Overvig et al. *Light: Science & Applications* (2024)13:28
https://doi.org/10.1038/s41377-023-01350-9

Official journal of the CIOMP 2047-7538
www.nature.com/lsa

ARTICLE Open Access

Spatio-temporal coupled mode theory for nonlocal metasurfaces

Adam Overvig[1], Sander A. Mann[1] and Andrea Alù[1,2]✉

Abstract

Diffractive nonlocal metasurfaces have recently opened a broad range of exciting developments in nanophotonics research and applications, leveraging spatially extended—yet locally patterned—resonant modes to control light with new degrees of freedom. While conventional grating responses are elegantly captured by temporal coupled mode theory, current approaches are not well equipped to capture the arbitrary spatial response observed in the nascent field of nonlocal metasurfaces. Here, we introduce spatio-temporal coupled mode theory (STCMT), capable of elegantly capturing the key features of the resonant response of wavefront-shaping nonlocal metasurfaces. This framework can quantitatively guide nonlocal metasurface design while maintaining compatibility with local metasurface frameworks, making it a powerful tool to rationally design and optimize a broad class of ultrathin optical components. We validate this STCMT framework against full-wave simulations of various nonlocal metasurfaces, demonstrating that this tool offers a powerful semi-analytical framework to understand and model the physics and functionality of these devices, without the need for computationally intense full-wave simulations. We also discuss how this model may shed physical insights into nonlocal phenomena in photonics and the functionality of the resulting devices. As a relevant example, we showcase STCMT's flexibility by applying it to study and rapidly prototype nonlocal metasurfaces that spatially shape thermal emission.

Introduction

Coupled mode theory (CMT) has become a powerful tool to model and understand the scattering response of complex resonant electromagnetic structures[1]. Central to CMT is the phenomenological identification of (i) the resonant modes supported by the system, and (ii) ports, or channels, that carry energy to and from the resonant scatterer. CMT provides an intuitive and accurate framework to study a wide range of resonant devices. In the spatial domain[2–4], it is commonly used to study waveguide modes and their coupling; while in the time domain temporal coupled mode theory (TCMT)[5,6] has found marked success in modeling Fano resonances and in the construction of scattering matrices. This technique can be applied both to localized resonances[7–9] and to extended resonances supported by metasurfaces, photonic crystal slabs and subwavelength gratings[10,11]. However, in TCMT the spatial distribution of the mode is ignored: the modal amplitude a is assumed to be function only of time. This feature limits TCMT to studying the spectral features of the device, while detailed spectro-spatial information of the modes themselves is not captured. For this reason, TCMT is primarily effective at studying infinitely periodic grating structures, while finite and spatially varying gratings are not straightforwardly modeled.

Yet, the spectro-spatial properties of photonic resonant systems are important in various settings, particularly when the optical energy is correlated across distances larger than the wavelength—i.e., when nonlocality, also known as spatial dispersion, cannot be neglected. In turn, these systems exhibit strong dependence of the optical response on the momentum of light[12,13]. In dielectric thin film platforms, the dispersion of guided modes provides an engineerable mechanism to control nonlocal phenomena[14–16]. In this

Correspondence: Andrea Alù (aalu@gc.cuny.edu)
[1]Photonics Initiative, Advanced Science Research Center, City University of New York, New York, NY 10031, USA
[2]Physics Program, Graduate Center of the City University of New York, New York, NY 10016, USA

© The Author(s) 2024
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

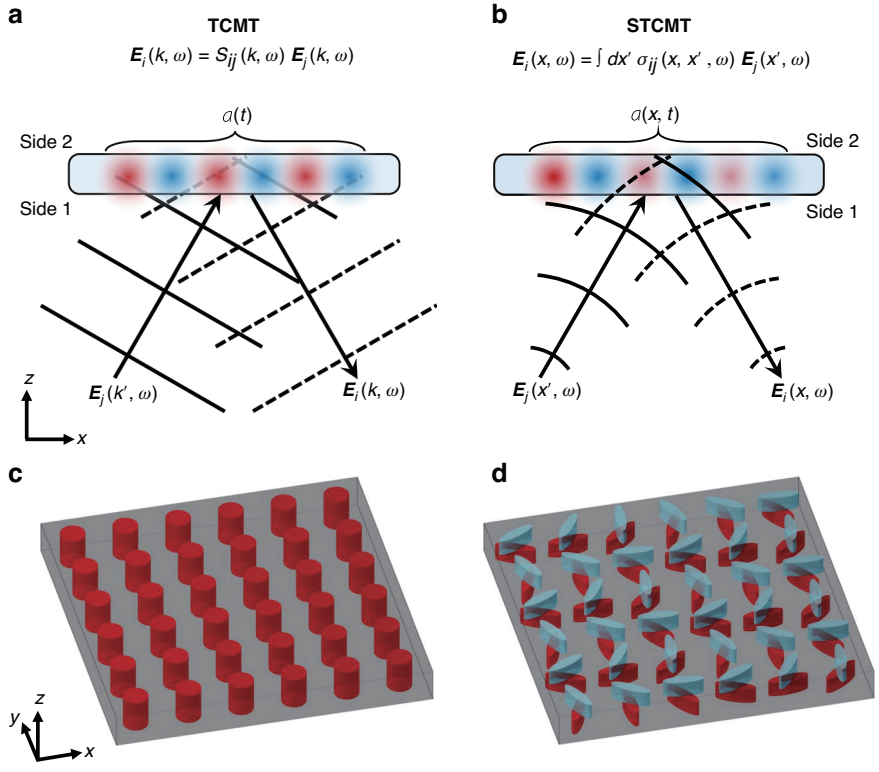

Fig. 1 Coupled mode theories. a TCMT is effective at modeling systems in which the complex modal amplitude $a(t)$ is spatially invariant, wherein the scattered field $\mathbf{E}_j$ is the product of the scattering matrix S_{ij} and the incident field $\mathbf{E}_i$. **b** STCMT is necessary when the complex modal amplitude $a(x,t)$ is spatially varying, wherein the scattered field $\mathbf{E}_j$ is the convolution of the scattering kernel σ_{ij} and the incident field $\mathbf{E}_i$. **c** Example conventional nonlocal metasurface with spatially invariant meta-unit cells. **d** Example diffractive nonlocal metasurface with spatially varying meta-unit cells, varying from left to right

context, nonlocal metasurfaces[17–19] have been garnering attention for their capability of enhancing the control of light by tailoring it in momentum space, of particular interest for signal processing[18], image differentiation[19,20], spectral[21] and chiral sensing[22], augmented reality applications[23] and for compactifying free space[24,25]. This class of devices employs subwavelength, transversely invariant structures, hence the spatial features of their extended resonant modes do not escape the near field of the device. As a result, TCMT remains effective, lumping the resonance into a spatially invariant modal parameter $a(t)$ [Fig. 1a].

Recent reviews[26], perspectives[27–29] and studies on the fundamental role of nonlocality in optics[30] suggest that embracing nonlocality is a budding area of inquiry and a compelling direction for the maturing field of metasurfaces. In particular, engineered nonlocalities can be combined with locally varying features to enable *diffractive nonlocal metasurfaces*[28]. These ultrathin optical devices support highly selective resonant responses combined with tailored spatial variations, which are capable of patterning the optical wavefront engaging the resonance[31–39]. Leveraging the high-Q physics of quasi-bound states in the continuum (q-BICs)[40], these devices support long-lived, spatially extended resonant states that couple to free-space with amplitude, phase and polarization properties spatially tailored across the aperture through symmetry-controlled perturbations. By spatially varying the perturbations with a uniform phase gradient, for instance, light can be anomalously diffracted with near-unity efficiency and circular polarization selectivity[33]. By nonperiodically tailoring the local perturbations across the device, the resonant response can efficiently engage arbitrarily tailored optical wavefronts of choice. As a remarkable example, a nonlocal metalens with hyperbolic phase distribution across the aperture can selectively engage only fields originating from its focal spot, while it remains transparent for other wavefront shapes[34]. However, since their response is spectro-spatial in a nontrivial way, it cannot be captured by TCMT. Rather, a new theoretical framework featuring a spatially varying modal parameter $a(x,t)$ is required [Fig. 1b]. Figure 1c depicts a conventional nonlocal metasurface whose meta-unit cells are subwavelength and transversely invariant, hence compatible with a TCMT description. Figure 1d on the contrary depicts an example diffractive nonlocal metasurface (based on ref. [33]) with a spatially varying orientation angle, incompatible with a TCMT description and requiring a generalized CMT model.

Here, we introduce a spatio-temporal coupled mode theory (STCMT) that extends TCMT to capture spatially varying resonant modes. We show excellent agreement with full-wave simulations of spatially selective diffractive nonlocal metasurfaces[34], demonstrating the validity and relevance of this model. Our theory introduces a *nonlocality length*, which quantifies the degree of nonlocality of a given device, and it establishes how this parameter directly controls the degree of spatial selectivity. We clarify two distinct regimes of operation for nonlocal photonic devices, governed by the numerical aperture of the anomalous diffraction encoded into the device, and we demonstrate the relation between spatial selectivity and the supported eigenmodes. While recent efforts on spatio-temporal coupled mode theory have focused on temporally varying systems[41] or conventional low-contrast grating systems[42], our work provides a broad and much-needed framework for the rapidly emerging field of spatially varying and finite nonlocal metasurfaces. Our STCMT elegantly captures the fundamental working mechanisms of spatially varying resonant photonic systems, provides physical insights, and enables rational designs and rapid, computationally efficient analysis of next-generation photonic systems.

Results

In the following, we develop STCMT by writing down the appropriate dynamical equations, determining the constraints stemming from time-reversal invariance, reciprocity and energy conservation, and then computing the scattering from a general nonlocal metasurface by appropriately specifying its parameters.

