optica

Scattering concentration bounds: brightness theorems for waves

HANWEN ZHANG,[1,2] CHIA WEI HSU,[1,3] (iD) AND OWEN D. MILLER[1,2,*] (iD)

[1]*Department of Applied Physics, Yale University, New Haven, Connecticut 06511, USA*
[2]*Energy Sciences Institute, Yale University, New Haven, Connecticut 06511, USA*
[3]*Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA*
*Corresponding author: owen.miller@yale.edu

The brightness theorem—brightness is nonincreasing in passive systems—is a foundational conservation law, with applications ranging from photovoltaics to displays, yet it is restricted to the field of ray optics. For general linear wave scattering, we show that power per scattering channel generalizes brightness, and we derive power-concentration bounds for systems of arbitrary coherence. The bounds motivate a concept of "wave étendue" as a measure of incoherence among the scattering-channel amplitudes and which is given by the rank of an appropriate density matrix. The bounds apply to nonreciprocal systems that are of increasing interest, and we demonstrate their applicability to maximal control in nanophotonics, for metasurfaces and waveguide junctions. Through inverse design, we discover metasurface elements operating near the theoretical limits. © 2019 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

https://doi.org/10.1364/OPTICA.6.001321

The "brightness theorem" states that optical radiance cannot increase in passive ray-optical systems [1]. It is a consequence of a phase-space conservation law for optical étendue, which is a measure of the spatial and angular spread of a bundle of rays and has had a wide-ranging impact: it dictates the upper bounds to solar-energy concentration [2,3] and fluorescent-photovoltaic efficiency [3], it is a critical design criterion for projectors and displays [4], and it undergirds the theory of nonimaging optics [5]. Yet a generalization to electromagnetic radiance is not possible, as coherent wave interference can yield dramatic radiance enhancements. A natural question is whether Maxwell's equations, and more general wave-scattering physics, exhibit related conservation laws.

In this paper, we develop analogous conservation laws for power flow through the scattering channels that comprise the bases of linear scattering matrices. By a density-matrix framework more familiar to quantum settings, we derive bounds on power concentration in scattering channels, determined by the coherence of the incident field. The ranks of the density matrices for the incoming and outgoing fields play the role of étendue, and maximal eigenvalues dictate maximum possible power concentration. For the specific case of a purely incoherent excitation of N incoming channels, power cannot be concentrated onto fewer than N outgoing channels, which in the ray-optical limit simplifies to the classical brightness theorem. In resonant systems described by temporal coupled-mode theory, the number of coupled resonant modes additionally restricts the flow of wave étendue through the system. The bounds require only passivity and apply to nonreciprocal systems. We discuss their ramifications in nanophotonics—for

the design of metasurfaces, waveguide multiplexers, random-media transmission, and more—while noting that the bounds apply more generally to scattering in acoustic, quantum, and other wave systems.

Background: Optical rays exist in a four-dimensional phase space determined by their position and momentum values in a plane transverse to their propagation direction. Optical étendue [5] denotes the phase-space volume occupied by a ray bundle. In ideal optical systems, phase-space evolution is governed by Liouville's theorem, and thus radiance and étendue are invariants of the propagation. A differential ray bundle propagating through area dA and solid angle dΩ, in a medium of refractive index n and tilted at an angle θ, has an étendue of $n^2\cos\theta dA d\Omega$. Figure 1(a) depicts étendue conservation in ray-optical systems and the consequent trade-off between spatial (dA) and angular (dΩ) concentration. Electromagnetic radiance is intensity per unit area per unit solid angle, which in ray optics is proportional to the flux per unit étendue. By étendue invariance, in tandem with energy conservation, ray-optical brightness cannot increase. In nonideal systems, étendue can decrease when rays are reflected or absorbed, but any such reduction is accompanied by power loss, and the theorem still applies.

Extending radiometric concepts such as radiance into wave systems with coherence, beyond ray optics, has been the subject of considerable study [6–15]. Wigner functions can represent generalized phase-space distributions in such settings, and they are particularly useful for "first-order optics," i.e., paraxial approximations, spherical waves, etc. Yet the Wigner function and similar approaches cannot simultaneously satisfy all necessary
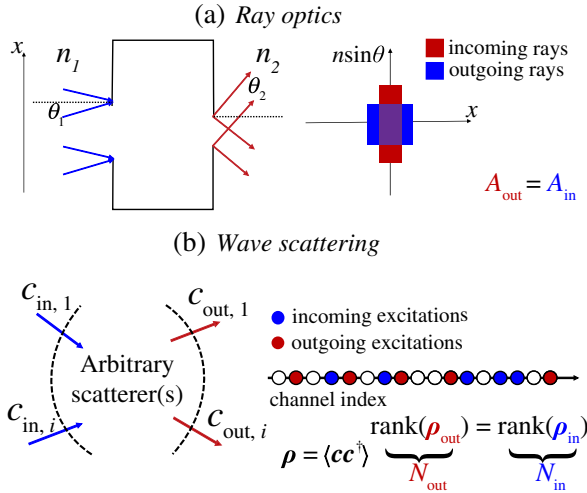
Fig. 1. (a) In ray optics, there is a trade-off in spatial and angular concentration of rays, by virtue of étendue conservation and the brightness theorem. (b) For general wave scattering, the scattering channels comprise the phase space. In ideal systems, the phase-space volumes are conserved: $A_{out} = A_{in}$ in (a), and $N_{out} = N_{in}$ in (b), where N denotes the number of excited channels (filled circles) or, more generally, the rank of the respective density matrix ρ.

properties of a generalized radiance [8,12,14]. This does not preclude the possibility for a Wigner-function-based brightness theorem—indeed, this represents an interesting open question [16]—but we circumvent the associated challenges by recognizing power transported on scattering channels as the "brightness" constraints in general wave-scattering systems.

Concentration bounds: Consider generic linear wave scattering in which some set of input waves ψ_{in} are coupled to a set of output waves ψ_{out} (in domains that may be overlapping or disjoint) through a scattering operator $\mathcal{S}$ as follows: $\psi_{out} = \mathcal{S}\psi_{in}$. We assume that the scattering process is not amplifying but does not have to be reciprocal or unitary. To describe the scattering process in a finite-dimensional basis, we adapt the formalism developed in Refs. [17–21]. As is well established in classical and quantum scattering theory [22–25], the operator $\mathcal{S}$ comprises two contributions: a "direct" (background) contribution from waves that travel from input to output without the scatterer present, which we denote with the operator $\mathcal{D}$, and a "scattered-field" contribution from waves that are scattered from input to output only in the presence of the scatterer, which we denote with the operator $\mathcal{T}$ (as in "T matrix" approaches [26–29]). A key insight of Refs. [17–21] is that the $\mathcal{T}$ operator is compact (in fact, it is a Hilbert–Schmidt operator, by the integrability of the squared Frobenius norm of its kernel), which means that one can accurately represent it by a *finite*-dimensional singular-value decomposition

$$\mathcal{T} = \mathcal{U}\Sigma\mathcal{V}^{\dagger}, \qquad (1)$$

where $\mathcal{U}$ and $\mathcal{V}$ define orthonormal bases under an appropriate inner product $\langle,\rangle$, and the restriction to finite dimensions is possible by retaining only those singular vectors corresponding to nonzero singular values, i.e., "well-coupled channels" [21]. The direct-process operator $\mathcal{D}$ is *not* necessarily compact—for example, $\mathcal{D}$ for scattering within a spherical domain is the identity operator [23,24,26]—and thus does not have the same natural

decomposition. Nevertheless, we can still project the input and output states onto $\mathcal{V}$ and $\mathcal{U}$, respectively. Such a representation will necessarily miss an infinite number of input states with a non-trivial direct-process contribution, but by definition those states will have no interaction with the scatterer, and thus they have no consequence on power-concentration bounds or on the definition of a wave étendue. We include the direct process at all in order to naturally incorporate interference effects between the direct and scattering processes. Thus, for any scattering problem, the columns of $\mathcal{V}$ and $\mathcal{U}$ define our scattering channels, within which our input and output waves can be decomposed, as follows:

$$\psi_{in} = \mathcal{V}\mathbf{c}_{in}, \qquad (2a)$$

$$\psi_{out} = \mathcal{U}\mathbf{c}_{out}, \qquad (2b)$$

where $\mathbf{c}_{in}$ and $\mathbf{c}_{out}$ are the vector coefficients of the excitations on these channels as shown in Fig. 1(b). The scattering *matrix* connects $\mathbf{c}_{in}$ to $\mathbf{c}_{out}$ and can be found by starting with the definition of the $\mathcal{S}$ operator, $\psi_{out} = \mathcal{U}\mathbf{c}_{out} = \mathcal{S}\psi_{in} = \mathcal{S}\mathcal{V}\mathbf{c}_{in}$, and then taking the inner product with $\mathcal{U}$ to find

$$\mathbf{c}_{out} = \underbrace{\langle U, SV \rangle}_{\mathbf{S}}\mathbf{c}_{in} = \mathbf{S}\mathbf{c}_{in}. \qquad (3)$$

We take our inner product to be a power normalization, such that $\mathbf{c}_{in}^{\dagger}\mathbf{c}_{in}$ and $\mathbf{c}_{out}^{\dagger}\mathbf{c}_{out}$ represent the total incoming and outgoing powers, respectively.

Perfectly coherent excitations allow for arbitrarily large modal concentration (e.g., through phase-conjugate optics [30,31]), but the introduction of incoherence incurs restrictions. To describe the coherence of incoming waves, we use a density matrix ρ_{in} [32] that is the ensemble average (hereafter denoted by $\langle \cdot \rangle$, over the source of incoherence) of the outer product of the incoming wave amplitudes, written as

$$\rho_{in} = \langle \mathbf{c}_{in}\mathbf{c}_{in}^{\dagger} \rangle. \qquad (4)$$

The incoherence of the outgoing channels is represented in the corresponding outgoing-wave density matrix

$$\rho_{out} = \langle \mathbf{c}_{out}\mathbf{c}_{out}^{\dagger} \rangle = \mathbf{S}\rho_{in}\mathbf{S}^{\dagger}. \qquad (5)$$

The matrices ρ_{in} and ρ_{out} represent density operators projected onto the $\mathcal{U}$ and $\mathcal{V}$ bases. Both matrices are Hermitian and positive semidefinite.

