{-1}(\mathbf{x},\omega) = \mathbb{R}^{-1}(\mathbf{x},\omega)$ when $\mathbf{x}\in\mathrm{R}$. If $\mathbb{V}(\mathbf{x},\omega)$ may only take on a single $3\times3$ matrix value distinct from $\mathbf{0}$, as is usually true when designing a photonic device composed of a single material, then equation (9), with $\boldsymbol{\chi}(\omega)$ denoting the electric susceptibility matrix of the material, may be further simplified by setting $\underline{\mathbb{R}}^{-1}(\mathbf{x},\omega) = \boldsymbol{\chi}^{-1}(\omega)$; see equation (11) below. The only quantity in equation (9) that is not typically known from the outset of a design problem is $\mathbf{J}(\mathbf{x},\omega)$. At the same time, because the physics of Maxwell's equations is incorporated through $\mathbb{G}_0$, the true behaviour of the material scattering potential $\boldsymbol{\chi}^{-1}(\omega)$ is incorporated through $\underline{\mathbb{R}}^{-1}(\mathbf{x},\omega)$, and the self-consistency of viewing $\mathbf{J}(\mathbf{x},\omega)$ as the electric polarization current density resulting from $\mathbf{E}^{\mathrm{i}}(\mathbf{x},\omega)$ is incorporated by association with equation (3), any vector field that respects equation (9) for all possible choices of $\mathbb{P}(\mathbf{x},\omega)$ actually defines an effective medium scattering structure[40] (a mix between the material properties of $\mathbb{V}(\mathbf{x},\omega)$ and the background).

***Optimization bounds.*** The observation that every constraint of the form given by equation (9) applies to any structure of a given material within $\Omega$ implies that a great number of common photonic objectives can be stated as quadratically constrained quadratic programmes (QCQPs)[41], and, in turn, bounded using standard relaxation techniques from optimization theory[97]. The connection between equation (9) and QCQPs is most easily seen by switching to a more compact notation. Making use of the fact that integration may be viewed as an inner product for fields or functions[120], equation (9) may be written in bra-ket form as

$$\langle\mathbf{T}|\mathbb{P}(\boldsymbol{\chi}^{-1} - \mathbb{G}_0)|\mathbf{T}\rangle = \langle\mathbf{T}|\mathbb{P}|\mathbf{E}^{\mathrm{i}}\rangle, \tag{10}$$

with $|\mathbf{T}\rangle = \frac{\mathrm{i}Z}{k_0}|\mathbf{J}\rangle = \mathbb{T}|\mathbf{E}^{\mathrm{i}}\rangle$, and $\langle\mathbf{A}|\mathbf{B}\rangle = \int \mathrm{d}\mathbf{x}\ \mathbf{A}^{\star}(\mathbf{x})\cdot\mathbf{B}(\mathbf{x})$ denoting the standard complex-conjugate inner product, leading to the quadratic constraint equations

$$\begin{aligned}&\mathrm{Im}(\langle\mathbf{E}^{\mathrm{i}}|\mathbb{P}|\mathbf{T}\rangle) - \langle\mathbf{T}|[(\boldsymbol{\chi}^{-1\dagger} - \mathbb{G}_0^{\dagger})\mathbb{P}]^{\mathrm{A}}|\mathbf{T}\rangle = 0,\\ &\mathrm{Re}(\langle\mathbf{E}^{\mathrm{i}}|\mathbb{P}|\mathbf{T}\rangle) - \langle\mathbf{T}|[(\boldsymbol{\chi}^{-1\dagger} - \mathbb{G}_0^{\dagger})\mathbb{P}]^{\mathrm{S}}|\mathbf{T}\rangle = 0.\end{aligned} \tag{11}$$

In these expressions, and the proceeding text, S superscripts will be used to mark the Hermitian (symmetric) part of the contained linear response function $\mathbf{M}^{\mathrm{S}} = (\mathbf{M} + \mathbf{M}^{\dagger})/2$, and A superscripts will be used to denote the antisymmetric part, $\mathbf{M}^{\mathrm{A}} = (\mathbf{M} - \mathbf{M}^{\dagger})/2\mathrm{i}$, so that, like a complex number, $\mathbf{M} = \mathbf{M}^{\mathrm{S}} + \mathrm{i}\mathbf{M}^{\mathrm{A}}$. Given the freedom in choosing $\mathbb{P}$, there is no difference between $\mathbb{P}$ and $\mathbb{P}^{\dagger}$.

When $\mathbb{P}$ is set to the domain identity $\mathbb{I}_{\Omega}$, the first relation of equation (11) is a statement of the conservation of 'resistive' power within the domain[37]: the power drawn by the polarization current from the field, the inner product $\mathrm{Im}(\langle\mathbf{E}^{\mathrm{i}}|\mathbf{T}\rangle)$, must equal the sum of the power lost by the polarization current to material absorption[121],

$$\langle\mathbf{T}|(\boldsymbol{\chi}^{-1\dagger})^{\mathrm{A}}|\mathbf{T}\rangle = \int_{\Omega}\mathrm{d}\mathbf{x}\ \mathbf{T}^{\star}(\mathbf{x})\cdot\frac{\mathrm{Im}[\mathbb{V}(\mathbf{x},\omega)]}{\|\mathbb{V}(\mathbf{x},\omega)\|^2}\cdot\mathbf{T}(\mathbf{x}) \tag{12}$$

and to outgoing radiation[122]

$$\begin{aligned}\langle\mathbf{T}|\mathbb{G}_0^{\mathrm{A}}|\mathbf{T}\rangle = \iint_{\Omega}&\mathrm{d}\mathbf{x}'\mathrm{d}\mathbf{x}\,\mathbf{T}^{\star}(\mathbf{x}')\cdot\mathrm{Im}[\mathbb{G}_0(\mathbf{x}',\mathbf{x},\omega)]\\ &\cdot\mathbf{T}(\mathbf{x}).\end{aligned} \tag{13}$$

The second relation contained in equation (11), as illustrated further below, makes the related statement that 'reactive' power must also be conserved[123,124] (see REFS[36,103,125]). Correspondingly, for any $\mathbb{P}$, the equalities of equation (11) are analogous to a complex generalization of Poynting's theorem (restricted to $\mathbb{P}$)[125]. These relations impose requirements on the characteristics of $\mathbb{V}$ and spatial extent of the domain $\Omega$ (through $\mathbb{G}_0$ which is limited to $\Omega$) needed to achieve resonant response. That is, depending on the material the device will be made of and the spatial volume that it may possibly occupy, there may be strong limits on the degree to which the amplitude and phase of $\mathbf{J}$ relative to $\mathbf{E}^{\mathrm{i}}$ may be tuned.

Recollecting that the field and induced polarization currents are linearly related by $|\mathbf{E}\rangle = |\mathbf{E}^{\mathrm{i}}\rangle + \frac{\mathrm{i}Z}{k_0}\mathbb{G}_0|\mathbf{J}\rangle$, it follows that any physical process described by quadratic field terms — including the fundamental time-averaged power-transfer quantities of absorption $k_0\langle\mathbf{E}|\boldsymbol{\chi}^{\mathrm{A}}|\mathbf{E}\rangle/2Z$, extraction $\mathrm{Re}(\langle\mathbf{E}^{\mathrm{i}}|\mathbf{J}\rangle)/2$ and scattering $Z\langle\mathbf{J}|\mathbb{G}_0^{\mathrm{A}}|\mathbf{J}\rangle/2k_0$, which rest as the basic figures of merit for the design of antennas[96,126,127], light trapping devices[87,128–132] and optoelectronic coupling[9,133–135] — can be considered as a quadratic objective. In this language, taking $\mathrm{f}_0(|\mathbf{T}\rangle)$ to denote some quadratic function of the polarization, and K a complete set of constraints — an independent set $\{\mathbb{P}_k\}_{k\in\mathrm{K}}$ of local ‘polarization’ projections matrices spanning the $n\times n$ vector space of matrices for an $n$ dimensional real-number computational representation — the goal of maximizing (or minimizing) any such objective through material structuring may be formulated as

$$\begin{aligned}
&\max_{|\mathbf{T}\rangle}\left(\min_{|\mathbf{T}\rangle}\right)\mathrm{f}_0(|\mathbf{T}\rangle)\\
&\text{such that }\ \forall k\in\mathrm{K} \qquad\qquad\qquad\qquad\qquad\qquad (14)\\
&\mathrm{Re}(\langle\mathbf{E}^{\mathrm{i}}|\mathbb{P}_k|\mathbf{T}\rangle)-\langle\mathbf{T}|[\mathbb{P}_k(\boldsymbol{\chi}^{-1\dagger}-\mathbb{G}_0^{\dagger})]^{\mathbb{S}}|\mathbf{T}\rangle=0,\\
&\mathrm{Im}(\langle\mathbf{E}^{\mathrm{i}}|\mathbb{P}_k|\mathbf{T}\rangle)-\langle\mathbf{T}|[\mathbb{P}_k(\boldsymbol{\chi}^{-1\dagger}-\mathbb{G}_0^{\dagger})]^{\mathbb{A}}|\mathbf{T}\rangle=0,
\end{aligned}$$

which is a QCQP. Because the enforcement of fewer constraints always leads to maxima of greater or equal value (minima of equal or smaller value) in any optimization, the imposition of any collection of constraints that can be formed from K may be used to construct a relaxed QCQP that contains the feasible set (the set of fields that respects every imposed constraint) of equation (14) — an optimization with maxima (minima) at least as large (small) as equation (14). Any bound on such a relaxed programme is necessarily a bound on equation (14). Hence, by applying any further relaxation, such as Lagrange duality or semi-definite programming[41], it is possible to obtain limits on the physically realizable values of $\mathrm{f}_0(|\mathbf{T}\rangle)$ that universally apply to any material structure within Ω (see REFS[36–41,103,136,137]). As highlighted below, the extent to which these limits incorporate various physical phenomena may be tuned by selecting, either by intuition or algorithm[37], the collection of constraints ($\mathbb{P}_k$ projections) that are concurrently imposed. Moreover, in contrast to many traditional approaches to limits where individual components of an expression are bounded and then subsequently summed or composed to form a global bound, the optimization framework of equation (14) properly describes interactions between multiple constraints.

Although no further refinements of equation (14) will be examined hereafter, it should be noted that this basic optimization bounds approach can, and in certain cases should, be extended in at least two meaningful ways. First, by moving to complex frequencies as described elsewhere[18,138], it appears possible to adapt equation (14) to treat many broadband and temporal problems. Detailed discussions of these modifications can be found elsewhere[39,139]. Second, in situations involving multiple incident fields or scattering objectives, including applications to multifunctional devices — design objectives such as optical multiplexing[140–142], meta-optic imaging components[143–146] and optical computing[147–149] — it is necessary to broaden the scope of the quadratic equalities included in equation (14) to properly account for the additional challenges presented by the need to engineer multiple field transformations within a common structure. A full account of these alterations is given in REF.[40].

