Maximal Free-Space Concentration of Light

Hyungki Shim,1, 2 Haejun Chung, 1 and Owen D. Miller 1

1Department of Applied Physics and Energy Sciences Institute, Yale University, New Haven, Connecticut 06511, USA
2Department of Physics, Yale University, New Haven, Connecticut 06511, USA

(Dated: May 28, 2019)

We derive upper bounds to free-space light concentration, mapping out the limits to maximum intensity for any spot size and optical beam-shaping device. For sub-diffraction-limited optical beams, our bounds suggest the possibility for orders-of-magnitude intensity enhancements compared to existing demonstrations, and we use inverse design to discover metasurfaces operating near these new limits. Our bounds require no assumptions about symmetry, scalar waves, or weak scattering, instead relying primarily on the transformation of a quadratic program via orthogonal-projection methods. The bounds and inverse-designed structures presented here can be useful for applications from imaging to 3D printing.

Free-space optical beams with large focal-point intensities and arbitrarily small spot sizes—below the diffraction limit—are a long-sought goal [1–3] for applications from imaging [4–8] to 3D printing [9, 10], for which nanostructured lenses have enabled recent experimental breakthroughs [11, 12]. In this Letter, we derive fundamental limits to free-space optical-beam concentration, revealing the maximum possible focal-point intensity (related to the well-known “Strehl ratio” [13, 14]) for any desired spot size. For waves incident from any region of space—generated by scattering structures, spatial light modulators, or light sources of arbitrary complexity—we show that the non-convex beam-concentration problem can be transformed to a quadratic program [15] with easily computable global optima. We also extend this approach to derive maximum intensity independent of the exit surface of an incident wave. Our bounds simplify to those derived by Fourier analysis of prolate spheroidal wave functions [16–18] in the scalar 1D limit. By honing in on the two essential degrees of freedom—the field intensity at the focal point, and its average over a ring at the desired spot size—optimal beam concentration can be further simplified to a rank-two optimization, resulting in analytical upper bounds in the far zone. For very small spot sizes \( G \) which are most desirable for transformative applications, we show that the focal-point intensity must decrease proportional to \( G^4 \), a dimension-independent scaling law that cannot be overcome through any form of wavefront engineering. The bounds have an intuitive interpretation: the ideal field profile at the exit surface of an optical beam-shaping device must have maximum overlap with the fields radiating from a dipole at the origin yet be orthogonal to the fields emanating from a current loop at the spot size radius. We compare theoretical proposals and experimental demonstrations to our bounds, and we find that there is significant opportunity for order-of-magnitude intensity enhancements at those small spot sizes. We use “inverse design” [19–22], a large-scale computational-optimization technique, to design metasurfaces that generate nearly optimal wavefronts and closely approach our general bounds.

It is now well-understood that the diffraction “limit,” which is a critical factor underpinning resolution limits in imaging [4–8], photolithography [23–25], etc. [26–29], is not a strict bound on the size of an optical focal spot, but rather a soft threshold below which beam formation is difficult in some generic sense (e.g., accompanied by high-intensity sidelobes). Although evanescent waves can be leveraged to surpass the diffraction limit [30–34], they require structuring in the near field. The possibility of sub-diffraction-limited spot sizes without near-field effects was recognized in 1952 by Toraldo di Francia [35]; stimulated by results on highly-directive antennas [36], he analytically constructed successively narrower beam profiles with successively larger sidelobe energies (i.e. energies outside the first zero), in a scalar, weak-scattering asymptotic limit. Subsequent studies [37–40] have connected the theory of sub-diffraction-limited beams to “super-oscillations” in Fourier analysis [41, 42], i.e. bandlimited functions that oscillate over length/time scales faster than the inverse of their largest Fourier component. For one- and two-dimensional scalar fields, superoscillatory wave solutions have been explicitly constructed [39, 43–45], and in the one-dimensional case energy-concentration bounds have been derived [18] by the theory of prolate spheroidal wave functions [16, 17]. For optical beams, the only known bounds to focusing (apart from bounds on energy density at a point without considering spot sizes [46, 47]) are those derived in Ref. [48, 49] (and recently in Ref. [50] albeit with a different optimization goal), which use special-function expansions and/or numerical-optimization techniques to discover computational bounds that apply for weakly scattering, rotationally symmetric filters in a scalar approximation. A bound that does not require weak scattering was developed in Ref. [49], but it still assumes rotational symmetry in a scalar diffraction theory.

The recent demonstrations [11, 12, 51–61] of complex wavelength-scale surface patterns focusing plane waves to sub-diffraction-limited spot sizes has inspired hope that the previous tradeoffs of large sidelobe energies or small focal-point intensities might be circumvented or ameliorated by strongly scattering media accounting for the vector nature of light [54, 57], as all previous [18, 39, 48, 49]...
asymptotic scaling relations and energy bounds require assumptions of rotational symmetry, weak scattering (except Ref. [49]), and scalar waves. Such possibilities are especially enticing in the context of the broader emergence of “metasurfaces” [62–64] enabling unprecedented optical response.

In this Article, we derive bounds on the maximum concentration of a light that do apply in the fully vectorial, strongly scattering regime, without imposing any symmetry constraints. Our derivation starts with the electromagnetic equivalence principle [65], which allows us to consider the effects of any scatterer/modulator/light source as effective currents on some exit surface (Sec. IA). The optimal beam-concentration problem is non-convex due to the requirement for a particular spot size, but we use standard transformations from optimization theory to rewrite the problem as a quadratic program amenable to computational solutions for global extrema. We subsequently bound the solution to the quadratic problem by a simpler and more general rank-two optimization (Sec. IA), and also develop bounds independent of exit surface via modal decomposition (Sec. IB). The rank-two bounds reduce to analytic expressions in the far zone, and we compare the ideal field profiles to various theoretical and experimental demonstrations in Sec. II. We show that there is still opportunity for orders-of-magnitude improvements, and design metasurfaces approaching our bounds (Sec. III). Finally, in Sec. IV, we discuss extensions of our framework to incorporate metrics other than focal-point intensity, near-field modalities, inhomogeneities, new point-spread functions, and more.

I. GENERAL BOUNDS

A. Aperture-Dependent Bounds

Consider a beam generated by almost any means, e.g., an incident wave passing through a scatterer with a complex structural profile [66–68], precisely controlled spatial light modulators [69–72], or a light source with a complex spatial emission profile [73–75]. The physics underlying the extent to which such a beam is concentrated spatially in free space is distilled to its essence by the electromagnetic equivalence principle [65]: the propagating fields are uniquely defined by their tangential values on any beam-generation exit surface, forming effective current solutions that encapsulate the entire complexity of the field-generation process. By this principle, the beam-focusing problem is equivalent to asking: what is the maximum spatial concentration of a beam generated by electric and magnetic surface currents radiating in free space? We depict this distillation of the problem in Fig. 1. We consider fields and currents at a single temporal frequency \( \omega (e^{-i\omega t} \text{ time evolution}) \), and simplify the expressions to follow by encapsulating the electric and magnetic fields \((\mathbf{E}, \mathbf{H})\) and currents \((\mathbf{K}_{\text{eff}}, \mathbf{N}_{\text{eff}})\) in 6-vectors \(\psi\) and \(\xi\), respectively:

\[
\psi = \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}, \quad \xi = \begin{pmatrix} \mathbf{K}_{\text{eff}} \\ \mathbf{N}_{\text{eff}} \end{pmatrix}.
\]

The fields \(\psi\) emanating from the effective currents \(\xi\) distributed across the “exit” surface \(A\) are given by the convolution of the currents with \(\Gamma\), the known \(6 \times 6\) free-space dyadic Green’s function [76]:

\[
\psi(x) = \int_A \Gamma(x, x')\xi(x').
\]

Thus, the currents comprise the degrees of freedom determining the beam shape. As illustrated in Fig. 1, finding the maximum focal intensity at a single point for any desired focal spot size now reduces to determining the optimal effective currents. We will assume equations such as Eq. (2) can be solved by any standard electromagnetic discretization scheme [77], and we will write the matrix versions with the same symbols but without position arguments. For example, \(\psi = \Gamma\xi\) is the matrix equivalent of Eq. (2), with \(\psi\) and \(\xi\) vectors and \(\Gamma\) a matrix. The total intensity at any point in free space, summing electric and magnetic contributions, is given by the squared norm of \(\psi\):

\[
I(x) = |\psi(x)|^2 = \int_A \int_A \xi^\dagger(x'')\Gamma\dagger(x, x'')\Gamma(x, x')\xi(x').
\]

We now formulate the maximal-concentration question as a constrained optimization problem. The ideal optical beam has maximum focal intensity at a point \((x = 0)\), zero field along some spot-size contour \(C\), and a total propagating power \(P\) not exceeding an input value of \(P_0\). Thus, the maximum focal intensity, and the ideal effective currents generating it, solve the optimization problem:

\[
\begin{aligned}
\text{maximize} \quad & I(x = 0) = \xi^\dagger\Gamma_0\Gamma_0\xi \\
\text{subject to} \quad & \psi(x)|_C = \Gamma_0\xi = 0 \quad \text{and} \quad P \leq P_0,
\end{aligned}
\]

where the “0” and \(C\) subscripts indicate that \(\Gamma\) and \(\psi\) are evaluated (in the appropriate basis) at the origin or at the spot-size contour, respectively. Attempting to directly solve Eq. (4) is infeasible: the \(\Gamma_0\dagger\Gamma_0\) matrix is positive semidefinite (which is nonconvex under maximization [78]), the equality constraint prevents the use of Rayleigh-quotient-based approaches [79], and the power constraint is difficult to write in a simple linear or quadratic form.

We can bypass the nonconvexity of the problem through multiple transformations. First, to simplify the power constraint, we replace it with a constraint on the intensity of the effective currents, normalized such that their total intensity is one: \(\xi^\dagger\xi = 1\). We seek the ideal beam, which has all of its intensity generating
power in the direction of the maximum-intensity spot, validating this replacement. Second, we subsume the equality constraint by projecting the currents $\xi$ onto the subspace of all currents that generate zero field on $\mathcal{C}$: 

$$\xi = \left( I - \Gamma^\dagger_C \left( \Gamma_C \Gamma^\dagger_C \right)^{-1} \Gamma_C \right) \nu = \mathbf{P} \nu,$$

where $I$ is the identity matrix, the second term is the orthogonal projection $\mathbf{P}$ rows of $\Gamma^\dagger_C$ onto the null space of $\Gamma_C$, and we have assumed any linearly dependent rows of $\Gamma_C$ have been removed such that the inverse of $\Gamma_C \Gamma^\dagger_C$ exists. By this projection, for any $\nu$, the equality constraint is satisfied, $\Gamma_C \xi = \Gamma_C \mathbf{P} \nu = \left( \Gamma_C - \Gamma_C \right) \nu = 0$.

Finally, we simplify the quadratic figure of merit, $\Gamma_0 \Gamma_0^\dagger$, by recognizing that $\Gamma_0$ is a $6 \times 6N$ matrix, where $N$ are the effective-current-basis degrees of freedom, and thus $\Gamma_0^\dagger \Gamma_0$ is a matrix with rank at most 6, as dictated by the polarizations of the electric and magnetic fields at the origin. Instead of incorporating all intensities, we could instead project the field at the origin onto an arbitrary six-component polarization vector $\mu$. The intensity at the origin in this polarization is given by $\nu^\dagger \mathbf{P} \Gamma_0^\dagger \mu^\dagger \Gamma_0 \mathbf{P} \nu$, where the inner matrix $\Gamma_0^\dagger \mu^\dagger \Gamma_0$ is now rank one. Rank-one quadratic forms are particularly simple, as evidenced here by the fact that we can define a vector $\gamma_\mu = \Gamma_0^\dagger \mu$ such that the intensity at the origin is simply $\nu^\dagger \mathbf{P} \gamma_\mu^\dagger \mathbf{P} \nu$.