We first describe the nature of the problem. Coupled mode theories are useful because they eliminate the full vectoral information of Maxwell's equations in favor of a few modal parameters and scattering coefficients involving a finite number of input and output ports. In STCMT, we seek to assign the spatial and temporal properties of the scattering into a single modal parameter $a(x,t)$. For simplicity, here we develop STCMT for a single mode; our approach may be generalized for multiple modes. While STCMT applies more broadly, it is instructive to consider a concrete system. As our archetypical example: a bound state in the continuum (BIC) is a bound mode in a nonlocal metasurface with a complex modal field $E_{BIC}(x,y,z)$ that can be treated as invariant to time (other than the harmonic response $\exp(-i\omega t)$); it is an eigenstate of a closed system. This closed system is typically infinitely periodic with subwavelength periodicity—the mode profile has subwavelength spatial details ("microstructure") that repeat indefinitely without change across the aperture of the device (it has no "macrostructure"). To access it from free space, it may be perturbed into a q-BIC with a modal field $E_{qBIC}(x,y,z,t)$, which is a quasi-normal mode of an open system that leaks to the radiation continuum and therefore will generally vary with time (e.g., decay away). In a weakly perturbed system, we approximate this field as[34]

$$E_{qBIC}(x,y,z,t) \approx a(t)E_{BIC}(x,y,z) \tag{1}$$

where the spatial part of the q-BIC is treated as identical to that of the BIC, while the $a(t)$ captures the time dependence introduced by the perturbation. In TCMT, after assuming that the characteristic time scale on which $a(t)$ varies is much slower than the resonant frequency ω_{res}, we may solve the scattering problem solely in terms of this modal parameter $a(t)$, or its Fourier counterpart $a(\omega)$. In other words, we simplify the scattering problem into solely solving for the "temporal macrostructure" of the system, which is sufficient to capture the spectral features of the q-BIC (i.e., Fano resonant line shapes).

Here, the coupled mode theory we seek will additionally capture the spatial properties of a mode in a *spatially varying system* in a parallel way. In such a system, the perturbation transforming the BIC to a q-BIC varies across the device and so Eq. (1) becomes

$$E_{qBIC}(x,y,z,t) \approx a(x,t)E_{BIC}(x,y,z) \tag{2}$$

where we now include the spatial macrostructure of the mode into $a(x,t)$. In close analogy to the temporal macrostructure in TCMT, in STCMT, we assume this macrostructure varies slowly compared to the microstructure, i.e., at length scales larger than the unit cell of a nonlocal metasurface. In this way, we may retain the interesting aspects of the spatially varying mode, while still ignoring the detailed microstructure of the modes. STCMT therefore naturally extends the phenomenology of TCMT to the spatial domain, capturing the *spectrospatial* behavior of finite and spatially varying nonlocal metasurfaces.

Dynamical equations

We obtain the spatio-temporal dynamical equations by appropriately manipulating and tailoring standard TCMT equations[5]. Under an $e^{-i\omega t}$ time convention, and for a single resonance, these equations are

$$\frac{\mathrm{d}a(t)}{\mathrm{d}t} = -i\Omega a(t) + \langle \kappa^* | s_+ \rangle \tag{3}$$

$$|s_-(t)\rangle = C|s_+(t)\rangle + a(t)|d\rangle \tag{4}$$

Equation (3) describes the dynamics of the complex modal amplitude $a(t)$ excited by an incoming wave $|s_+(t)\rangle$, where $|a(t)|^2$ is normalized as the energy stored in the q-BIC per unit area at time t and $\langle s_+|s_+\rangle$ is the incident intensity. The in-coupling vector $|\kappa\rangle$ contains the coupling coefficients to

the mode from free-space excitation, and $\Omega = \omega_0 - i\gamma$ is a complex frequency, with ω_0 being the (resonant) modal frequency and γ being the total scattering rate. Equation (4) describes the dynamics of the outgoing (scattered) waves $|s_-\rangle$ (also normalized such that $\langle s_-|s_-\rangle$ is the outgoing intensity) as the interference of the "background" scattering, described by the matrix C, and the mode leaking to free space, described by the out-coupling vector $|d\rangle$ (see Supplementary Section S1 for additional background).

In contrast, the dynamical equations for a device with space-varying modal properties must inherently depend on the nature of the mode itself. In other words, we need at the outset to know the form of the dispersion of the mode. We expect, for instance, the phenomenology of a traveling wave to differ from that of a standing wave, and hence we expect distinct dynamical equations. A general approach to capture this dependence is a Taylor expansion for the resonant frequency in terms of momentum k:

$$\Omega(k) = \left(\omega_0 + ck + \frac{b}{2}k^2 + \ldots\right) - i\left(\gamma_0 + \gamma_1 k + \frac{\gamma_2}{2}k^2 \ldots\right) \tag{5}$$

where k is the transverse momentum, ω_0, c, b are the coefficients for the real part of the resonant frequency, and $\gamma_0, \gamma_1, \gamma_2$ are the coefficients for the imaginary part of the resonant frequency. For instance, a traveling wave is captured by the linear term with speed c, while a standing wave with parabolic dispersion, such as near a band edge, is captured by the coefficient b. As detailed in Supplementary Section S2, transforming Eqs. (3) and (4) into the frequency domain, including its momentum dependence and the expansion (5), and then transforming to the space-time domain, yields the dynamical equations (up to second order)

$$\frac{\mathrm{d}a(x,t)}{\mathrm{d}t} + i(\omega_0 - i\gamma_0)a(x,t) + i(c - i\gamma_1)\frac{\mathrm{d}a(x,t)}{\mathrm{d}x} + i(b - i\gamma_2)\frac{\mathrm{d}^2a(x,t)}{\mathrm{d}x^2} = \int \mathrm{d}x' \langle \kappa^*(x,x')|s_+(x')\rangle \tag{6}$$

$$|s_-(x)\rangle = \int \mathrm{d}x' (C(x,x')|s_+(x')\rangle + a(x)|d(x,x')\rangle) \tag{7}$$

Equations (6) and (7) are the dynamical equations for STCMT, including spatial dispersion in both the *coupling* properties (e.g., convolution terms on the right-hand sides) and the *modal* properties (spatial derivative terms on the left-hand sides). We note that this is can be readily extended to higher order expansions if needed, but in the most common scenarios the coefficients in Eq. (6) are more than sufficient. For instance, a pure traveling wave is characterized by a first-order spatial derivative, with Eq. (6) simplifying to a form akin to the one-way wave equation (see ref. [42]); a standing wave at a band edge, modeled parabolically, is characterized by a second-order spatial derivative, with Eq. (6) simplifying to a Schrödinger equation (as we explore below). The coefficients c, b therefore quantify the spatial dispersion of two modes of different type, each of which can be captured by STCMT. Similarly, the radiative decay profile, captured by the coefficient $\gamma_{1,2}$, also describes a nonlocal effect.

Next, we specialize our general STCMT to capture the functionalities demonstrated by recent diffractive nonlocal metasurfaces[31–37]. These devices leverage q-BIC modes, whose Q-factors are sufficiently large such that the background varies slowly compared to the resonant features[5]. In their infinitely periodic implementation, they abide a TCMT description with $c = \gamma_1 = \gamma_2 = 0$ near normal incidence: they are well captured by a parabolic band with a Q-factor and resonant polarization state that is not a function of the incident angle. Their design assumes that the scattering of a q-BIC can be spatially tailored by locally changing a small symmetry-breaking perturbation, and that distant perturbations do not affect the local polarization or phase[31]. Hence, the nonlocality of these devices is assumed to be purely *modal*, or captured by the spatial derivatives in the left-hand side, while the *coupling* to and from the mode is spatially instantaneous (local):

$$\begin{aligned} C(x,x') &\propto \delta(x-x') \\ |\kappa(x,x')\rangle &\propto \delta(x-x') \\ |d(x,x')\rangle &\propto \delta(x-x') \end{aligned} \tag{8}$$

Together, these assumptions yield the dynamical equations

$$\frac{\mathrm{d}a(x,t)}{\mathrm{d}t} + i(\omega_0 - i\gamma)a(x,t) + ib\frac{\mathrm{d}^2a(x,t)}{\mathrm{d}x^2} = \langle \kappa^*(x)|s_+(x,t)\rangle \tag{9}$$

$$|s_-(x,t)\rangle = C(x)|s_+(x,t)\rangle + a(x,t)|d(x)\rangle \tag{10}$$

where we simplify the notation to $\gamma = \gamma_0$. Supplementary Section S2b derives the same set of equations using the leading order terms of a cosine series expansion.

Scattering Kernel and Green's function

In TCMT, the scattering matrix is obtained by assuming time-harmonic solutions, solving Eq. (3) for $a(t)$, and then inserting the result into Eq. (4) to yield[5]

$$|s_-(\omega)\rangle = S(\omega)|s_+(\omega)\rangle \tag{11}$$

$$S(\omega) = C(\omega) + \frac{|d\rangle\langle\kappa^*|}{i(\Omega - \omega)} \tag{12}$$

In comparison, in STCMT we must solve Eq. (9) before inserting it into Eq. (10). To do so, we recognize that, when no source is present, we obtain a Schrödinger

equation with complex potential $V = \omega_0 - i\gamma$, recognizable by the rearrangement:

$$i\frac{\mathrm{d}a(x,t)}{\mathrm{d}t} = -\frac{b}{2}\frac{\mathrm{d}^2 a(x,t)}{\mathrm{d}x^2} + Va(x,t) \tag{13}$$

Compared to the conventional Schrödinger equation, we have $= 1$ and $b = 1/m$, consistent with the expectation that the band curvature plays the role of an inverse effective mass. We also note that non-Hermiticity (a complex valued potential) is expected of an open system. This well-studied equation abides solutions via the Propagator, or Green's function (Supplementary Section S3).

In practice, we are most often interested in the response to monochromatic waves. In this case, Eq. (9) becomes

$$i(\omega_0 - \omega - i\gamma)a(x,\omega) + ib\frac{\mathrm{d}^2 a(x,\omega)}{\mathrm{d}x^2} = \langle \kappa^*(x)|s_+(x,\omega)\rangle \tag{14}$$

and we seek Green's function satisfying

$$i(\omega_0 - \omega - i\gamma)G(x,x',\omega) + ib\frac{\mathrm{d}^2 G(x,x',\omega)}{\mathrm{d}x^2} = \delta(x - x') \tag{15}$$

which yields the modal amplitude

$$a(x,\omega) = \int \mathrm{d}x' G(x,x',\omega)\langle \kappa^*(x')|s_+(x',\omega)\rangle \tag{16}$$

Insertion of Eq. (16) into the space-frequency form of Eq. (10) yields the scattering equation

$$|s_-(x,\omega)\rangle = \int \mathrm{d}x' \sigma(x,x',\omega)|s_+(x',\omega)\rangle \tag{17}$$

$$\sigma(x,x',\omega) = C(x,\omega)\delta(x - x') + G(x,x',\omega)|d(x)\rangle\langle \kappa^*(x')| \tag{18}$$

where the scattering kernel σ correlates input positions x' with output positions x, and where the closed form of Green's function is

$$G(x,x',\omega) = \frac{i}{b}\xi(\omega)\exp\left(-\frac{|x - x'|}{\xi(\omega)}\right) \tag{19}$$

where

$$\xi(\omega) = \sqrt{\frac{ib/2}{\gamma + i(\omega_0 - \omega)}} \tag{20}$$

The real part of ξ_0 takes an especially simple form at the band edge

$$\xi_0 = \mathrm{Re}[\xi(\omega_0)] = \sqrt{\mathrm{b}/\gamma} = \sqrt{\mathrm{b}\tau_\mathrm{r}} \tag{21}$$

which we call the *nonlocality length*, the characteristic distance that light travels in-plane before scattering out when engaging with the band-edge mode (a quantitative measure of the nonlocality of the metasurface). Since nonlocality is equivalently defined as a dependence on the incident wavevector, no optical interface is truly local. However, as a rule of thumb, if the nonlocality length is larger than the wavelength in the surrounding media, the resulting device may be considered nonlocal, since its nonlocality cannot be neglected (see Supplementary Section S4 for a comparison of a nonlocal metasurface to a bare interface between two media).