## Representative scattering limits

To build intuition and provide context, the ensuing section reviews three increasingly complex tutorial applications of equation (14) to set fundamental upper bounds on optical response. Beginning with the conservation of resistive power, focusing on the equivalent problems of maximizing thermal emission or net absorption, the modal characteristics of $\mathbb{G}_0^{\mathrm{A}}$ in relation to equation (11) are shown to reproduce familiar asymptotic results from quasistatics and ray optics. However, because the optimization framework does not rely on the validity of such approximate forms, the limits are also seen to be meaningfully applicable to intermediate and wavelength-scale regimes hitherto inaccessible to bounds calculations. Next, by further imposing that reactive power be conserved, simultaneously enforcing the two constraints in equation (11) through $\mathbb{P}=\mathbb{I}_{\Omega}$, limits on achievable scattering cross-sections are found to anticipate conditions on the size of the design domain under which resonant response is possible for a given material choice. Finally — as exemplified through calculations of bounds on scattering cross-sections, radiative emission from a dipolar source in the near field of body, and power splitting — the set of integral relations contained in the relaxations of equation (14) has the effect of defining the degree to which the physics of scattering theory is enforced, and correspondingly, the number and types of integral constraints imposed in calculating optimization bounds function as complements to the different number and types of optimization degrees of freedom that may be used in structural optimization. For proper comparison against realizable devices, constraints must be selected in a manner that ‘resolves’ the wave physics of the problem: for example, accurate limits on phenomena dominated by rapidly decaying (evanescent) fields require a greater number of local constraints. In almost all of these representative applications, objective values obtained through structural ‘topology’ (or ‘density’) optimization are found to come within an order of magnitude of their determined limit values[19].

***Resistive power conservation.*** As a first application of equation (14), we discuss how the conservation of resistive power sets an upper bound on thermal radiation and, by reciprocity, angle-integrated absorption. At a microscopic level, thermal emission is the result of stochastically fluctuating electrical currents in matter[150,151], with the precise relationship between temperature, energy dissipation and field fluctuations in an object in equilibrium determined by the fluctuation–dissipation theorem[152]. The basis for such a relation may be intuitively understood from Brownian motion[153]. A particle travelling in a fluid experiences a dissipation of its net velocity due to collisions with the constituent particles of the surrounding fluid. Complementarily, these collisions impart momentum, causing fluctuations in the position of the particle about its average location, $\langle x^2\rangle = 2Dt$, where $t$ is the elapsed time and $D$ is the diffusion coefficient, which by the Stokes–Einstein relation $D = k_B T/\gamma$ is inversely proportional to the drag (dissipation) coefficient $\gamma$ (REF.[154]). An analogous relation is seen in the Nyquist formula for Johnson noise, $\langle V^2 \rangle = 4Rk_B T d\nu$, where $V$ is the voltage between the terminals of an open circuit (such as a conductive wire), $R$ the electrical resistance, and $d\nu$ a frequency interval[155]. The fluctuation–dissipation theorem generalizes and formalizes these observations. For the electromagnetic settings considered here[112],

$$\underbrace{\{J_i(\mathbf{x},\omega)J_j^{\star}(\mathbf{x}',\omega')\}_T}_{\text{fluctuation}} = \frac{\omega\varepsilon_0}{2\pi}\coth\left(\frac{\hbar\omega}{2k_B T}\right)\underbrace{\boldsymbol{\chi}_{ij}^{A}(\mathbf{x},\omega)}_{\text{dissipation}}\,\delta_{ij}\,\delta(\mathbf{x}-\mathbf{x}')\delta(\omega-\omega'), \tag{15}$$

where $\{\ldots\}$ denotes a thermal ensemble average: fluctuations in the current density are point correlated and proportional to the dissipative part of the electric susceptibility (the optical conductivity).

Exploiting this relation and the incoherent nature of the fluctuations, the net emitted power may be expressed as a sum over independent radiative channels. Generally, the instantaneous power emitted by a current source is

$$P_{\text{rad}} = -\int_{R} \mathbf{J}(\mathbf{x},t)\cdot\mathbf{E}(\mathbf{x},t) \tag{16}$$

where the minus sign results from the convention of emitted power. Switching to the spectral domain (the convention used here is $\mathbf{J}(\mathbf{x},t) = \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\mathbf{J}(\mathbf{x},\omega)$), and assuming that the collection of fluctuating dipolar sources distributed throughout the body satisfies equation (15), the thermal power radiated by a body held at a constant temperature $T$ (see REFS[114,122,156]) is given by

$$\{P_{\text{rad}}\}_T = \int_0^{\infty} d\omega\, \Pi(\omega,T)\underbrace{\frac{2}{\pi}\text{Tr}\left[\mathbb{G}_0^{A}(\mathbb{T}^{A} - \mathbb{T}^{\dagger}\mathbb{G}_0^{A}\mathbb{T})\right]}_{\Phi(\omega)} \tag{17}$$

where $\Pi(\omega,T) = \frac{\hbar\omega}{\exp(\hbar\omega/(k_B T)) - 1}$ is the Planck thermal occupation function, and $\Phi(\omega)$ the corresponding angle-integrated spectral transfer function (absorption or emission) of the body; the Tr symbol denotes a trace over both the position and polarization indices of the dipole sources, which together is the complete set of indices of the enclosed operators.

In the breakup of equation (17), the $\mathbb{T}^{A} - \mathbb{T}^{\dagger}\mathbb{G}_0^{A}\mathbb{T}$ term contained in $\Phi$ constitutes an algebraic description of absorption, equation (12), expressed as the subtraction of radiated power $\mathbb{T}^{\dagger}\mathbb{G}_0^{A}\mathbb{T}$, equation (13), from the total extracted (extinction) power $\mathbb{T}^{A}$. This association is not accidental: as a consequence of reciprocity, evaluating the trace over a (delocalized) basis of waves incident on the body changes the interpretation of $\Phi$ from the net emitted power due to dipolar sources within the object (thermal emission) to the net power absorbed in the body due to incident plane waves (angle-integrated absorption), but the algebraic form of $\Phi$ remains unaltered. Because $\mathbb{G}_0^{A}$ describes how outgoing radiation carries power away from an object into the surrounding environment[118,122], a natural basis in which to evaluate the trace is the eigenmode expansion

$$\mathbb{G}_0^{A} = \sum_n \rho_n |\mathbf{Q}_n\rangle\langle\mathbf{Q}_n|, \tag{18}$$

with each of the radiative coefficients $\rho_n$ non-negative by passivity. Setting $|\mathbf{T}_n\rangle = \mathbb{T}|\mathbf{Q}_n\rangle$ ($|\mathbf{J}_n\rangle = (-ik_0/Z)|\mathbf{T}_n\rangle$ is the resulting electric polarization current density in the object resulting from the $n$th radiative mode), $\Phi$ becomes

$$\Phi = \frac{2}{\pi}\sum_n \rho_n\left(\text{Im}[\langle\mathbf{Q}_n|\mathbf{T}_n\rangle] - \langle\mathbf{T}_n|\mathbb{G}_0^{A}|\mathbf{T}_n\rangle\right). \tag{19}$$

Even without the imposition of a single constraint, the form of equation (19) places fairly strong restrictions on the extent to which the net absorption (emission) cross-section of an object can be enhanced compared with its geometric cross-section[122]. To optimize absorption, it is clear from equation (19) that each radiative mode must generate a strong polarization: $\text{Im}[\langle\mathbf{Q}_n|\mathbf{T}_n\rangle]$ is the extracted power. However, the generation of these currents necessarily leads to radiative losses, $\langle\mathbf{T}_n|\mathbb{G}_0^{A}|\mathbf{T}_n\rangle$, which grow relatively in strength as the size of the domain increases through the growth of the $\rho_n$ radiative coupling coefficients[86,122].

As a first example of optimization bounds, we analyse the maximization of $\Phi$ subject to the constraint that resistive power is conserved:

$$\begin{aligned} &\max_{\{|\mathbf{T}_n\rangle\in\Omega\}} \frac{2}{\pi}\sum_n \rho_n\left(\text{Im}[\langle\mathbf{Q}_n|\mathbf{T}_n\rangle] - \langle\mathbf{T}_n|\mathbb{G}_0^{A}|\mathbf{T}_n\rangle\right)\\ &\text{such that } \forall n\\ &\text{Im}[\langle\mathbf{Q}_n|\mathbf{T}_n\rangle] - \langle\mathbf{T}_n|(\boldsymbol{\chi}^{-1\dagger} - \mathbb{G}_0^{\dagger})^{A}|\mathbf{T}_n\rangle = 0, \end{aligned} \tag{20}$$

effectively the simplest version of equation (14). Owing to the form of $\Phi$, the only difference between the objective and constraint in equation (20) is the material loss term $||\mathbf{T}_n||^2/\zeta$, with the factor $\zeta \equiv ||\chi||^2/\text{Im}[\chi]$ quantifying the maximum magnitude that the polarization current density can achieve relative to the incident electric field[157]. Simply, to maintain equilibrium, the net (integrated) power extracted by the object at each frequency must be perfectly balanced by the sum of two possible loss mechanisms: the absorption of power into material degrees of freedom ($\boldsymbol{\chi}^{-1\dagger A}$), here considered to be an infinitely large thermal bath[158], and power re-radiated (scattered or reflected) back into the ambient environment ($\mathbb{G}_0^{A}$).