The above transformations yield the equivalent but now tractable optimization problem:

$$\begin{align*}
\text{maximize} & \quad \nu^\dagger \mathbf{P} \gamma_\mu^\dagger \mathbf{P} \nu \\
\text{subject to} & \quad \nu^\dagger \mathbf{P} \nu \leq 1,
\end{align*}$$

(5)

where $\nu$ represents arbitrary effective currents, $\mathbf{P}$ projects them to satisfy the zero-field condition, and $\gamma_\mu$ represents the conjugate transpose of the Green’s function from the effective currents to the maximum-intensity point. Equation (5) is equivalent to a Rayleigh-quotient maximization, and the solution is therefore given by the largest eigenvalue and corresponding eigenvector of the generalized eigenproblem $\mathbf{P} \gamma_\mu^\dagger \mathbf{P} \nu = \lambda \mathbf{P} \nu$. Here, because $\gamma_\mu \gamma_\mu^\dagger$ is rank one, it is straightforward to show (SM) that the solution can be written analytically, with maximal eigenvector $\nu = \mathbf{P} \gamma_\mu / \| \mathbf{P} \gamma_\mu \|$ and maximal eigenvalue of $\gamma_\mu \mathbf{P} \gamma_\mu^\dagger$. Reinserting the transformed variable definitions from above, the optimal (unnormalized) effective currents are given by $\xi_{\text{opt}} = \Gamma_0^\dagger \mu - \Gamma^\dagger_C \left( \Gamma_C \Gamma^\dagger_C \right)^{-1} \Gamma_C \Gamma_0^\dagger \mu$.

Then, we have that the $\mu$-polarized intensity at the origin, for any wavefront-shaping device in any configuration, is bounded above by the expression

$$I \leq \mu^\dagger \left[ \Gamma_0 \Gamma_0^\dagger - \Gamma_0^\dagger \Gamma_C \left( \Gamma_C \Gamma^\dagger_C \right)^{-1} \Gamma_C \Gamma_0 \right] \mu.$$  

(6)

Equation (6) represents a first key theoretical result of our work. Although it may have an abstract appearance, it is a decisive global bound to the optimization problem, requiring only evaluation of the known free-space dyadic Green’s function at the maximum-intensity point, the zero-field contour, and the effective-current exit surface. The matrix in the square brackets is a $6 \times 6$ matrix, whose largest eigenvector represents the optimal polarization. And the structure of Eq. (6) has simple physical intuition: the maximum intensity of an unconstrained beam would simply focus as much of the effective-current radiation to the origin, as dictated by $\Gamma_0 \Gamma_0^\dagger$, but the constraint requiring zero field on $\mathcal{C}$ necessarily reduces the intensity by an amount proportional to the projection of the spot-size field ($\Gamma_C$) on the field at the origin ($\Gamma_0$).

The transformations leading to Eq. (6) are exact, requiring no approximations nor simplifications. Thus the optimal fields, given by $\xi_{\text{opt}}$, are theoretically achievable Maxwell-equation solutions, and the bound of Eq. (6) is tight: no smaller upper bound is possible. We find that

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Our framework establishes maximum light concentration for any zero-field contour, such as a circle. We derive two bounds: one that incorporates the shape of the exit aperture (enabling comparison with the well-known Strehl ratio), while otherwise independent of the beam-generation method, and a second that requires only a modal basis and is independent of aperture.}
\end{figure}
using a spectral basis [81] for the zero-field contour and simple collocation [81] for the aperture plane suffice for rapid convergence and numerical evaluation of Eq. (6) within seconds on a laptop.

To simplify the upper bound and gain further physical intuition, we can leverage the fact that the high-interest scenario is small, sub-diffraction-limited spot sizes. The zero-field condition, i.e., $\Gamma C \xi = 0$ in Eq. (4), is typically a high-rank matrix due to the arbitrarily large number of degrees of freedom in discretizing the zero-field contour. Yet in a spectral basis, such as Fourier modes on a circular contour or spherical harmonics on a spherical surface, for small spot sizes it will be the lowest-order mode, polarized along the $\mu$ direction, that is most important in constraining the field. If we denote the basis functions as $\phi_i$, then $\phi_0$ would be the prime determinant of the zero-field constraint at small spot sizes. Instead of constraining the entire field to be zero along the zero-field contour, then, if we only constrain the zeroth-order, $\mu$-polarized mode, we will lessen the bound but gain the advantage that the zero-field constraint is now of the form $(\phi_0^T \Gamma C) \xi = 0$, a vector-vector product with rank one. Then the previous analysis can be applied, with the replacement $\Gamma C \rightarrow \phi_0^T \Gamma C$. We can introduce two new fields, physically motivated below, by the definitions

$$\psi_0 = \Gamma_0^T \mu$$

$$\psi_1 = \Gamma_C^T \phi_0.$$  \hspace{1cm} (7)

$$\psi_1 = \Gamma_C^T \phi_0.$$  \hspace{1cm} (8)

Given these two fields, algebraic manipulations (SM) lead to an upper bound on the maximum intensity,

$$I \leq |\psi_0^T \psi_0 - |\psi_0^T \psi_1|^2/\psi_1^T \psi_1|,$$  \hspace{1cm} (9)

where the bound comprises a first term that denotes the intensity of a spot-size-unconstrained beam, while the second term accounts for the reduction due to imposition of the spot size constraint.

The fields of Eqs. (7,8) can intuitively explain the bound of Eq. (9). Whereas $\Gamma_0$ and $\Gamma_C$ generate fields in the focusing region from currents in the aperture plane, $\Gamma_0^T$ and $\Gamma_C^T$ generate fields in the aperture plane from the focusing region. By reciprocity [65], which relates $\Gamma(x,x')$ to $\Gamma(x',x)$, the field $\psi_0 = \Gamma_0^T \mu$ is related to the field emanating from dipolar sources at the focal spot back to the aperture plane (it is the conjugate of that field, with the signs of magnetic sources and fields reversed—reciprocity flips the signs of off-diagonal matrices of $\Gamma$). Similarly, $\psi_1 = \Gamma_C^T \phi_0$ is related to the field emanating from the zero-field region back to the aperture plane. As illustrated in Fig. 2, the bound of Eq. (9) states that the maximum focal-spot intensity is given by the norm of the first field (focal point to aperture) minus the overlap of that field with the second (zero-field region to aperture). The smaller a desired spot size is, the closer these fields are to each other, increasing their overlap and reducing the maximum intensity possible. This intuition is furthered by considering the optimal effective currents that would achieve the bound of Eq. (9), which are given by (SM):

$$\xi_{\text{opt}} = \xi_0 \left[ 1 - \frac{|\psi_0^T \psi_1|^2}{|\psi_1^T \psi_1|} \right].$$  \hspace{1cm} (10)

where $\xi_0 = 1/\sqrt{|\psi_0^T \psi_0 - |\psi_0^T \psi_1|^2|}$ is a normalization factor such that $\xi_{\text{opt}}^T \xi_{\text{opt}} = 1$. Equation (10) demonstrates that the ideal field on the exit surface should maximize overlap with $\psi_0$ while being orthogonal to $\psi_1$. For small spot sizes, these two fields are almost identical, resulting in a significantly reduced maximum intensity.

### B. Modal-Decomposition Bounds

Alternatively, one might ask about maximum spatial concentration of light independent of exit surface, simply enforcing the condition that the light field comprises propagating waves. For example, in a plane, what combination of plane waves (or any other modal basis [82]) offers maximum concentration? In this case, the formulation is very similar to that of Eq. (4), except that now the field $\psi$ is given as a linear combination of modal fields: $\psi = U c$, where $U$ is a modal basis matrix (after appropriate discretization) and $c$ is a vector of modal-decomposition coefficients. By analogy with $\Gamma_0$ and $\Gamma_C$, we can define the field at zero and on the zero-field contour in the modal basis as $U_0$ and $U_C$, respectively. Then, the bounds of Eqs. (6,9) and the definitions of Eqs. (7,8) apply directly to the modal-decomposition case.
with the replacements $\Gamma_0 \rightarrow U_0$ and $\Gamma_C \rightarrow U_C$. For completeness, we can write here the general bounds:

$$I \leq \mu \left[ U_0 U_0^\dagger - U_0 U_C^\dagger \left( U_C U_C^\dagger \right)^{-1} U_C U_0^\dagger \right] \mu. \quad (11)$$

Equation (11) represents the second key general theoretical result: as for Eq. (6), it appears abstract, but it is a simple-to-compute global bound on the intensity via the $6 \times 6$ matrix in square brackets, which again has clear physical intuition as the maximum unconstrained intensity (from $U_0 U_0^\dagger$) minus the projection of that field onto the representation of a constant field along the zero contour projected onto the modal basis (the second term).

The bound of Eq. (11) applies generally to any modal basis and zero-field contour. For the prototypical case of plane-wave modes and a circular zero-field contour in the plane, one can find a semi-analytical expression for Eq. (11), with ideal field profiles shown in Fig. 3. If one defines the maximal unconstrained intensity as $I_0$ (which is $3k^2/16\pi$ given appropriate normalizations) then, as we show in the SM, the maximum focusing intensity for spot size $R$ is given by a straightforward thorough tedious combination of zeroth, first, and second-order Bessel functions; in the small-spot-size limit ($kR \ll 1$), the asymptotic bound is

$$I \leq \frac{13}{13824\pi} (kR)^4 I_0. \quad (12)$$

The maximum intensity must fall of at least as the fourth power of spot size, identical to the dependence of the aperture-dependent bounds in the far field, which has important ramifications for practical design, as we show in the next section.

Our bounds share a common origin with those of an “optical eigenmode” approach [83]: the quadratic nature of power and momentum flows in electromagnetism. A key difference appears to be the choice of figure of merit, as well as the purely computational nature of the optical-eigenmode approach [83, 84], using computational projections onto numerical subspaces. Above, we have shown that orthogonal projections and physically-motivated Fourier decompositions lead to analytical and semi-analytical bound expressions.

We show in the SM that in the case of scalar waves in one dimension that our modal-decomposition bounds coincide exactly with those derived by a combination of Fourier analysis and interpolation theory [18, 85]. In fact, if in the 1D case one were to stack $U_0$ and $U_C$ in a single matrix and multiply by its conjugate transpose then the resulting matrix, $\begin{pmatrix} U_0 U_0^\dagger & U_0 U_C^\dagger \\ U_C U_0^\dagger & U_C U_C^\dagger \end{pmatrix}$, is exactly the matrix of sinc functions that defines the eigenvector for which prolate spheroidal wave functions (PSWFs) are the eigenvalues [86]. Thus, our modal-basis approach can be understood as a vector-valued, multi-dimensional generalization of the PSWF-based Fourier analysis of minimum-energy superoscillatory signals.

**II. OPTICAL BEAMS IN THE FAR ZONE**

The bounds of Eqs. (6,9) allow arbitrary shapes for the exit surface and the zero-field contour. The prototypical case of interest, for many applications across imaging and 3D printing, for example, involves a beam of light shaped or created within a planar aperture, or more generally within any half space where the exit surface can be chosen to be a plane, and propagating along one direction, with concentration measured by the spot size in a transverse two-dimensional plane. Hence the exit surface is an aperture plane and the zero-field contour is a spot-size circle. For ease of comparison, a dimensionless concentration metric known as the “Strehl ratio” [13, 14] quantifies focusing in the far zone of such beams, where diffraction effects cannot be accounted for in the normalization.