Physical constraints

Before proceeding with the application of STCMT to practical metasurface devices, we must determine the relevant physical constraints to our system, including time-reversal invariance, reciprocity and conservation of energy. In TCMT, these constraints impose the well-known requirements[5]

$$\langle d|d\rangle = 2\gamma \tag{22}$$

$$|\kappa\rangle = |d\rangle \tag{23}$$

$$C|d^*\rangle = -|d\rangle \tag{24}$$

These results hold, for instance, in infinitely periodic devices supporting q-BICs (see Supplementary Section S1). As detailed in Supplementary Section S5, we find that in our STCMT all these conditions hold *locally*, i.e.,

$$\langle d(x)|d(x)\rangle = 2\gamma \tag{25}$$

$$|\kappa(x)\rangle = |d(x)\rangle \tag{26}$$

$$C(x)|d^*(x)\rangle = -|d(x)\rangle \tag{27}$$

We stress that this result is not obvious a priori and, indeed, these do not hold for the general case described by Eqs. (6) and (7). Future work may explore systems with nonlocal coupling (i.e., devices for which Eq. (8) do not hold), in which case Eqs. (26) and (27) are no longer valid.

The significance of this result for modeling diffractive nonlocal metasurfaces should not go underappreciated, as it confers a major simplification to model this class of systems and enables the introduction of a semi-analytical model to describe them. Because the *coupling* is local, each unit cell of a nonlocal metasurface may be modeled as if it were in an infinite array, allowing easy correlation between geometric and phenomenological degrees of freedom (compatible with TCMT in the momentum-frequency domain). Then, a given space-varying nonlocal

metasurface is treated by specifying a spatial profile of the coupling coefficients, wherein the correlations across the surface are captured by a shift-invariant Green's function (*modal* nonlocality). This design flow is compatible with the foundational design principles of *local* metasurfaces[43], wherein a library of pre-computed optical scatterers may be arrayed across the surface as if they operate independently of their nearest neighbors. This ubiquitous design approach is sometimes called the "local phase approximation". Yet, we find that, when Eq. (8) holds, the same design principles are applicable even in the much more complex scenario discussed in this work, considering *nonlocal* responses due to a guided mode.

Eigenwaves

Of central interest to nonlocal metasurfaces is the principal eigenwave, which is the wavefront that a nonlocal metasurface selects for: maximal reflectance is achieved when the eigenwave is incident, and the reflected wave is the eigenwave's phase conjugated copy [55]. Using $\rho = \sigma_{11}$ as the reflection kernel for light incident from side 1, the eigenwave from side 1 is solution to the relation

$$E^*_{eig}(x,\omega) = \int \mathrm{d}x' \rho(x,x',\omega) E_{eig}(x',\omega) \tag{28}$$

or equivalently

$$E_{eig}(x,\omega) = \int \mathrm{d}x' \rho^*(x,x',\omega) E^*_{eig}(x',\omega) \tag{29}$$

Combing Eqs. (28) and (29) we find

$$E_{eig}(x,\omega) = \int \mathrm{d}x' \int \mathrm{d}x'' \rho^*(x,x',\omega)\rho(x',x'',\omega) E_{eig}(x'',\omega) \tag{30}$$

suggesting that the eigenwave of interest is the eigenvector of

$$R_{eig} E_{eig}(x,\omega) = \int \mathrm{d}x' \int \mathrm{d}x'' \rho^*(x,x',\omega)\rho(x',x'',\omega) E_{eig}(x'',\omega) \tag{31}$$

that has the largest eigenvalue R_{eig}. We will show that STCMT can be used to explicitly calculate the eigenwaves of a given nonlocal metasurface. Equation (31) also implies the existence of higher-order eigenwaves that reflect to their time-reversed copy but with non-maximal reflectance and eigenwaves for frequencies other than the band-edge frequency ω_0.

Diffractive nonlocal metasurfaces

In this section, we apply STCMT to illustrative examples, confirming that our analytical framework reproduces the results of full-wave simulations, quantitatively captures the underlying physics and enables rational designs for relevant functionalities. Schematically depicted in Fig. 2a, we explore systems based on refs. [33] and[34], composed of four elliptical inclusions controlling the scattering (polarization angles ϕ and θ) at the top and bottom interfaces simultaneously and independently [Fig. 2b]. Figure 2c, d shows the calculated reflectance and reflected phase for an example meta-unit based on this system, computed using the finite difference time domain (FDTD) method. Figure 2e, f depict the TCMT modeling of the infinitely periodic meta-unit (developed in Supplementary Section S1b), with (in units such that the speed of light is $c_0 = 1$) the band edge frequency $\omega_0 = 2\pi/\lambda_{be}$ with $\lambda_{be} = 1.6104\,\mu m$, the band curvature $b = 0.021\,\mu m$, and the lifetime $\tau_r = 250\mu m$. In particular, we use the chiral q-BIC system in refs. [33] and[34], in which the response was shown to: (i) have near-zero background reflection, (ii) exclusively engage circularly polarized light of one handedness, and (iii) decay with a geometric phase of 2α for a meta-unit characterized by in-plane orientation angle α. In a scalar basis of the selected spin we have

$$|d(x)\rangle = i\sqrt{\gamma}\begin{bmatrix} \exp[i2\alpha(x)] \\ -i\exp[-i2\alpha(x)] \end{bmatrix} \tag{32}$$

where the first element is the coefficient scattering to side 1 and the second element is the coefficient scattering to side 2. For a system based on the meta-units in Fig. 2 (and refs. [33,34]), this yields [from Eqs. (18) and (19)]

$$\sigma(x,x',\omega) = \delta(x-x')\begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix} + i\tfrac{\gamma}{b}\xi(\omega)\exp\left(-\tfrac{|x-x'|}{\xi(\omega)}\right)\begin{bmatrix} e^{i2\alpha(x)}e^{i2\alpha(x')} & -ie^{i2\alpha(x)}e^{-i2\alpha(x')} \\ -ie^{-i2\alpha(x)}e^{i2\alpha(x')} & -e^{-i2\alpha(x)}e^{-i2\alpha(x')} \end{bmatrix} \tag{33}$$

For clarity, we detail the behavior for light incident from side 1, separating the σ into a complex reflection kernel $\rho = \sigma_{11}$ and a complex transmission kernel $\tau = \sigma_{21}$:

$$\begin{aligned} \rho(x,x',\omega) &= i\tfrac{\gamma}{b}\xi(\omega)\exp\left(-\tfrac{|x-x'|}{\xi(\omega)}\right)e^{i2\alpha(x)}e^{i2\alpha(x')} \\ \tau(x,x',\omega) &= -i\delta(x-x') + \tfrac{\gamma}{b}\xi(\omega)\exp\left(-\tfrac{|x-x'|}{\xi(\omega)}\right)e^{-i2\alpha(x)}e^{i2\alpha(x')} \end{aligned} \tag{34}$$

The other kernels σ_{12} and σ_{22} are treated equivalently. Figure 2g, h shows the magnitude and phase of σ_{11} for the meta-unit corresponding to Fig. 2e, f. Note that in this case the phase angle $\alpha(x) = \alpha$ is constant, meaning that this metasurface can be handled with conventional