For inputs defined by some ρ_{in}, how much power can flow into a single output channel, or more generally into a linear combination given by a unit vector $\hat{\mathbf{u}}$? If we denote $\hat{\mathbf{u}}^{\dagger}\mathbf{c}_{out}$ as $\mathbf{c}_{out,\hat{\mathbf{u}}}$, then the power through $\hat{\mathbf{u}}$ is $\langle |\mathbf{c}_{out,\hat{\mathbf{u}}}|^{2} \rangle = \hat{\mathbf{u}}^{\dagger}\rho_{out}\hat{\mathbf{u}} = \hat{\mathbf{u}}^{\dagger}\mathbf{S}\rho_{in}\mathbf{S}^{\dagger}\hat{\mathbf{u}}$. The quantity $\langle |\mathbf{c}_{out,\hat{\mathbf{u}}}|^{2} \rangle$ is a quadratic form in ρ_{in}, such that its maximum value is dictated by its largest eigenvalue [33], λ_{max}, leading to the inequality

$$\langle |\mathbf{c}_{out,\hat{\mathbf{u}}}|^{2} \rangle \leq \lambda_{max}(\rho_{in})(\hat{\mathbf{u}}^{\dagger}\mathbf{S}\mathbf{S}^{\dagger}\hat{\mathbf{u}}). \qquad (6)$$

To bound the term in parentheses, $\hat{\mathbf{u}}^{\dagger}\mathbf{S}\mathbf{S}^{\dagger}\hat{\mathbf{u}}$, we consider *coherent* scattering for a new input: $\mathbf{c}_{in} = \mathbf{S}^{\dagger}\hat{\mathbf{u}}$. For this input field, the incoming power is $\hat{\mathbf{u}}^{\dagger}\mathbf{S}\mathbf{S}^{\dagger}\hat{\mathbf{u}}$, while the outgoing power in the unit vector $\hat{\mathbf{u}}$ is $|\hat{\mathbf{u}}^{\dagger}\mathbf{c}_{out}|^{2} = (\hat{\mathbf{u}}^{\dagger}\mathbf{S}\mathbf{S}^{\dagger}\hat{\mathbf{u}})^{2}$. Enforcing the inequality that the outgoing power in $\hat{\mathbf{u}}$ must be no larger than the (coherent) total incoming power, we immediately have the identity $\hat{\mathbf{u}}^{\dagger}\mathbf{S}\mathbf{S}^{\dagger}\hat{\mathbf{u}} \leq 1$. (We provide an alternative proof in Supplement 1.) Inserting into Eq. (6), we arrive at the bound

$$\langle |\mathbf{c}_{out,\hat{\mathbf{u}}}|^{2} \rangle \leq \lambda_{max}(\rho_{in}). \qquad (7)$$

Equation (7) is a key theoretical result of this paper. It states that for a system whose incoming power flow and coherence are described by a density matrix $\boldsymbol{\rho}_{\text{in}}$, the maximum concentration of power is the largest eigenvalue of that density matrix. For a coherent input (akin to quantum-mechanical "pure states" [34]), there is a single nonzero eigenvalue, equal to 1, such that all of the power can be concentrated into a single channel. For equal incoherent excitation of N independent incoming states, the density matrix is diagonal with all nonzero eigenvalues equal to $1/N$, in which case

$$\langle|\mathbf{c}_{\text{out},\hat{\mathbf{u}}}|^2\rangle \leq \frac{1}{N}. \tag{8}$$

Equation (8) is less general than Eq. (7) but provides intuition and is a closer generalization of the ray-optical brightness theorem. Since the average output power per independent state must be less than or equal to $1/N$, at least N independent outgoing states must be excited, or a commensurate amount of power must be lost to dissipation. In reciprocal systems, this bound follows from reversibility. In Supplement 1, we prove that Eq. (8) simplifies to the ray-optical brightness theorem for continuous plane-wave channels in homogeneous media.

Just like the ray-optical brightness theorem [1,35], our scattering-channel bounds can alternatively be understood as a consequence of the second law of thermodynamics. If it were possible to concentrate incoherent excitations of multiple channels, then one could filter out all other channels and create a scenario with a cold body on net sending energy to a warm body [20]. The partially coherent case is not as physically intuitive, but the application of such thermodynamic reasoning could be applied to the modes that diagonalize $\boldsymbol{\rho}_{\text{in}}$, and then a basis transformation would yield Eq. (7).

Wave étendue: Equations (7) and (8) imply that the incoherent excitation of N inputs cannot be fully concentrated to fewer than N outputs. This motivates the identification of "wave étendue" as the number of incoherent excitations on any subset of channels (incoming, outgoing, etc.). For a density matrix $\boldsymbol{\rho}$, one can count independence by the matrix *rank* and define étendue = rank($\boldsymbol{\rho}$).

To understand the evolution of wave étendue through the scattering process, we reconsider the singular-value decomposition (SVD) of Eq. (1). The matrix Σ is a square matrix with dimensions $N \times N$, where N is the number of well-coupled pairs of input and output scattering channels. Since the singular values are nonzero, we know that Σ is full rank. The scattering matrix **S** is the sum of Σ and the direct-process matrix, and its rank will be N minus the number of coherent perfect absorber (CPA) states, $N - N_{\text{CPA}}$, where the CPA states arise if the direct process exactly cancels a scattered wave, yielding perfect absorption [36,37]. (Technically, these may be partial-CPA states, exhibiting perfect cancellation of the direct fields only on the range of the $\mathcal{T}$ operator.) The density matrix ρ_{in} is a representation of the incoming excitations on the basis $\mathcal{V}$ of Eq. (1), and thus it cannot have rank greater than N itself. By the relation $\boldsymbol{\rho}_{\text{out}} = \mathbf{S}\boldsymbol{\rho}_{\text{in}}\mathbf{S}^{\dagger}$ and the matrix-product inequality rank(AB) $\leq$ min(rank(A), rank(B)) (Ref. [33]), the rank of $\boldsymbol{\rho}_{\text{out}}$ must lie within bounds given by the rank of $\boldsymbol{\rho}_{\text{in}}$ minus the number of CPA states and the rank of $\boldsymbol{\rho}_{\text{in}}$ itself as follows:

$$\text{rank}(\boldsymbol{\rho}_{\text{in}}) - N_{\text{CPA}} \leq \text{rank}(\boldsymbol{\rho}_{\text{out}}) \leq \text{rank}(\boldsymbol{\rho}_{\text{in}}). \tag{9}$$

Equation (9) defines the maximum diversity possible in the evolution of wave étendue in linear scattering systems. For lossless systems—or more generally any system without CPA states—we must have $N_{\text{CPA}} = 0$, in which case Eq. (9) is a conservation law stating that the density-matrix rank is always conserved. (In Supplement 1 we show that this simplifies to the classical wave-étendue conservation law in the ray-optics limit.) Figure 1(b) depicts this rank-defined (channel-counting) definition of wave étendue. In wave-scattering systems, phase space is defined by distinct scattering channels, without recourse to the position and momentum unique to free-space states.

Metasurface design: To probe the channel-concentration bounds, we consider control of diffraction orders through complex metasurfaces for potential applications such as augmented-reality optics [38,39] and photovoltaic concentrators [40–42]. Figure 2(a) depicts a designable gradient refractive-index profile with a period of 2λ and a thickness of 0.5λ. (Such an element could be one unit cell within a larger, non-periodic
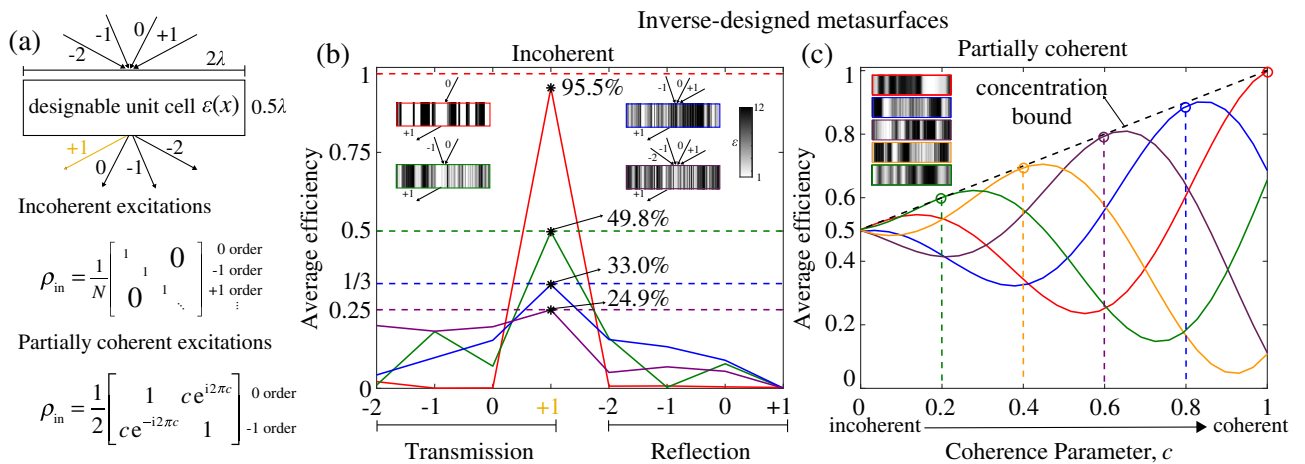

Fig. 2. (a) Periodic metasurface element to be designed for maximal power in the +1 transmission diffraction order (yellow). We consider incoherent excitations among the four incident orders, with a diagonal density matrix, as well as partially coherent excitations between the 0 and −1 order, represented by an off-diagonal term with coherence parameter c. Inverse-designed metasurfaces closely approaching the coherence- and channel-dependent bounds are shown in (b) for incoherent excitations among up to four channels, and in (c) for partially coherent excitations between two channels. [Designs in (c) are all optimal for the fully incoherent case because $\boldsymbol{\rho}_{\text{in}}$ is a constant multiple of the identity matrix. This should not be considered a generic phenomenon when excitation powers are unevenly distributed.]

metasurface [43–46].) For incoherent excitation of N diffraction orders, Eq. (8) dictates that the maximum average efficiency of concentrating light into a single output order (+1) cannot be greater than $1/N$ [dashed lines in Fig. 2(c)]. For s-polarized light incoherently incident from orders 0 (red); –1, 0, (green); –1, 0, +1 (blue); and –2, –1, 0, +1 (purple) (20 deg angle of incidence for the zeroth order), we use adjoint-based "inverse design" [47–54] (Supplement 1) to discover optimal refractive-index profiles of the four metasurfaces shown in Fig. 2(b). (Broader angular control and binary refractive-index profiles could be generated through standard optimization augmentations [50,51], but here we emphasize the brightness-theorem consequences.) The transmission spectrum was computed by the Fourier modal method [55] with a freely available software package [56]. In Fig. 2(b), as the number of incoherent channels excited increases from 1 to 4, the average efficiency of the optimal structures decreases from 95.5% to 24.9%. (In Supplement 1 we show that optimal blazed gratings fall far short of the bounds.) We also probe the effects of partial coherence by varying the coherence between two input orders, per the density matrix in Fig. 2(a). By Eq. (7), maximum concentration is determined by the largest eigenvalue of ρ_{in}, which is $1 - c/2$, where c is the coherence parameter. Figure 2(c) shows inverse-designed structures for $c = 0.2, 0.4, 0.6, 0.8, 1$, with unique structures optimizing the response depending on the coherence of the excitation. All of the structures maximize efficiency in the incoherent $c = 0$ case because the eigenvalues of the density matrix are degenerate, and thus transmission of any state is optimal.