In the far zone, with the focusing–aperture distance much larger than the aperture radius and the wavelength of light, the six electric and magnetic polarizations decouple, reducing the response for any one to a scalar problem. As we show in the SM, for any aperture-plane polarization, the focal-point field $\psi_0$ of Eq. (7) is proportional to $e^{-ikz}/z$, for propagation direction $z$ and wavenumber $k = \omega/c$, while the zero-contour field $\psi_1$ of Eq. (8) is proportional to the same factor multiplied by the zeroth-order Bessel function $J_0$, i.e., $\psi_1 \sim J_0(kr\rho_0/z)e^{-ikz}/z$, where $\rho_0$ is the spot-size radius and $r$ is the radial position in the aperture plane. These are the Green’s-function solutions and require no assumptions about the symmetries of the optimal fields. The evaluation of the overlap integrals $\psi_0^\dagger \psi_0$, $\psi_0^\dagger \psi_1$, and $\psi_1^\dagger \psi_1$ in the aperture plane are integrals of constants and Bessel functions. For any
shape of aperture, we can find an analytical bound on the maximal focusing intensity by evaluating the bound for the circumscribing circle of radius $R$. Performing the integrals (SM), Eq. (9) becomes

$$I \leq \frac{k^2 R^2}{16\pi z^2} - \frac{1}{4\pi \rho_0^2} \frac{[J_1(kR\rho_0/z)]^2}{[J_0(kR\rho_0/z)]^2 + [J_1(kR\rho_0/z)]^2}.$$  

(13)

Equation (13) provides a general bound for any aperture-focus separation distance $z$ and spot-size radius $\rho_0$. The dependence on $kR/z$ and related quantities is characteristic of any far-zone beam, and can be divided out for a separation-distance-independent bound. The Strehl ratio accounts for this dependence in circularly-symmetric beams by dividing the focal-point intensity by that of an Airy disk, which is the diffraction-limited pattern produced by a circular aperture. Within the Strehl ratio is a normalized spot-size radius, $\eta_0 = kR\rho_0/z$, which equals the Airy-pattern spot size multiplied by a normalized spot size $G$ between 0 and 1. We can generalize the Strehl definition beyond the Airy pattern: instead, divide the maximum intensity, Eq. (9), by the intensity of an unconstrained focused beam (without the zero-field condition), which is simply $\psi_1^\dagger \psi_1$ (which conforms to the usual Airy definition for a circular aperture).

Thus $S = I/I_{\text{max}} = 1 - |\psi_0^\dagger \psi_1|^2 / \sqrt{\psi_0^\dagger \psi_0 \psi_1^\dagger \psi_1}$. By this definition, the Strehl ratio $S_{\text{max}}$ of the optimal-intensity beam of Eq. (13) is given by

$$S_{\text{max}} = 1 - \frac{4}{\eta_0^4} \frac{[J_1(\eta_0)]^2}{[J_0(\eta_0)]^2 + [J_1(\eta_0)]^2}.$$  

(14)

As $\eta_0$ decreases below the first zero of $J_1$, i.e., the diffraction-limited Airy spot size, the second term of Eq. (14) increases, reducing $S_{\text{max}}$. Although we arrived at Eqs. (13,14) from Eq. (9), the bound derived from loosening the constraints and solving the rank-two optimization problem, our numerical results show that in the far zone, the full-rank optimization problem of Eq. (5) that is bounded above by Eq. (6) has exactly the same solution (the equivalence is not exact for non-circular apertures, but even then the discrepancy practically vanishes for spot size $G \ll 1$). Physically, this means that in the far zone, maximally focused beams are symmetric under rotations around the propagation axis, for any spot size, such that only their first Fourier coefficient is nonzero on the spot-size ring. From a design perspective, this equivalence implies that the bound of Eq. (14) is physically achievable, and that the corresponding Maxwell field exhibits the largest possible intensity for a given spot size.

Figure 4(a) plots the intensity bound, Eq. (9), for a variety of exit-aperture shapes and from the near zone to the far zone, with a generic spot size $G = 0.55$. The bound is scaled by $z^2$, the square of the aperture-focus distance, to account for the quadratic power decay. In each case the far-zone bound is larger than that of the near zone or mid zone, suggesting that the far-zone bounds of Eq. (13) and Eq. (14) may be global bounds at any distance. For each case, the maximum intensity is bounded above by the bound for the circumscribing circle (solid black line), while the optimal field profiles are highly dependent on the aperture shapes (inset images).

Figure 4(b) plots the far-zone bound $S_{\text{max}}$ as a function of spot size $G$ (blue curve), and compares various theoretical results from the literature [12, 56, 84, 87–90]. (Many of the references include experimental results; for fair comparison and to exclude experimental errors, we use either their simulated $G$ and $S$ values, or reconstruct them with our own simulations as detailed in the SM.) For relatively large spot sizes ($G > 0.7$), theoretical proposals for amplitude [87] or phase [88] filters can closely approach the limits, though embedded in the proposals is a weak-scattering assumption that may be difficult to achieve in practice. (The key reason they fall short of the bound is that they do not allow for multiple scattering to redistribute energy in the exit plane.) For small spot sizes, on the other hand, the maximum Strehl ratio decreases rapidly. A Taylor expansion of Eq. (14) reveals the asymptotic bound (SM),

$$S_{\text{max}} = \eta_0^4/192, \quad \eta_0 \ll 1,$$  

(15)

which represents a severe restriction—halving the spot size costs a sixteenfold decrease in maximum focal intensity. This fundamental limit suggests that extremely small spot sizes are impractical, from both power-consumption and fabrication-tolerance perspectives. The quartic dependence is independent of dimensionality (in the SM we show that the same dependence arises for a focal sphere, as well as for focal points in 2D problems) and can be explained generally: for small enough spot sizes, the zero-contour field will always have a maximum at the origin and thus all odd powers in a Taylor expansion around the origin must be zero. The first non-constant field dependence in the expansion is quadratic, and since the overlap quantities in the intensity bound are themselves quadratic in the field, the general intensity dependence on spot size always results in an $\sim \eta_0^4$ scaling law.

Perhaps the most important region of the figure is for intermediate values of spot size ($0.1 \lesssim G \lesssim 0.7$), where it is possible to meaningfully shrink the spot size below the diffraction limit without an overwhelming sacrifice of intensity. This is the region that recent designs [12, 56, 84, 89, 90] have targeted (especially $0.5 < G < 0.7$) with a variety of approaches, including super-oscillatory lenses/needles and optimized pupil filters. Yet as seen in Fig. 4(a), these designs mostly fall dramatically short of the bounds. The best result by this metric is the “optimized pupil filter” of Ref. [84], whose quadratic-programming approach comes within a factor of 5 of the bound, and demonstrates the utility of computational-design approaches for maximum intensity. The other designs fall short of the bounds by factors of 100-1000X or more, offering the possibility for significant improvement by judicious design of the diffractive optical...
Amplitude-only filter
Phase-only filter
Optimized pupil filter
$S_{\text{max}}$, Eq. (13)
$\approx 3\times$
$\approx 5\times$
Super-oscillatory lens
Super-oscillatory optical needle
Amplitude-modulated mask

FIG. 4: (a) Intensity bounds for various aperture shapes (normalized to far-zone bound, Eq. (13)). The bounds, scaled by $z^2$ for aperture–focus distance $z$, are largest in the far zone, where the optimal field profiles (inset images) are highly dependent on the aperture shape. Equation (13) is the general bound of the circumscribing circle for each shape, represented in the dashed line. (b) Maximum Strehl ratio (Eq. (14)) compared to previous designs, including amplitude [87]/phase [88]-only pupil filters, optimized pupil filters [84], a super-oscillatory lens [12], a super-oscillatory optical needle [56], a supercritical lens [89], and an amplitude-modulated mask [90]. Inset: Our bound (blue), despite allowing for arbitrary diffractive optical elements, is smaller than bound of Ref. [48] (black dashed), which requires rotation symmetry in the scalar and weak-scattering limit.

We can also characterize the effective currents on the aperture that achieve the bound in Eq. (14). For spot sizes close to the diffraction limit ($G = 1$), the currents are maximally concentrated around the aperture rim and decreases towards the center, where the amplitude is close to zero. This is because small, localized spots require large transverse wavevector components, which originate from the currents around the rim. As spot size decreases, those edge currents are partially redistributed to the center, to create the interference effects giving rise to sub-diffraction-limited spots.

As discussed above, our bounds, both in the general case of Eq. (6) and in the optical-beam case of Eq. (13), are tight in the sense that they are achievable by fields that are solutions of Maxwell’s equations, given by $\psi = \Gamma_{\text{opt}}$. Yet as shown in Fig. 4, theoretical designs for sub-diffraction-limited beams have fallen far short of the bounds. A natural question, then, is whether realistic material patterning and designs can generate the requisite fields to achieve the bounds?

### III. INVERSE-DESIGNED METASURFACES

We use “inverse design” to discover refractive-index profiles that can approach the concentration bounds. Inverse design [19–22] is a large-scale computational-optimization technique, mathematically equivalent to backpropagation in neural networks [91–93], that enables rapid computation of sensitivities with respect to arbitrarily many structural/material degrees of freedom. Given such sensitivities, standard optimization techniques [94] such as gradient descent (employed here) can be used to discover locally optimal structures, often exhibiting orders-of-magnitude better performance [95, 96] than structures with few parameters design by hand or brute force.

For some target wavelength $\lambda$, we consider metasurfaces with widths (diameters) ranging from 10$\lambda$ to 23$\lambda$, with thicknesses of 1.9$\lambda$, equivalent to films with thicknesses on the order of 1$\mu$m and widths in the dozens of microns for visible-frequency light. We consider two-dimensional scattering (i.e. metasurfaces that are translation-invariant along one dimension), to reduce the...
computational cost and demonstrate the design principle. Dimensionality has only a small effect on the bounds; in the SM, we show that the 2D equivalent of Eq. (14) is

\[ S^2_{\text{max}} = 1 - \frac{4}{\eta_0 \sin(2\eta_0) + 2\eta_0}. \]  

(16)

For the design variables, we allow the permittivity at every point on the metasurface to vary (i.e., “topology optimization” [19, 21]), and we generate two types of designs (depicted in Fig. 5): binary metasurfaces, in which the permittivity must take one of two values (chosen here as 1 and 12), and grayscale metasurfaces, in which the permittivity can vary smoothly between two values as in gradient-index optics [97].

The optical figure of merit that we design to maximize is not exactly the zero-point intensity of Eq. (4), as the zero-field constraint is difficult to implement numerically. Instead, for a desired spot size \( G \), we subtract a constant times the field intensity at the points \( \pm G \) away from the origin:

\[ F = |\psi(0)|^2 - \lambda \left( |\psi(-G)|^2 + |\psi(G)|^2 \right). \]  

(17)

This is a penalty method [98] that can enforce arbitrarily small field intensities (with sufficiently accurate simulations) by increasing the constant \( \lambda \) over the course of the optimization. We take the electric field polarized out of the plane, such that \( \psi \) can be simplified to a scalar field solution of the Helmholtz equation, which we solve by the finite-difference time-domain (FDTD) method [99, 100]. Adjoint-based sensitivities are computed for every structural iteration via two computations: the “direct” fields propagating through the metasurface, and “adjoint” fields that emanate from the maximum-intensity and zero-field locations, with phases and amplitudes of the exciting currents determined by the derivatives of Eq. (17) with respect to the field.

Figure 5 depicts the results of many design optimizations. Figure 5(a) compares the bound of Eq. (16) (black; the nearly identical 3D bound is in grey) to the computed Strehl ratio of unique optimal designs at spot sizes ranging from 0.21 to 1; strikingly, the designed metalenses closely approach the bound for all spot sizes, with the best designs achieving 90% of the maximum possible intensity. In Fig. 5(b,c,d) three specific designs are shown alongside the resulting field profiles in their focal planes. The intensity does not perfectly reach zero but is forced to be significantly smaller than the peak intensity through the penalty constant \( \lambda \) in Eq. (17). It is difficult to explain exactly how the computationally designed metasurface patterns achieve nearly optimal focusing; for spot sizes close to 1, the variations in material density suggest an effective gradient-index-like profile that offers lens-like phase variations across the device width, though the scattering effects of the front and rear surfaces renders such explanations necessarily incomplete. The depicted design with \( G = 0.21 \) exhibits to our knowledge the smallest spot size of any theoretical proposal to date.