Applying the relaxation of Lagrange duality (see elsewhere[41,97,103,159,160]), the optimal objective value of equation (20) can be expressed as

$$\Phi_{\text{opt}} = \frac{1}{2\pi}\sum_n \begin{cases} 1 & \zeta \geq \dfrac{1}{\rho_n} \\ \dfrac{4\zeta\rho_n}{(1+\zeta\rho_n)^2} & \zeta < \dfrac{1}{\rho_n} \end{cases} \tag{21}$$

$$\equiv \frac{1}{2\pi}\sum_n \begin{cases} 1 & \tau_{\text{abs},n} \geq \tau_{\text{rad},n} \\ \dfrac{4\tau_{\text{abs},n}\tau_{\text{rad},n}}{(\tau_{\text{rad},n}+\tau_{\text{abs},n})^2} & \tau_{\text{abs},n} < \tau_{\text{rad},n} \end{cases}, \tag{22}$$

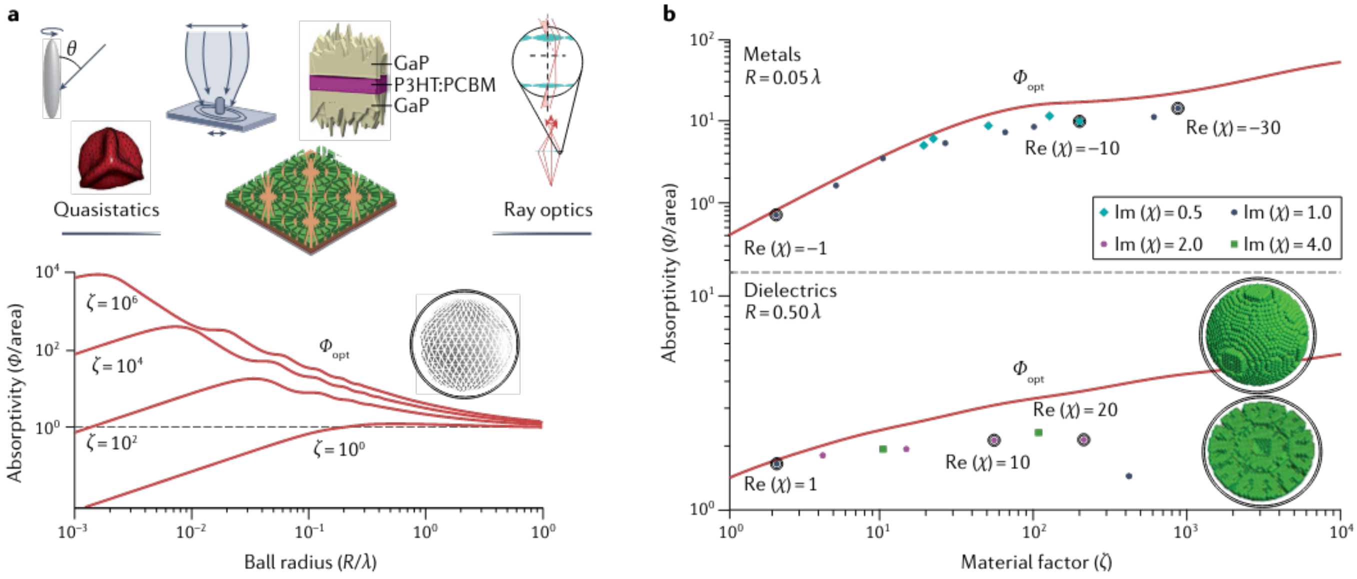

Fig. 4 | **Bounds on angle-integrated absorption based on the conservation of resistive power. a** | Illustration of the results of equation (20) as a function of the size (radius $R/\lambda$) of a spherically bounded design volume for several values of the material factor $\zeta = \|\chi\|^2/\mathrm{Im}(\chi)$. Notably, in transitioning between very small design volumes, $R \ll \lambda$, and very large design volumes, $R \gg \lambda$, equation (20) is seen to smoothly blend the familiar $\propto V$ and $\propto A$ limits of quasistatic and ray-optic (blackbody) approximations. Above panel **a**, a collection of representative use cases of radiative absorption is included to give a sense of the range of length scales covered below. Working from left to right, these images represent electrically small antennae with $\theta$ the incident angle of incoming radiation (reproduced with permission from REF.[183], Royal Society of Publishing), metallic nanoparticles (reproduced with permission from REF.[121], APS), nanostructured solar cells (reproduced with permission from REF.[135], under a Creative Commons licence CC BY 4.0, REF.[88], PNAS, REF.[130], ACS), and lensing (reproduced with permission from REF.[184], IOP). Panel **b** depicts a comparison of $\Phi_{opt}$ against performance values obtained by computational design methods for two set domain sizes, $R/\lambda = 1/20$ for metals ($\mathrm{Re}(\chi) < 0$) and $R/\lambda = 1/2$ for dielectrics ($\mathrm{Re}(\chi) > 0$). Unless additional constraints are included, findings for dielectric materials limited to small domains, and findings for metals in large domains, exhibit substantially larger disagreement. PCBM, phenyl-C61-butyric acid methyl ester; P3HT, poly(3-hexylthiophene). The lower part of panel **a** is adapted with permission from REF.[122], APS; panel **b** is adapted with permission from REF.[122], APS.

with radiative lifetime $\tau_{\mathrm{rad},n} \equiv 1/\rho_n$ and material absorption lifetime $\tau_{\mathrm{abs},n} \equiv \zeta$ under a coupled-mode analysis (see Supplementary Information). The surprising simplicity of $\Phi_{opt}$ as arising from a sum over independent channel contributions follows from the fact that, by imposing only the resistive power conservation constraint, the optimal bound polarization currents end up proportional to the radiative states (the eigenbasis of $\mathbb{G}_0^{\mathrm{A}}$), $|\mathbf{T}_n\rangle = c_n|\mathbf{Q}_n\rangle$, with the maximum bound polarization response $\|c_n\| \leq \min\{\frac{1}{2\rho_n}, \zeta\}$ in each channel.

A trio of plots of $\Phi_{opt}$, bounding angle-integrated absorption (equivalently emission) for bodies enclosed in a spherical ball of radius $R$, is shown in FIG. 4. Beyond the exact quantitative predictions appearing in panel **a**, and the excellent agreement with computationally designed structures appearing in panel **b** (see REFS[108,122] for details), it is seen that the mere conservation of net resistive power is sufficient for equation (20) to inherently reproduce fundamental quasistatic and blackbody behaviour. In the limit of a small design volume, $\zeta\rho_n \ll 1$ ($\tau_{\mathrm{abs},n} \ll \tau_{\mathrm{rad},n}$) for all $n$, $\Phi_{opt}$ is seen to exhibit a volumetric scaling consistent with the assumption that the magnitude of all generated polarization currents can grow as large as material loss allows: as the volume grows, so does the available power in each channel, and hence so should the polarization response. However, owing to the necessary coupling of these currents with radiative states, volumetric growth cannot persist indefinitely. Eventually, in each index of equation (19), growth in $\rho_n$ (decay in $\tau_{\mathrm{rad},n}$) causes radiative losses to discordantly overwhelm net extracted power if the magnitude of $|\mathbf{T}_n\rangle$ (the material lifetime $\tau_{\mathrm{abs},n}$) becomes too large, leading the associated channel (index) to enter the saturation condition of equation (22), visible in FIG. 4 as the onset of steps. As an increasing number of channels saturate, volumetric scaling begins to asymptotically transition to area scaling, regardless of the supposed value of $\zeta$. Directly, the blackbody limit equating absorption and geometric cross-sections appears out of equation (20), irrespective of assumed material properties.

The behaviours of the $\rho_n$ radiative expansion coefficients also have interesting implications for antenna design[95,161]. For any excitation contained within Ω, the largest radiative coefficient $\rho_n$ in equation (18) sets a lower bound on the radiative lifetime, $\tau_{\mathrm{rad}}$, and consequently a lower bound on the radiative quality factor $Q_{\mathrm{r}} = \omega\tau_{\mathrm{rad}}/2$. In the limit of a deeply subwavelength design volume, the largest $\rho_n$ (the dipole coefficient) scales $\propto (k_0R)^3$, setting a lower bound on the radiative quality factor $\propto (k_0R)^{-3}$ (REFS[162–164]). This lower bound on $Q_r$, known as the Chu limit ($Q_{\mathrm{rad}} \geq \frac{1}{k_0R} + \frac{1}{(k_0R)^3}$ (REF.[164]), correspondingly sets an upper limit on antenna bandwidth that restricts processing speed[165].

Finally, as suggested above, the sum expression for $\Phi_{opt}$ given by equation (22) is completely analogous to a modal description of absorption by non-interacting excitations and, as one might expect, under this analogy the channel saturation condition of $\zeta\rho_n = 1$ is exactly the rate-matching condition $\tau_{\mathrm{rad},n} = \tau_{\mathrm{abs},n}$ between radiative and material absorption lifetimes. Consequently, it is accurate to interpret equation (22) as a model wherein an idealized object independently extracts power from each of the radiative (multipole) states in a physically optimal way, with $\rho_n$ and $\zeta$ setting limits on the associated coupled-mode radiative and dissipation rates for every channel. As discussed in greater detail below, this perspective implicitly assumes that resonant response is achievable in each individual channel, and that attaining resonant response in one channel has no implications for any other (a consequence of ignoring constraints imposed by the conservation of reactive power). Neither assumption is typically true in practice.

***Total power conservation.*** Given the remarkable agreement observed in FIG. 4b between the absorption characteristics of structures obtained by computational techniques and the associated bounds, it is natural to wonder whether related conclusions can be drawn for all basic scattering quantities; namely, to what extent are the results of current inverse methods explained by the basic necessity of conserving power?

Recalling that the power scattered from an initial electric field $|\mathbf{E}^{\mathrm{i}}\rangle$ at a single frequency $\omega$ is[102,103]

$$P_{\mathrm{sct}} = \frac{k_0}{2Z}\langle \mathbf{E}^{\mathrm{i}}|[\mathbb{T}^{\mathrm{A}} - \mathbb{T}^{\dagger}(\mathbb{V}^{-1\dagger})^{\mathrm{A}}\mathbb{T}]|\mathbf{E}^{\mathrm{i}}\rangle, \tag{23}$$

the problem of maximizing the scattering cross-section of an object contained within a design volume $\Omega$ becomes the optimization statement

$$\begin{aligned}
&\max_{|\mathrm{T}\rangle \in \Omega} \frac{k_0}{2Z}[\mathrm{Im}(\langle \mathbf{E}^{\mathrm{i}}|\mathbf{T}\rangle) - \langle \mathbf{T}|\boldsymbol{\chi}^{-1\dagger\mathrm{A}}|\mathbf{T}\rangle]\\
&\text{such that}\\
&\mathrm{Im}(\langle \mathbf{E}^{\mathrm{i}}|\mathbf{T}\rangle) - \langle \mathbf{T}|(\boldsymbol{\chi}^{-1\dagger} - \mathbb{G}_0^{\dagger})^{\mathrm{A}}|\mathbf{T}\rangle = 0,\\
&\mathrm{Re}(\langle \mathbf{E}^{\mathrm{i}}|\mathbf{T}\rangle) - \langle \mathbf{T}|(\boldsymbol{\chi}^{-1\dagger} - \mathbb{G}_0^{\dagger})^{\mathrm{S}}|\mathbf{T}\rangle = 0,
\end{aligned} \tag{24}$$

where $|\mathbf{E}^{\mathrm{i}}\rangle$ is the electric field of an incident plane wave and, as before, $|\mathbf{T}\rangle = \mathbb{T}|\mathbf{E}^{\mathrm{i}}\rangle$. If only the conservation of resistive power is imposed, the solution of equation (24) closely mirrors the coupled-mode expression for integrated emission given by equation (22). However, as is clear from a comparison of the solid (enforcing total power conservation) and dashed (enforcing only resistive power conservation) lines of FIG. 5 (particularly in panel **b**), the inclusion of the global reactive power constraint, the final line of equation (24), leads to substantially tighter limits, with the cost that the solution of equation (24) does not have a simple semi-analytic form (except for the single channel asymptotic examined in the Supplementary Information). The physical mechanism underlying these differences is the appearance of phase information. Paralleling the well-known response characteristics of a simple harmonic oscillator, when only the conservation of resistive power is enforced there is a bound on the relative magnitude of $|\mathbf{T}\rangle$ set by material absorption and radiative emission. All information about the relative phase offset of the response, partially set by $\mathrm{Re}[\chi(\omega)]$, is ignored. Once the need to conserve reactive power is taken into account, physical restrictions are placed on both the relative magnitude and phase of the response[36,103,125]. These restrictions can either limit, or even exclude, resonant response in certain situations. Here, by resonant response, we mean that, under sufficiently small changes to the material properties of a design, there is a field observable that scales roughly $\propto \zeta = \|\chi(\omega)\|^2/\mathrm{Im}[\chi(\omega)]$.