Étendue transmission: An important related scenario to consider is one in which the direct (background) process is ignored, with the focus solely on interactions with scatterers. Instead of the input and output fields considered above, the relevant decomposition is instead into *incident* and *scattered* fields. Using the same terminology as in the input-output scattering operator, we can connect the incident and scattered fields by a $\mathcal{T}$ operator [26–29] $\psi_{scat} = \mathcal{T}\psi_{inc}$. Again, as shown in Refs. [17–21], the $\mathcal{T}$ operator is compact and defines incoming- and scattered-wave bases by its SVD, Eq. (1). Furthermore, to align with various applications described below, we will specify a set of $N_{trans} \leq \text{rank}\,\mathcal{U}$ desirable "transmission" channels that are a subset of the scattered-field channels defined by $\mathcal{U}$. To understand transmission flow into these channels, we will define our finite-dimensional $\mathbf{T}$ matrix as the restriction of $\mathcal{T}$ onto this subset of scattered-field channels. The matrix $\mathbf{T}$ connects the incoming-field channels to the transmission channels. The "transmission" terminology, partially meant to avoid further overload of the word

"scattering," is intended simply to represent the flow of energy through a system, enabled by interactions with a scatterer. For a planar or periodic scatterer, both reflected and transmitted waves would be part of this generalized "transmission" process, as long as they differ from the direct free-space process.

We define "étendue transmission" as the number of incoherent excitations that can successfully be transmitted through scatterer interactions onto the transmission channels. Equation (8) dictates that at least N output channels are excited for N orthogonal inputs, and indeed this result is proven in the incoherent case in Ref. [20] through an SVD of the $\mathbf{T}$ matrix (denoted therein by "$\mathbf{S}$"). If the number of transmission channels, N_{trans}, is less than $\text{rank}(\rho_{inc})$, where ρ_{inc} is the incident-wave density matrix, then the incoherent excitations cannot all be concentrated onto the transmission channels, and some power must necessarily be scattered into other scattering channels.

Resonance-assisted transmission, in which resonances couple the incident and transmission channels, introduces an additional constraint: the number of resonant modes (resonances) M coupled to the relevant channels. Resonant modes are not scattering channels; instead, they are the quasi-normal modes (QNMs) of the scatterer, subject to outgoing boundary conditions. (Quasi-normal modes have been extensively studied and applied to various scattering systems for the last decade [57–59], and in the limit of closed systems and self-adjoint Maxwell operators they reduce to conventional guided and standing-wave modes [60].) We consider systems where the interaction with resonant modes can be described by temporal coupled mode theory (TCMT) [61–63], wherein the scattering process is encoded in an $M \times M$ matrix Ω, comprising the real and imaginary parts of the resonant-mode resonant frequencies, and a matrix $\mathbf{K}$, denoting channel–mode coupling. In TCMT, the T matrix for the resonance-assisted transmission component is (Supplement 1) $\mathbf{T} = -i\mathbf{K}_{trans}(\Omega - \omega)^{-1}\mathbf{K}_{inc}^T$, where $\mathbf{K}_{trans}$ and $\mathbf{K}_{inc}$ are the $N_{trans} \times M$ and $N_{inc} \times M$ submatrices of $\mathbf{K}$ denoting modal couplings to the transmission and incident channels, respectively.

The maximum (average) power flow into a single transmission output channel is subject to the bounds of Eqs. (7) and (8), now in terms of the density matrix ρ_{inc}. The matrix ρ_{trans} equals $\mathbf{T}\rho_{inc}\mathbf{T}^\dagger$. By recursive application of the matrix-rank inequality used above, we can see that

$$\text{rank}(\rho_{trans}) \leq \min(\text{rank}(\rho_{inc}), M, N_{trans}). \quad (10)$$

The number of orthogonal outputs is less than or equal to the *minimum* of the numbers of incident inputs, resonant modes, and transmission channels. As depicted in Fig. 3, transmission channels and resonant modes act like apertures [35] in restricting the flow of étendue through a system.

We may also consider total transmission onto all N_{trans} transmission channels, i.e., $\sum_i \langle |\mathbf{c}_{trans,i}|^2 \rangle$. Since the transmission onto a single output is bounded above by $\lambda_{max}(\rho_{inc})$, the total power is bounded above by the sum of the first $\text{rank}(\rho_{trans})$ eigenvalues (Supplement 1) as follows:

$$\sum_i \langle |\mathbf{c}_{trans,i}|^2 \rangle \leq \sum_{i=1}^{\min(\text{rank}(\rho_{inc}), M, N_{trans})} \lambda_i, \quad (11)$$

where the eigenvalues are indexed in descending order. For incoherent excitation of the N_{inc} channels, $\lambda_i(\rho_{inc}) = 1/N_{inc}$ for all i, and the term on the right of Eq. (11) simplifies to $\min(N_{inc}, M, N_{trans})/N_{inc}$. In resonance-assisted transmission
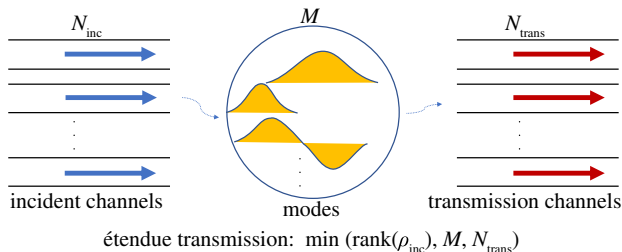

Fig. 3. Étendue, defined as the rank of wave-scattering density matrices, is restricted in resonance-assisted transmission processes by the number of transmission channels and channel-coupled resonances in the process.
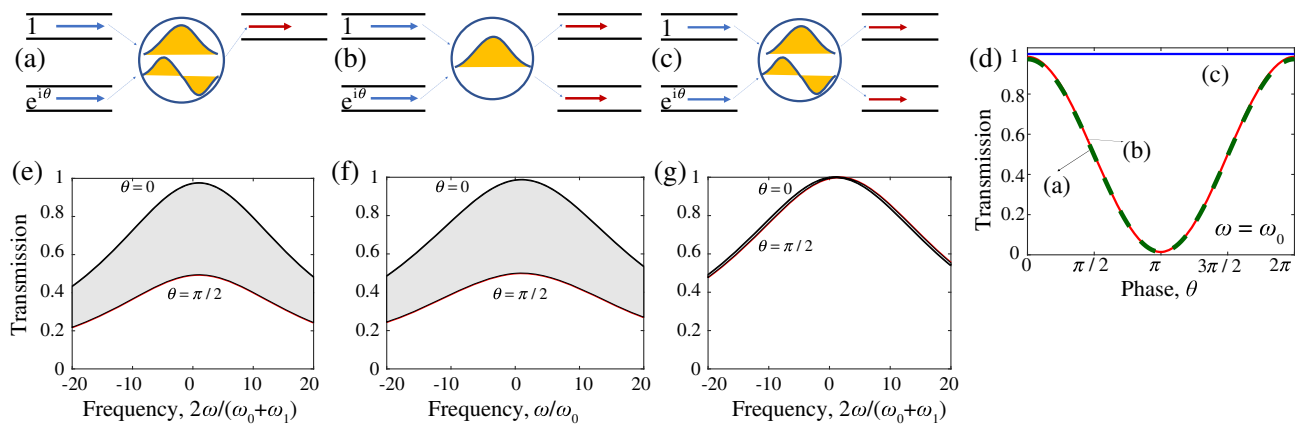
Fig. 4. Robustness of waveguide junctions is susceptible to étendue restrictions. For two input channels, we consider (a) one output, (b) one mode, and (c) no restrictions. (e)–(g) Transmission for (a)–(c) with input phase angles in $\theta = [0, \pi/2]$. (d) Transmission as a function of phase, on resonance. Case (c) is designed to be almost perfectly insensitive to phase; such designs are impossible in cases (a) and (b).

scenarios, Eq. (11) represents an additional constraint on power flow: in addition to the number of output channels, total power flow is further constrained by the number of distinct resonant modes that interact with them. Étendue restrictions anywhere in the transmission process necessarily generate scattering to channels other than the desired transmission channels.

We apply the resonance-assisted-transmission bounds to TCMT models of waveguide multiplexers as depicted in Fig. 4. There has been significant interest [52,64–68] in the design of compact junctions for routing light. In Figs. 4(a)–4(c), we consider "input" waveguides and "output" waveguides coupled by a resonant scattering system. For two input waveguides, we consider three scenarios: (a) one output and two resonances, (b) one resonance and two outputs, and (c) two resonances and two outputs. In each case, a highly controlled coherent excitation can, through appropriate design of the resonator, yield perfect transmission on resonance at the output port. But a *robust* design, impervious to noise or other incoherence, may be required, and such noise would introduce incoherence that is subject to the bound of Eq. (11). In each case, we optimize TCMT model parameters (cf. Supplement 1) to maximize transmission for all phase differences between two inputs. Device (c) maintains perfect transmission, whereas devices (a) and (b) are highly sensitive to noise as predicted by the restrictions to étendue flow.

The channel-concentration bounds and wave-étendue concept generalize classical ray-optical ideas to general wave scattering. In addition to the nanophotonic design problems considered here, there are numerous potential applications. First, they resolve how to incorporate polarization into ray-optical étendue, showing unequivocally that polarizing unpolarized light requires doubling classical étendue, an uncertain conjecture in display design [69,70]. Moreover, the natural incorporation of nonreciprocity into the bounds is of particular relevance given the emerging interest in nonreciprocal photonics [71–73] and acoustics [74], and it places constraints on many of these systems (extensions to time-modulated TCMT systems should be possible). Another additional application space is in random-scattering theory [25,75–77]. For opaque optical media comprising random scatterers, there is significant interest both in controlling the scattering channels (e.g., wavefront shaping) as well as studying the effects of partial coherence on scattering and absorption [78].

Our work shows that fully coherent excitations are optimal for maximal concentration on the fewest possible states. A related question that remains open is whether partial coherence might be optimal for total transmission. The framework developed herein may lead to fundamental limits to control in such systems.

See Supplement 1 for supporting content.