IV. SUMMARY AND EXTENSIONS

We have established bounds on the maximal concentration of free-space vector electromagnetic waves. We derive bounds for any desired zero-field contour, either incorporating an aperture as in Eqs. (6,9,14) or dependent only on a modal basis, as in Eqs. (6,12). By a suitable transformation of the light-concentration problem, we obtain analytic bounds in multiple regimes (small spot sizes, far zone, etc.), providing insight into the ideal excitation field as well as revealing a dimension-independent quartic spot-size scaling law. Using inverse design, we have theoretically proposed optimal metasurface designs approaching these bounds.
Looking forward, there are a number of related questions and application areas where this approach can be applied. One natural assumption to probe is the requirement for zero field on a spot size contour. As we show in the SM, however, loosening the constraint such that the average field intensity along the spot-size contour is small, rather than zero, can lead at most to only modest increases in focal intensity. One may also be interested in metrics other than focal-point intensity. One common metric, a minimal-energy metric for a fixed focal-point intensity [18, 85], in fact is equivalent to focal-point maximization metric under constraints of fixed energy, as can be shown by deriving the Lagrangian function for each. Other metrics, potentially focusing on minimizing energy within a specific region (e.g., the field of view), can be seamlessly incorporated into the approach developed here.

We have considered the medium through which light propagates to be free space (or any homogeneous material, which simply modifies the speed of light in the medium), but our results in fact apply directly to any inhomogeneous background by using the corresponding Green’s function in Eq. (6). For near-field imaging with the image plane near some scattering medium, our approach can identify the optimal resolution. It can also potentially be applied to random media [101], where the Green’s function can be appropriately modified through ensemble averaging, to potentially identify optimal concentration within complex disordered media.

Finally, we can expect that the bounds here can be used as a family of potential point-spread functions across imaging applications. Various emerging techniques at the intersection of quantum optics, metrology, and parameter estimation theory [102, 103] suggest the possibility for imaging with resolution improvements beyond the classical Rayleigh limit and Airy disk. A tandem of the subdiffraction-limited point-spread functions provided here with quantum measurement theory may yield even further improvements.

V. ACKNOWLEDGMENTS

We thank Michael Fiddy and Lawrence Domash for helpful conversations. H. S., H. C., and O. D. M. were partially supported by DARPA and Triton Systems under award number 140D6318C0011, and were partially supported by the Air Force Office of Scientific Research under award number FA9550-17-1-0093.

[1] C. W. McCutchen, Superresolution in Microscopy and the Abbe Resolution Limit, Journal of the Optical Society of America 57, 1190 (1967).
[2] E. H. K. Stelzer, Beyond the diffraction limit?, Nature 417, 806 (2002).
[3] N. I. Zheludev, What diffraction limit?, Nature Materials 7, 420 (2008).
[4] M. G. L. Gustafsson, Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution, Proceedings of the National Academy of Sciences 102, 13081 (2005).
[5] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olchvry, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, Imaging Intra-cellular Fluorescent Proteins at Nanometer Resolution, Science 313, 1642 (2006).
[6] M. J. Rust, M. Bates, and X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM), Nature Methods 3, 793 (2006).
[7] S. W. Hell, Far-field optical nanoscopy, Science 316, 1153 (2007).
[8] B. Huang, W. Wang, M. Bates, and X. Zhuang, Three Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy, Science 319, 810 (2008).
[9] H. Lipson and M. Kurman, Fabricated: The new world of 3D printing (John Wiley & Sons, 2013).
[10] H. N. Chia and B. M. Wu, Recent advances in 3D printing of biomaterials, Journal of Biological Engineering 9, 4 (2015).
[11] F. M. Huang and N. I. Zheludev, Super-resolution without evanescent waves, Nano Letters 9, 1249 (2009).
[12] E. T. Rogers, J. Lindberg, T. Roy, S. Savo, J. E. Chad, M. R. Dennis, and N. I. Zheludev, A super-oscillatory lens optical microscope for subwavelength imaging, Nature Materials 11, 432 (2012).
[13] K. Strehi, Uber luftschlieren und zonenfehler, Zeitschrift für Instrumentenkunde 22, 213 (1902).
[14] M. Born and E. Wolf, Principles of optics: electromagnetic theory of propagation, interference and diffraction of light (Elsevier, 2013).
[15] D. P. Bertsekas, Nonlinear programming, 3rd ed. (Athena Scientific, Belmont, MA, 2016).
[16] D. Slepian, Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty — I, Bell System Technical Journal 40, 43 (1961).
[17] H. J. Landau and H. O. Pollak, Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty — II, Bell System Technical Journal 40, 65 (1961).
[18] P. J. Ferreira and A. Kempf, Superscillations: Faster than the Nyquist rate, IEEE Transactions on Signal Processing 54, 3732 (2006).
[19] J. S. Jensen and O. Sigmund, Topology optimization for nano-photons, Laser and Photonics Reviews 5, 308 (2011).
[20] O. D. Miller, Photonic Design: From Fundamental Solar Cell Physics to Computational Inverse Design, Ph.D. thesis, University of California, Berkeley (2012).
[21] M. P. Bendsoe and O. Sigmund, Topology optimization: theory, methods, and applications (Springer Science & Business Media, 2013).
[22] S. Molevary, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, Inverse design in nanophotonics, Nature Photonics 12, 659 (2018).
[23] H. Kawata, J. M. Carter, A. Yen, and H. I. Smith, Optical projection lithography using lenses with numerical apertures greater than unity, Microelectronic Engineering 9, 31 (1989).

[24] T. Ito and S. Okazaki, Pushing the limits of lithography, Nature 406, 1027 (2000).

[25] T. W. Odom, V. R. Thalladi, J. C. Love, and G. M. Whitesides, Generation of 30-50 nm structures using easily fabricated, composite PDMS masks, Journal of the American Chemical Society 124, 12112 (2002).

[26] D. Rugar, Resolution beyond the diffraction limit in the acoustic microscope: A nonlinear effect, Journal of Applied Physics 56, 1338 (1984).

[27] G. Indebetouw, Y. Tada, J. Rosen, and G. Brooker, Scanning holographic microscopy with resolution exceeding the Rayleigh limit of the objective by superposition of off-axis holograms, Applied Optics 46, 993 (2007).

[28] K. Kumar, H. Duan, R. S. Hegde, S. C. Koh, J. N. Wei, and J. K. Yang, Printing colour at the optical diffraction limit, Nature Nanotechnology 7, 557 (2012).

[29] R. Collier, Optical holography (Elsevier, 2013).

[30] N. Fang, H. Lee, C. Sun, and X. Zhang, Sub-Diffraction-Limited Optical Imaging with a Silver Superlens, Science 308, 534 (2005).

[31] R. L. Kostelak, J. S. Weiner, E. Betzig, T. D. Harris, and J. K. Trautman, Breaking the Diffraction Barrier: Optical Microscopy on a Nanometric Scale, Science 251, 1468 (2006).

[32] R. Merlin, Radiationless Electromagnetic Interference: Evanescent-Field Lenses and Perfect Focusing, Science 317, 927 (2007).

[33] X. Zhang and Z. Liu, Superlenses to overcome the diffraction limit, Nature Materials 7, 435 (2008).

[34] G. Ma, X. Fan, F. Ma, J. De Rosny, P. Sheng, and M. Fink, Towards anti-causal Green’s function for three-dimensional sub-diffraction focusing, Nature Physics 14, 608 (2018).

[35] G. Toraldo di Francia, Super-gain antennas and optical resolving power, Il Nuovo Cimento 9, 426 (1952).

[36] S. A. Schelkunoff, A mathematical theory of linear arrays, The Bell System Technical Journal 22, 80 (1943).

[37] M. Berry, Faster than Fourier in quantum coherence and reality Celebration of the 60th Birthday of Yakir Aharonov ed JS Anandan and JL Safko (1994).

[38] M. V. Berry, Evanescent and real waves in quantum billiards and Gaussian beams, Journal of Physics A: Mathematical and General 27, L391 (1994).

[39] M. V. Berry and S. Popescu, Evolution of quantum superoscillations and optical superresolution without evanescent waves, Journal of Physics A: Mathematical and General 39, 6965 (2006).

[40] J. Lindberg, Mathematical concepts of optical superresolution, Journal of Optics 14, 083001 (2012).

[41] A. Kempf and P. J. Ferreira, Unusual properties of superoscillating particles, Journal of Physics A: Mathematical and General 37, 12067 (2004).

[42] Y. Aharonov, F. Colombo, I. Sabadini, D. C. Struppa, and J. Tollaksen, Some mathematical properties of superoscillations, Journal of Physics A: Mathematical and Theoretical 44, 365304 (2011).

[43] A. M. H. Wong and G. V. Eleftheriades, Adaptation of Schelkunoff’s Superdirective Antenna Theory for the Realization of Superoscillatory Antenna Arrays, IEEE Antennas and Wireless Propagation Letters 9, 315 (2010).

[44] I. Chremmos and G. Fikioris, Superscillations with arbitrary polynomial shape, Journal of Physics A: Mathematical and Theoretical 48, 265204 (2015).

[45] M. K. Smith and G. J. Gbur, Construction of arbitrary vortex and superoscillatory fields, Optics Letters 41, 4979 (2016).

[46] I. M. Bassett, Limit to concentration by focusing, Optica Acta 33, 279 (1986).

[47] C. J. R. Sheppard and K. G. Larkin, Optimal concentration of electromagnetic radiation, Journal of Modern Optics 41, 1495 (1994).

[48] T. R. M. Sales and G. M. Morris, Fundamental limits of optical superresolution, Optics Letters 22, 582 (1997).

[49] H. Liu, Y. Yan, Q. Tan, and G. Jin, Theories for the design of diffractive superresolution elements and limits of optical superresolution, Journal of the Optical Society of America A 19, 2185 (2002).

[50] K. S. Rogers, K. N. Bourdakos, G. H. Yuan, S. Mahajan, and E. T. F. Rogers, Optimising superoscillatory spots for far-field super-resolution imaging, Optics Express 26, 8095 (2018).

[51] F. M. Huang, N. Zhedudev, Y. Chen, and F. Javier Garcia De Abajo, Focusing of light by a nanohole array, Applied Physics Letters 90, 091119 (2007).

[52] F. M. Huang, Y. Chen, F. J. Garcia De Abajo, and N. I. Zheludev, Optical super-resolution through superoscillations, Journal of Optics A: Pure and Applied Optics 9, 8285 (2007).

[53] M. R. Dennis, A. C. Hamilton, and J. Courtial, Superoscillation in speckle patterns, 33, 2976 (2008).

[54] X. Wang, J. Fu, X. Liu, and L.-M. Tong, Subwavelength focusing by a micro/nanofiber array, Journal of the Optical Society of America A 26, 1827 (2009).

[55] K. Kitamura, S. Sakai, and S. Noda, Sub-wavelength focal spot with long depth of focus generated by radially polarized, narrow-width annular beam, Optics Express 18, 4518 (2010).

[56] E. T. Rogers, S. Savo, J. Lindberg, T. Roy, M. R. Dennis, and N. I. Zheludev, Super-oscillatory optical needle, Applied Physics Letters 102, 031108 (2013).

[57] L. Huang, X. Chen, H. Mühlenbernd, H. Zhang, S. Chen, B. Bai, Q. Tan, G. Jin, K. W. Cheah, C. W. Qiu, J. Li, T. Zentgraf, and S. Zhang, Three-dimensional optical holography using a plasmonic metasurface, Nature Communications 4, 2808 (2013).

[58] K. Huang, H. Ye, J. Teng, S. P. Yeo, B. Luk’yanchuk, and C. W. Qiu, Optimization-free superoscillatory lens using phase and amplitude masks, Laser and Photonics Reviews 8, 152 (2014).

[59] F. Qin, K. Huang, J. Wu, J. Jiao, X. Luo, C. Qiu, and M. Hong, Shaping a subwavelength needle with ultra-long focal length by focusing azimuthally polarized light, Scientific Reports 5, 9977 (2015).

[60] A. M. Wong and G. V. Eleftheriades, Superoscillations without Sidebands: Power-Efficient Sub-Diffraction Imaging with Propagating Waves, Scientific Reports 5, 8449 (2015).