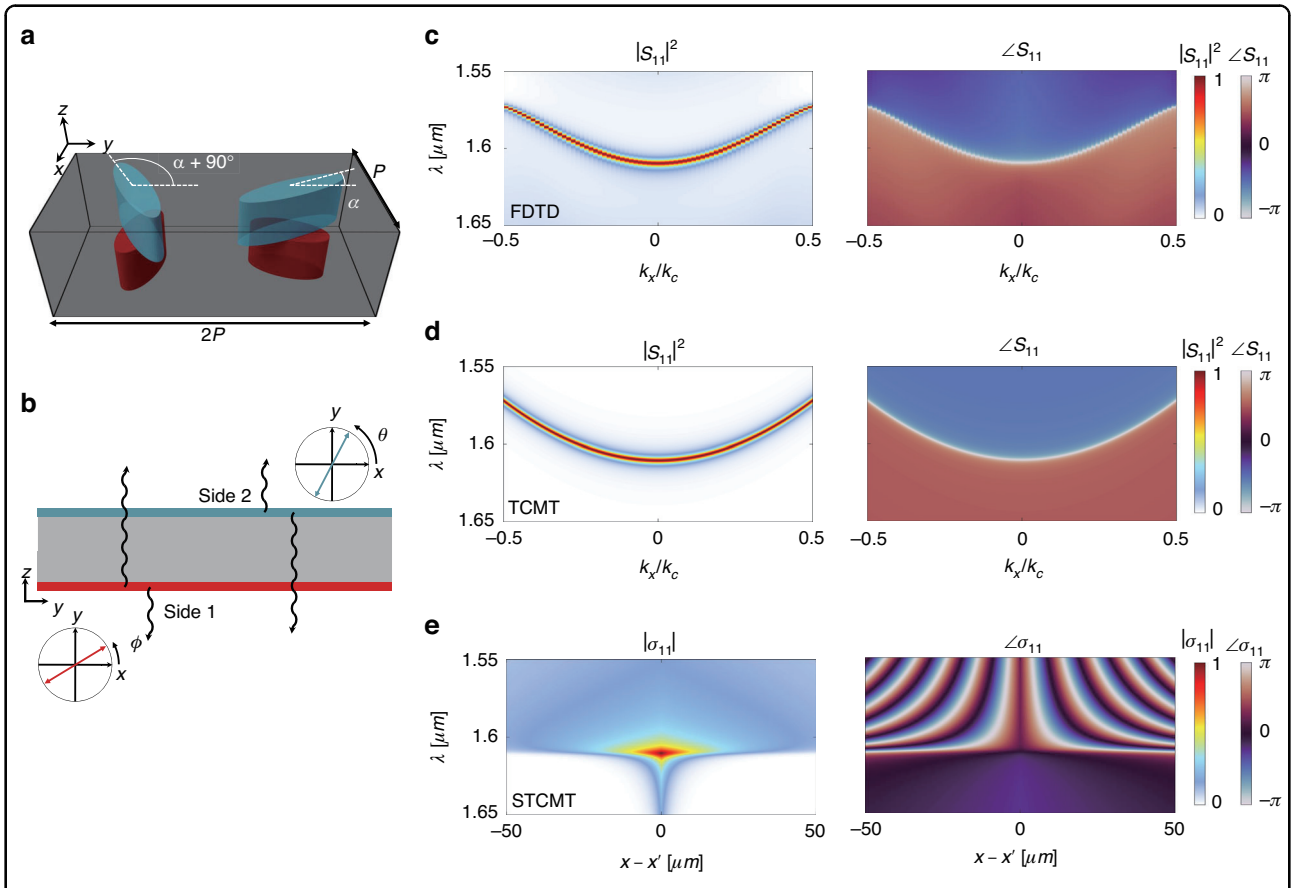

Fig. 2 Modeling periodic nonlocal metasurfaces with FDTD, TCMT, and STCMT. a Prototypical unit cell for a diffractive nonlocal metasurface, containing four elliptical inclusions. The orientation angle α is the key parameter controlling the geometric phase (the orientation angles of the bottom (red) ellipses are not independent of α). The lattice parameter a satisfies $2P<\lambda_0$, eliminating all but 1 diffraction order on each side. **b** Schematic of the coupled mode theory approximation of the scattering in (a), wherein independent scattering events occur at the interfaces, each of which may generally couple to both side 1 (down) and side 2 (up). **c** Reflectance and reflection phase in the momentum-frequency response of an example meta-unit, computed by Finite Difference Time Domain (FDTD). **d** Reflectance and reflection phase in the momentum-frequency response of an example meta-unit, modeled using TCMT. **e** Amplitude and phase of σ_{11} in the space-frequency domain of the same meta-unit, a key object for modeling this system using STCMT

TCMT. However, in the following examples, the angle is spatially varied and TCMT is not suitable.

Linear phase gradient: analytical model

An archetypal metasurface functionality is anomalously deflection using a linear phase gradient. Here, we consider a gradient in the geometric angle following

$$2\alpha(x) = -k_G x \tag{35}$$

In this case, the nonlocal kernels take the form

$$\begin{aligned}\rho(x,x',\omega) &= i\tfrac{\gamma}{b}\xi(\omega)\exp\left(-\tfrac{|x-x'|}{\xi(\omega)}\right)e^{-ik_G x}e^{-ik_G x'}\\ \tau(x,x',\omega) &= -i\delta(x-x') + \tfrac{\gamma}{b}\xi(\omega)\exp\left(-\tfrac{|x-x'|}{\xi(\omega)}\right)e^{ik_G x}e^{-ik_G x}\end{aligned} \tag{36}$$

When $k_G = 0$, we recover the nonlocal reflection kernel without phase gradient, which we denote with the subscript 0:

$$\begin{aligned}\rho_0(x,x',\omega) &= i\tfrac{\gamma}{b}\xi(\omega)\exp\left(-\tfrac{|x-x'|}{\xi(\omega)}\right)\\ \tau_0(x,x',\omega) &= -i\delta(x-x') + \tfrac{\gamma}{b}\xi(\omega)\exp\left(-\tfrac{|x-x'|}{\xi(\omega)}\right)\end{aligned} \tag{37}$$

Since a phase gradient provides a spatially constant momentum kick, it is natural to study these devices in momentum space. In particular, we seek the scattering matrix

$$S(k,k',\omega) = \begin{bmatrix} S_{11}(k,k',\omega) & S_{12}(k,k',\omega)\\ S_{21}(k,k',\omega) & S_{22}(k,k',\omega)\end{bmatrix} \tag{38}$$

where k' and k are the wavevectors of the incoming and outgoing plane waves, respectively; the indices refer to the top and bottom sides of the metasurface. We then relate this scattering matrix to σ through the mixed Fourier

transform (see Supplementary Section S6)

$$S_{ij}(k,k',\omega) = \int \mathrm{d}x \int \mathrm{d}x' \sigma_{ij}(x,x',\omega)\exp(ik'x')\exp(-ikx) \tag{39}$$

Using this relation, the scattering components for a metasurface without phase gradient are

$$\begin{aligned} S^0_{11}(k,k',\omega) &= \frac{-\delta(k-k')}{1-i\left(\omega-\omega_0-\frac{b}{2}k^2\right)\tau_r} \\ S^0_{21}(k,k',\omega) &= \delta(k-k')\left[-i+\frac{i}{1-i(\omega-\omega_0-\frac{b}{2}k^2)\tau_r}\right] \\ S^0_{22}(k,k',\omega) &= S^0_{11}(k,k',\omega) \\ S^0_{12}(k,k',\omega) &= S^0_{21}(k,k',\omega) \end{aligned} \tag{40}$$

Notably, this result is identical to conventional TCMT but with the explicit addition of a Dirac δ term enforcing scattering only when $k = k'$ (conservation of momentum in a specular process).

Meanwhile, the scenario with nonzero k_G in terms of Eq. (37) yields

$$\begin{aligned} \rho(x,x',\omega) &= \rho_0(x-x',\omega)e^{-ik_Gx}e^{-ik_Gx'} \\ \tau(x,x',\omega) &= \tau_0(x-x',\omega)e^{ik_Gx}e^{-ik_Gx'} \end{aligned} \tag{41}$$

which implies, due to the basic properties of the Fourier transform,

$$\begin{aligned} S_{11}(k,k',\omega) &= S^0_{11}(k+k_G,k'-k_G,\omega) \\ S_{21}(k,k',\omega) &= S^0_{21}(k-k_G,k'-k_G,\omega) \\ S_{22}(k,k',\omega) &= S^0_{11}(k-k_G,k'+k_G,\omega) \\ S_{12}(k,k',\omega) &= S^0_{12}(k+k_G,k'+k_G,\omega) \end{aligned} \tag{42}$$

Hence, our first conclusion from STCMT is that the impact of a nonlocal phase gradient is to shift the scattering matrices in k-space, in both the *input* and *output* spaces. In particular, in the reflection elements $[S_{11}(k,k',\omega)$ and $S_{22}(k,k',\omega)]$ the shift in input and output momentum is equal and opposite, producing anomalous reflection. Since we have nonzero scattering only when the Dirac delta function's argument is 0, from side 1 we see that reflection occurs only when $k = k' + 2k_G$, and from side 2 only when $k = k' - 2k_G$. Hence, our results capture the fact that the anomalous reflection is associated with the $m = \pm 2$ diffraction orders[31], and the direction of momentum kick imparted by the nonlocal metasurface depends the direction of incidence[28].

Linear phase gradient: numerical simulations

We now validate our analytical model numerically for $k_G = 2\pi/W$ and $W = 6.4\,\mu m$, focusing here on the band-edge frequency (Supplementary Section S7 studies frequencies off the band edge). It is useful here to cast STCMT in matrix form, detailed in Supplementary Section S8, which makes the scattering kernels discrete. We write them as P and T (capital ρ and τ) and the input electric field as a column vector $\mathbf{E}_{in}$, where each entry is the field at a discretized position x, and we compute the reflected $\mathbf{E}_r$ and transmitted $\mathbf{E}_t$ fields with

$$\begin{aligned} \mathbf{E}_r &= \mathrm{P}\mathbf{E}_{in} \\ \mathbf{E}_t &= \mathrm{T}\mathbf{E}_{in} \end{aligned} \tag{43}$$

The main diagonal of P and T correspond to $x = x'$, i.e., the local response, while the off-diagonal components contain the correlations across distant positions, i.e., the nonlocal response. Finally, given the finite nature, boundary conditions must be specified; Supplementary Section S9 details both periodic and radiative boundary conditions. Supplementary Section S10 describes the extension of this procedure to 2D metasurfaces.

Using this discretized form of STCMT, Fig. 3 compares the scenario without a phase gradient [Fig. 3a–d] to the case with a phase gradient [Fig. 3e–h]. The former is identical to the structures described in Fig. 2., while the latter differs only in that the orientation angles of the ellipses varies spatially. Figure 3a schematically shows resonant reflection at the band-edge frequency, $\omega = \omega_0$, of the typical parabolic band diagrams in Fig. 3b, depicting specular reflection at normal incidence when incident from either side (S_{11} and S_{22}). Figure 3c shows the corresponding amplitude and phase of P and T, which, using Eq. (39), yield the scattering matrix elements in Fig. 3d. In this case, specular transmission arises except at normal incidence from either side, wherein specular reflection occurs (tracked by the red circle and blue triangle from sides 1 and 2, respectively).