REFERENCES

1. R. W. Boyd, *Radiometry and the Detection of Optical Radiation* (Wiley, 1983).
2. H. Ries, "Thermodynamic limitations of the concentration of electromagnetic radiation," J. Opt. Soc. Am. **72**, 380–385 (1982).
3. G. Smestad, H. Ries, R. Winston, and E. Yablonovitch, "The thermodynamic limits of light concentrators," Sol. Energy Mater. **21**, 99–111 (1990).
4. M. S. Brennesholtz and E. H. Stupp, *Projection Displays*, 2nd ed. (Wiley, 2008).
5. R. Winston, J. C. Miñano, and P. G. Benitez, *Nonimaging Optics* (Elsevier, 2005).
6. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University, 1995).
7. A. Walther, "Radiometry and coherence," J. Opt. Soc. Am. **58**, 1256–1259 (1968).
8. A. T. Friberg, "On the existence of a radiance function for finite planar sources of arbitrary states of coherence," J. Opt. Soc. Am. **69**, 192–198 (1979).
9. R. G. Littlejohn and R. Winston, "Corrections to classical radiometry," J. Opt. Soc. Am. A **10**, 2024–2037 (1993).
10. R. G. Littlejohn and R. Winston, "Generalized radiance and measurement," J. Opt. Soc. Am. A **12**, 2736–2743 (1995).
11. M. A. Alonso, "Radiometry and wide-angle wave fields. I. Coherent fields in two dimensions," J. Opt. Soc. Am. A **18**, 902–909 (2001).
12. M. A. Alonso, "Radiometry and wide-angle wave fields III: partial coherence," J. Opt. Soc. Am. A **18**, 2502–2511 (2001).
13. M. E. Testorf, B. M. Hennelly, and J. Ojeda-Castañeda, *Phase-space Optics: Fundamentals and Applications* (McGraw-Hill, 2010).

14. M. A. Alonso, "Wigner functions in optics: describing beams as ray bundles and pulses as particle ensembles," Adv. Opt. Photon. **3**, 272–365 (2011).

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

16. L. Apresyan and Y. A. Kravtsov, *Radiation Transfer* (Routledge, 2019).

17. D. A. Miller, "Fundamental limit for optical components," J. Opt. Soc. Am. B **24**, A1–A18 (2007).

18. D. A. Miller, "Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths," Appl. Opt. **39**, 1681–1699 (2000).

19. D. A. Miller, L. Zhu, and S. Fan, "Universal modal radiation laws for all thermal emitters," Proc. Natl. Acad. Sci. USA **114**, 4336–4341 (2017).

20. D. A. Miller, "All linear optical devices are mode converters," Opt. Express **20**, 23985–23993 (2012).

21. D. A. Miller, "Waves, modes, communications and optics," arXiv:1904.05427 (2019).

22. R. G. Newton, *Scattering Theory of Waves and Particles* (Springer Science & Business Media, 2013).

23. C. Mahaux and H. A. Weidenmüller, "Shell-model approach to nuclear reactions," in *Soft Matter* (1969).

24. D. Jalas, A. Petrov, M. Eich, W. Freude, S. Fan, Z. Yu, R. Baets, M. Popović, A. Melloni, J. D. Joannopoulos, M. Vanwolleghem, C. R. Doerr, and H. Renner, "What is—and what is not—an optical isolator," Nat. Photonics **7**, 579–582 (2013).

25. S. Rotter and S. Gigan, "Light fields in complex media: mesoscopic scattering meets wave control," Rev. Mod. Phys. **89**, 015005 (2017).

26. J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics* (Pearson Harlow, 2014), Vol. **185**.

27. P. Waterman, "Matrix formulation of electromagnetic scattering," Proc. IEEE **53**, 805–812 (1965).

28. M. Mishchenko and P. Martin, "Peter waterman and t-matrix methods," J. Quantum Spectrosc. Radiat. Transfer **123**, 2–7 (2013).

29. M. I. Mishchenko, L. D. Travis, and A. A. Lacis, *Scattering, Absorption, and Emission of Light by Small Particles* (Cambridge University, 2002).

30. A. Yariv, "Phase conjugate optics and real-time holography," IEEE J. Quantum Electron. **14**, 650–660 (1978).

31. A. Yariv, *Quantum Electronics*, 3rd ed. (Wiley, 1989).

32. S. Withington and J. A. Murphy, "Modal analysis of partially coherent submillimeter-wave quasi-optical systems," IEEE Trans. Anntenas Propag. **46**, 1651–1659 (1998).

33. R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. (Cambridge University, 2013).

34. L. D. Landau and E. Lifshitz, *Course of Theoretical Physics Vol. 3 Quantum Mechanics* (Pergamon, 1958).

35. J. Chaves, *Introduction to Nonimaging Optics* (CRC Press, 2016).

36. Y. Chong, L. Ge, H. Cao, and A. D. Stone, "Coherent perfect absorbers: time-reversed lasers," Phys. Rev. Lett. **105**, 053901 (2010).

37. D. G. Baranov, A. Krasnok, T. Shegai, A. Alù, and Y. Chong, "Coherent perfect absorbers: linear control of light with light," Nat. Rev. Mater. **2**, 17064 (2017).

38. O. Cakmakci and J. Rolland, "Head-worn displays: a review," J. Disp. Technol. **2**, 199–216 (2006).

39. T. Levola, "Diffractive optics for virtual reality displays," J. Soc. Inf. Disp. **14**, 467–475 (2006).

40. J. S. Price, X. Sheng, B. M. Meulblok, J. A. Rogers, and N. C. Giebink, "Wide-angle planar microtracking for quasi-static microcell concentrating photovoltaics," Nat. Commun. **6**, 6223 (2015).

41. M. A. Shameli and L. Yousefi, "Absorption enhancement in thin-film solar cells using an integrated metasurface lens," J. Opt. Soc. Am. B **35**, 223–230 (2018).

42. Z. Lin, B. Groever, F. Capasso, A. W. Rodriguez, and M. Lončar, "Topology-optimized multilayered metaoptics," Phys. Rev. Appl. **9**, 044030 (2018).

43. N. Yu and F. Capasso, "Flat optics with designer metasurfaces," Nat. Mater. **13**, 139–150 (2014).

44. D. Lin, P. Fan, E. Hasman, and M. L. Brongersma, "Dielectric gradient metasurface optical elements," Science **345**, 298–302 (2014).

45. A. Arbabi, Y. Horie, M. Bagheri, and A. Faraon, "Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission," Nat. Nanotechnol. **10**, 937–943 (2015).

46. M. Khorasaninejad and F. Capasso, "Metalenses: versatile multifunctional photonic components," Science **358**, eaam8100 (2017).

47. A. Jameson, L. Martinelli, and N. A. Pierce, "Optimum aerodynamic design using the Navier-Stokes equations," Theor. Comput. Fluid Dyn. **10**, 213–237 (1998).

48. O. Sigmund and J. Søndergaard Jensen, "Systematic design of photonic band-gap materials and structures by topology optimization," Philos. Trans. R. Soc. London, Ser. A **361**, 1001–1019 (2003).

49. J. Lu, S. Boyd, and J. Vucković, "Inverse design of a three-dimensional nanophotonic resonator," Opt. Express **19**, 10563–10570 (2011).

50. J. S. Jensen and O. Sigmund, "Topology optimization for nano-photonics," Laser Photon. Rev. **5**, 308–321 (2011).

51. O. D. Miller, "Photonic design: from fundamental solar cell physics to computational inverse design," Ph.D. thesis (University of California, 2012).

52. C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," Opt. Express **21**, 21693–21701 (2013).

53. V. Ganapati, O. D. Miller, and E. Yablonovitch, "Light trapping textures designed by electromagnetic optimization for subwavelength thick solar cells," IEEE J. Photovoltaics **4**, 175–182 (2014).

54. N. Aage, E. Andreassen, B. S. Lazarov, and O. Sigmund, "Giga-voxel computational morphogenesis for structural design," Nature **550**, 84–86 (2017).

55. M. Moharam and T. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).

56. V. Liu and S. Fan, "S4: a free electromagnetic solver for layered periodic structures," Comput. Phys. Commun. **183**, 2233–2244 (2012).

57. C. Sauvan, J.-P. Hugonin, I. Maksymov, and P. Lalanne, "Theory of the spontaneous optical emission of nanosize photonic and plasmon resonators," Phys. Rev. Lett. **110**, 237401 (2013).

58. P. Lalanne, W. Yan, K. Vynck, C. Sauvan, and J.-P. Hugonin, "Light interaction with photonic and plasmonic resonances," Laser Photon. Rev. **12**, 1700113 (2018).

59. P. Lalanne, W. Yan, A. Gras, C. Sauvan, J.-P. Hugonin, M. Besbes, G. Demésy, M. Truong, B. Gralak, F. Zolla, and A. Nicolet, "Quasinormal mode solvers for resonators with dispersive materials," J. Opt. Soc. Am. A **36**, 686–704 (2019).

60. J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light* (Princeton University, 2011).

61. S. Fan, W. Suh, and J. D. Joannopoulos, "Temporal coupled-mode theory for the Fano resonance in optical resonators," J. Opt. Soc. Am. A **20**, 569–572 (2003).

62. H. A. Haus, *Waves and Fields in Optoelectronics* (Prentice-Hall, 1984).

63. W. Suh, Z. Wang, and S. Fan, "Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities," IEEE J. Quantum Electron. **40**, 1511–1518 (2004).

64. Y. Ding, J. Xu, F. Da Ros, B. Huang, H. Ou, and C. Peucheret, "On-chip two-mode division multiplexing using tapered directional coupler-based mode multiplexer and demultiplexer," Opt. Express **21**, 10376–10382 (2013).

65. R. Ji, L. Yang, L. Zhang, Y. Tian, J. Ding, H. Chen, Y. Lu, P. Zhou, and W. Zhu, "Five-port optical router for photonic networks-on-chip," Opt. Express **19**, 20258–20268 (2011).

66. A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vucković, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," Nat. Photonics **9**, 374–377 (2015).

67. B. Shen, P. Wang, R. Polson, and R. Menon, "An integrated-nanophotonics polarization beamsplitter with 2.4×2.4 μm^2 footprint," Nat. Photonics **9**, 378–382 (2015).

68. S. Fan, P. R. Villeneuve, J. D. Joannopoulos, and H. Haus, "Channel drop tunneling through localized states," Phys. Rev. Lett. **80**, 960–963 (1998).

69. S. A. Lerner and B. Dahlgrenn, "Etendue and optical system design," Proc. SPIE **6338**, 633801 (2006).

70. F. Fournier and J. Rolland, "Design methodology for high brightness projectors," J. Display Technol. **4**, 86–91 (2008).

71. K. Fang, Z. Yu, and S. Fan, "Realizing effective magnetic field for photons by controlling the phase of dynamic modulation," Nat. Photonics **6**, 782–787 (2012).

72. L. D. Tzuang, K. Fang, P. Nussenzveig, S. Fan, and M. Lipson, "Non-reciprocal phase shift induced by an effective magnetic flux for light," Nat. Photonics **8**, 701–705 (2014).

73. D. L. Sounas and A. Alù, "Non-reciprocal photonics based on time modulation," Nat. Photonics **11**, 774–783 (2017).

74. S. A. Cummer, J. Christensen, and A. Alù, "Controlling sound with acoustic metamaterials," Nat. Rev. Mater. **1**, 16001 (2016).