[61] A. M. Wong and G. V. Eleftheriades, Broadband superoscillation brings a wave into perfect three-dimensional focus, Physical Review B 95, 075148 (2017).

[62] A. V. Kildishev, A. Boltasseva, and V. M. Shalaev, Planar photonics with metasurfaces, Science 339, 1232009.
11

[63] N. Yu and F. Capasso, Flat optics with designer metasurfaces, Nature Materials 13, 139 (2014).

[64] M. Khorasaninejad, W. T. Chen, R. C. Devlin, J. Oh, A. Y. Zhu, and F. Capasso, Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging, Science 352, 1190 (2016).

[65] J. A. Kong, Theory of electromagnetic waves (Wiley-Interscience, New York, 1975).

[66] U. Levy, M. Abashin, K. Ikeda, A. Krishnamoorthy, J. Cunningham, and Y. Fainman, Inhomogeneous dielectric metamaterials with surface-variant polarizability, Physical Review Letters 98, 243901 (2007).

[67] Z. Wei, Y. Long, Z. Gong, H. Li, X. Su, and Y. Cao, Highly efficient beam steering with a transparent metasurface, Optics Express 21, 10739 (2013).

[68] S. Keren-Zur, O. Avayu, L. Michaeli, and T. Ellenbogen, Nonlinear Beam Shaping with Plasmonic Metasurfaces, ACS Photonics 3, 117 (2016).

[69] A. M. Weiner, Femtosecond pulse shaping using spatial light modulators, Review of Scientific Instruments 71, 1929 (2000).

[70] N. Chattrapiban, E. A. Rogers, D. Cofield, W. T. Hill, III, and R. Roy, Generation of nondiffracting Bessel beams by use of a spatial light modulator, Optics Letters 28, 2183 (2003).

[71] C.-S. Guo, X.-L. Wang, W.-J. Ni, H.-T. Wang, and J. Ding, Generation of arbitrary vector beams with a spatial light modulator and a common path interferometric arrangement, Optics Letters 32, 3549 (2007).

[72] L. Zhu and J. Wang, Arbitrary manipulation of spatial amplitude and phase using phase-only spatial light modulators, Scientific Reports 4, 7441 (2014).

[73] P. Lodahl, A. F. Van Driel, I. S. Nikolaev, A. Irman, K. Overgaag, D. Vanmaekelbergh, and W. L. Vos, Controlling the dynamics of spontaneous emission from quantum dots by photonic crystals, Nature 430, 654 (2004).

[74] M. Ringler, A. Schwemer, M. Wunderlich, A. Nichit, K. Kürzinger, T. A. Klar, and J. Feldmann, Shaping Emission Spectra of Fluorescent Molecules with Single Plasmonic Nanoresonators, Physical Review Letters 100, 203002 (2008).

[75] J. Bleuse, J. Claudon, M. Creasey, N. S. Malik, J.-M. Gérard, I. Maksymov, J.-P. Hugonin, and P. Lalanne, Inhibition, Enhancement, and Control of Spontaneous Emission in Photonic Nanowires, Physical Review Letters 106, 103601 (2011).

[76] W. C. Chew, Waves and fields in inhomogeneous media (IEEE press, 1995).

[77] J.-M. Jin, Theory and computation of electromagnetic fields (John Wiley & Sons, 2011).

[78] S. Boyd and L. Vandenberghe, Convex optimization (Cambridge university press, 2004).

[79] R. A. Horn and C. R. Johnson, Matrix analysis, 2nd ed. (Cambridge university press, 2013).

[80] L. N. Trefethen and D. Bau III, Numerical linear algebra (SIAM, 1997).

[81] J. P. Boyd, Chebyshev and Fourier spectral methods, 2nd ed. (Dover, New York, 2001).

[82] U. Levy, S. Derevyanko, and Y. Silberberg, Light Modes of Free Space, Progress in Optics 61, 237 (2016).

[83] M. Mazilu, J. Baungartl, S. Kosmeier, and K. Dholakia, Optical Eigenmodes: exploiting the quadratic nature of the light-matter interaction, Optics Express 19, 933 (2011).

[84] S. Kosmeier, M. Mazilu, J. Baungartl, and K. Dholakia, Enhanced two-point resolution using optical eigenmode optimized pupil functions, Journal of Optics 13, 105707 (2011).

[85] L. Levi, Fitting a bandlimited signal to given points, IEEE Transactions on Information Theory 11, 372 (1965).

[86] D. Slepian, Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty—V: The Discrete Case, Bell System Technical Journal 57, 1371 (1978).

[87] P. N. Gundu, E. Hack, and P. Rastogi, ‘Apodized superresolution’ - Concept and simulations, Optics Communications 249, 101 (2005).

[88] D. M. de Juana, J. E. Oti, V. F. Canales, and M. P. Cagigal, Design of superresolving continuous phase filters, Optics Letters 28, 607 (2003).

[89] F. Qin, K. Huang, J. Wu, J. Teng, C. W. Qin, and M. Hong, A Supercritical Lens Optical Label-Free Microscopy: Sub-Diffraction Resolution and Ultra-Long Working Distance, Advanced Materials 29, 1602721 (2017).

[90] X. H. Dong, A. M. H. Wong, M. Kim, and G. V. Eleftheriades, Superresolution far-field imaging of complex objects using reduced superoscillating ripples, Optica 4, 1126 (2017).

[91] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, Nature 323, 533 (1986).

[92] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation 1, 541 (1989).

[93] P. J. Werbos, The Roots of Backpropagation (John Wiley & Sons, Inc., 1994).

[94] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd ed. (Springer, New York, NY, 2006).

[95] C. M. Lalau-Keraiy, S. Bhargava, O. D. Miller, and E. Yablonovitch, Adjoint shape optimization applied to electromagnetic design, Optics Express 21, 21693 (2013).

[96] V. Ganapati, O. D. Miller, and E. Yablonovitch, Light trapping textures designed by electromagnetic optimization for subwavelength thick solar cells, IEEE Journal of Photovoltaics 4, 175 (2014).

[97] E. Marchand, Gradient index optics (Academic, New York, 1978).

[98] D. P. Bertsekas, Constrained optimization and Lagrange multiplier methods (Academic press, 2014).

[99] A. Taflove, A. Oskooi, and S. G. Johnson, Advances in FDTD computational electrodynamics: photons and nanotechnology (Artech house, 2013).

[100] A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson, Meep: A flexible free-software package for electromagnetic simulations by the FDTD method, Computer Physics Communications 181, 687 (2010).

[101] A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, Controlling waves in space and time for imaging and focusing in complex media, Nature Photonics 6, 283 (2012).

[102] M. Tsang, R. Nair, and X. M. Lu, Quantum theory of superresolution for two incoherent optical point sources,
Physical Review X 6, 031033 (2016).

[103] S. Zhou and L. Jiang, Modern description of Rayleigh’s criterion, Physical Review A 99, 013808 (2019).
I. MODAL-DECOMPOSITION BOUNDS

In this section we expand on the modal-decomposition bounds discussed in the main text, and we derive the analytical bounds for a planar decomposition. We start with the same optimization problem as in the main text,

\[
\max_{\mu, \psi} |\mu |^2 \psi(x_c)|^2
\]
subject to
\[
\psi(x_c) = 0,
\]
\[
\int |\psi(x)|^2 \leq 1,
\]  
(1)

where \( \mu \) is a unit six-vector oriented along the direction of optimal polarization.

A. General Bounds

Now we want to consider the case in which the field \( \psi \) is expanded in some modal basis. We will index the modal basis with a continuous variable \( k \), which is the Fourier wavevector in a 2D plane or 3D spherical surface, and a discrete variable \( i \), which can correspond e.g. to a polarization index, in which case the field can be decomposed into modes \( u_i(k; x) \) with coefficients \( \tilde{\psi}_i(k) \):

\[
\psi(x) = \sum_i \int \tilde{\psi}_i(k) u_i(k; x) \, dk.
\]  
(2)
Similarly, we assume the modes are normalized such that
\[ \int u_i^\dagger(k; x) u_j(k'; x) \, dx = \delta(k - k') \delta_{ij}, \] (3)
in which case it is straightforward to integrate \( u_j(k'; x) \) against Eq. (2) to determine the modal coefficients \( \tilde{\psi}_i(k) \):
\[ \tilde{\psi}_i(k) = \int u_i^\dagger(k; x) \psi(x) \, dx. \] (4)

To enforce the zero-field condition in Eq. (1), we will follow the same procedure as in the main text, enforcing the condition against some discrete set of basis functions (a different set that are defined only on the contour) \( \phi_i(x) \). To derive an analytical bound we will require only that the field be zero as measured against the first basis function, i.e.,
\[ \int_{C} \phi_0^\dagger(x) \psi(x) \, dx = 0. \] (5)
If we replace \( \psi \) with its modal decomposition, we can rewrite this zero-field condition as
\[ \sum_i \int dk \left[ \int_{C} dx \phi_0^\dagger(x) u_i(k; x) \right] \tilde{\psi}_i(k) = 0. \] (6)
The inner integral is the conjugate transpose of the inverse decomposition of \( \phi_0 \), and is thus the conjugate transpose of the \( i \)-polarized modal coefficients of \( \phi_0 \), which we will call \( \tilde{\phi}_0, i \), in analogy with Eq. (4). Then the zero-field condition is simply \( \sum_i \int \tilde{\phi}_0, i(k) \tilde{\psi}_i(k) \, dk = 0 \). Making the modal-decomposition replacement everywhere in Eq. (1), it can be written:
\[
\begin{align*}
\text{maximize} & \quad \left| \sum_i \int \tilde{\psi}_i(k) u_i(k; x_0) \, dk \right|^2 \\
\text{subject to} & \quad \sum_i \int dk \tilde{\phi}_0, i(k) \tilde{\psi}_i(k) = 0, \\
& \quad \int \left| \tilde{\psi}(k) \right|^2 \, dk \leq 1,
\end{align*}
\] (7)
where in the zero-field constraint we take \( \phi_0 = 0 \) for points not on the zero-field contour \( C \). Just as in the main text, we can simplify the optimization problem by defining two scalar fields
\begin{align*}
\psi_0, i(k) &= \mu^T u_i^*(k; x_0) \\
\psi_1(k) &= \tilde{\phi}_0, i(k),
\end{align*}
(8)
where an asterisk denotes complex conjugation. Then, by any standard discretization scheme, we can write Eq. (7) as the finite-dimensional problem
\[
\begin{align*}
\text{maximize} & \quad \left| \psi_0^\dagger \psi_0 \right|^2 \\
\text{subject to} & \quad \psi_1^\dagger \tilde{\psi} = 0, \\
& \quad \tilde{\psi}^\dagger \tilde{\psi} \leq 1.
\end{align*}
\] (9)
As shown in the main text, this optimization problem is bounded above by the value,
\[ I \leq \psi_0^\dagger \psi_0 - \frac{\left| \psi_1^\dagger \psi_0 \right|^2}{\psi_1^\dagger \psi_1}. \] (10)

B. Planar bounds

The prototypical scenario in which one would want to apply the modal-decomposition bounds is to an optical beam in which light is focused at a single point, with a spot size defined by a circle (ring) in some transverse plane, and
with the modes comprising all propagating plane waves. Then the modes are:

\[ u_m(k; x) = \frac{\hat{\varepsilon}_m}{2\pi} e^{i\varepsilon_m \cdot x} \]  

(11)

where we use \( k \) to denote the wavevector of each plane wave, and \( m \) is the modal index that additionally accounts for the two polarizations \( \hat{\varepsilon}_m \) per wavevector. The modal fields of Eq. (11) incorporate both the electric and magnetic fields as long as \( \hat{\varepsilon}_m \) is a six-vector, and given the \( 1/2\pi \) prefactor they satisfy the orthogonality relation of Eq. (3). (Note that the \( 1/2\pi \) prefactor is really the square of two \( 1/\sqrt{2\pi} \) prefactors in the unitary angular-frequency Fourier transforms in two dimensions.) Given the modes, we just need to compute \( \psi_0(k) \) and \( \psi_1(k) \) from Eq. (8), from which the intensity bound is given by Eq. (10). The field at the origin (max-intensity point) is

\[ \psi_{0,i}(k) = \frac{\hat{\varepsilon}_i^1}{2\pi} \mu. \]  