In contrast, in the presence of a phase gradient the device retroreflects light at the resonance frequency, as shown in Fig. 3c. From side 1, the band edge mode is shifted in k-space by k_G, meaning that the resonance only occurs for $k' = k_G$:

$$\omega(k') = \omega_0 + \frac{b}{2}(k'-k_G)^2 \tag{44}$$

The shift is in the opposite direction for light from side 2 [Fig. 3f]. This shift, as well as the anomalous reflection, is encoded in P and T within the phase information [Fig. 3g], and the corresponding scattering matrix at the band-edge frequency [Fig. 3h] confirms that the resonant mode occurs at $k' = k_G$ and $k = -k_G$ when light is incident from side 1 (red circle), while it occurs at $k' = -k_G$ and $k = k_G$ when light is incident from side 2 (blue triangle). Note that this behavior is consistent with the requirements stemming from reciprocity: the scattering matrices must be symmetric about the main diagonal, which

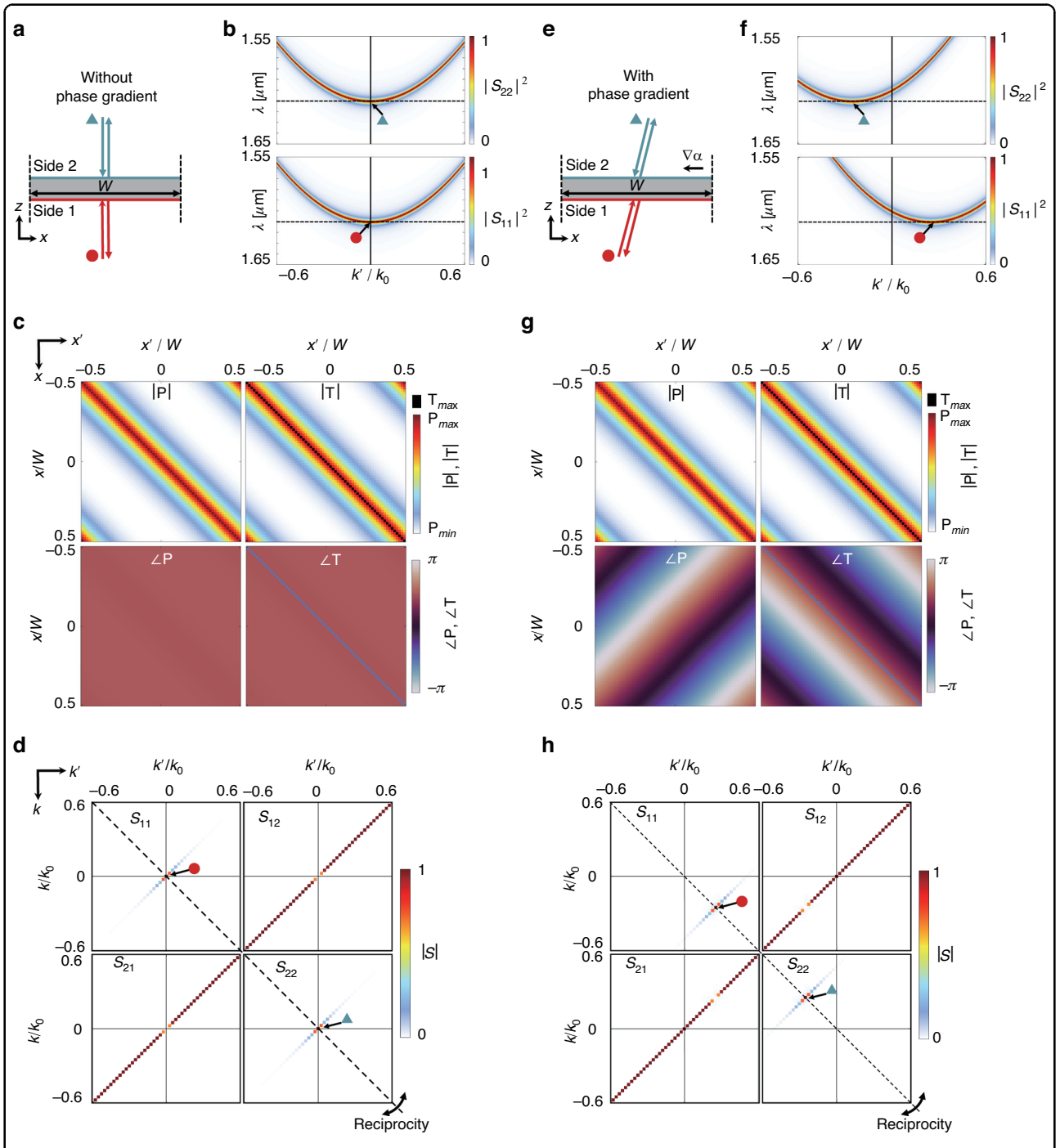

Fig. 3 Analysis of phase gradient devices. a Schematic of the resonant response of a nonlocal phase gradient device with $k_G = 0$ at the band-edge frequency. **b** Reflectance maps of S_{11} and S_{22} as a function of incident momentum k' for the device in (**a**), where the dashed line indicates the band-edge frequency. **c** Nonlocal kernel matrices for the device in (**a**) at the band-edge frequency (note that T is identical to P except along the main diagonal, where its magnitude is much higher). **d** Scattering matrix for the device in (**a**), showing unity specular reflectance only at normal incidence. **e**–**h** Same as in (**a**–**d**) but for a device with $k_G = 2\pi/W$ operated at the band-edge frequency. Note that for sufficient resolution, (**d**, **h**) are computed using the mixed Fast Fourier Transform of a version of (**c**, **g**) including 10 metasurface periods. The red circle and blue triangle track the band-edge response from below and above respectively, in each case

corresponds to $k = -k'$, or retroreflection (see also Supplementary Section S5). Hence, we see that the action of the phase gradient is to shift the band-edge response along the retroreflection condition by an amount $(k_G, -k_G)$ in S_{11}. The corresponding shift in S_{22} is $(-k_G, k_G)$, consistent with conservation of energy [the dip

and peak occurring at the same k' in S_{12} and S_{22}, and likewise for S_{11} and S_{21}]. That is, the behavior of a nonlocal phase gradient is to *tilt* the response in k-space [Fig. 3e] rather than add a unidirectional momentum kick irrespective of incident direction: it is a vertically asymmetric phenomenon, therefore requiring vertically asymmetry in the structure to achieve, irrespective of polarization[28,44]. The scattered fields corresponding to the peaks in (h) are shown in Fig. S4 of the supplementary materials.

Nonlocal metalens

Next, we treat nonlocal metalenses—an archetypal example of spatially varying nonlocal metasurfaces of broad relevance for applications [Fig. 4a]. We first analytically study the eigenwaves of nonlocal metalenses, and then numerically demonstrate agreement with the spatial selectivity demonstrated in ref. [34]. Our validation with full-wave simulations then motivates us to employ STCMT to rapidly and thoroughly study a variety of nonlocal metalenses, shedding light on their operation. In particular, we demonstrate that, as a function of the band structure b, the Q-factor $Q = \omega_0 \tau_r$, and the numerical aperture of the metalens NA, a nonlocal metalens can transition between two regimes of operation: wavefront-selective and wavefront-shaping. Supplementary Section S11 extends the following to frequencies off the band edge.

In particular, we study nonlocal metalenses described by

$$2\alpha(x) = -k_0\sqrt{x^2 + f^2} \tag{45}$$

Using relation (31), the eigenwave of the metalens is given by

$$R_{eig}E_{eig}(x,\omega) = \int_{-W/2}^{W/2} \mathrm{d}x' \int_{-W/2}^{W/2} \mathrm{d}x'' \rho_0^*(x - x', \omega) \rho_0(x' - x'', \omega) e^{ik_0\sqrt{x^2+f^2}} e^{-ik_0\sqrt{(x'')^2+f^2}} E_{eig}(x'', \omega) \tag{46}$$

By inspection, as $W \to \infty$ (i.e., ignoring boundary effects), we have

$$E_{eig}(x,\omega) = e^{ik_0\sqrt{x^2+f^2}} \tag{47}$$

because in this case

$$R_{eig} = \int_{-W/2}^{W/2} \mathrm{d}x' \int_{-W/2}^{W/2} \mathrm{d}x'' \rho_0^*(x - x', \omega) \rho_0(x' - x'', \omega) = 1 \tag{48}$$

which is identical to the spatially invariant q-BIC modeled by the subscripted kernels. For finite W and radiative boundaries, Eq. (48) does not yield a perfectly unity result —to be expected for a finite device (see Supplementary Section S9). Nevertheless, the idealized solution[45] captures the underlying phenomenon: this nonlocal metalens is selective to a point source set at $z_0 = -f$, just as described in ref. [34].

However, unlike ref. [34], here we may specify the non-ideal eigenwaves for finite structures, including radiative boundary conditions, e.g., using the discrete matrix form of our STCMT. The nonlocal reflection matrix for this scenario is seen in Fig. 4b, c, from which the eigenwave E_{eig} is computed, depicted in Fig. 4d overlaid with the reflected wave ($\mathrm{P}E_{eig}$), which is the phase conjugated copy of the input as expected. We also may numerically compute the reflectance due to point sources incident from side 1 at position (x_0, z_0) [Fig. 4a]:

$$E_{in}(x,\omega) = \exp\left(ik_0\sqrt{(x - x_0)^2 + z_0^2}\right) \tag{49}$$

Using the metasurface introduced in ref. [34], we numerically compute the reflectance Fig. 4e as a function of wavelength and z_0 (while keeping $x_0 = 0$) and find remarkable agreement with the results from full-wave simulation in Fig. 4f. The peak reflectance in the STCMT is near-unity, while full-wave simulations predict peak reflectance near 70%, suggesting that the metalens has room for optimization in its design. Regardless, our STCMT can clearly capture the response of the nonlocal metalens.

Next, we explore the response of the system in Fig. 2 as a function of b, Q, and numerical aperture, $\mathrm{NA} = n\frac{W/2}{\sqrt{(W/2)^2+(f/2)^2}}$, where n is the index of refraction of the surrounding medium. The NA characterizes the range of deflection angles encoded into the q-BIC, and hence it also characterizes the range of resonant wavelengths across the device for normally incident light. As the NA increases, the part of the nonlocal metalens resonating at a given frequency is reduced, and the reflectance also diminishes—i.e., spatial selectivity increases with NA. This feature is confirmed in Fig. 5a–d, which show the spectral reflectance as a function of z_0 for four values of NA. When $\mathrm{NA} = 0.8$, representing a very large range of deflection angles, only point sources placed near $z_0 = -f$ reflect substantial power at the band-edge frequency. That is, point sources near the eigenwave experience large reflectivity, while point sources away from the eigenwave are non-resonantly transmitted. When the NA lowers, we observe an increase in reflectance away from the eigenwave condition, e.g., for large $|z_0|$, which approaches a plane wave in the limit that $z_0 \to -\infty$. So while when $\mathrm{NA} = 0.8$ we have a wavefront-selective device, in contrast, when $\mathrm{NA} = 0.1$, we have a wavefront-shaping