Most notably, as confirmed by FIG. 5b, when confined to a spherical subwavelength domain, the largest possible scattering cross-section that can be achieved by structuring a dielectric material is exactly the $\propto V^2$ Rayleigh scattering of a ball[166] (see Supplementary Information and REF.[103] for further details).

The inclusion of reactive power also implies considerable alterations to the mathematical model and interpretation of equation (24) as compared with equation (20). Precisely, the sum form of $\Phi_{\mathrm{opt}}$ given by equation (22) arises because both the objective and constraint of equation (20) are diagonalized in an eigenbasis of $\mathbb{G}_0^{\mathrm{A}}$. When further physical constraints are imposed (such as global reactive power or the local constraints introduced in the next set of examples) simultaneous diagonalization is rarely possible. In part indicating the richness of the physics being described, the symmetric and antisymmetric components of the Green's function, $\mathbb{G}_0^{\mathrm{S}}$ and $\mathbb{G}_0^{\mathrm{A}}$, do not share a common eigenbasis[38,167]. Hence, whenever response characteristics are not dominated by the conservation of resistive power, it is generally not possible to describe scattering phenomena in terms of an orthogonal basis of weakly interacting modes. Once both aspects of the Green's function are included, radiative channels are mixed both among themselves and with non-radiative states. Relatedly, although the content discussed in the Supplementary Information stands as an exception, there is typically no simple way to use optimization bounds as means of doing parameter extraction for coupled mode theory.

***Local constraints.*** Extrapolating from the last two examples, enforcing that total power must be conserved is found to capture almost all relevant physical effects that limit achievable scattering characteristics for propagating waves — far-field applications such as maximizing plane-wave absorption, thermal radiation[122] and scattering cross-sections[36,37,103,168]. However, when applied to objectives governed by rapidly varying (evanescent) fields or multiple length scales, these coarse characterizations of the integral-wave relations may miss important aspects of the problem, leading to bounds with little connection to reality[38,39]. Specifically, contrasting equation (24) with the system of equations that must be solved to approximate Maxwell's equations numerically, it is clear that imposing global power conservation (a total of two constraints) cannot possibly capture all relevant wave features: the use of $\mathbb{P} = \mathbb{I}_{\Omega}$ in equation (11) only guarantees that scattering theory is true on average over the volume of the domain. At any spatial point, the solution resulting from an optimization problem statement such as equation (24) may violate equation (6), so long as these violations cancel each other when integrated over the entire scattering domain. To remedy

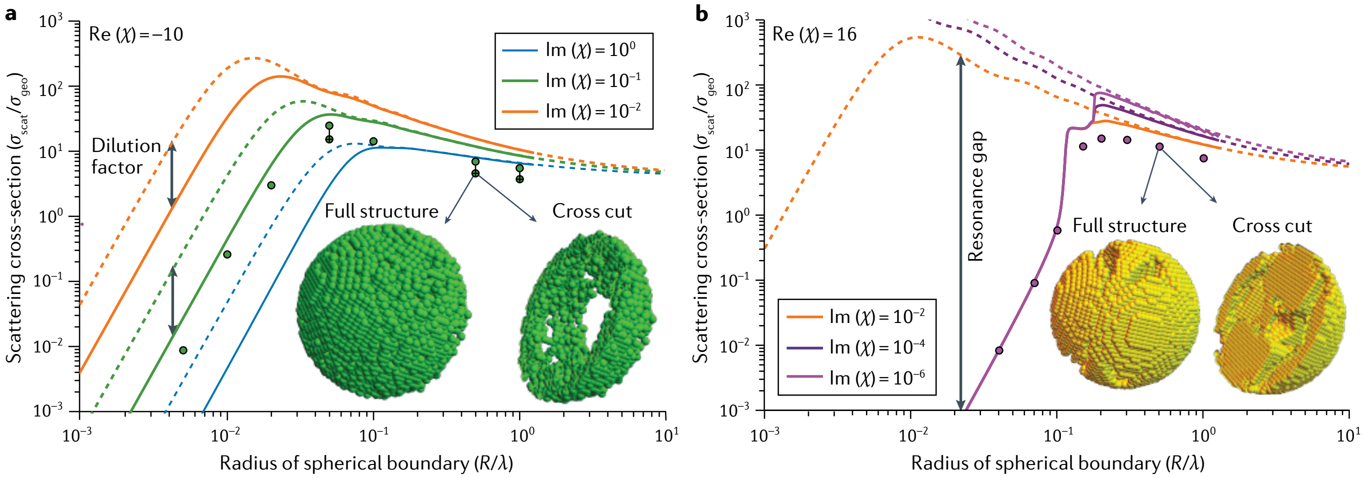

Fig. 5 | **Bounds on scattering cross-sections based on the conservation of total power. a,b** | Results of equation (24) as a function of the size (radius $R/\lambda$) of a spherically bounded design volume for two representative values of $\mathrm{Re}[\chi(\omega)]$ near optical frequencies (a near plasmonic metal, **a**, and a strong dielectric, **b**). Dashed lines result from imposing only the conservation of resistive power; solid lines result from additionally imposing the conservation of reactive power. The dots appearing in both panels mark scattering cross-sections achieved in actual geometries discovered by numeric (inverse) design, for $\chi = -10 + \mathrm{i}10^{-1}$ and $\chi = 16 + \mathrm{i}10^{-6}$ respectively. For the metal structures in **a**, aligned cross-hatched dots result from binarizing the discovered permittivity profiles, which are otherwise allowed to take on 'greyscale' values[19]. Two sample structures are shown as insets, with the plane wave incident from the more solid side of both designs (from the left in **a**, from the right in **b**) and aligned along the left–right symmetry axis. Again, in the limit that $R/\lambda \gg 1$ or $R/\lambda \ll 1$ the calculated bounds approach the scaling predictions of ray optics (geometric cross-sections) and quasistatic scattering ($\propto V^2$), irrespective of the electric susceptibility. The inclusion of reactive power conservation for strong metals ($\mathrm{Re}[\chi(\omega)] \ll -3$) confined to small design volumes ($R/\lambda \ll 1$) shows that the structuring needed to achieve resonant scattering away from the plasmon condition of $\mathrm{Re}[\chi(\omega)] = -3$ reduces achievable material scaling characteristics — the appearance of 'dilution factors'. The effect of total power conservation on obtained limits for dielectric materials confined to subwavelength domains is more radical — causing the appearance of 'resonance gaps' — leading to the general conclusion that scattering cross-section enhancements surpassing about 200 should not be expected for near optical frequencies. Panels **a**, **b** are reproduced with permission from REF.[103], under a Creative Commons licence CC BY 4.0.

this issue, additional physics in the form of localized constraints incorporating higher spatial resolution can be included (see FIG. 6).

Owing to the nature of the relaxation techniques commonly used to solve optimization bounds (see REFS[38,41]), the spatial oscillation of such local violations of 'true' physics tend to track the spatial oscillations of the incident field, as seen in FIG. 6b, and accordingly, by imposing localized projections ($\mathbb{P} = \mathbb{I}_{\mathrm{P}}$ where $\mathrm{P} \subset \Omega$) in equation (11) targeting a specific region of violation, it is usually possible to guide optimization limits towards increasingly physical characteristics: the vacuum Green's function $\mathbb{G}_0$ does not propagate all information equally, but rather blurs rapid field fluctuations as the point of observation moves away from the source. As such, for each design problem there are characteristic lengths below which differences between field solutions have no pragmatic relevance. If independent constraints enforce 'averaged' physics on a grid finer than the smallest of these length scales, then the solution of equation (14) should differ little from what is realizable in practice[169]. Directly, the number and distribution of local constraints can be viewed as 'tuning knobs' incorporating physics at the expense of computational complexity, with the caveat that the bounds are generally evaluated by convex relaxation instead of the exact solution to equation (14).

These ideas are exemplified by the minimum-radius limits on a 2D power splitter, distributing power from an incident wave equally between $2n+1$ cylindrical wave channels, depicted in FIG. 6a (REF.[39]). In asserting only the conservation of global power, the minimum diameter asymptotes to $\sim\lambda/3$, a size at which it begins to become possible for a non-physical response oscillating near $\lambda/\sqrt{\mathrm{Re}(\chi)+1}$ to satisfy global power conservation while maintaining large local violations. Conversely, when local constraints are added, this asymptote disappears, and the required radius (suggesting increasing device complexity) begins to grow steadily with the number of power divisions desired. In FIG. 6b, the strict use of global constraints similarly suggests that shortly after attaining a resonance criteria of $R \gtrsim \lambda/5$, by structuring a material with $\mathrm{Re}(\chi) = 16$ within a spherical ball of radius $R$, it is possible to enhance the scattering cross-section inversely proportional to material loss ($\mathrm{Im}(\chi)$). No such scaling is found in computationally synthesized structures, marked in the figure by dots, and under the imposition of eight (evenly spaced) radial shell constraints this non-physical feature all but vanishes.

Stemming from the need to properly describe subwavelength field characteristics, local constraints are also generally required to formulate relevant limits on near-field phenomena. Following FIG. 6c — bounds on the maximization of radiative Purcell enhancement for a dipolar current source separated from an arbitrary device by a distance $d$, again contained within a spherical ball of radius $R$ — when only global constraints are considered enhancement is seen to scale $\propto 1/\mathrm{Im}\chi$ when $d \ll \lambda$ (light lines), as is characteristic of material loss limited resonant response[65]. Upon inclusion of localized constraints defined over concentric spherical shells (dark lines), the maximal radiation is observed to drop by several orders of magnitude and no longer grows with decreasing Im$\chi$, confirming the challenge of achieving optimal polarization fields capable of resonantly scattering evanescent fields into propagating waves. Conversely, even with many more localizing shells (see REF.[38] for details), limits on larger domains and stronger dielectrics show no such saturation and exhibit $\propto d^{-3}$ diverging growth — implying that arbitrarily high angular momentum states can be out-coupled with nearly fixed efficiency. Because of the resolution arguments stated above, if additional non-symmetric localized constraints were to be imposed, particularly in the region of the design volume nearest the dipole, tighter limits may be anticipated.