(12)

Given the electric/magnetic symmetry in the modal fields there is no advantage to maximizing the electric or magnetic fields or some linear combination, and thus we can choose the unit vector \( \mu \) arbitrarily to align with the electric field along a nominal \( \hat{x} \) direction. We will introduce the six-vector \( \hat{x} = (\hat{x}, 0) \), giving

\[ \psi_{0,i}(k) = \frac{\hat{\varepsilon}_i^1 \hat{x}}{2\pi}. \]  

(13)

Meanwhile, \( \psi_1 \) is given by the Fourier transform of the function \( \psi_0 \) that is a basis function for the zero-field condition. In a plane, one would want zero field on a circle of radius \( R \), in which case a sensible basis against which to measure the field on the circle would be the complex exponentials, i.e.

\[ \phi_n(x) = \hat{\varepsilon}_n e^{in\theta} \]  

(14)

where \( \theta \) is the angle at a location on the circle, \( n \) is a non-negative integer, and \( \hat{\varepsilon}_n \) is a polarization six-vector. (Note that technically \( \phi \) should have two indices, one for the radial variations and one for polarization, but we will not include that since we will only consider a single basis function.) Here we will choose the same polarization for the first basis function, i.e., \( \phi_0(x) = \hat{x} \). Then we can write \( \psi_1 \) as

\[ \psi_1(k) = \tilde{\phi}_0(k) \]  

\[ = \int u_i^1(k; x) \phi_0(x) \, dx \]  

\[ = \frac{\hat{\varepsilon}_i^1 \hat{x}}{2\pi} \int_{r=R} e^{-ik \cdot x} \, dx. \]  

(15)

The integral in Eq. (15) can be simplified by transforming both \( x \) and \( k \) to polar coordinates: \( x = r \cos \theta, y = r \sin \theta, k_x = \rho \cos \theta', \) and \( k_y = \rho \sin \theta' \). Then the integral is

\[ R \int_0^{2\pi} e^{iR\rho (\cos \theta \cos \theta' + \sin \theta \sin \theta')} \, d\theta = R \int_0^{2\pi} e^{iR\rho \cos \theta} \, d\theta \]  

\[ = 2\pi R J_0(R\rho), \]  

(16)

where \( J_0 \) is the zeroth-order Bessel function, and the second equality follows because the endpoints of an integral can be shifted arbitrarily when integrating a periodic function over its period. The field \( \psi_1 \) is the two-dimensional Fourier transform of a ring, which is the zeroth-order Bessel function. The polarization term \( \hat{\varepsilon}_i^1 \hat{x} \) is slightly more complicated and we will not solve for it just yet, leaving

\[ \psi_1(k) = (\hat{\varepsilon}_i^1 \hat{x}) R J_0(R\rho). \]  

(17)

To determine the bound, we need to derive three quantities: \( \psi_0^\dagger \psi_0, \psi_0^\dagger \psi_1, \) and \( \psi_1^\dagger \psi_1 \). First is \( \psi_0^\dagger \psi_0 \):

\[ \psi_0^\dagger \psi_0 = \sum_i \int |\psi_{0,i}(k)|^2 \, dk \]  

\[ = \frac{1}{4\pi^2} \int \int \left[ \sum_i (\hat{\varepsilon}_i^1 \hat{x})^2 \right] \, dk_x dk_y \]
To determine the polarization-dependent term, we can use the fact that the two electric-field polarizations for every wavevector are orthogonal to the wavevector itself. Thus, \((\mathbf{k} \cdot \hat{x})^2 + \sum_i (\mathbf{e}_i^\dagger \hat{x})^2 = 1\), since the total \(x\)-component of three orthogonal unit vectors must 1, independent of their orientation. Rearranging, we have that \(\sum_i (\mathbf{e}_i^\dagger \hat{x})^2 = 1 - (\mathbf{k} \cdot \hat{x})^2\), and we can simplify the integral

\[
\int \int \left[ 1 - \frac{k_z^2}{k^2} \right] dk_x dk_y = \int_0^k \int_0^{2\pi} \left[ 1 - \frac{\rho^2 \cos^2 \theta}{k^2} \right] \rho \, d\rho \, d\theta.
\]

Thus we have

\[
\psi_0^\dagger \psi_0 = \frac{3k^2}{16\pi}.
\]

Next, we determine the value of \(\psi_1^\dagger \psi_1\). This term has precisely the same \(\sum_i (\mathbf{e}_i^\dagger \hat{x})^2\) term in it, so we will use the same procedure as above to simplify the integral to:

\[
\psi_1^\dagger \psi_1 = \frac{R}{2\pi} \int_0^k \int_0^{2\pi} \rho J_0(R\rho) \left[ 1 - \frac{\rho^2 \sin^2 \theta}{k^2} \right] \, d\rho \, d\theta.
\]

The first integral is given by \(\int \int \rho J_0(R\rho) = 2\pi k J_1(kR)/R\). The second term is given by \(\int \rho^3 J_0(R\rho) = (\pi k^2 / R^2) [2J_2(kR) - kR J_3(kR)]\). Hence, the total term is given by

\[
\psi_1^\dagger \psi_1 = k \left[ \frac{J_1(kR)}{2} - \frac{J_2(kR)}{kR} + \frac{1}{2} J_3(kR) \right].
\]

We can use the recurrence relation \(J_3(x) = 4J_2(x)/2 - J_1(x)\) to rewrite this term as

\[
\psi_1^\dagger \psi_1 = \frac{R^2}{2\pi} \int_0^k \int_0^{2\pi} \rho J_0(R\rho)^2 \left[ 1 - \frac{\rho^2 \sin^2 \theta}{k^2} \right] \, d\rho \, d\theta.
\]

Finally, we determine the value of \(\psi_1^\dagger \psi_1\), which again has the same polarization term, and is given by

\[
\psi_1^\dagger \psi_1 = \pi k^2 R^2 \left\{ [J_0(kR)]^2 + [J_1(kR)]^2 \right\}
\]

The first integral is given by

\[
2\pi \int_0^k R\rho [J_0(R\rho)]^2 \, d(R\rho) = \pi k^2 R^2 \left\{ [J_0(kR)]^2 + [J_1(kR)]^2 \right\}
\]

The second integral is given by

\[
\frac{\pi}{R^2} \int_0^{kR} v^3 [J_0(v)]^2 \, dv = \pi k^4 R^2 \left\{ \frac{1}{6} [J_0(kR)]^2 + \frac{1}{3} \frac{J_0(kR) J_1(kR)}{kR} + \frac{1}{6} \left( 1 - \frac{2}{k^2 R^2} \right) [J_1(kR)]^2 \right\}
\]

Hence \(\psi_1^\dagger \psi_1\) is given by

\[
\psi_1^\dagger \psi_1 = \pi k^2 R^2 \left\{ \frac{5}{6} [J_0(kR)]^2 + \frac{5}{6} [J_1(kR)]^2 - \frac{J_0(kR) J_1(kR)}{3kR} + \frac{[J_1(kR)]^2}{3k^2 R^2} \right\}.
\]

Finally, we can write the bound:

\[
I \leq \frac{3k^2}{16\pi} - \frac{1}{\pi R^2} \left\{ \frac{5}{6} [J_0(kR)]^2 + \frac{5}{6} [J_1(kR)]^2 - \frac{J_0(kR) J_1(kR)}{3kR} + \frac{[J_1(kR)]^2}{3k^2 R^2} \right\}.
\]
For small spot sizes $R$, i.e. $kR \ll 1$, a Taylor expansion of the bound gives

$$I \leq \frac{13}{13824\pi} k^6 R^4 \quad kR \ll 1.$$  \hfill (28)

We can divide this bound by $I_0 = \psi_0^\dagger \psi_0$ to get a bound normalized by the unconstrained optimal value,

$$\frac{I}{I_0} \leq \frac{13}{2592} k^4 R^4 \quad kR \ll 1.$$  \hfill (29)

Finally, the one assumption we made to ultimately get to Eqs. (27)–(29) was that the optimal polarization is in the $\hat{x}$ direction. By symmetry, there is no difference between $\hat{x}$, $\hat{y}$, or any linear combination thereof. The perpendicular polarization, $\hat{z}$, would intuitively reduce performance (losing e.g. normal-incidence waves from the basis), but we can verify this intuition. The first term, $\psi_0^\dagger \psi_0$, is now given by

$$\psi_0^\dagger \psi_0 = \frac{1}{4\pi^2} \int \int \left[ \sum_i (\epsilon^\dagger_i \epsilon_i)^2 \right] \, dk_x \, dk_y = \frac{1}{4\pi^2} \int \int \left[ 1 - \frac{k_z^2}{k^2} \right] \, dk_x \, dk_y = \frac{1}{2\pi k^2} \int_0^k \rho^3 \, d\rho = \frac{k^2}{8\pi}.$$  \hfill (30)

The second term, $\psi_1^\dagger \psi_1$, is given by

$$\psi_0^\dagger \psi_1 = \frac{R}{k^2} \int_0^k \rho^3 J_0(R\rho) \, d\rho = \frac{1}{R} \left[ 2J_2(kR) - kR J_3(kR) \right].$$  \hfill (31)

Lastly, the third term is given by

$$\psi_1^\dagger \psi_1 = \frac{2\pi R^2}{k^2} \int_0^k \rho^3 [J_0(R\rho)]^2 \, d\rho \, d\theta = \frac{\pi}{3} \left[ (kR)^2 J_0(kR)^2 + 2(kR)J_0(kR)J_1(kR) + (k^2 R^2 - 2)J_1(kR)^2 \right].$$  \hfill (32)

Putting these terms together, the $z$-polarized bound is

$$I \leq \frac{k^2}{8\pi} - \frac{3}{\pi R^2} \frac{[2J_2(kR) - (kR)J_3(kR)]^2}{(kR)^2 J_0(kR)^2 + 2(kR)J_0(kR)J_1(kR) + (k^2 R^2 - 2)J_1(kR)^2}.$$  \hfill (33)

For small $kR$, the bound simplifies to

$$I \leq \frac{k^6 R^4}{2304\pi},$$  \hfill (34)

which is smaller than the equivalent bound for the $x$-polarization, Eq. (28). Indeed, one can verify computationally that the general bound for the $z$-polarization, Eq. (33), is always smaller than the equivalent bound for $x$-polarization, Eq. (27), generally by about a factor of 2. Hence, the transverse polarization is optimal, and the bounds of Eqs. (27)–(29) are the global intensity bounds.