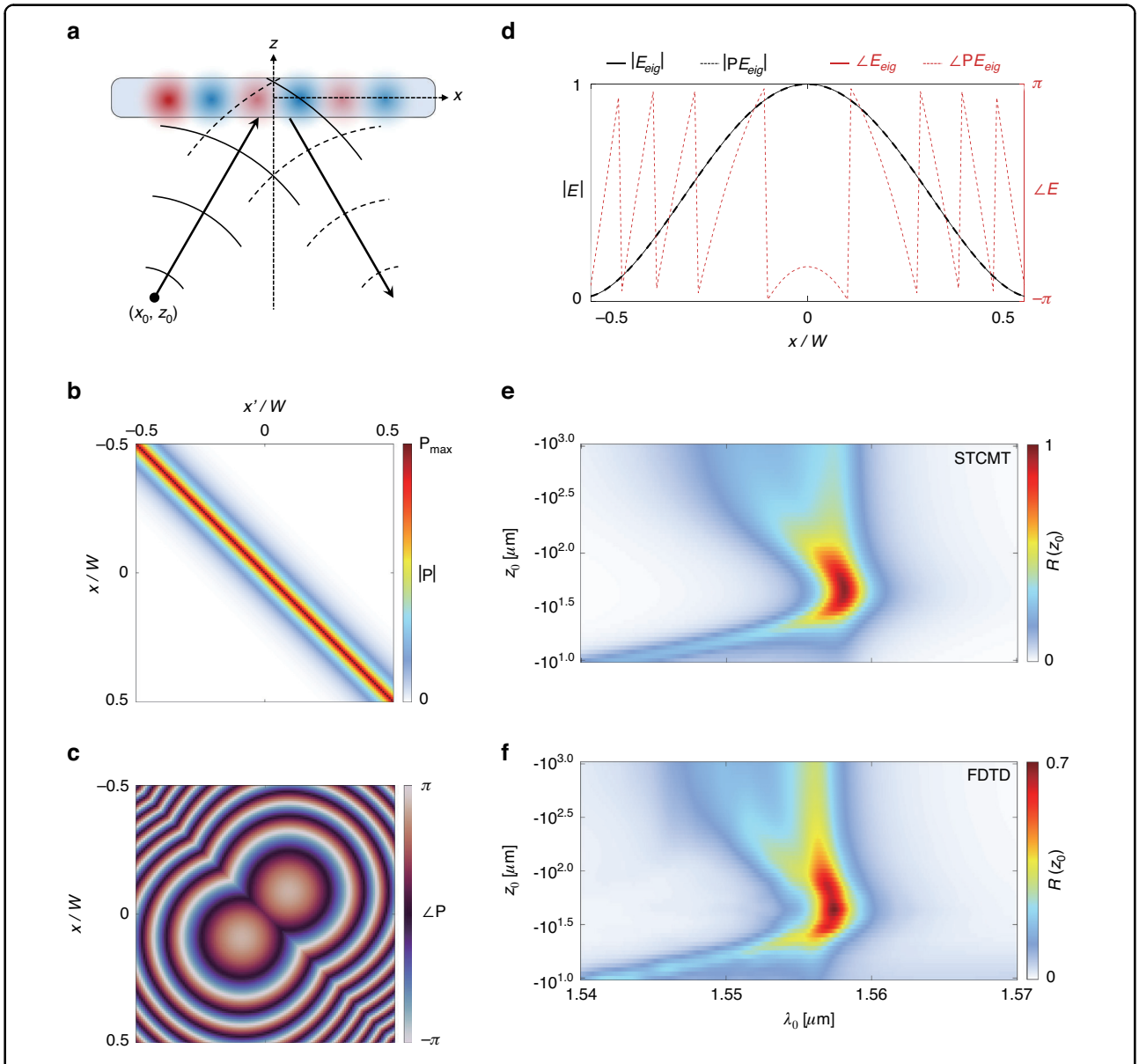

Fig. 4 Nonlocal metalens. a Nonlocal metalens selective to a point source placed at (x_0, z_0). Magnitude (**b**) and phase (**c**) of the nonlocal reflection kernel P for a metalens of width $W = 50\,\mu m$ with q-BIC parameters $b = 0.032\,\mu m, \lambda_0 = 1.558\,\mu m$, and $\tau_r = 297\,\mu m$, and NA = 0.48, matched to the system in ref. [34]. **d** Eigenwave and reflected wave, showing unchanged amplitude and conjugated phase profiles upon reflection. The reflectance due to excitation by point sources placed at locations z_0 on the optical axis $(x, y) = (0, 0)$ for (**e**) STCMT model and (**f**) full-wave simulations based on the system reported in ref. [34]

device, abiding large reflectance under a larger range of incident conditions.

Importantly, the transition from wavefront-selective to wavefront-shaping is a function of the degree of locality. When ξ_0 is substantially larger than λ_0, the device is highly selective, i.e., it only acts as a wavefront-shaping device for very low NA. But when ξ_0 is comparable to λ_0, the device is not selective, i.e., it acts as a wavefront-shaping device for a large range of NA. We find (Supplementary Section S12) that the wavefront-shaping regime is satisfied when

$$\mathrm{NA} < \sqrt{\frac{2/k_0}{|b|Q}} = \frac{1}{\sqrt{2}\pi}\frac{\lambda_0}{\xi_0} \tag{50}$$

We therefore may quantitatively conclude that the degree of spatial selectivity is due to the degree of non-locality of the q-BIC. Supplementary Section S13 further quantifies this feature by discussing the expansion in the

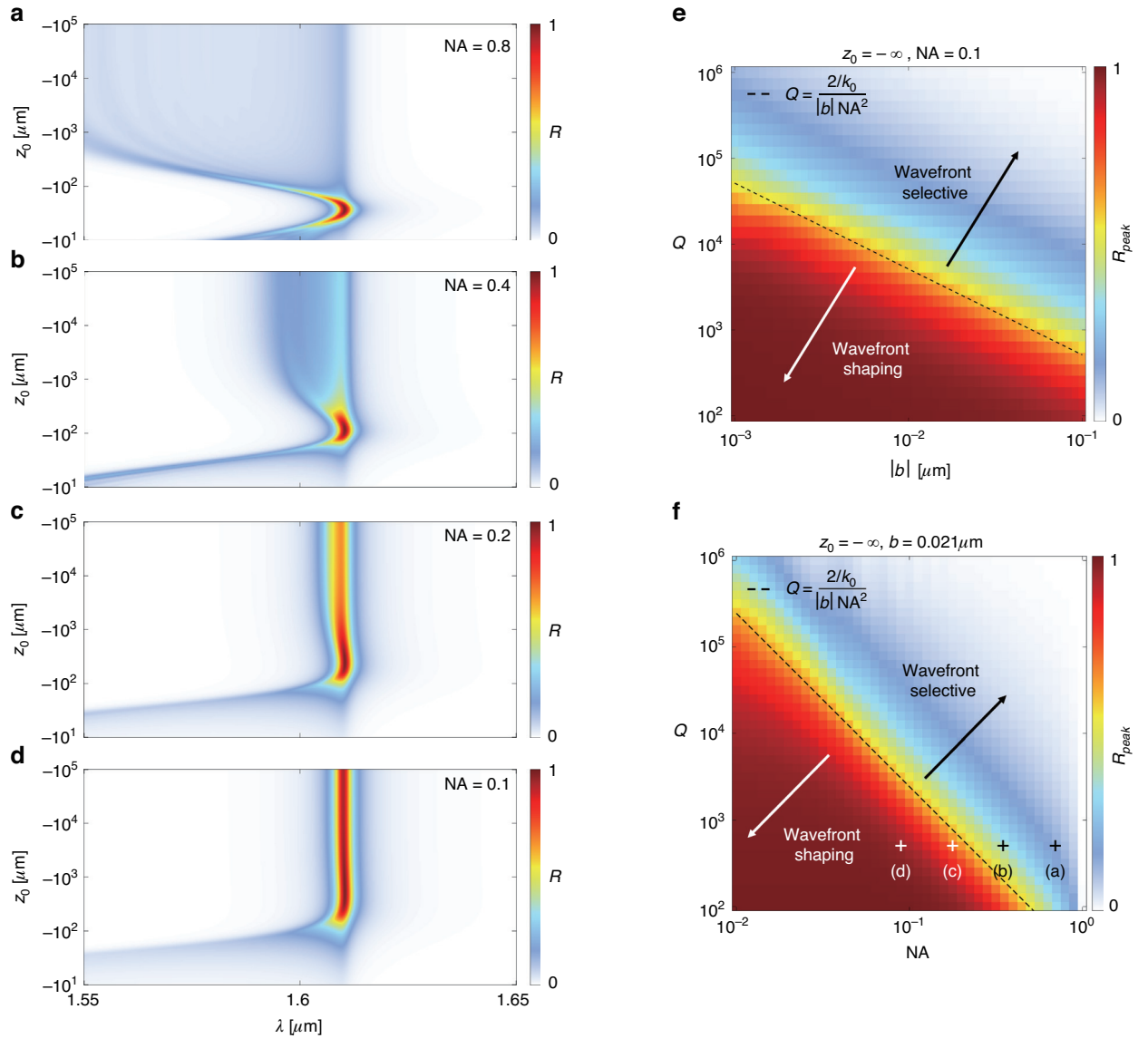

Fig. 5 Selectivity of a nonlocal metalens. a–d Reflectance for a device with $W = 75\,\mu m$ and varying NA, with q-BIC parameters $b = 0.021\,\mu m$, $\lambda_0 = 1.6104\,\mu m$, and $\tau_r = 250\,\mu m$, due to excitation by point sources placed at locations z_0 on the optical axis $(x, y) = (0, 0)$. **e** Peak reflectance due to a normally incident plane wave for a device with NA = 0.1 as a function of b and Q. **f** Peak reflectance due to a normally incident plane wave for a device with b = 0.021 μm as a function of NA and Q

eigenbasis of the device, an analysis related to modal expansions in optics[46].

We next explore devices with varying b, Q, and NA to demonstrate the transition between wavefront-selective and wavefront-shaping regimes. Figure 5e shows the peak reflectance for a normally incident plane wave $z_0 = -\infty$ for a device with NA = 0.1 as a function of b and Q: when they are both large (a highly nonlocal device), the plane wave is transmitted unperturbed through the device, indicating wavefront-selectivity. When they are both small (a highly local device), the plane wave is anomalously reflected with near-unity efficiency, and the device is wavefront-shaping. The transition Q-factor for a given value of b is overlaid on Fig. 5e by rearranging Eq. (50), showing good agreement between the numerical calculations and the physical arguments outlined here. Additionally, Fig. 5f shows a device with $b = 0.021\,\mu m$ as a function of NA and Q, showing similar agreement in the transition between the two regimes.

Thermal metasurfaces

To further showcase the applicability of STCMT, we now model thermal metasurfaces[35,44], which provide a powerful platform to shape emission from metasurfaces.