## Outlook

As exemplified by the preceding discussion and representative examples, the nascent development of a methodology combining physical constraints with optimization theory has already proved to be a versatile and fruitful framework for understanding and computing electromagnetic limits. Still, challenges and opportunities remain.

Although this Review focuses on photonics, the core ideas behind the method are by no means restricted to phenomena governed by Maxwell's equations, and have direct analogues in other domains of wave physics, such as acoustics and quantum mechanics. Even more broadly, the general concept of obtaining performance limits by rigorously bounding a simpler problem through relaxed physics is applicable to essentially any field of engineering. (See elsewhere[139] for applications to quantum control).

As it currently stands, the framework requires physical objectives to be formulated as quadratic functions of the polarization fields. This excludes various photonic problems, such as nonlinear processes (the Kerr effect, for example) or objectives which may be more naturally expressed as eigenproblems (such as the maximization of bandgaps[170] or the engineering of Dirac points[171]). More work is needed to explore alternative approaches to these problems and broaden the applicability of the framework.

Another avenue for further research, particularly relevant to device applications, is the incorporation of fabrication constraints into limit calculations. The limit framework currently does not take device fabricability into account, in contrast to inverse design where there are methods that impose minimum feature size and/or connectivity requirements to produce structures amenable to nanofabrication[172–175]. Adding fabrication constraints into the framework may lead to tighter limits for practical engineering scenarios.

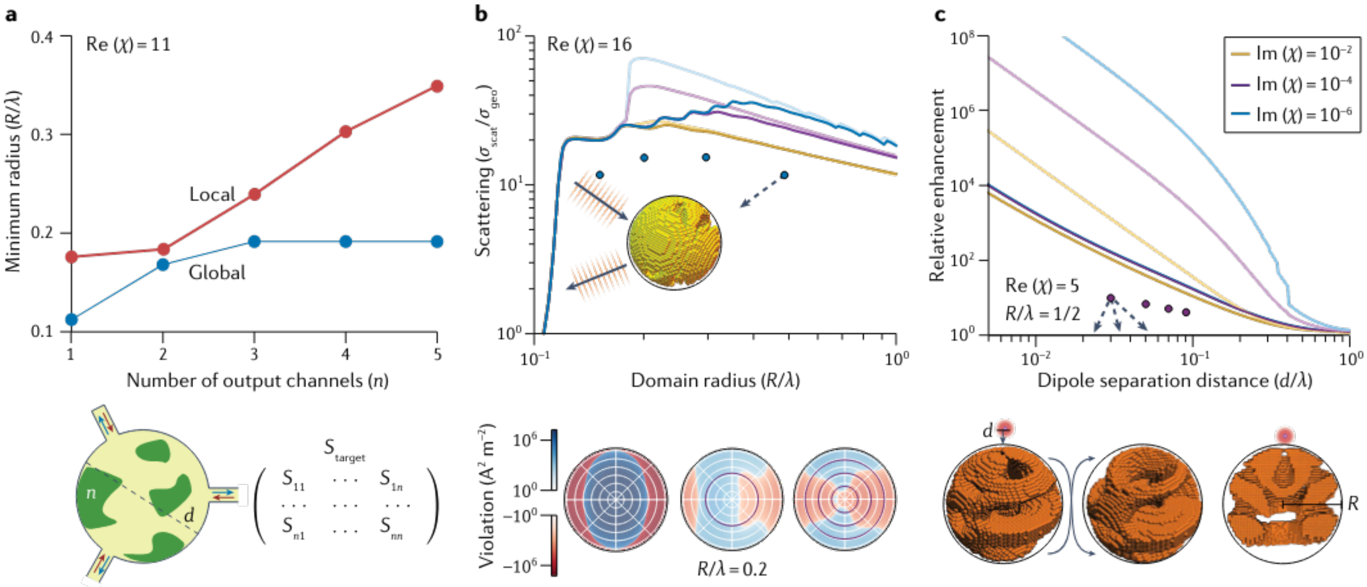

Fig. 6 | **Impact of localized constraints. a–c** | The figure shows upper bounds under the enforcement of local constraints on (**a**) the minimum radius $R$ of a cylindrical design domain necessary to encode a 'power-splitting' $\mathbb{S}$ matrix, evenly distributing power from a single plane-wave input into $2n+1$ outgoing radiative channels; (**b**) plane-wave scattering as described in the previous subsection; and (**c**) radiative Purcell enhancement for an electric dipole source in the near field of an object. Lighter-coloured lines in **b** and **c** result from imposing the global conservation of resistive and reactive power. Like-coloured darker lines are obtained by enforcing local conservation constraints over radial shell subdomains (varying distribution with application). As in previous plots, dots in **b** and **c** mark actual enhancement values achieved by structures discovered by computational methods. Views of one such design are given below **c**. The logarithmic colour maps included below **b** indicate the local violation of the conservation of reactive power for one, two and four evenly spaced shell subregion constraints, based on the optimal polarization field extracted from the solution to the Lagrange duality relaxation of equation (14). It can be seen that introducing additional subregion constraints lowers the overall magnitude of the violations compared with the case where only constraints over the entire design region are included. However, pointwise violations remain, since the constraints are not enforced at every point, but integrated over the volume of the shell subregions. The extent to which the relaxed solution approaches a physically feasible optimal solution of equation (14) with the introduction of more detailed constraints is a topic of ongoing research (see Outlook section). The schematic shown below **a** conceptualizes the engineering of a device to realize a particular scattering matrix. In all three panels, additional constraints lead to increasingly realistic field features and limit values. Panel **a** is adapted with permission from REF.[39], APS; panel **b** is adapted from REF.[38], under a Creative Commons licence CC BY 4.0; panel **c** is adapted from REF.[38], under a Creative Commons licence CC BY 4.0.

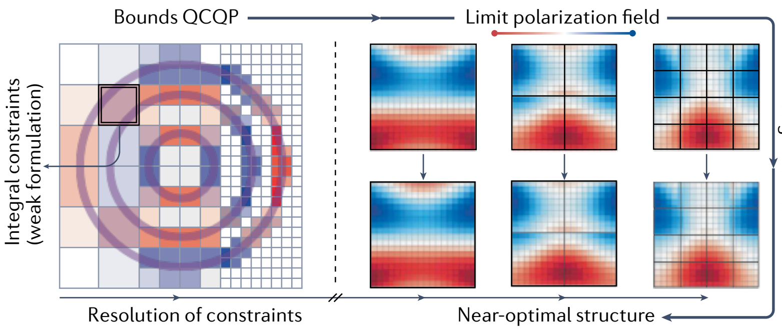

Fig. 7 | **From performance bounds to structural optimization.** Through the common denominator of spatial divisions, a complementary relationship exists between inverse design and performance bounds. In inverse design, the non-convex optimization problem of structural optimization is heuristically approximated by particular designs representing local optima, with resulting device performance generally increasing with additional structural degrees of freedom. In optimization bounds, a coarse-grained version of the same optimization problem is bounded by exploiting some convex relaxation in order to determine a value that must be respected by all allowed designs, with resulting limits generally becoming tighter under the use of additional constraints. Consequently, as the number and resolution of the local constraints going into a bound quadratically constrained quadratic programme (QCQP) increases, its solution field typically becomes an increasingly better approximation of the polarization field of a globally optimal structure. This observation suggests a potential, presently unexplored, workflow for the computational design of photonic devices: evaluate the limits of an application to a tightness wherein the approximate field distribution can be extracted and used as a starting point for the discovery of near optimal structures. Two artistic interpretations of this workflow are depicted by the left- and right-hand sides of the figure. On the left, the true extent to which the limit polarization field (red–white–blue colours) is determined is shown to converge to a simplistic underlying optimal structure with increasing constraint resolution. On the right, the polarization fields are representative of what would realistically come out of computations — fully resolved fields and structure converging towards some final optimal structure as the resolution of the imposed constraints is improved.

It is not yet understood whether, or under what conditions, the convex relaxation techniques that are used to obtain optimization limits can be guaranteed to solve their associated QCQPs; for example, whether the limit is equal to the true QCQP solution or just some larger (smaller) value. Non-affine equality constraints such as the conservation of resistive and reactive power are non-convex, and QCQPs with non-convex constraints are not generally thought to be solvable by any convex relaxation[41,176]. However, a vast majority of investigations have found that calculated limit fields are in fact optimal solutions of the initial optimization problem statement[35,36,38], with inclusion of large numbers of local constraints only resulting in numerical ill-conditioning. Such guarantees are not merely a theoretical exercise. So long as the underlying QCQP is actually solved, the addition of increasingly fine localized constraints in the bounds computations leads to progressively better approximations to an optimal realizable polarization field, and this information could be leveraged to great effect in inverse design. For instance, it could be used as a starting point for adjoint optimization to recover a near optimal structure (FIG. 7), or as a guide for the design parameters that should be considered to possibly realize improved performance characteristics. Furthermore, it is probably not necessary to enforce local constraints down to the computational pixel level to make use of these potentially powerful connections; a coarse distribution may be enough to extract an approximate optimal structure. Indeed, the onset of ill-conditioning with finer local constraints suggests that, depending on the design problem, there is a characteristic length scale beyond which more detailed structuring becomes unnecessary.

On a related note, although the bounds computation is convex, it is not necessarily easy to solve numerically. Besides ill-conditioning resulting from the imposition of large numbers of local constraints, numerical instabilities also occur in systems with low loss (such as semi-transparent media). To address this, it may be possible to formulate alternate constraints better suited to handle low-loss/lossless design problems; one example is given by REF.[168], although the proposed constraints appear to provide non-trivial limits only for small design regions and low dielectric contrasts.

The large scale of most practical photonics problems also poses a challenge: current demonstrations of the framework are restricted to 2D[39–41,168] or highly symmetric domains in 3D, exploiting efficient spectral basis representations[35,36,177] to limit the size of the associated system matrices. Beyond these early proof-of-concept explorations, there is much room for development of standardized packages for evaluating limits (possibly in conjunction with inverse methods) on 3D photonic devices by exploiting general-purpose techniques — numerical Maxwell solvers using localized bases — and open-source optimization methods[178].