### C. 1D Line: Equivalence to PSWF Approach

Consider a two-dimensional scalar beam-propagation problem in which one wants to concentrate light as tightly as possible on a cross-sectional line. Then the general approach from the main text still applies, with Eqs. (27)–(28) of this SM still providing bounds; to derive the 1D bounds, we simply need to identify $\psi_0$ and $\psi_1$. (There is no longer an $i$ index since it is a scalar problem.) The modal fields are identical to those of the three-dimensional case, but now without any polarization factor and with only one factor of $1/\sqrt{2\pi}$ due to the single dimension:

$$u(k; x) = \frac{1}{\sqrt{2\pi}} e^{ikx},$$  \hfill (35)

leading to $\psi_0$ and $\psi_1$ values of

$$\psi_0(k) = \frac{1}{\sqrt{2\pi}}$$  \hfill (36)

$$\psi_1(k) = \frac{1}{2} \int u^*(k; x) [\delta(x - R) + \delta(x + R)] \, dx = \frac{\cos(kR)}{\sqrt{2\pi}}.$$  \hfill (37)
Then the three bound terms are
\[ \psi_0^\dagger \psi_0 = \frac{k}{\pi} \int_{-k}^{k} |\psi_0|^2 \, dk \] (38)
\[ \psi_0^\dagger \psi_1 = \frac{1}{2\pi} \int_{-k}^{k} \cos(kR) \, dk = \frac{\sin(kR)}{\pi R} \] (39)
\[ \psi_1^\dagger \psi_1 = \frac{1}{2\pi} \int_{-k}^{k} \cos^2(kR) \, dk = \frac{k}{2\pi} \left[ 1 + \frac{\sin(2kR)}{2kR} \right] \] (40)

Inserting these three terms into the bound of Eq. (10), we have
\[ I \leq \frac{k}{\pi} \left[ 1 - 2 \frac{\operatorname{sinc}^2(kR)}{1 + \operatorname{sinc}(2kR)} \right] \] (41)

If we denote the unconstrained maximum intensity as \( I_0 \), then
\[ I_0 = \frac{k}{\pi} \] (42)

and the ratio of \( I_{\text{max}} \) to \( I_0 \) is
\[ \frac{I_{\text{max}}}{I_0} = 1 - 2 \frac{\operatorname{sinc}^2(kR)}{1 + \operatorname{sinc}(2kR)} \] (43)

We can show that these bounds, Eqs. (41,43), are equivalent to those that would be derived by the classical prolate-spheroidal-wave-function approach that dates to the work of Slepian, Polak, and others [1, 2]. Apparently the most general bounds for the scalar 1D problem are those derived in Ref. [3], although they primarily tackle the case in which one wants a signal that varies from +1 to -1 over some arbitrarily small distance. Here we will apply their analysis to the imaging problem, in which one wants a large intensity at the origin and then zeros at symmetric points around the focus. In following their approach (and those of earlier authors [4, 5]), instead of maximizing the intensity at the origin, we will find the signal with minimum total energy such that the field amplitude at the origin equals one. (The two problems are equivalent, with solutions that are inverses of each other.)

The statement of the problem is as follows:
\[ \min_{\tilde{\psi}(k)} \int |\tilde{\psi}(k)|^2 \, dk \] subject to \[ \psi(-R) = 0 \]
\[ \psi(0) = 1 \]
\[ \psi(R) = 0 \].

We can write this in a more general form, inserting the specific values later, as
\[ \min_{\tilde{\psi}(k)} \int |\tilde{\psi}(k)|^2 \, dk \] subject to \[ \psi(x_i) = \psi_i, \quad i = 0, 1, 2 \]

where \( x_0 = -R, x_1 = 0, x_2 = R \), and \( \psi_0 = 0, \psi_1 = 1, \) and \( \psi_2 = 0 \). The field constraint can be written
\[ \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(k)e^{ikx_i} \, dk = \psi_i. \] (46)

The Lagrangian for this problem is
\[ \mathcal{L} = \int |\tilde{\psi}(k)|^2 \, dk + \sum_i \lambda_i \left[ \operatorname{Re} \left( \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(k)e^{ikx_i} \, dk \right) - \psi_i \right], \] (47)

where the \( \lambda_i \) are the Lagrange multipliers, and we can drop the constraints on the imaginary part of \( \tilde{\psi}(k) \), which simply go to zero in the minimization. Then the extremal condition can be found either by independently varying
the real and imaginary parts of $\tilde{\psi}(k)$, or, more conveniently, by using the CR calculus [6] and formally treating $\tilde{\psi}(k)$ and $\tilde{\psi}^*(k)$ as independent variables. In the latter case, we can only look at the variations in $\tilde{\psi}^*(k)$, which give

$$\frac{\partial L}{\partial \tilde{\psi}^*(k)} = \tilde{\psi}(k) + \sum_i \lambda_i \frac{1}{2\sqrt{2\pi}} e^{-ikx_i} = 0. \quad (48)$$

Thus we have

$$\tilde{\psi}(k) = \sum_i \lambda_i e^{-ikx_i}, \quad (49)$$

where we use the freedom in the Lagrange multipliers to absorb the negative sign and constant prefactor. We can find the (renormalized) Lagrange multipliers by inserting this expression into the field constraint, Eq. (46), giving

$$\frac{1}{\sqrt{2\pi}} \sum_j \lambda_j \int e^{ik(x_i - x_j)} dk = \psi_i. \quad (50)$$

A straightforward calculation shows that $\int e^{ik(x_i - x_j)} = 2k \text{sinc}(k(x_i - x_j))$, such that we can rewrite this expression as

$$\sum_j \underbrace{\text{sinc}(k(x_i - x_j))}_M \lambda_j = \sqrt{\frac{\pi}{2k}} \psi_i, \quad (51)$$

where we have defined the matrix $M$ and vector $a$ such that this can be succinctly written

$$M \lambda = a. \quad (52)$$

The Lagrange multipliers are the solution of this $3 \times 3$ matrix equation, which then define the optimal $\tilde{\psi}(k)$ by Eq. (49). Note that $M$ is precisely the matrix that defines the eigenequation for the discrete prolate spheroidal wave functions [2, 3]. The total energy in the minimum-energy solution is then given by

$$\int |\tilde{\psi}(k)|^2 dk = \sum_{ij} \lambda_i \lambda_j \int e^{ik(x_i - x_j)} dk$$

$$= 2k \sum_{ij} \lambda_i \lambda_j \text{sinc}(k(x_i - x_j))$$

$$= 2k X^T M \lambda$$

$$= 2k a^T M^{-1} a. \quad (53)$$

Now we can use the fact that only the second element of $a$ is nonzero (since $\psi_0$ and $\psi_2$ are 0, to write

$$\int |\tilde{\psi}(k)|^2 dk = \frac{\pi}{k} (M^{-1})_{22}, \quad (54)$$

where $(M^{-1})_{22}$ refers to the second row and second column of the matrix $M^{-1}$. Since we know the $x_i$ values are $-R, 0, R$, the matrix $M$ is given by

$$M = \begin{pmatrix} 1 & \text{sinc}(kR) & \text{sinc}(2kR) \\ \text{sinc}(kR) & 1 & \text{sinc}(2kR) \\ \text{sinc}(2kR) & \text{sinc}(kR) & 1 \end{pmatrix}. \quad (55)$$

The value of $(M^{-1})_{22}$ is

$$(M^{-1})_{22} = \frac{1}{|M|} \begin{vmatrix} M_{11} & M_{13} \\ M_{31} & M_{33} \end{vmatrix}. \quad (56)$$

The two determinants are given by

$$|M| = 1 + 2 \text{sinc}^2(kR) [\text{sinc}(2kR) - 1] - \text{sinc}^2(2kR), \quad (57)$$
The ratio of Eq. (58) to Eq. (57) can be shown to be
\[
1 - 2 \frac{\sin^2(kR)}{1 + \sin(2kR)}.
\]
Hence the minimum energy is given by exactly this expression multiplied by \( \pi/k \), as seen from Eq. (54). This is the minimum energy for a unit amplitude; by contrast, the maximum amplitude for a fixed energy (equal to 1) is given by the inverse of the minimum-energy solution (as can be shown by rescaling). Thus the PSWF approach gives a maximum amplitude of
\[
I \leq \frac{k}{\pi} \left[ 1 - \frac{\sin^2(kR)}{1 + \sin(2kR)} \right],
\]
which is exactly the same as that predicted by our approach.
Hence, we have shown that our approach indeed reduces exactly to the PSWF-based Fourier-analysis bound in the 1D scalar limit.

II. APERTURE-DEPENDENT BOUNDS

A. Optimal solutions

Starting from the rank-two optimization problem given in the main text,

\[
\begin{align*}
\text{maximize} & \quad \nu^\dag P \gamma \mu \gamma \mu^\dag P \nu \\
\text{subject to} & \quad \nu^\dag P \nu \leq 1,
\end{align*}
\]

(61)

where \( \mu \) is a six-component polarization vector and \( P \) is a projection matrix, which is Hermitian and idempotent (\( PP = P \)) such that \( \xi = \left( I - \Gamma C \left( \Gamma C \Gamma^\dag C \right)^{-1} \Gamma C \right) \nu = P \nu \). Equation (61) is equivalent to a Rayleigh-quotient optimization, which is maximized by the largest eigenvalue of the generalized eigenproblem

\[
P \gamma \mu \gamma \mu^\dag P \nu = \lambda P \nu,
\]

(62)

whose eigenvector is given by \( \nu = P \gamma \mu / \| P \gamma \mu \| \) (satisfying the constraint in Eq. (61) with largest magnitude), with the nonzero eigenvalue \( \gamma_\mu \mu \gamma \mu = \gamma \mu \left( 1 - \Gamma C \left( \Gamma C \Gamma^\dag C \right)^{-1} \Gamma C \right) \gamma \mu \), using the property that \( P \nu = \nu \). It is guaranteed to be the only nonzero eigenvalue, since \( P \gamma \mu \gamma \mu^\dag P \) is a rank one matrix, and any other eigenvectors would have to be orthogonal to \( P \gamma \mu \) such that the matrix times the eigenvector is necessarily zero. Using the definition \( \gamma \mu = \Gamma^\dag 0 \mu \), we can write the upper bound to intensity \( I \) in terms of the dyadic Green’s functions \( \Gamma \) [7]:

\[
I \leq \mu^\dag \left[ \Gamma 0 \Gamma^\dag 0 - \Gamma 0 \Gamma^\dag C \left( \Gamma C \Gamma^\dag C \right)^{-1} \Gamma C \Gamma^\dag 0 \right] \mu.
\]

(63)

If we loosen the zero-field constraint and instead constrain only the zeroth-order, \( \mu \)-polarized mode (\( \phi^\dag 0 \Gamma C \xi = 0 \) where \( \phi_0 \) is a 6-vector), we can simplify the upper bound on intensity. By replacing \( \Gamma C \rightarrow \phi^\dag 0 \Gamma C \), Eq. (63) reduces to

\[
I \leq \mu^\dag \left[ \Gamma 0 \Gamma^\dag 0 - \Gamma 0 \Gamma^\dag C \left( \phi^\dag 0 \Gamma C \phi_0 \right)^{-1} \phi^\dag 0 \Gamma C \Gamma^\dag 0 \right] \mu.
\]

(64)

Defining \( \psi_0 = \gamma \mu = \Gamma^\dag 0 \mu \) and \( \psi_1 = \Gamma^\dag C \phi_0 \), Eq. (64) reduces to an expression in terms of only the two new fields \( \psi_0 \) and \( \psi_1 \),

\[
I \leq \psi_0^\dag \psi_0 - \frac{|\psi_0^\dag \psi_1|^2}{\psi_0^\dag \psi_0}.
\]

(65)
To find the Strehl ratio, we need to divide Eq. (65) by the diffraction-limited (i.e. no zero-field constraint) intensity, which is simply $\psi_0^\dagger \psi_0$ (corresponding to the projection matrix $P$ being an identity matrix). Hence the upper bound to the Strehl ratio, which we denote $S_{\text{max}}$, is

$$S_{\text{max}} = 1 - \frac{|\psi_0^\dagger \psi_1|^2}{\psi_0^\dagger \psi_0 \psi_1^\dagger \psi_1}. \quad (66)$$

The (normalized) optimal effective current is $\xi_{\text{opt}} = P_\nu = P \psi_0 / \| P \psi_0 \|$ (where $P = I - \psi_1^\dagger \psi_1 / \psi_1^\dagger \psi_1$),

$$\xi_{\text{opt}} = \xi_0 \left[ \psi_0 - \frac{\psi_0^\dagger \psi_1}{\psi_1^\dagger \psi_1} \psi_1 \right]. \quad (67)$$

where $\xi_0 = 1/\sqrt{\psi_0^\dagger \psi_0 - |\psi_0^\dagger \psi_1|^2 / \psi_1^\dagger \psi_1}$ such that $\xi_0^\dagger \xi_0 = 1$. (the same analysis holds for the general bound in Eq. (63), but for simplicity we have shown the Strehl ratio bound and normalized optimal currents for the simpler bound of Eq. (65)).