75. A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, "Controlling waves in space and time for imaging and focusing in complex media," Nat. Photonics **6**, 283–292 (2012).

76. C. W. Hsu, S. F. Liew, A. Goetschy, H. Cao, and A. D. Stone, "Correlation-enhanced control of wave focusing in disordered media," Nat. Phys. **13**, 497–502 (2017).

77. C. W. Hsu, A. Goetschy, Y. Bromberg, A. D. Stone, and H. Cao, "Broadband coherent enhancement of transmission and absorption in disordered media," Phys. Rev. Lett. **115**, 223901 (2015).

78. F. Bigourdan, R. Pierrat, and R. Carminati, "Enhanced absorption of waves in stealth hyperuniform disordered media," Opt. Express **27**, 8666–8682 (2019).

Scattering concentration bounds: brightness theorems for waves: supplementary material

HANWEN ZHANG[1,2], **CHIA WEI HSU**[1,3], **AND OWEN D. MILLER**[1,2,*]

[1]*Department of Applied Physics, Yale University, New Haven, Connecticut 06511, USA*
[2]*Energy Sciences Institute, Yale University, New Haven, Connecticut 06511, USA*
[3]*Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA*
[*]*Corresponding author: owen.miller@yale.edu*

This document provides supplementary information to "Scattering concentration bounds: brightness theorems for waves," https://doi.org/10.1364/OPTICA.6.001321.

CONTENTS

1. DERIVING THE CLASSICAL BRIGHTNESS THEOREM FROM ITS WAVE-SCATTERING GENERALIZATION

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

In ray optics, the wavelength is taken to zero, such that even an infinitesimal area is arbitrarily large relative to the wavelength. Thus we can consider the differential area as the "box" (actually square) into which the states must fit with periodic boundary condition. If we take $\Delta A = \Delta x \Delta y$ (with z as the propagation direction), the states that satisfy periodic bound-

ary conditions on ΔA are plane waves with k_x and k_y taking integral multiples of $2\pi/\Delta x$ and $2\pi/\Delta y$, and k_z fixed by the frequency and specific values of k_x and k_y. Thus we can write non-normalized electric-field states as

$$\mathbf{E}_i = e^{i\mathbf{k}_i \cdot \mathbf{r}}. \tag{S1}$$

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

$$\mathbf{E}_i = \sqrt{\frac{2Z_0}{\Delta A}} e^{i\mathbf{k}_i \cdot \mathbf{r}}. \tag{S2}$$

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

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

$$N = 2\frac{\Delta\mathbf{k}_\perp}{4\pi^2/\Delta A} = \frac{1}{2\pi^2}\Delta A \, k_\perp \Delta k_\perp \Delta\phi. \tag{S3}$$

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

Eq. (S3), we obtain

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

(S4)

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

Now reconsider power flowing from area dA_1 and solid angle $d\Omega_1$ at the entrance plane. The number of channels that are (incoherently) excited is given by

$$N_1 = \frac{\omega^2}{2\pi^2 c^2} n_1^2 \cos\theta_1 dA_1 d\Omega_1.$$

(S5)

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

$$P_{\text{out},i} \leq \frac{1}{N_1} = \frac{1}{N_2},$$

(S6)

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

We could have equivalently derived the usual concentration-ratio bounds on optical concentrators, and by standard ray-optical arguments one can see that the étendue derived in Eq. (S4) generalizes to non-ideal ray-optical systems as usual.

Justification of the choice of periodic boundary condition

Here we provide a mathematical proof of the fact that our periodic basis set can represent *any* function, in the appropriate normed space. Before going into the mathematical details, we can provide the intuition: on any finite domain of size L, a periodic-boundary Fourier-series expansion can accurately represent a function everywhere except at its boundary. In the limit that $L \to \infty$ (or equivalently $\lambda \to 0$), the function values exactly at the boundary no longer matter, and the periodic-boundary representation approximates *any* function arbitrarily well. This idea is well-understood in Fourier analysis; now we apply it to the physical optics problem at hand.

Suppose we start with an arbitrary wave field $f(x)$ that varies along a one-dimensional plane along dimension x. (The 2D-plane generalization is trivial, simply multiplying the 1D results.) This function can be represented by its Fourier transform as

$$f(x) = \int dk_x f(k_x) e^{ik_x x}.$$

(S7)

In our ray-optics derivation, we use the approximate representation

$$f_N(x) = \sum_{n=-N}^{N} c_n e^{i\frac{2\pi}{L} nx},$$

(S8)

where

$$N = \text{floor}\left(\frac{kL}{2\pi}\right) = \text{floor}\left(\frac{L}{\lambda}\right).$$

(S9)

Our claim is that the error between $f(x)$ and $f_N(x)$, measured in the L^2 norm, goes to zero in the ray optics limit ($L/\lambda \to \infty$, or $N \to \infty$), i.e.

$$e_N = \|f(x) - f_N(x)\|_2^2 = \int_{-L/2}^{L/2} dx \, |f(x) - f_N(x)|^2 \to 0. \text{ (S10)}$$

Proof: Given Eq. (S7), it is straightforward to find the coefficients c_n that are the optimal Fourier-series coefficients of $f_N(x)$ that most closely approximate $f(x)$:

$$c_n = \frac{1}{L} \int_{-L/2}^{L/2} dx f(x) e^{-i\frac{2\pi}{L} nx} = \int_{-L/2}^{L/2} dx f(x) e^{-ik_n x} \quad \text{(S11)}$$

where $k_n = \frac{2\pi}{L} n$. Then we can insert the Fourier-transform definition of $f(x)$ into Eq. (S11) and integrate to find that the c_n are given by

$$\begin{aligned} c_n &= \frac{1}{L} \int dk_x \int_{-L/2}^{L/2} dx \, e^{i(k_x - k_n)x} f(k_x) \\ &= \int dk_x \, \text{sinc}\left((k_x - k_n)\frac{L}{2}\right) f(k_x), \end{aligned}$$

(S12)

where $\text{sinc}(x) = \frac{\sin(x)}{x}$. Then we substitute c_n into Eq. (S8) to obtain $f_N(x)$ as

$$f_N(x) = \int dk_x \left\{ \sum_{n=-N}^{N} \text{sinc}\left((k_x - k_n)\frac{L}{2}\right) e^{ik_n x} \right\} f(k_x). \text{ (S13)}$$

We can directly insert this expression for $f_N(x)$ into the error expression, Eq. (S10), and carry out the integration over x to obtain

$$\begin{aligned} e_N &= L \int dk_x' \int dk_x f(k_x) f^*(k_x') \left\{ \text{sinc}\left((k_x - k_x')\frac{L}{2}\right) \right. \\ &\left. - \sum_{n=-N}^{N} \text{sinc}\left((k_x - k_n)\frac{L}{2}\right) \text{sinc}\left((k_x' - k_n)\frac{L}{2}\right) \right\}, \end{aligned}$$

(S14)

with the help of the orthogonality of Fourier series, $\int_{-L/2}^{L/2} dx \, e^{i(k_n - k_n')x} = \delta_{nn'} L$.

What remains to be shown is that the whole expression in the curly bracket goes to zero,

$$\begin{aligned} &\text{sinc}\left((k_x - k_x')\frac{L}{2}\right) \\ &\quad - \sum_{n=-N}^{N} \text{sinc}\left((k_x - k_n)\frac{L}{2}\right) \text{sinc}\left((k_x' - k_n)\frac{L}{2}\right) \\ &= \text{sinc}\left((k_x - k_x')\frac{L}{2}\right) \\ &\quad - \sum_{n=-N}^{N} \text{sinc}\left(k_x \frac{L}{2} - n\pi\right) \text{sinc}\left(k_x' \frac{L}{2} - n\pi\right) \text{(S15)} \\ &\longrightarrow 0, \end{aligned}$$

as N goes to ∞. This can be proven by invoking the Mittag-Leffler expansion of $\cot(z)$,

$$\cot(z) = \sum_{n=-\infty}^{\infty} \frac{1}{z - n\pi}, \quad \text{for all complex number} \quad z, \quad \text{(S16)}$$

which can be found in complex-analysis textbooks [6].

In order to use the expansion, we rewrite the double sinc sum by partial fraction. Hence,

$$\sum_{n=-N}^{N} \text{sinc}\left(k_x \frac{L}{2} - n\pi\right) \text{sinc}\left(k_x' \frac{L}{2} - n\pi\right)$$

$$= \sum_{n=-N}^{N} \underbrace{\sin(k_x \frac{L}{2} - n\pi)}_{=\sin(k_x \frac{L}{2})(-1)^n} \underbrace{\sin(k_x' \frac{L}{2} - n\pi)}_{=\sin(k_x' \frac{L}{2})(-1)^n} \frac{1}{k_x \frac{L}{2} - n\pi} \frac{1}{k_x' \frac{L}{2} - n\pi}$$

$$= \sin(k_x \frac{L}{2}) \sin(k_x' \frac{L}{2}) \frac{1}{k_x \frac{L}{2} - k_x' \frac{L}{2}} \qquad (S17)$$

$$\times \underbrace{\sum_{n=-N}^{N} \left(\frac{1}{k_x' \frac{L}{2} - n\pi} - \frac{1}{k_x \frac{L}{2} - n\pi} \right)}_{\longrightarrow \cot(k_x' \frac{L}{2}) - \cot(k_x \frac{L}{2}) \quad \text{as} \quad N \to \infty}$$

$$\longrightarrow \frac{\sin\left((k_x - k_x')\frac{L}{2}\right)}{k_x \frac{L}{2} - k_x' \frac{L}{2}} = \text{sinc}\left((k_x - k_x')\frac{L}{2}\right), \qquad (S18)$$

as N goes to ∞, which is the ray optics limit that $\frac{L}{\lambda}$ goes to ∞. It is worth mentioning that the convergence is uniform, as convergence of Eq. (S16) is uniformly for all complex z. Consequently, for any bounded $f(k)$, we can bring the N limit into the k_x and k_x' integral in Eq. (S14), proving that

$$\lim_{N\to\infty} e_N = 0. \qquad (S19)$$

To conclude, we have proven that the discrete plane wave basis in the ray-optics limit is capable of representing arbitrary wave forms. More abstractly, what we have proven is a well-known fact — that a Fourier series is able to represent any integrable function on a closed interval in the L^2 norm sense [5]. The L^2 norm is the physically meaningful one here, since observable power/momentum flows are defined by inner products of the appropriate wave/field functions.