Published online 1 July 2022

1. Yariv, A. & Yeh, P. *Photonics: Optical Electronics in Modern Communications* (Oxford Univ. Press, 2006).
2. Oh, S.-H. et al. Nanophotonic biosensors harnessing van der Waals materials. *Nat. Commun*. https://www.nature.com/articles/s41467-021-23564-4 (2021).
3. Zhang, S. et al. Metasurfaces for biomedical applications: imaging and sensing from a nanophotonics perspective. *Nanophotonics* **10**, 259–293 (2021).
4. Garnett, E. C., Ehrler, B., Polman, A. & Alarcon-Llado, E. Photonics for photovoltaics: advances and opportunities. *ACS Photonics* **8**, 61–70 (2020).
5. Brunner, D., Marandi, A., Bogaerts, W. & Ozcan, A. Photonics for computing and computing for photonics. *Nanophotonics* **9**, 4053–4054 (2020).
6. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* https://www.nature.com/articles/s41566-020-00754-y (2021).
7. Dory, C. et al. Inverse-designed diamond photonics. *Nat. Commun.* **10**, 3309 (2019).
8. Chakravarthi, S. et al. Inverse-designed photon extractors for optically addressable defect qubits. *Optica* **7**, 1805–1811 (2020).
9. Liu, K., Sun, S., Majumdar, A. & Sorger, V. J. Fundamental scaling laws in nanophotonics. *Sci. Rep.* **6**, 37419 (2016).
10. Joannopoulos, J. D., Johnson, S. G., Winn, J. N. & Meade, R. D. *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, 2008).
11. Vahala, K. J. Optical microcavities. *Nature* **424**, 839–846 (2003).
12. Khurgin, J. B. How to deal with the loss in plasmonics and metamaterials. *Nat. Nanotechnol.* **10**, 2–6 (2015).
13. Jacob, Z. et al. Engineering photonic density of states using metamaterials. *Appl. Phys. B* **100**, 215–218 (2010).
14. Sreekanth, K. V., Krishna, K. H., De Luca, A. & Strangi, G. Large spontaneous emission rate enhancement in grating coupled hyperbolic metamaterials. *Sci. Rep.* **4**, 6340 (2014).
15. Popov, V., Lavrinenko, A. V. & Novitsky, A. Operator approach to effective medium theory to overcome a breakdown of Maxwell garnett approximation. *Phys. Rev. B* **94**, 085428 (2016).
16. Schneider, P.-I. et al. Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction. *ACS Photonics* **6**, 2726–2733 (2019).
17. Lalau-Keraly, C. M., Bhargava, S., Miller, O. D. & Yablonovitch, E. Adjoint shape optimization applied to electromagnetic design. *Opt. Express* **21**, 21693 (2013).
18. Liang, X. & Johnson, S. G. Formulation for scalable optimization of microcavities via the frequency-averaged local density of states. *Opt. Express* **21**, 30812–30841 (2013).
19. Christiansen, R. E. & Sigmund, O. Inverse design in photonics by topology optimization: tutorial. *J. Opt. Soc. Am. B* **38**, 496–509 (2021).
20. Liu, D., Tan, Y., Khoram, E. & Yu, Z. Training deep neural networks for the inverse design of

nanophotonic structures. *ACS Photonics* **5**, 1365–1369 (2018).

21. Jiang, J., Chen, M. & Fan, J. A. Deep neural networks for the evaluation and design of photonic devices. *Nat. Rev. Mater.* https://www.nature.com/articles/s41578-020-00260-1 (2021).
22. Wang, F., Christiansen, R. E., Yu, Y., Mørk, J. & Sigmund, O. Maximizing the quality factor to mode volume ratio for ultra-small photonic crystal cavities. *Appl. Phys. Lett.* **113**, 241101 (2018).
23. Albrechtsen, M. et al. Nanometer-scale photon confinement inside dielectrics. Preprint at https://arxiv.org/abs/2108.01681 (2021).
24. Betzig, E., Lewis, A., Harootunian, A., Isaacson, M. & Kratschmer, E. Near field scanning optical microscopy (NSOM): development and biophysical applications. *Biophys. J.* **49**, 269–279 (1986).
25. Sánchez, E. J., Novotny, L. & Xie, X. S. Near-field fluorescence microscopy based on two-photon excitation with metal tips. *Phys. Rev. Lett.* **82**, 4014–4017 (1999).
26. Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
27. Vicidomini, G., Bianchini, P. & Diaspro, A. STED super-resolved microscopy. *Nat. Methods* **15**, 173–182 (2018).
28. Bates, M., Jones, S. A. & Zhuang, X. Stochastic optical reconstruction microscopy (STORM): a method for superresolution fluorescence imaging. *Cold Spring Harbor Protocols* http://cshprotocols.cshlp.org/content/2013/6/pdb.top075143 (2013).
29. Shockley, W. & Queisser, H. J. Detailed balance limit of efficiency of p–n junction solar cells. *J. Appl. Phys.* **32**, 510–519 (1961).
30. López, A. L. & Andreev, V. M. *Concentrator Photovoltaics*, Vol. 130 (Springer, 2007).
31. Henry, C. H. Limiting efficiencies of ideal single and multiple energy gap terrestrial solar cells. *J. Appl. Phys.* **51**, 4494–4500 (1980).
32. Luque, A. & Martí, A. Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels. *Phys. Rev. Lett.* **78**, 5014 (1997).
33. Guo, Y., Cortes, C. L., Molesky, S. & Jacob, Z. Broadband super-Planckian thermal emission from hyperbolic metamaterials. *Appl. Phys. Lett.* **101**, 131106 (2012).
34. Thompson, D. et al. Hundred-fold enhancement in far-field radiative heat transfer over the blackbody limit. *Nature* **561**, 216–221 (2018).
35. Molesky, S., Venkataram, P. S., Jin, W. & Rodriguez, A. W. Fundamental limits to radiative heat transfer: theory. *Phys. Rev. B* **101**, 035408 (2020).
36. Gustafsson, M., Schab, K., Jelinek, L. & Capek, M. Upper bounds on absorption and scattering. *New J. Phys.* **22**, 073013 (2020).
37. Kuang, Z., Zhang, L. & Miller, O. D. Maximal single-frequency electromagnetic response. *Optica* **7**, 1746–1757 (2020).
38. Molesky, S., Chao, P. & Rodriguez, A. W. Hierarchical mean-field $\mathbb{T}$ operator bounds on electromagnetic scattering: Upper bounds on near-field radiative purcell enhancement. *Phys. Rev. Res.* **2**, 043398 (2020).
39. Kuang, Z. & Miller, O. D. Computational bounds to light–matter interactions via local conservation laws. *Phys. Rev. Lett.* **125**, 263607 (2020).
40. Molesky, S. et al. $\mathbb{T}$-operator limits on optical communication: metaoptics, computation, and input–output transformations. *Phys. Rev. Res.* **4**, 013020 (2022).
41. Angeris, G., Vučković, J. & Boyd, S. Heuristic methods and performance bounds for photonic design. *Opt. Express* **29**, 2827–2854 (2021).
42. Brillouin, L. *Wave Propagation and Group Velocity*, Vol. 8 (Academic, 2013).
43. Schulz-DuBois, E. Energy transport velocity of electromagnetic propagation in dispersive media. *Proc. IEEE* **57**, 1748–1757 (1969).
44. Loudon, R. The propagation of electromagnetic energy through an absorbing dielectric. *J. Phys. A* **3**, 233 (1970).
45. Yaghjian, A. D. Internal energy, Q-energy, Poynting's theorem, and the stress dyadic in dispersive material. *IEEE Trans. Antennas Propag.* **55**, 1495–1505 (2007).
46. Glasgow, S., Ware, M. & Peatross, J. Poynting's theorem and luminal total energy transport in passive dielectric media. *Phys. Rev. E* **64**, 046610 (2001).
47. Welters, A., Avniel, Y. & Johnson, S. G. Speed-of-light limitations in passive linear media. *Phys. Rev. A* **90**, 023847 (2014).
48. Tucker, R., Pei-Cheng, K. & Chang-Hasnain, C. Slow-light optical buffers: capabilities and fundamental limitations. *J. Lightw. Technol.* **23**, 4046–4066 (2005).
49. Liu, C., Dutton, Z., Behroozi, C. H. & Hau, L. V. Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* **409**, 490–493 (2001).
50. Hau, L. V., Harris, S. E., Dutton, Z. & Behroozi, C. H. Light speed reduction to 17 metres per second in an ultracold atomic gas. *Nature* **397**, 594–598 (1999).
51. Yariv, A., Xu, Y., Lee, R. K. & Scherer, A. Coupled-resonator optical waveguide: a proposal and analysis. *Opt. Lett.* **24**, 711 (1999).
52. Soljacic, M. et al. Photonic-crystal slow-light enhancement of nonlinear phase sensitivity. *J. Opt. Soc. Am. B* **19**, 2052 (2002).