STCMT enables rapid study and prototyping of focused incoherent light emission with respect to ξ_0 and the lens design parameters. To tackle this problem, we extend the STCMT formulation by (i) accounting for polarization ports in a reflection- and absorption-only scattering problem, (ii) incorporating a spatially varying local metasurface response in the background $C(x)$, and (iii) adding absorption loss, distinguishing the radiative scattering rate γ_r and the nonradiative scattering (loss) rate γ_{nr}. Here, the scattering matrix S_{ij} is indexed by linear polarizations $i, j \in x, y$, and the background scattering matrix has full birefringence with a fast axis oriented at $\theta(x)$:

$$C(x) = \begin{bmatrix} \cos[2\theta(x)] & \sin[2\theta(x)] \\ \sin[2\theta(x)] & -\cos[2\theta(x)] \end{bmatrix} \tag{51}$$

Meanwhile, for perturbations governing the q-BIC defined by an orientation angle α, the coupling coefficients are[44]

$$|d(x)\rangle = \sqrt{\gamma_r}\left\{ \begin{bmatrix} \cos[2\alpha(x)] \\ \sin[2\alpha(x)] \end{bmatrix} - i \begin{bmatrix} \cos[2\theta(x) - 2\alpha(x)] \\ \sin[2\theta(x) - 2\alpha(x)] \end{bmatrix} \right\} \tag{52}$$

Finally, to account for the presence of loss, we make the simple adjustment

$$\xi(\omega) = \sqrt{\frac{ib/2}{\gamma_r + \gamma_{nr} + i(\omega_0 - \omega)}} \tag{53}$$

For clarity, we may also separate the scattering kernel into its local and nonlocal components

$$\sigma(x, x', \omega) = \rho_{loc}(x) + \rho_{nonloc}(x, x', \omega) \tag{54}$$

where

$$\rho_{loc}(x) = C(x)\delta(x - x') \tag{55}$$

$$\rho_{nonloc}(x, x', \omega) = G(x, x', \omega)|d(x)\rangle\langle d^*(x')| \tag{56}$$

Then, assuming that the power not reflected is absorbed by the device, the absorption of a given input wave can be simply computed as

$$\mathrm{A}(\omega) = 1 - \langle s_+|\sigma^*(x, x', \omega)\sigma(x, x', \omega)|s_+\rangle \tag{57}$$

while the corresponding thermal emission is described by the time-reversed wave, following the universal modal radiation laws[47].

As an archetypal example, Fig. 6 studies a thermal metalens that preferentially emits to a focal point f with a circular polarization, encoded using the profiles[35]

$$\theta(x) = \pm \frac{2\pi}{\lambda_0}\sqrt{x^2 + f^2} \tag{58}$$

$$2\alpha(x) = \theta(x) \pm \frac{\pi}{4} \tag{59}$$

for the chosen polarization $[1 \quad \mp i]/\sqrt{2}$. Figure 6a shows the nonlocal and local scattering kernels, where the four quadrants correspond to the input and output polarization states in an x, y basis. In contrast to the lossless, scalar nonlocal metalens of the previous section, the local kernel in Fig. 6a encodes a varying birefringent axis along the main diagonals of each quadrant [Eq. (51) with Eq. (58)]. Then, we compute the absorption of the metalens due to excitation by point sources placed at positions (x_0, z_0) above the metasurface. Figure 6b shows the spatial absorption maps for three example Q-factors operating at the band-edge frequency, showing that the focal spot tightens as the lifetime of the q-BIC grows. This directly corresponds to an increase in the nonlocality length $\xi_0 = \sqrt{b\tau_r}$, i.e., the metasurface becomes increasingly wavefront selective; hence, increasing b has the same effect at the band-edge frequency. Supplementary Section S14 discusses the response as a function of frequency.

We next rapidly compute the full width at half maximum (FWHM), Δx of the thermal metalens response as a function of Q-factor and NA. Figure 6c shows how ξ_0 and the peak absorption A_{peak} depend on Q. Notably, A_{peak} drops off around $\xi_0 = W$, i.e., when the nonlocality length is comparable to the aperture size (width) of the lens. This stems from incomplete interference: if the finite size of the lens is too small, the q-BIC leaks out the sides of the metalens instead of building up a complete Fano resonance. Figure 6d shows the absorption at the focal plane $z_0 = f$ as a function of Q, showing a continuous trend consistent with Fig. 6d when $\xi_0 < W$. However, when $\xi_0 > W$ we notice instead that the FWHM plateaus above a certain Q [Fig. 6e]. Below that Q-factor, Δx is determined by the target NA together with Q; above that Q-factor, Δx is determined only by the NA. From this, we learn that there is a threshold Q above which the focal spot does not continue to tighten; instead, increasing the Q-factor only reduces the intensity of the response. These results suggest, intuitively, that to improve the response further (i.e., reduce Δx while maintaining $A_{peak} \approx 1$) we must increase the metalens aperture (W) while also increasing the coherence across the aperture (ξ_0).

Discussion

TCMT is the foundational model and framework for studying infinite, periodic resonant flat optics. In analogy, here we have introduced STCMT as the

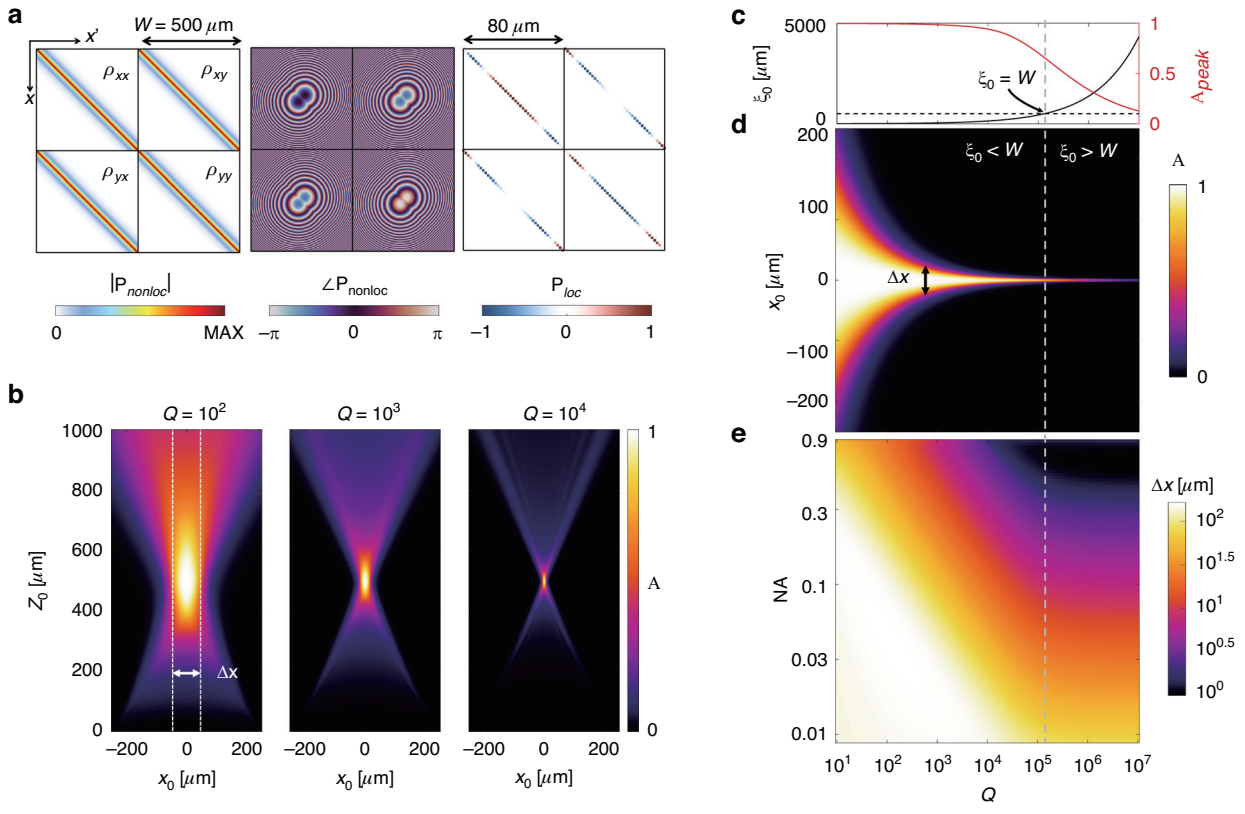

Fig. 6 Thermal metalenses. a Components of the nonlocal reflection kernel for a thermal metalens. (left) Magnitude and (middle) phase of the nonlocal component of the nonlocal reflection kernel P_{nonloc} at the band edge frequency of a thermal metalens that is $W = 500\,\mu m$ wide, focusing to a point $f = 500\,\mu m$ away with $\lambda_0 = 4\,\mu m$. (Right) The corresponding local reflection kernel P_{loc} (with only the central $80\,\mu m$ depicted for visual clarity). **b** Computed absorption for sample point sources placed at positions (x_0, z_0) due to a thermal lens with $b = 0.15\,\mu m$ for three example Q-factors. **c** Nonlocality $\xi_0 = \sqrt{b\tau_r}$ and peak absorption A_{peak} [i.e., for a source placed at $(x_0, z_0) = (0, f)$] as a function of Q. The horizontal black dashed line marks a length of W; the vertical gray dashed line marks the Q at which $\xi_0 = W$. **d** Absorption for point sources placed in the focal plane at $z_0 = f$ as a function of Q. **e** Full width at half maximum Δx of the response at the focal plane as a function of Q and numerical aperture NA, showing two regimes of behavior

framework to model space-varying resonant flat optics. Given the importance of TCMT in understanding and modeling Fano resonant metasurfaces, we believe STCMT is a foundational tool for designing nonlocal meta-optics before implementing a concrete geometry. STCMT qualitatively and quantitatively clarifies the classification of nonlocal metasurfaces into two functionalities: wavefront-selective and wavefront-shaping nonlocal metasurfaces. In practice, this allows rapid classification and prototyping of a system to match the requirements of the application. More generally, STCMT enables computation of the idealized performance of a given system, accounting for finite size, space-varying properties, and degrees of freedom manipulatable across the surface. In other words, it sets goalposts (upper bounds on performance) to be reached for in practical implementation under a given set of assumptions (geometry, number of layers, loss, size, reciprocity, etc.).

In addition to its practical advantages, STCMT captures and provides insight into several novel aspect of recent efforts on nonlocal metasurfaces. We clarified that (i) the two hallmarks of a nonlocal phase gradient (shift of band structure in input momentum, and anomalous reflection) are simple consequences of a mixed Fourier transformation in the context of STCMT; (ii) a symmetric response after such mixed Fourier transformation is enforced by reciprocity, revealing that a near-unity efficiency nonlocal phase gradient is a vertically asymmetric phenomenon, in turn revealing the requirement of vertical asymmetry in the structure; and (iii) that spatial selectivity is intrinsically tied to the eigenwave to a degree controlled by the nonlocality length ξ_0 of the q-BIC (which in turn is a quantitative measure of the lateral interaction length, thereby setting the minimum metasurface size required to observe a Fano response, as seen in Fig. 6c).