2. ALTERNATIVE PROOF FOR THE BRIGHTNESS-CONCENTRATION INEQUALITY

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

3. COUPLED-MODE THEORY FOR ÉTENDUE TRANSMISSION

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

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

are related by the coupled-mode equations:

$$i(\mathbf{\Omega} - \omega)\mathbf{a} = \mathbf{K}^T \mathbf{c}_{\text{in}} \qquad (S20)$$

$$\mathbf{c}_{\text{out}} = \mathbf{C}\mathbf{c}_{\text{in}} + \mathbf{K}\mathbf{a}, \qquad (S21)$$

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

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

$$\mathbf{c}_{\text{out}} = [\underbrace{\mathbf{C}}_{\text{direct process}} - \underbrace{i\mathbf{K}(\mathbf{\Omega} - \omega)^{-1}\mathbf{K}^T}_{\text{resonance assisted process}}]\mathbf{c}_{\text{in}},$$
$$(S22)$$

where the term in square brackets is the scattering matrix.

We are interested in the *resonance assisted* transmission from N_{inc} to an orthogonal subset of N_{trans} scattering channels. So we are invited to write Eq. (S22) as

$$\underbrace{\mathbf{c}_{\text{out}} - \mathbf{C}\mathbf{c}_{\text{in}}}_{\mathbf{c}_{\text{out}}^{\text{reson}}} = -i\mathbf{K}(\mathbf{\Omega} - \omega)^{-1}\mathbf{K}^T \mathbf{c}_{\text{in}}. \qquad (S23)$$

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

$$\mathbf{c}_{\text{trans}} = \mathbf{T}\mathbf{c}_{\text{inc}}, \qquad (S24)$$

can be written as

$$\mathbf{T} = -i\mathbf{K}_{\text{trans}}(\mathbf{\Omega} - \omega)^{-1}\mathbf{K}_{\text{inc}}^T. \qquad (S25)$$

With the T-matrix definition of Eq. (S24), we can analyze resonance assisted power flow into the transmission channels due to incoherent excitations of the incident channels with the same matrix-trace approach used for $\mathbf{c}_{\text{inc}}$, $\mathbf{c}_{\text{out}}$, and the scattering matrix in the main text. As a result, the average power in transmission channel i is given by

$$\langle |\mathbf{c}_{\text{trans},i}|^2 \rangle = \mathbf{u}_i^\dagger \mathbf{T}\rho_{\text{inc}}\mathbf{T}^\dagger \mathbf{u}_i. \qquad (S26)$$

Let us consider the total average power (equal to the average total power) transmitted onto all transmission channels, which

is simply the sum of Eq. (S26) over i:

$$\sum_i \langle |\mathbf{c}_{\text{trans},i}|^2 \rangle = \sum_i \mathbf{u}_i^\dagger \mathbf{T} \boldsymbol{\rho}_{\text{inc}} \mathbf{T}^\dagger \mathbf{u}_i \tag{S27}$$

$$= \text{Tr}\left(\mathbf{T} \boldsymbol{\rho}_{\text{inc}} \mathbf{T}^\dagger \sum_i \mathbf{u}_i \mathbf{u}_i^\dagger \right) \tag{S28}$$

$$= \text{Tr}\left(\mathbf{T} \boldsymbol{\rho}_{\text{inc}} \mathbf{T}^\dagger \right) = \text{Tr}\left(\mathbf{T}^\dagger \mathbf{T} \boldsymbol{\rho}_{\text{inc}} \right) \tag{S29}$$

$$= \text{Tr}\left(\mathbf{T}^\dagger \mathbf{T} \boldsymbol{\rho}_{\text{inc}} \sum_i \mathbf{v}_i \mathbf{v}_i^\dagger \right) \tag{S30}$$

$$= \sum_i \lambda_i \mathbf{v}_i^\dagger \mathbf{T}^\dagger \mathbf{T} \mathbf{v}_i \,, \tag{S31}$$

where $\mathbf{v}_i$ is the eigenvector of $\boldsymbol{\rho}_{\text{inc}}$ with eigenvalue λ_i (we label those eigenvalues in decreasing order, i.e. $\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots$). The summation over i includes all necessary channels so that $\sum_i \mathbf{u}_i \mathbf{u}_i^\dagger$ and $\sum_i \mathbf{v}_i \mathbf{v}_i^\dagger$ are both identity. However, the rank of $\mathbf{T}$ restricts the meaningful inclusion of $\mathbf{v}_i$ – the number of $\mathbf{v}_i$ such that $\mathbf{T}\mathbf{v}_i \neq 0$ is at most $\text{rank}(\mathbf{T})$ as a consequence of the rank-nullity theorem[2]. Besides, since there are at most $\text{rank}(\boldsymbol{\rho}_{\text{inc}})$ number of $\mathbf{v}_i$s are relevant, we can thus bound Eq. (S31) by

$$\sum_i \langle |\mathbf{c}_{\text{trans},i}|^2 \rangle \leq \sum_{i=1}^{U} \lambda_i \mathbf{v}_i^\dagger \mathbf{T}^\dagger \mathbf{T} \mathbf{v}_i \,, \tag{S32}$$

where $U = \min\left(\text{rank}(\boldsymbol{\rho}_{\text{inc}}), \text{rank}(\mathbf{T}) \right)$.

In the main text, we used a coherent-scattering example to show that for each individual i, $\mathbf{u}_i^\dagger \mathbf{S} \mathbf{S}^\dagger \mathbf{u}_i \leq 1$. Exactly the same proof, but with the inc/trans channels replacing the in/out channels, leads to the same inequality for $\mathbf{T}$, i.e. $\mathbf{u}_i^\dagger \mathbf{T} \mathbf{T}^\dagger \mathbf{u}_i \leq 1$ and $\mathbf{v}_i^\dagger \mathbf{T}^\dagger \mathbf{T} \mathbf{v}_i \leq 1$. Thus all of the eigenvalues of $\mathbf{T}\mathbf{T}^\dagger$ and $\mathbf{T}^\dagger\mathbf{T}$ thus both less than or equal to one (and they are greater than or equal to zero because $\mathbf{T}\mathbf{T}^\dagger$ and $\mathbf{T}^\dagger\mathbf{T}$ are positive semidefinite). This reduce Eq. (S31) into

$$\sum_i \langle |\mathbf{c}_{\text{trans},i}|^2 \rangle \leq \sum_{i=1}^{U} \lambda_i \,. \tag{S33}$$

We can still do better by bounding the rank of $\mathbf{T}$ using the fact that $\text{rank}(AB) \leq \min\left(\text{rank}(A), \text{rank}(B) \right)$. We write again the coupled-mode expression for $\mathbf{T}$, Eq. (S25), now enumerating the number of rows and columns of each matrix:

$$\mathbf{T} = -i \underbrace{\mathbf{K}_{\text{trans}}}_{N_{\text{trans}} \times M} \underbrace{(\boldsymbol{\Omega} - \omega)^{-1}}_{M \times M} \underbrace{\mathbf{K}_{\text{inc}}^T}_{M \times N_{\text{inc}}} \,. \tag{S34}$$

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

$$\text{rank}(\mathbf{T}) \leq \min\left(N, M, N_{\text{trans}} \right). \tag{S35}$$

Finally, combining Eq. (S33), Eq. (S35) and the fact $\text{rank}(\boldsymbol{\rho}_{\text{in}}) \leq N_{\text{inc}}$ gives

$$\sum_i \langle |\mathbf{c}_{\text{trans},i}|^2 \rangle \leq \sum_{i=1}^{\min(\text{rank}(\boldsymbol{\rho}_{\text{inc}}), M, N_{\text{trans}})} \lambda_i, \tag{S36}$$

where eigenvalues are placed in decreasing order with $\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots$. In the case of equal, incoherent excitations (for which all of the eigenvalues of $\boldsymbol{\rho}_{\text{inc}}$ are $1/N_{\text{inc}}$):

$$\sum_i \langle |\mathbf{c}_{\text{trans},i}|^2 \rangle \leq \frac{\min\left(N_{\text{inc}}, M, N_{\text{trans}} \right)}{N_{\text{inc}}} \,. \tag{S37}$$

4. COUPLING CONSTANTS OF COUPLE-MODE MODELS OF WAVEGUIDE COMBINERS

In this section, we list the coupling constants engineered to maximize transmission of waveguide combiners in Fig. 4. A random sweep on these parameters are done and those achieving a full transmission are selected. As aforementioned, $\mathbf{C}$ is chosen to be $-\mathbf{I}$ in Eq. (S21). The $\boldsymbol{\Omega}$ in Eq. (S20) can be further split into a diagonal Hermitian part $\boldsymbol{\Omega}_H$ and an anti-Hermitian part $-i\boldsymbol{\Gamma}$, i.e.

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}_H - i\boldsymbol{\Gamma} \,. \tag{S38}$$

Due to the constraint,

$$2\boldsymbol{\Gamma} = \mathbf{K}^\dagger \mathbf{K} \,, \tag{S39}$$

it is sufficient to specify only $\boldsymbol{\Omega}_H$ and $\mathbf{K}$ to fully determine the coupled mode equations in Eqs. (S20,S21). For simplicity, in all three cases, $\mathbf{K}$ is chosen so that $\boldsymbol{\Gamma}$ is diagonal, meaning the resonance modes are orthogonal.

The numerical value of $\mathbf{K}$ and $\boldsymbol{\Omega}_H$ are as follows:
Waveguide combiner (a): two channels to one channel with two resonances:

$$\boldsymbol{\Omega}_H = \begin{pmatrix} \omega_0 & 0 \\ 0 & \omega_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{K} = \begin{pmatrix} 3.58 & 2.81 \\ 2.81 & -3.58 \\ 4.09 & 0 \end{pmatrix} \,. \tag{S40}$$

Waveguide combiner (b): two channels to two channels with one resonance:

$$\boldsymbol{\Omega}_H = \omega_0 = 1 \quad \text{and} \quad \mathbf{K} = \begin{pmatrix} 3.58 \\ 2.81 \\ 4.09 \\ 1.93 \end{pmatrix} \,. \tag{S41}$$

Waveguide combiner (c): two channels to two channels with two resonances:

$$\boldsymbol{\Omega}_H = \begin{pmatrix} \omega_0 & 0 \\ 0 & \omega_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{K} = \begin{pmatrix} 3.58 & 2.81 \\ 2.81 & -3.58 \\ 4.09 & 1.93 \\ 1.93 & -4.09 \end{pmatrix} \,. \tag{S42}$$

5. PERMITTIVITY DATA OF OPTIMIZED METASURFACE UNIT CELLS AND INVERSE DESIGN METHODOLOGY

In this section we provide the permittivity ϵ data of the unit cells of metasurfaces in Fig. 2 of the main text. All structures have unit cell of size 2λ and thickness 0.5λ, where λ is the wavelength of incident excitations. The permittivity ϵ is constant in the vertical dimension and are discretized into 100 equal-sized grids in the horizontal dimension. The leftmost grid is labelled as 1 and the rightmost one is labelled as 100. In the following table, we tabulate the numerical value of ϵ from grid 1 to grad 100 for all nine structures, represented by the color they appear in Fig. 2(b) and (c). A blazed grating of the same unit cell dimension, shown in Fig. ??, is optimized as a comparison to metasurfaces in Fig.
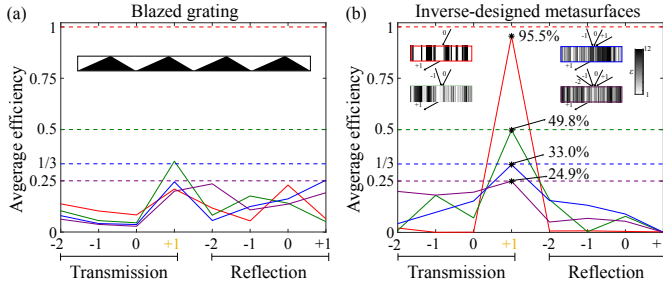
Fig. S1. (a)An optimized blazed grating (vertex angle 24.8°) falls far short of the concentration limit, compared with Fig. 2(c) in the main text. (b)A copy of Fig. 2(c) in the main text.