53. Povinelli, M. L., Johnson, S. G. & Joannopoulos, J. D. Slow-light, band-edge waveguides for tunable time delays. *Opt. Express* **13**, 7145–7159 (2005).
54. Miller, D. A. B. Fundamental limit for optical components. *J. Opt. Soc. Am. B* **24**, A1–A18 (2007).
55. Miller, D. A. B. Fundamental limit to linear one-dimensional slow light structures. *Phys. Rev. Lett.* **99**, 203903 (2007).
56. Fleury, R., Monticone, F. & Alù, A. Invisibility and cloaking: origins, present, and future perspectives. *Phys. Rev. Applied* **4**, 037001 (2015).
57. Pendry, J. B. Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
58. Hashemi, H., Oskooi, A., Joannopoulos, J. D. & Johnson, S. G. General scaling limitations of ground-plane and isolated-object cloaks. *Phys. Rev. A* **84**, 023815 (2011).
59. Maier, S. A. *Plasmonics: Fundamentals and Applications* (Springer, 2007).
60. Lee, K. K., Avniel, Y. & Johnson, S. G. Rigorous sufficient conditions for index-guided modes in microstructured dielectric waveguides. *Opt. Express* **16**, 9261 (2008).
61. Rechtsman, M. C. & Torquato, S. Method for obtaining upper bounds on photonic band gaps. *Phys. Rev. B* **80**, 155126 (2009).
62. Vardeny, Z. V., Nahata, A. & Agrawal, A. Optics of photonic quasicrystals. *Nat. Photonics* **7**, 177–187 (2013).
63. Yu, S., Qiu, C.-W., Chong, Y., Torquato, S. & Park, N. Engineered disorder in photonics. *Nat. Rev. Mater.* **6**, 226–243 (2021).
64. Pick, A. et al. General theory of spontaneous emission near exceptional points. *Opt. Express* **25**, 12325 (2017).
65. Miller, O. D. et al. Fundamental limits to optical response in absorptive systems. *Opt. Express* **24**, 3329–3364 (2016).
66. Barnett, S. M. & Loudon, R. Sum rule for modified spontaneous emission rates. *Phys. Rev. Lett.* **77**, 2444–2446 (1996).
67. Scheel, S. Sum rule for local densities of states in absorbing dielectrics. *Phys. Rev. A* **78**, 013841 (2008).
68. Markvart, T. The thermodynamics of optical étendue. *J. Opt. A* **10**, 015008 (2007).
69. Ries, H. Thermodynamic limitations of the concentration of electromagnetic radiation. *J. Opt. Soc. Am.* **72**, 380–385 (1982).
70. Zhang, H., Hsu, C. W. & Miller, O. D. Scattering concentration bounds: brightness theorems for waves. *Optica* **6**, 1321–1327 (2019).
71. Chung, H. & Miller, O. D. High-na achromatic metalenses by inverse design. *Opt. Express* **28**, 6945–6965 (2020).
72. Banerji, S. et al. Imaging with flat optics: metalenses or diffractive lenses? *Optica* **6**, 805–810 (2019).
73. Chen, W. T. et al. A broadband achromatic metalens for focusing and imaging in the visible. *Nat. Nanotechnol.* https://www.nature.com/articles/s41565-017-0034-6 (2018).
74. Lin, Z. & Johnson, S. G. Overlapping domains for topology optimization of large-area metasurfaces. *Opt. Express* **27**, 32445 (2019).
75. Shrestha, S., Overvig, A. C., Lu, M., Stein, A. & Yu, N. Broadband achromatic dielectric metalenses. *Light Sci. Appl.* https://www.nature.com/articles/s41377-018-0078-x (2018).
76. Wang, S. et al. A broadband achromatic metalens in the visible. *Nat. Nanotechnol.* **13**, 227–232 (2018).
77. Presutti, F. & Monticone, F. Focusing on bandwidth: achromatic metalens limits. *Optica* **7**, 624 (2020).
78. Kirchhoff, G. in *Von Kirchhoff bis Planck*, 131–151 (Springer, 1978).
79. Onnes, H. K. & Ehrenfest, P. Simplified deduction of the formula from the theory of combinations which Planck uses as the basis of his radiation-theory. *Proc. KNAW* **17**, 870–873 (1914).
80. Robitaille, P.-M. Kirchhoff's law of thermal emission: 150 years. *Prog. Phys.* **4**, 3–13 (2009).
81. Miller, D. A., Zhu, L. & Fan, S. Universal modal radiation laws for all thermal emitters. *Proc. Natl Acad. Sci. USA* **114**, 4336–4341 (2017).
82. Ellis, A., McCarthy, M., Al Khateeb, M., Sorokina, M. & Doran, N. Performance limits in optical communications due to fiber nonlinearity. *Adv. Opt. Photonics* **9**, 429–503 (2017).
83. Mizuno, K. et al. A black body absorber from vertically aligned single-walled carbon nanotubes. *Proc. Natl Acad. Sci. USA* **106**, 6044–6047 (2009).
84. Yoon, J. et al. Broadband epsilon-near-zero perfect absorption in the near-infrared. *Sci. Rep.* **5**, 22941 (2015).
85. Magdi, S., Ji, D., Gan, Q. & Swillam, M. A. Broadband absorption enhancement in organic solar cells using refractory plasmonic ceramics. *J. Nanophotonics* **11**, 016001 (2017).
86. Venkataram, P. S., Molesky, S., Jin, W. & Rodriguez, A. W. Fundamental limits to radiative heat transfer: the limited role of nanostructuring in the near-field. *Phys. Rev. Lett.* **124**, 013904 (2020).
87. Yablonovitch, E. Statistical ray optics. *Journal of the Optical Society of America* **72**, 899–907 (1982).
88. Yu, Z., Raman, A. & Fan, S. Fundamental limit of nanophotonic light trapping in solar cells. *Proc. Natl Acad. Sci. USA* **107**, 17491–17496 (2010).
89. Dienerowitz, M., Mazilu, M. & Dholakia, K. Optical manipulation of nanoparticles: a review. *J. Nanophotonics* **2**, 021875 (2008).
90. Macchi, A., Veghini, S. & Pegoraro, F. Light sail acceleration reexamined. *Phys. Rev. Lett.* **103**, 085003 (2009).
91. Kenneth, O. & Klich, I. Opposites attract: a theorem about the Casimir force. *Phys. Rev. Lett.* **97**, 160401 (2006).
92. Venkataram, P. S., Molesky, S., Chao, P. & Rodriguez, A. W. Fundamental limits to attractive and repulsive Casimir–Polder forces. *Phys. Rev. A* **101**, 052115 (2020).
93. Gustafsson, M. & Nordebo, S. Optimal antenna currents for Q, superdirectivity, and radiation patterns using convex optimization. *IEEE Trans. Antennas Propag.* **61**, 1109–1118 (2013).
94. Gustafsson, M. & Capek, M. Maximum gain, effective area, and directivity. *IEEE Trans. Antennas Propag.* **67**, 5282 (2019).
95. Capek, M., Gustafsson, M. & Schab, K. Minimization of antenna quality factor. *IEEE Trans. Antennas Propag.* **65**, 4115–4123 (2017).
96. Capek, M. et al. Optimal planar electric dipole antennas: searching for antennas reaching the fundamental bounds on selected metrics. *IEEE Antennas Propag. Mag.* **61**, 19–29 (2019).
97. Boyd, S. & Vandenberghe, L. *Convex Optimization* (Cambridge Univ. Press, 2004).
98. Zhao, Q., Zhang, L. & Miller, O. D. Minimum dielectric-resonator mode volumes. Preprint at https://arXiv.org/abs/2008.13241 (2020).
99. Goh, H. & Alù, A. Nonlocal scatterer for compact wave-based analog computing. *Phys. Rev. Lett.* **128**, 073201 (2022).
100. Bérenger, J.-P. On the Huygens absorbing boundary conditions for electromagnetics. *J. Comput. Phys.* **226**, 354–378 (2007).
101. Lindell, I. V. & Sihvola, A. *Boundary Conditions in Electromagnetics* (Wiley, 2019).
102. Tsang, L., Kong, J. A. & Ding, K.-H. Scattering of electromagnetic waves: theories and applications, Vol. 27 (Wiley, 2004).
103. Molesky, S., Chao, P., Jin, W. & Rodriguez, A. W. Global $\mathbb{T}$ operator bounds on electromagnetic scattering: upper bounds on far-field cross sections. *Phys. Rev. Res.* **2**, 033172 (2020).
104. Sun, L. & Chew, W. C. A novel formulation of the volume integral equation for electromagnetic scattering. *Waves Random Complex Media* **19**, 162–180 (2009).
105. Samokhin, A. B. *Integral Equations and Iteration Methods in Electromagnetic Scattering* (de Gruyter, 2013).
106. Costabel, M., Darrigrand, E. & Sakly, H. The essential spectrum of the volume integral operator in electromagnetic scattering by a homogeneous body. *C. R. Math.* **350**, 93–197 (2012).
107. Polimeridis, A. G., Reid, M. T. H., Johnson, S. G., White, J. K. & Rodriguez, A. W. On the computation of power in volume integral equation formulations. *IEEE Trans. Antennas Propag.* **63**, 611–620 (2014).