While we demonstrated the utility of STCMT for understanding and studying diffractive nonlocal

metasurfaces, this represents a unified framework for metasurfaces *with or without* the assumption of locality. For instance, the local assumption is recovered if only the main diagonal of the nonlocal matrices is populated, and the effects of nearest neighbors are included when more elements than just the main diagonal are nonzero. As discussed further in Supplementary Section S15, our framework may enable future work explicitly taking into account neighbor-neighbor (nonlocal) interactions. Additionally, as showcased in Fig. 6, our framework may include both local and nonlocal metasurface behaviors simultaneously, as required for thermal emission control[35]. Future directions may extend our results to study the resonant dynamics of several spectrally overlapping nonlocal modes[6], leaky-wave metasurfaces based on traveling wave q-BICs[47], the effects due to partial coherence[48], and even nonreciprocal behavior[49]. Finally, Supplementary Section S16 discusses the limitations of the present study. STCMT elegantly captures the essential physics of these systems, and enables rapid, analytical and numerical study and design of next-generational meta-optical system. These emerging platforms are capable of unprecedent control of light for applications such as augmented reality, secure optical communications, compact optical sources, and nonlinear and quantum optics.

Materials and methods

The calculations for Figs. 3–6 were carried out using the matrix form of STCMT, described in the supplementary information. The calculations for Fig. 2 were carried out using the analytical forms of TCMT and STCMT, described in the supplementary information. The full-wave simulations were carried out using Lumerical FDTD solutions. Example codes of the matrix form of STCMT are provided in the supplementary information.

Acknowledgements
This work was supported by the Air Force Office of Scientific Research and the Simons Foundation. We thank Nanfang Yu and Stephanie Malek for discussions on the regime of wavefront-shaping metasurfaces.

Author contributions
A.O. performed the calculations. A.O. and S.M. carried out the derivations and analysis. A.A. supervised the research. All authors contributed to designing the study, discussing the results, and preparing the manuscript.

Data availability
All data needed to evaluate the conclusions in this study are presented in the paper and in the supplementary materials.

Conflict of interest
The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41377-023-01350-9.

Received: 7 August 2023 Revised: 18 November 2023 Accepted: 29 November 2023

Published online: 24 January 2024

References

1. Haus, H. A. *Waves and Fields in Optoelectronics* (Prentice-Hall, 1984).
2. Haus, H. A. & Huang, W. Coupled-mode theory. *Proc. IEEE* **79**, 1505–18 (1991).
3. Paddon, P. & Young, J. F. Two-dimensional vector-coupled-mode theory for textured planar waveguides. *Phys. Rev. B* **61**, 2090–101 (2000).
4. Johnson, S. G. et al. Adiabatic theorem and continuous coupled-mode theory for efficient taper transitions in photonic crystals. *Phys. Rev. E* **66**, 066608 (2002).
5. Fan, S., Suh, W. & Joannopoulos, J. D. Temporal coupled-mode theory for the Fano resonance in optical resonators. *J. Optical Soc. Am. A* **20**, 569–73 (2003).
6. Suh, W., Wang, Z. & Fan, S. H. Temporal coupled-mode theory and the presence of non orthogonal modes in lossless multimode cavities. *IEEE J. Quantum Electron.* **40**, 1511–8 (2004).
7. Ruan, Z. C. & Fan, S. H. Temporal coupled-mode theory for Fano resonance in light scattering by a single obstacle. *J. Phys. Chem. C* **114**, 7324–9 (2010).
8. Sweeney, W. R., Hsu, C. W. & Stone, A. D. Theory of reflectionless scattering modes. *Phys. Rev. A* **102**, 063511 (2020).
9. Verslegers, L. et al. From electromagnetically induced transparency to super-scattering with a single structure: a coupled-mode theory for doubly resonant structures. *Phys. Rev. Lett.* **108**, 083902 (2012).
10. Zhou, M. et al. Inverse design of metasurfaces based on coupled-mode theory and adjoint optimization. *ACS Photonics* **8**, 2265–73 (2021).
11. Alpeggiani, F. et al. Quasinormal-mode expansion of the scattering matrix. *Phys. Rev. X* **7**, 021035 (2017).
12. Landau, L. D. & Lifshitz, E. M. *Electrodynamics of Continuous Media* 2nd edn. (Butterworth-Heinemann, 1984).
13. Agarwal, G. S., Pattanayak, D. N. & Wolf, E. Electromagnetic fields in spatially dispersive media. *Phys. Rev. B* **10**, 1447–75 (1974).
14. Lemarchand, F. et al. Study of the resonant behaviour of waveguide gratings: increasing the angular tolerance of guided-mode filters. *J. Opt. A: Pure Appl. Opt.* **1**, 545–51 (1999).
15. Magnusson, R., Shokooh-Saremi, M. & Wang, X. Dispersion engineering with leaky mode resonant photonic lattices. *Opt. Express* **18**, 108–16 (2010).
16. Nguyen, H. S. et al. Symmetry breaking in photonic crystals: on-demand dispersion from flatband to Dirac cones. *Phys. Rev. Lett.* **120**, 066102 (2018).
17. Monticone, F., Valagiannopoulos, C. A. & Alù, A. Parity-time symmetric nonlocal metasurfaces: all-angle negative refraction and volumetric imaging. *Phys. Rev. X* **6**, 041018 (2016).
18. Silva, A. et al. Performing mathematical operations with metamaterials. *Science* **343**, 160–3 (2014).
19. Kwon, H. et al. Nonlocal metasurfaces for optical signal processing. *Phys. Rev. Lett.* **121**, 173004 (2018).
20. Zhou, Y. et al. Flat optics for image differentiation. *Nat. Photonics* **14**, 316–23 (2020).
21. Tittl, A. et al. Imaging-based molecular barcoding with pixelated dielectric metasurfaces. *Science* **360**, 1105–9 (2018).
22. Hong, J. et al. Nonlocal metasurface for circularly polarized light detection. *Optica* **10**, 134–41 (2023).
23. Song, J. H. et al. Nonlocal metasurfaces for spectrally decoupled wavefront manipulation and eye tracking. *Nat. Nanotechnol.* **16**, 1224–30 (2021).
24. Guo, C., Wang, H. W. & Fan, S. H. Squeeze free space with nonlocal flat optics. *Optica* **7**, 1133–8 (2020).
25. Reshef, O. et al. An optic to replace space and its application towards ultra-thin imaging systems. *Nat. Commun.* **12**, 3512 (2021).
26. Shastri, K. & Monticone, F. Nonlocal flat optics. *Nat. Photonics* **17**, 36–47 (2023).
27. Barton, D. III et al. High-*Q* nanophotonics: sculpting wavefronts with slow light. *Nanophotonics* **10**, 83–88 (2020).
28. Overvig, A. & Alù, A. Diffractive nonlocal metasurfaces. *Laser Photonics Rev.* **16**, 2100633 (2022).
29. Chai, R. H. et al. Emerging planar nanostructures involving both local and nonlocal modes. *ACS Photonics* **10**, 2031–44 (2023).
30. Miller, D. A. B. Why optics needs thickness. *Science* **379**, 41–45 (2023).
31. Overvig, A. C., Malek, S. C. & Yu, N. Multifunctional nonlocal metasurfaces. *Phys. Rev. Lett.* **125**, 017402 (2020).
32. Lawrence, M. et al. High quality factor phase gradient metasurfaces. *Nat. Nanotechnol.* **15**, 956–61 (2020).

33. Overvig, A., Yu, N. F. & Alù, A. Chiral quasi-bound states in the continuum. *Phys. Rev. Lett.* **126**, 073001 (2021).
34. Overvig, A. C. & Alù, A. Wavefront-selective fano resonant metasurfaces. *Adv. Photonics* **3**, 026002 (2021).
35. Overvig, A. C., Mann, S. A. & Alù, A. Thermal metasurfaces: complete emission control by combining local and nonlocal light-matter interactions. *Phys. Rev. X* **11**, 021050 (2021).
36. Overvig, A. C. et al. Observation of a polarization-agnostic geometric phase in nonlocal metasurfaces. Preprint at https://arXiv.org/abs/2302.13215 (2023).
37. Malek, S. C. et al. Multifunctional resonant wavefront-shaping meta-optics based on multilayer and multi-perturbation nonlocal metasurfaces. *Light Sci. Appl.* **11**, 246 (2022).
38. Lin, L. et al. Universal narrowband wavefront shaping with high quality factor meta-reflect-arrays. *Nano Lett.* **23**, 1355–62 (2023).
39. Hail, C. U. et al. High quality factor metasurfaces for two-dimensional wavefront manipulation. Preprint at https://arxiv.org/abs/2212.05647 (2022).
40. Koshelev, K. et al. Asymmetric metasurfaces with high-*Q* resonances governed by bound states in the continuum. *Phys. Rev. Lett.* **121**, 193903 (2018).
41. Dana, B., Lobachinsky, L. & Bahabad, A. Spatiotemporal coupled-mode theory in dispersive media under a dynamic modulation. *Opt. Commun.* **324**, 165–7 (2014).
42. Bykov, D. A. & Doskolovich, L. L. Spatiotemporal coupled-mode theory of guided mode resonant gratings. *Opt. Express* **23**, 19234–41 (2015).
43. Yu, N. F. et al. Light propagation with phase discontinuities: generalized laws of reflection and refraction. *Science* **334**, 333–7 (2011).
44. Nolen, J. R. et al. Arbitrarily polarized and unidirectional emission from thermal metasurfaces. Preprint at https://arxiv.org/abs/2301.12301 (2023).
45. Huang, H. Q. et al. Leaky-wave metasurfaces for integrated photonics. *Nat. Nanotechnol.* **18**, 580–8 (2023).
46. Miller, D. A. B. Waves modes, communications, and optics: a tutorial. *Adv. Opt. Photonics* **11**, 679–825 (2019).
47. Miller, D. A. B., Zhu, L. X. & Fan, S. H. Universal modal radiation laws for all thermal emitters. *Proc. Natl Acad. Sci. USA* **114**, 4336–41 (2017).
48. Wolf, E. *Introduction to the Theory of Coherence and Polarization of Light* (Cambridge University Press, 2007).
49. Mann, S. A., Sounas, D. L. & Alù, A. Nonreciprocal cavities and the time-bandwidth limit. *Optica* **6**, 104–10 (2019).