2(c) in the main text. The performance of the blazed grating falls far short of the theoretical limit.

In order to design such metasurfaces with 100 design parameters (permitivity ϵ in each grid), the adjoint method is adopted for computing the gradient of an objective function with respect to each design parameter efficiently. The objective function in our case is the sum of transmission power of various orders. Once the gradient is known, standard optimization technique, such as steepest descent, can be used to optimize the objective function to a desirable value. The adjoint method allows the gradient computation to be done in two simulations for each optimization iteration — a direct simulation to obtain the electric field $\mathbf{E}_{\mathrm{dir}}$ everywhere in the designable region and a adjoint to obtain the $\mathbf{E}_{\mathrm{adj}}$ everywhere. Then the gradient at each position is proportional to the imaginary part of $\mathbf{E}_{\mathrm{dir}} \cdot \mathbf{E}_{\mathrm{adj}}$. The source of the adjoint simulation is often the time-reversed field to be optimized, which is obtained in the direct simulation.

With all that said, in the following section we provide a detailed derivation of the above procedures described. Suppose we have a generic figure of merit $f(\mathbf{E}, \mathbf{H})$ and given some initial design with material distribution $\epsilon(\mathbf{r})$ and $\mu(\mathbf{r})$ to be optimized. We need to first solve the "direct" problem and obtain the current value of figure of merit f, $\mathbf{E}_{\mathrm{dir}}$ and $\mathbf{H}_{\mathrm{dir}}$ everywhere . To see what is the next step to go, we perturb the design by $\delta\epsilon(\mathbf{r})$ and $\delta\mu(\mathbf{r})$ in the designable region, then the response of f due to this material variation is given by [3]

$$\delta f = 2\mathrm{Re}\left\{ \int (\mathrm{d}\mathbf{r})\delta\epsilon(\mathbf{r})\mathbf{E}_{\mathrm{dir}} \cdot \mathbf{E}_{\mathrm{adj}} + \delta\mu(\mathbf{r})\mathbf{H}_{\mathrm{dir}} \cdot \mathbf{H}_{\mathrm{adj}} \right\}, \quad (S43)$$

where the integral is over the designable region. The key is to calculate $\mathbf{E}_{\mathrm{adj}}$ and $\mathbf{H}_{\mathrm{adj}}$ generated by the adjoint sources $\mathbf{P}_{\mathrm{adj}}$ and $\mathbf{M}_{\mathrm{adj}}$ respectively. This can be done by solving the "adjoint" problem (to the "direct" one) described in the following.

By the reciprocity of Green's functions, the adjoint sources are of the form [3]

$$\mathbf{P}_{\mathrm{adj}} = \frac{\partial f}{\partial \mathbf{E}} \quad \text{and} \quad \mathbf{M}_{\mathrm{adj}} = -\frac{\partial f}{\partial \mathbf{H}}, \quad (S44)$$

hence,

$$\mathbf{J}_{\mathrm{adj}}^{E} = -\mathrm{i}\omega\frac{\partial f}{\partial \mathbf{E}} \quad \text{and} \quad \mathbf{J}_{\mathrm{adj}}^{M} = \mathrm{i}\omega\frac{\partial f}{\partial \mathbf{H}}, \quad (S45)$$

where $\mathbf{J}_{\mathrm{adj}}^{E}$ and $\mathbf{J}_{\mathrm{adj}}^{M}$ are the adjoint electric current and magnetic current respectively. Next, we expand any field by power-orthogonal modes. This form of expansion is particularly conve-

nient for the case of metasurfaces. We define

$$\Theta = \begin{pmatrix} & -\mathbf{n}\times \\ \mathbf{n}\times & \end{pmatrix}, \quad (S46)$$

so that the modal orthogonality relation can be written as

$$\int \mathrm{d}\mathbf{S} \cdot \begin{pmatrix} \mathbf{E}_i \\ \mathbf{H}_i \end{pmatrix}^{\dagger} \Theta \begin{pmatrix} \mathbf{E}_j \\ \mathbf{H}_j \end{pmatrix} = 2\mathrm{Re}\left\{ \int \mathrm{d}S\mathbf{n} \cdot \mathbf{E}_i \times \mathbf{H}_j^* \right\} = 4P_i\delta_{ij}, \quad (S47)$$

where P_i carries the meaning of time averaged power of the ith mode. The integral is over the surface through which the energy flow and $\mathbf{n}$ is the unit normal vector to that surface. In this basis, we have

$$\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \sum_k a_k \begin{pmatrix} \mathbf{E}_k \\ \mathbf{H}_k \end{pmatrix}, \quad (S48)$$

with

$$\begin{aligned} a_k &= \frac{1}{4P_k} \int \mathrm{d}\mathbf{S} \cdot \begin{pmatrix} \mathbf{E}_k \\ \mathbf{H}_k \end{pmatrix}^{\dagger} \Theta \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} \\ &= \frac{1}{4P_k} \int \mathrm{d}\mathbf{S} \cdot \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}^{T} \Theta \begin{pmatrix} \mathbf{E}_k \\ \mathbf{H}_k \end{pmatrix}^{*}. \end{aligned} \quad (S49)$$

If the geometry in the region where our figure of merit is evaluated is not altered, the modes in the expansion Eq. (S48) remains the same. The only thing changes is the expansion amplitude a_k. Hence, it is more convenient to write sources $\mathbf{P}$ and $\mathbf{M}$ as

$$\mathbf{P}_{\mathrm{adj}} = \sum_k \frac{\partial f}{\partial a_k} \underbrace{\frac{\partial a_k}{\partial \mathbf{E}}}_{-\frac{1}{4P_k}\mathbf{n}\times\mathbf{H}_k^*} \quad \text{and} \quad \mathbf{M}_{\mathrm{adj}} = -\sum_k \frac{\partial f}{\partial a_k} \underbrace{\frac{\partial a_k}{\partial \mathbf{H}}}_{\frac{1}{4P_k}\mathbf{n}\times\mathbf{E}_k^*}, \quad (S50)$$

where the $\mathbf{E}$ and $\mathbf{H}$ derivative follows from Eq. (S49). Combine Eq. (S45) and Eq. (S50), we have

$$\begin{pmatrix} \mathbf{J}_{\mathrm{adj}}^{E} \\ \mathbf{J}_{\mathrm{adj}}^{M} \end{pmatrix} = \sum_k \frac{\mathrm{i}\omega}{4P_k} \frac{\partial f}{\partial a_k} \begin{pmatrix} \mathbf{n} \times \mathbf{H}_k^* \\ \mathbf{n} \times \mathbf{E}_k^* \end{pmatrix}. \quad (S51)$$

By the equivalence principle [4],

$$\begin{pmatrix} \mathbf{J}^{E} \\ \mathbf{J}^{M} \end{pmatrix} = \begin{pmatrix} \mathbf{n} \times \mathbf{H}_{\mathrm{eqv}} \\ -\mathbf{n} \times \mathbf{E}_{\mathrm{eqv}} \end{pmatrix}, \quad (S52)$$

we conclude that the incident fields of the "adjoint" problem is

$$\begin{pmatrix} \mathbf{E}_{\mathrm{adj}} \\ \mathbf{H}_{\mathrm{adj}} \end{pmatrix} = -\sum_k \frac{\mathrm{i}\omega}{4P_k} \frac{\partial f}{\partial a_k} \underbrace{\begin{pmatrix} \mathbf{E}_k^* \\ -\mathbf{H}_k^* \end{pmatrix}}_{\text{time-reversed pair}}. \quad (S53)$$

Recover its correct dimensions, we obtain the final expression:

$$\delta f = \frac{\omega}{2}\mathrm{Im}\left\{ \int \mathrm{d}V\delta\epsilon(\mathbf{r})\mathbf{E}_{\mathrm{dir}} \cdot \mathbf{E}_{\mathrm{adj}} + \delta\mu(\mathbf{r})\mathbf{H}_{\mathrm{dir}} \cdot \mathbf{H}_{\mathrm{adj}} \right\}, \quad (S54)$$

with incident fields for the "adjoint" problem as

$$\begin{pmatrix} \mathbf{E}_{\mathrm{adj}} \\ \mathbf{H}_{\mathrm{adj}} \end{pmatrix} = \sum_k \frac{\partial f}{\partial a_k} \frac{1}{P_k} \underbrace{\begin{pmatrix} \mathbf{E}_k^* \\ -\mathbf{H}_k^* \end{pmatrix}}_{\text{time-reversed pair}}. \quad (S55)$$

When the structure is discretized into small grids, where the ith grid has volume V_i and material constant denoted by ϵ_i and μ_i. If the variation of $\delta\epsilon$ and $\delta\mu$ is constant for each grid, Eq. (S54) reduces to a grid-wise expression

$$\frac{\delta f}{\delta \epsilon_i} = \frac{\omega}{2} \text{Im} \left\{ \int_{V_i} \mathrm{d}V \, \mathbf{E}_{\text{dir}} \cdot \mathbf{E}_{\text{adj}} \right\}, \tag{S56}$$

$$\frac{\delta f}{\delta \mu_i} = \frac{\omega}{2} \text{Im} \left\{ \int_{V_i} \mathrm{d}V \, \mathbf{H}_{\text{dir}} \cdot \mathbf{H}_{\text{adj}} \right\}. \tag{S57}$$

These two quantities serve as the gradient of the figure of merit f with respect to material constants, which can be used in gradient-based optimization algorithms.

REFERENCES

1. Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, 2nd ed. (Cambridge University Press, New York, NY, 2013) p. 235.
2. Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, 2nd ed. (Cambridge University Press, New York, NY, 2013) p. 6.
3. Owen Dennis Miller, *Photonic Design: From Fundamental Solar Cell Physics to Computational Inverse Design*, Ph.D. thesis, University of California, Berkeley (2012).
4. Allen Taflove, Ardavan Oskooi, and Steven G Johnson, *Advances in FDTD computational electrodynamics: photonics and nanotechnology* (Artech house, 2013).
5. Elias M Stein, and Rami Shakarchi, *Fourier analysis: an introduction* (Princeton University Press, 2011).
6. Edmund T Whittaker, and George N Watson, *A course of modern analysis* (Cambridge university press, 1996).