108. Polimeridis, A. G. et al. Fluctuating volume-current formulation of electromagnetic fluctuations in inhomogeneous media: incandescence and luminescence in arbitrary geometries. *Phys. Rev. B* **92**, 134202 (2015).
109. Liu, Q. S., Sun, S. & Chew, W. C. A potential-based integral equation method for low-frequency electromagnetic problems. *IEEE Trans. Antennas Propag.* **66**, 1413–1426 (2018).
110. Lippmann, B. A. & Schwinger, J. Variational principles for scattering processes. *Phys. Rev.* **79**, 469 (1950).
111. Lanczos, C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Natl Bureau Standards* **45**, 255–282 (1950).
112. Novotny, L. & Hecht, B. *Principles of Nano-optics* (Cambridge Univ. Press, 2012).
113. Kanwal, R. P. *Linear Integral Equations* (Springer, 2013).
114. Krüger, M., Bimonte, G., Emig, T. & Kardar, M. Trace formulas for nonequilibrium Casimir interactions, heat radiation, and heat transfer for arbitrary objects. *Phys. Rev. B* **86**, 115423 (2012).
115. Dyson, F. J. The *S* matrix in quantum electrodynamics. *Phys. Rev.* **75**, 1736 (1949).
116. Gell-Mann, M. & Goldberger, M. The formal theory of scattering. *Phys. Rev.* **91**, 398 (1953).
117. Van Kampen, N. S-matrix and causality condition. I. Maxwell field. *Phys. Rev.* **89**, 1072 (1953).
118. Landau, L. D. & Lifshitz, E. M. *Statistical Physics*, Vol. 5 (Elsevier, 2013).
119. Rudin, W. *Functional Analysis* (McGraw-Hill Education, 1991).
120. Rudin, W. *Real and Complex Analysis* (McGraw-Hill Education, 2006).
121. Miller, O. D. et al. Fundamental limits to extinction by metallic nanoparticles. *Phys. Rev. Lett.* **112**, 123903 (2014).
122. Molesky, S., Jin, W., Venkataram, P. S. & Rodriguez, A. W. $\mathbb{T}$-operator bounds on angle-integrated absorption and thermal radiation for arbitrary objects. *Phys. Rev. Lett.* **123**, 257401 (2019).
123. Chew, W. C. A new look at reciprocity and energy conservation theorems in electromagnetics. *IEEE Trans. Antennas Propag.* **56**, 970–975 (2008).
124. Valagiannopoulos, C. A. & Alú, A. The role of reactive energy in the radiation by a dipole antenna. *IEEE Trans. Antennas Propag.* **63**, 3736–3741 (2015).
125. Jackson, J. D. *Classical Electrodynamics* (AAPT, 1999).
126. Vercruysse, D. et al. Directional fluorescence emission by individual v-antennas explained by mode expansion. *ACS Nano* **8**, 8232–8241 (2014).
127. Shahpari, M. & Thiel, D. V. Fundamental limitations for antenna radiation efficiency. *IEEE Trans. Antennas Propag.* **66**, 3894–3901 (2018).
128. Siegel, R. & Spuckler, C. M. Refractive index effects on radiation in an absorbing, emitting, and scattering laminated layer. *J. Heat Transfer* **115**, 194–200 (1993).
129. Yu, Z., Raman, A. & Fan, S. Thermodynamic upper bound on broadband light coupling with photonic structures. *Phys. Rev. Lett.* **109**, 173901 (2012).
130. Callahan, D. M., Munday, J. N. & Atwater, H. A. Solar cell light trapping beyond the ray optic limit. *Nano Lett.* **12**, 214–218 (2012).
131. Mokkapati, S. & Catchpole, K. Nanophotonic light trapping in solar cells. *J. Appl. Phys.* **112**, 101101 (2012).
132. Miroshnichenko, A. E. & Tribelsky, M. I. Ultimate absorption in light scattering by a finite obstacle. *Phys. Rev. Lett.* **120**, 033902 (2018).
133. Niv, A., Gharghi, M., Gladden, C., Miller, O. D. & Zhang, X. Near-field electromagnetic theory for thin solar cells. *Phys. Rev. Lett.* **109**, 138701 (2012).
134. Miller, O. D. & Yablonovitch, E. Photon extraction: the key physics for approaching solar cell efficiency limits. *Proc. SPIE* **8808**, 880807 (2013).
135. Xu, Y., Gong, T. & Munday, J. N. The generalized Shockley–Queisser limit for nanostructured solar cells. *Sci. Rep.* **5**, 13536 (2015).
136. Schab, K. et al. Trade-offs in absorption and scattering by nanophotonic structures. *Opt. Express* **28**, 36584–36599 (2020).
137. Capek, M., Jelinek, L. & Masek, M. Fundamental bounds for multi-port antennas. In *Proc. 15th European Conference on Antennas and Propagation (EuCAP)*, https://doi.org/10.23919/EuCAP51087.2021.9411454 (IEEE, 2021).
138. Shim, H., Fan, L., Johnson, S. G. & Miller, O. D. Fundamental limits to near-field optical response over any bandwidth. *Phys. Rev. X* **9**, 011043 (2019).
139. Zhang, H., Kuang, Z., Puri, S. & Miller, O. D. Conservation-law-based global bounds to quantum optimal control. *Phys. Rev. Lett.* **127**, 110506 (2021).
140. Li, J. et al. Recent progress in mode-division multiplexed passive optical networks with low modal crosstalk. *Opt. Fiber Technol.* **35**, 28–36 (2017).
141. Yang, Z. et al. Density-matrix formalism for modal coupling and dispersion in mode-division multiplexing communications systems. *Opt. Express* **28**, 18658–18680 (2020).
142. Feng, C. et al. Wavelength-division-multiplexing (WDM)-based integrated electronic–photonic switching network (EPSN) for high-speed data processing and transportation. *Nanophotonics* **9**, 4579–4588 (2020).
143. Staude, I., Pertsch, T. & Kivshar, Y. S. All-dielectric resonant meta-optics lightens up. *ACS Photonics* **6**, 802–814 (2019).
144. Lin, Z. et al. End-to-end inverse design for inverse scattering for imaging and polarimetry. Preprint at https://arxiv.org/abs/2006.09145 (2020).
145. Phan, T. et al. High-efficiency, large-area, topology-optimized metasurfaces. *Light Sci. Appl.* **8**, 1–9 (2019).
146. Christiansen, R. E. et al. Fullwave Maxwell inverse design of axisymmetric, tunable, and multi-scale multi-wavelength metalenses. *Opt. Express* **28**, 33854–33868 (2020).
147. Estakhri, N. M., Edwards, B. & Engheta, N. Inverse-designed metastructures that solve equations. *Science* **363**, 1333–1338 (2019).
148. Li, L. et al. Intelligent metasurface imager and recognizer. *Light Sci. Appl.* **8**, 1–9 (2019).
149. Rajabalipanah, H., Abdolali, A., Iqbal, S., Zhang, L. & Cui, T. J. How do space-time digital metasurfaces serve to perform analog signal processing? Preprint at https://arXiv.org/abs/2002.06773 (2020).
150. Kravtsov, Y. A., Rytov, S. M. & Tatarskiĭ, V. Statistical problems in diffraction theory. *Sov. Phys. Usp.* **18**, 118 (1975).
151. Rytov, S. M., Kravtsov, Y. A. & Tatarskii, V. I. *Principles of Statistical Radiophysics 2. Correlation Theory of Random Processes* (Springer, 1988).
152. Kubo, R. The fluctuation–dissipation theorem. *Rep. Prog. Phys.* **29**, 255 (1966).
153. Mörters, P. & Peres, Y. *Brownian Motion*, Vol. 30 (Cambridge Univ. Press, 2010).
154. Marconi, U. M. B., Puglisi, A., Rondoni, L. & Vulpiani, A. Fluctuation–dissipation: response theory in statistical physics. *Phys. Rep.* **461**, 111–195 (2008).
155. Landauer, R. Johnson–Nyquist noise derived from quantum mechanical transmission. *Physica D* **38**, 226–229 (1989).
156. Bimonte, G., Emig, T., Kardar, M. & Krüger, M. Nonequilibrium fluctuational quantum electrodynamics: heat radiation, heat transfer, and force. *Annu. Rev. Condens. Matter Phys.* **8**, 119–143 (2017).
157. Miller, O. D., Johnson, S. G. & Rodriguez, A. W. Shape-independent limits to near-field radiative heat transfer. *Phys. Rev. Lett.* **115**, 204302 (2015).
158. Xu, H. J., Xing, Z. B., Wang, F. & Cheng, Z. Review on heat conduction, heat convection, thermal radiation and phase change heat transfer of nanofluids in porous media: fundamentals and applications. *Chem. Eng. Sci.* **195**, 462–483 (2019).
159. Beck, A. & Eldar, Y. C. Strong duality in nonconvex quadratic optimization with two quadratic constraints. *SIAM J. Optim.* **17**, 844–860 (2006).
160. Angeris, G., Vučković, J. & Boyd, S. P. Computational bounds for photonic design. *ACS Photonics* **6**, 1232 (2019).
161. Gustafsson, M., Capek, M. & Schab, K. Tradeoff between antenna efficiency and Q-factor. *IEEE Trans. Antennas Propag.* **67**, 2482–2493 (2019).
162. Chu, L. J. Physical limitations of omni-directional antennas. *J. Appl. Phys.* **19**, 1163–1175 (1948).
163. Harrington, R. F. Effect of antenna size on gain, bandwidth, and efficiency. *J. Res. Natl Bureau Standards D* **64**, 1 (1960).
164. McLean, J. S. A re-examination of the fundamental limits on the radiation Q of electrically small antennas. *IEEE Trans. Antennas Propag.* **44**, 672 (1996).
165. Capek, M. & Jelinek, L. Optimal composition of modal currents for minimal quality factor Q. *IEEE Trans. Antennas Propag.* **64**, 5230–5242 (2016).
166. Hulst, H. C. & van de Hulst, H. C. *Light Scattering by Small Particles* (Courier, 1981).
167. Harrington, R., Mautz, J. & Chang, Y. Characteristic modes for dielectric and magnetic bodies. *IEEE Trans. Antennas Propag.* **20**, 194–198 (1972).
168. Trivedi, R. et al. Bounds for scattering from absorptionless electromagnetic structures. *Phys. Rev. Applied* **14**, 014025 (2020).
169. Angeris, G., Vučković, J. & Boyd, S. Convex restrictions in physical design. *Sci. Rep.* **11**, 1–10 (2021).
170. Men, H., Lee, K. Y., Freund, R. M., Peraire, J. & Johnson, S. G. Robust topology optimization of three-dimensional photonic-crystal band-gap structures. *Opt. Express* **22**, 22632–22648 (2014).
171. Lin, Z. et al. Topology-optimized dual-polarization Dirac cones. *Phys. Rev. B* **97**, 081408 (2018).
172. Guest, J. K., Prĕvost, J. H. & Belytschko, T. Achieving minimum length scale in topology optimization using nodal design variables and projection functions. *Int. J. Numer. Methods Eng.* **61**, 238–254 (2004).
173. Zhou, M., Lazarov, B. S., Wang, F. & Sigmund, O. Minimum length scale in topology optimization by geometric constraints. *Computer Methods Appl. Mech. Eng.* **293**, 266–282 (2015).
174. Lazarov, B. S., Wang, F. & Sigmund, O. Length scale and manufacturability in density-based topology optimization. *Arch. Appl. Mech.* **86**, 189–218 (2016).
175. Li, Q., Chen, W., Liu, S. & Tong, L. Structural topology optimization considering connectivity constraint. *Struct. Multidiscipl. Optim.* **54**, 971–984 (2016).
176. Aaronson, S. Guest column: NP-complete problems and physical reality. *ACM SIGACT News* **36**, 30–52 (2005).
177. Jelinek, L., Gustafsson, M., Capek, M. & Schab, K. Fundamental bounds on the performance of monochromatic passive cloaks. *Opt. Express* **29**, 24068 (2021).
178. Johnson, S. G. et al. The NLopt nonlinear optimization package. Version 2.6.2 (2019); http://github.com/stevengj/nlopt
179. Sebbag, Y., Talker, E., Naiman, A., Barash, Y. & Levy, U. Demonstration of an integrated nanophotonic chip-scale alkali vapor magnetometer using inverse design. *Light Sci. Appl.* **10**, 54 (2021).
180. Kudyshev, Z. A., Kildishev, A. V., Shalaev, V. M. & Boltasseva, A. Machine-learning-assisted metasurface design for high-efficiency thermal emitter optimization. *Appl. Phys. Rev.* **7**, 021407 (2020).
181. Fleury, R., Soric, J. & Alù, A. Physical bounds on absorption and scattering for cloaked sensors. *Phys. Rev. B* **89**, 045122 (2014).
182. Miller, D. A. B. Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths. *Appl. Opt.* **39**, 1681–1699 (2000).
183. Gustafsson, M., Sohl, C. & Kristensson, G. Physical limitations on antennas of arbitrary shape. *Proc. R. Soc. A* **463**, 2589 (2007).
184. Hamilton, A. C. & Courtial, J. Metamaterials for light rays: ray optics without wave-optical analog in the ray-optics limit. *New J. Phys.* **11**, 013042 (2009).

### Acknowledgements

The authors acknowledge support from the National Science Foundation under the Emerging Frontiers in Research and Innovation (EFRI) programme, grant no. EFMA-1640986, the Cornell Center for Materials Research (MRSEC) through award no. DMR-1719875, and the Defense Advanced Research Projects Agency (DARPA) under grant agreements no. HR00112090011, no. HR00111820046 and no. HR0011047197. R.K.D. acknowledges financial support from the Princeton Presidential Postdoctoral Research Fellowship.

### Author contributions

The authors contributed equally to all aspects of the article.

### Competing interests

The authors declare no competing interests.

### Peer review information

*Nature Reviews Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.

### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Supplementary information

The online version contains supplementary material available at https://doi.org/10.1038/s42254-022-00468-w.

© Springer Nature Limited 2022