Grid Number	Red	Green	Blue	Purple
	Permittivity ϵ			
1	11.89	7.922	6.601	11.39
2	11.89	9.241	5.928	8.304
3	6.309	2.162	9.373	12
4	6.309	2.866	2.437	9.75
5	11.92	1	2.641	10.39
6	11.92	1.752	3.947	8.941
7	1.122	9.385	8.295	7.522
8	1.122	7.499	4.863	2.762
9	1.102	11.8	3.239	4.076
10	1.102	5.583	5.46	6.249
11	1.088	11.9	5.003	9.985
12	1.088	8.991	6.793	11.97
13	1.086	12	6.307	9.947
14	1.086	9.933	8.678	6.409
15	1.08	11.27	6.4	10.81
16	1.08	10.97	11.46	5.345
17	1.06	11.68	2.831	5.306
18	1.06	11.84	1.311	2.27
19	11.86	12	1.081	1.036
20	11.86	3.006	1.214	6.295
21	1.083	3.88	3.766	9.55
22	1.083	1.151	3.966	7.194
23	1.12	2.929	2.799	4.837
24	1.12	2.164	6.935	7.654
25	1.142	2.22	5.21	6.807
26	1.142	4.144	6.708	8.086
27	11.94	9.396	8.646	9.314
28	11.94	10.45	9.447	9.668
29	11.93	12	1.272	8.637
30	11.93	11.59	1.016	4.561
31	11.95	11.08	7.303	9.999
32	11.95	8.987	1.188	1.809
33	11.96	10.81	5.516	7.839
34	11.96	7.072	5.624	11.81
35	11.95	2.775	9.345	9.258
36	11.95	4.248	5.329	10.3
37	1.094	9.242	4.408	7.411
38	1.094	6.303	7.479	9.25
39	11.84	4.283	8.235	7.464
40	11.84	6.691	5.801	1.981

Grid Number	Red	Green	Blue	Purple
	Permittivity ϵ			
41	9.231	4.691	3.92	1
42	9.231	1.011	1.797	1.319
43	1.082	1.152	6.734	5.255
44	1.082	1.072	3.442	7.487
45	1.103	1.008	1.475	2.789
46	1.103	1.006	9.828	8.16
47	1.128	2.284	4.833	7.361
48	1.128	1.877	7.704	8.3
49	1.131	1.055	10.98	5.845
50	1.131	2.286	7.713	2.966
51	1.115	2.537	10.44	8.221
52	1.115	2.906	8	2.168
53	1.14	11.87	10.85	2.438
54	1.14	5.203	10.67	2.178
55	6.249	4.319	7.427	2.906
56	6.249	1.7	2.668	6.846
57	9.584	3.795	6.294	2.725
58	9.584	2.538	7.808	8.729
59	11.87	3.437	6.719	6.055
60	11.87	5.222	6.74	1.75
61	11.92	3.329	11.19	3.582
62	11.92	1.498	10.84	9.64
63	11.94	2.441	9.482	6.714
64	11.94	5.191	10.84	2.886
65	1.083	7.84	7.965	4.687
66	1.083	4.523	11.55	7.332
67	11.71	4.739	9.329	2.485
68	11.71	9.656	9.427	9.137
69	1.093	3.601	12	3.6
70	1.093	1.047	9.771	11.16
71	1.118	5.856	11.51	3.705
72	1.118	1	11.69	8.737
73	1.138	1.139	8.158	2.504
74	1.138	3.608	11.52	3.502
75	11.84	1.52	9.547	1.809
76	11.84	4.929	12	7.815
77	10.04	1.283	12	9.838
78	10.04	4.892	11.3	9.231
79	1.064	2.854	12	8.865
80	1.064	1.812	12	10.38

Grid Number	Red	Green	Blue	Purple
		Permittivity ϵ		
81	1.073	5.041	11.98	8.879
82	1.073	5.184	11.48	11.94
83	1.064	4.975	9.275	11.33
84	1.064	2.693	2.784	11.07
85	11.87	11.7	2.508	5.28
86	11.87	7.643	4.763	9.666
87	8.271	10.95	8.947	3.77
88	8.271	11.7	10.33	2.205
89	11.48	4.405	10.2	7.662
90	11.48	3.888	4.861	6.221
91	11.88	7.989	6.849	6.63
92	11.88	6.988	1.549	9.155
93	11.87	1.438	1.429	10.6
94	11.87	5.304	1.429	6.188
95	1.085	9.64	7.258	10.06
96	1.085	1.796	5.252	8.24
97	1.101	1.106	2.026	9.773
98	1.101	4.945	5.581	12
99	1.086	10.08	1.927	10.13
100	1.086	12	3.408	4.715

Table S1. Permittivity value of every grid point of the four metasurfaces in Fig. 2(b).

Grid Number	Red	Blue	Purple	Yellow	Green
			Permittivity ϵ		
1	7.548	11.76	10.33	5.319	2.667
2	6.51	11.72	10.6	2.436	1.902
3	5.003	8.872	8.305	3.825	1.941
4	3.994	10.95	8.504	9.406	2.273
5	3.516	11.14	4.964	9.629	2.575
6	3.279	11.47	6.554	11.69	3.744
7	3.361	11.43	7.313	11.54	4.91
8	5.218	11.65	4.648	10.3	6.271
9	7.594	11.72	4.625	11.08	7.473
10	9.32	11.75	3.695	11.22	9.204
11	10.39	11.62	3.419	11.21	9.069
12	10.63	11.75	3	10.45	8.726
13	10.9	10.44	2.565	11.19	7
14	10.79	6.259	1.627	10.88	6.062
15	10.35	5.46	2.419	10.98	4.065
16	8.859	1.405	1.327	8.252	5.774
17	6.678	1.627	1.432	1.511	2.957
18	5.141	1.531	1.866	7.184	3.153
19	4.571	1.272	1.622	4.697	3.562
20	4.363	1.323	1.067	7.794	3.969
21	5.412	1.395	1.049	9.355	4.899
22	7.32	2.085	1.082	11.43	5.808
23	9.546	2.75	2.212	11.04	6.244
24	10.62	3.388	1.08	11.41	7.843
25	11.12	4.359	1.566	10.74	7.186
26	11.25	5.939	5.879	6.003	6.844
27	11.24	7.008	7.23	3.399	6.006
28	11.13	5.632	5.488	1.66	4.507
29	10.88	7.903	6.349	1.488	2.733
30	10.48	4.988	7.377	1.744	2.723
31	9.931	8.198	10.4	1.409	5.077
32	7.629	5.222	9.879	2.492	5.116
33	6.943	4.609	7.52	3.651	7.973
34	8.429	4.359	6.119	8.259	9.132
35	9.335	3.637	2.806	5.748	8.091
36	10.13	7.12	4.074	7.071	8.269
37	10.31	6.297	3.358	8.606	6.003
38	10.55	3.424	1.256	8.709	4.619
39	10.52	5.281	1.609	10.07	3.256
40	10.57	5.094	1.643	10.34	2.414

	Red	Blue	Purple	Yellow	Green
Grid Number			Permittivity ϵ		
41	10.6	3.568	2.134	11.83	2.485
42	10.58	3.08	5.434	11.81	3.315
43	10.42	4.86	7.515	11.78	6.19
44	9.998	5.014	9.496	11.8	7.499
45	9.851	2.855	11.42	11.73	8.852
46	9.685	3.787	9.191	11.86	6.373
47	9.555	3.951	11.59	11.78	9.789
48	9.546	6.04	7.952	8.847	9.597
49	9.776	7.333	6.995	7.929	8.868
50	9.774	8.408	6.26	4.926	9.366
51	9.415	5.725	6.841	5.954	9.978
52	8.933	8.773	11.06	3.835	10.5
53	8.586	8.53	11.26	3.844	11.24
54	8.816	5.195	11.58	5.161	11.19
55	9.194	4.942	11.89	3.807	10.59
56	10.05	6.663	10.94	6.555	9.29
57	10.75	7.651	10.86	4.949	7.496
58	11.08	5.35	11.95	3.269	6.432
59	11.14	7.022	11.97	4.007	4.887
60	11.11	4.497	11.97	5.665	4.246
61	10.79	7.725	8.733	3.005	5.592
62	9.882	4.187	2.17	1.672	5.903
63	7.491	5.22	2.979	5.38	7.234
64	4.765	5.925	6.16	3.873	7.658
65	2.963	5.684	5.086	2.422	10.03
66	2.321	3.596	10.26	1.918	9.757
67	2.221	4.134	11.61	2.036	11.07
68	2.056	2.415	11.59	3.571	10.92
69	1.961	2.357	9.007	2.757	11.09
70	1.83	2.799	11.59	4.709	11.19
71	1.76	6.662	11.93	2.843	11.16
72	1.718	4.448	11.68	4.454	11.53
73	1.694	9.081	11.84	3.486	11.32
74	1.676	7.638	11.81	3.824	11.37
75	1.677	10.68	11.86	4.051	11.39
76	1.686	10.62	11.91	3.41	10.96
77	1.714	10.33	11.95	3.051	11.41
78	1.75	11.04	11.86	3.028	11.32
79	1.803	10.66	11.91	2.075	10.9
80	1.895	11.06	11.95	10.31	10.37

	Red	Blue	Purple	Yellow	Green
Grid Number			Permittivity ϵ		
81	1.979	10.95	5.021	6.687	9.682
82	2.1	11.26	4.714	7.621	9.162
83	2.195	11.7	1.117	6.089	8.026
84	2.423	11.54	2.396	6.826	5.468
85	2.549	11.67	2.281	8.962	4.681
86	2.769	11.35	3.909	6.953	3.953
87	2.897	11.15	4.568	11.58	3.826
88	2.941	8.333	3.977	11.35	3.519
89	2.952	3.502	7.863	10.73	3.396
90	2.884	3.468	8.549	10.7	4.025
91	2.887	1.864	7.358	11.82	3.814
92	3.153	1.978	11.06	11.84	4.737
93	3.389	2.869	10.58	10.25	5.439
94	4.322	4.798	9.657	11.1	6.352
95	5.968	3.351	10.15	10.62	8.625
96	7.621	3.65	9.488	10.04	8.412
97	8.799	7.751	11.96	9.177	7.233
98	9.374	8.295	10.58	4.901	6.993
99	9.461	7.933	9.879	2.688	4.756
100	9.068	9.613	9.359	3.652	3.318

Table S2. Permittivity value of every grid point of the five metasurfaces in Fig. 2(c).