B. Far-zone asymptotic solutions

Encoding the radiation from electric and magnetic currents in free space, the dyadic Green’s function $\Gamma$ [7] simplifies considerably in the far zone—the aperture and focal spot sizes are negligible compared to the distance between aperture and focal spot, such that the distance between any point on the spot and aperture $\approx z$.

Recalling the definition of $\psi_0$ and $\psi_1$ from the main text and adopting a Fourier basis (with coordinates on the aperture denoted by $x'$),

$$\psi_0 = \Gamma^\dagger(x = 0, x') \mu, \quad \psi_1 = \int_{\text{ring}} \Gamma^\dagger(x, x') \mu \quad (68)$$

In the far zone, the aperture-focus distance $z$ greatly exceeds the physical dimensions of the aperture and focal spot, such that $\psi_0$ can be written as

$$\psi_0 = \left( \frac{ke^{ikz}}{4\pi iz} \right)^* \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \mu = \frac{Ae^{-ikz}}{z} \tilde{\mu} \quad (69)$$

where $A \equiv \frac{4k}{\pi c}$ is fixed for a given wavelength and $\tilde{\mu} \equiv K \mu$ encoding the directions of $\psi_0$ and $\psi_1$ (same direction as $\psi_0$ in far zone).

To determine $\psi_1$, we need to keep the correction term from aperture and focal spot size in the exponent, since even small changes in the exponent can affect the field significantly. Denoting the coordinate on the aperture and spot-size ring as $(x', y')$ and $(x, y)$ respectively (with respective radius of $r$ and $\rho_0$), the exact distance between the two points is \( \sqrt{(x - x')^2 + (y - y')^2 + z^2} = z \sqrt{1 - 2 \frac{xx' + yy'}{zz'} + \frac{r_0^2 + r^2}{zz'}} \). Inserting this expression in the exponent appearing in $\psi_1$,

$$\psi_1 = \frac{A}{z} \tilde{\mu} \int_{\text{ring}} e^{-ikz \sqrt{1 - 2 \frac{xx' + yy'}{zz'} + \frac{r_0^2 + r^2}{zz'}}} \ d \phi. \quad (70)$$

In the far zone, we can neglect the last quadratic term in the exponent above. Since $z \gg \rho_0, r$, we can Taylor expand the exponential, $\sqrt{1 - 2 \frac{xx' + yy'}{zz'} \approx 1 - \frac{xx' + yy'}{zz'}}$. Then, we switch to polar coordinates with the angle on the aperture and focal spot defined as $\theta$ and $\phi$ respectively. It follows that $xx' + yy' = r_0 \cos(\theta - \phi)$ so that

$$\psi_1 = \frac{Ae^{-ikz}}{z} \tilde{\mu} \int_0^{2\pi} \rho_0 e^{-ik\rho_0 \cos(\theta - \phi)} d\phi. \quad (71)$$
Using the reciprocity-based intuition explained in the main text, we know that the (fictitious) current sources placed on the focal spot are circularly symmetric. This means we can set \( \theta = 0 \), as \( \psi_1 \) will only depend on the radial coordinate of the aperture. From the integral representation of the Bessel function of the first kind \([8]\), \( J_0(u) = \frac{1}{\pi} \int_0^{2\pi} e^{-i u \cos \varphi} \, d\varphi \), the final expression for \( \psi_2 \) only depends on the radial coordinate:

\[
\psi_1 = \frac{A e^{-ikz}}{z} 2\pi \rho_0 J_0 \left( \frac{k \rho_0}{z} \right) \tilde{\mu} = 2\pi A \rho_0 J_0 \left( \frac{k \rho_0}{z} \right) e^{-ikz} \tilde{\mu} .
\]  

(72)

Having calculated \( \psi_0 \) and \( \psi_1 \), it is now straightforward to derive the bound on maximal focusing intensity, Strehl ratio and the optimal effective current. Using the relations \( \int_0^u u' J_0(u') \, du' = u J_1(u) \), \( \int_0^u u' [J_0(u')]^2 \, du' = u^2 \left\{ [J_0(u)]^2 + [J_1(u)]^2 \right\} / 2 \) and rescaling the limits of integration from 0 to 1 (with the aperture radius \( R \) taken out of the limits), the maximum intensity is given by inserting Eqs. \((69,72)\) into Eq. \((65)\) (we choose units such that \( c = 1 \)):

\[
I \leq \frac{\pi R^2 |A|^2}{z^2} - \frac{2\pi R^2 |A|^2}{z^2} \left[ \int_0^1 \frac{J_0(k R \rho_0 r/z)}{r} \frac{dr}{J_0(k R \rho_0 r/z)} r \frac{dr}{J_0(k R \rho_0 r/z)} \right]^2 \frac{k^2 R^2}{16 \pi^2} \left[ \frac{J_1(k R \rho_0/z)^2}{J_0(k R \rho_0/z)^2 + [J_1(k R \rho_0/z)]^2} \right] .
\]

(73)

Defining a normalized spot-size radius \( \eta_0 = \frac{k R \rho_0}{z} \), the Strehl ratio in Eq. \((66)\) reduces to:

\[
S_{\text{max}} = 1 - 4 \frac{\eta_0^4}{(1 - \eta_0^2)^2} \left( 1 - \eta_0^2 \right) \left( 1 - \eta_0^2 \right) \frac{J_1(\eta_0)}{J_0(\eta_0)^2 + [J_1(\eta_0)]^2} = \frac{\eta_0^4}{192} \quad \eta_0 \ll 1 .
\]

(74)

An asymptotic bound of \( S_{\text{max}} \) for small spot sizes \( \eta_0 \ll 1 \) turns out to be quartic in \( \eta_0 \) to lowest order, as we show below. Invoking the power series representation of \( J_0(\eta_0) \) and \( J_1(\eta_0) \) and keeping up to quartic order in \( \eta_0 \),

\[
S_{\text{max}} = 1 - 4 \frac{\eta_0^4}{(1 - \eta_0^2)^2} \left( 1 - \eta_0^2 \right) \left( 1 - \eta_0^2 \right) \frac{J_1(\eta_0)}{J_0(\eta_0)^2 + [J_1(\eta_0)]^2} = \frac{\eta_0^4}{192} \quad \eta_0 \ll 1 .
\]

(75)

Similarly, inserting the far-zone fields of Eqs. \((69,72)\) into Eq. \((67)\) \((r \) denotes the radial coordinate on the aperture),

\[
\xi_{\text{opt}}(r) = \xi_0 \frac{A e^{-ikz}}{z} \left[ 1 - 2 \frac{\eta_0}{J_0(\eta_0)} \frac{J_1(\eta_0)}{J_0(\eta_0)^2 + [J_1(\eta_0)]^2} \right] \tilde{\mu} .
\]

(76)

Based on Eq. \((76)\), we can also deduce the optimal field profile that achieves the bound in Eq. \((74)\). Defining the normalized radial coordinate on the focal plane as \( \eta = \frac{k R \rho}{z} \) (where \( \rho \) is the focal-plane radial coordinate),

\[
|\psi(\eta)| = \psi_0 \left[ \frac{J_1(\eta)}{\eta} - 2 \frac{\eta_0}{J_0(\eta_0)} \frac{J_1(\eta_0)}{J_0(\eta_0)^2 + [J_1(\eta_0)]^2} \int_0^1 J_0(\eta_0 r) J_0(\eta_0) r \frac{dr}{J_0(\eta_0)} \right] \frac{\eta_0^2 - \eta^2}{\eta_0^2 - \eta^2} = 0 \quad \eta = \eta_0
\]

(77)

where \( \psi_0 \) is a normalization factor. Near the diffraction limit, the ideal field profile resembles an Airy disk pattern, with a sharply focused center peak and negligible sidelobe intensity. As the spot size decreases, the sidelobe intensity begins to increase at the expense of focal intensity.

### III. FAR-ZONE BOUNDS FOR FOCAL SPHERE

Instead of maximizing intensity for a given focal spot, we can consider a focal sphere centered around the peak intensity. While the same reasoning as in Sec. II B holds, we modify \( \psi_1 \) in Eq. \((68)\) to involve integration around a sphere of radius \( \rho_0 \) (we also adopt the zeroth-order spherical harmonic basis, which is a constant).

With \( z_0 \) denoting the \( z \)-coordinate on the sphere relative to the peak intensity, the distance between points on the aperture and focal sphere is \( \sqrt{(x - x')^2 + (y - y')^2 + (z + z_0)^2} = z \sqrt{1 - 2 \frac{x x' + y y' - z z_0}{z^2} + \frac{z_0^2 + r^2}{z^2}} \). Dropping the quadratic terms in the far zone, we can Taylor expand the exponential, \( \sqrt{1 - 2 \frac{x x' + y y' - z z_0}{z^2} + \frac{z_0^2 + r^2}{z^2}} \approx 1 - 2 \frac{x x' + y y' - z z_0}{z^2} \). Switching to spherical polar coordinates with the sphere parametrized by \( (\rho_0, \theta, \phi) \) (the angle on the aperture denoted by \( \theta_{\text{ap}} \),...
we have that \( xx' + yy' - zz_0 = r \rho_0 \cos(\theta_{ap} - \phi) - z \rho_0 \cos \theta \). In analogy to Eq. (71), we arrive at the following expression for \( \psi_1 \):

\[
\psi_1 = \frac{Ae^{-ikz}}{z} \hat{\mu} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \rho_0 e^{-ikr_0} \left[ r \cos(\theta_{ap} - \phi) - z \rho_0 \cos \theta \right] \sin \theta \, d\theta \, d\phi. \tag{78}
\]

Again, reciprocity tells us that \( \theta_{ap} \) can be set to 0. This enables us to do the \( \phi \)-integral:

\[
\psi_1 = \frac{2\pi Ae^{-ikz}}{z} \hat{\mu} \int_{0}^{\pi} \rho_0 e^{ikr_0} \cos \theta J_0 \left( \frac{kr_0}{z} \sin \theta \right) \sin \theta \, d\theta. \tag{79}
\]

Equation (79) is a general expression valid in the far zone. The derivations for maximum intensity, Strehl ratio, optimal currents and field profile mimic those laid out in Sec. II B, but the form of Eq. (79) renders them somewhat messier and less intuitive. For the sake of clarity (and also of interest for highly localized, subwavelength focal sphere), we instead focus on the scaling behavior of Strehl ratio for extremely small spot sizes. In such a regime (\( \eta_0 \ll 1 \)), we can further expand \( J_0 \) in the integrand, which results in spherical bessel functions \( j_0 \) and \( j_1 \):

\[
\psi_1 = \frac{2\pi Ae^{-ikz}}{z} \hat{\mu} \int_{0}^{\pi} \rho_0 e^{ikr_0} \cos \theta \left[ 1 - \frac{1}{4} \left( \frac{kr_0}{z} \sin \theta \right)^2 \right] \sin \theta \, d\theta = \frac{2\pi Ae^{-ikz}}{z} \hat{\mu} \left[ 2j_0(kr_0) - \frac{kr_0r_0}{z^2} j_1(kr_0) \right]. \tag{80}
\]

Inserting the above expression into Eq. (66) gives us Strehl ratio (after rescaling the limits of integration):

\[
S_{\text{max}} = 1 - \frac{\int_{0}^{1} \left[ j_0(kr_0) - \frac{kr_0r_0}{z^2} j_1(kr_0) \right]^2 \, dr \, dz}{\int_{0}^{1} \left[ j_0(kr_0) - \frac{kr_0r_0}{z^2} j_1(kr_0) \right]^2 \, dz} = 1 - \frac{2 \left[ j_0(kr_0) - \frac{kr_0r_0}{z^2} j_1(kr_0) \right]^2}{2[j_0(kr_0)]^2 - \frac{kr_0r_0}{z^2} j_0(kr_0) j_1(kr_0) + \frac{kr_0r_0}{z^2} [j_1(kr_0)]^2}. \tag{81}
\]

To simplify the analysis, we make another assumption that the focal sphere has subwavelength radius such that \( kr_0 < 1 \), allowing us to Taylor expand the spherical Bessel functions. Denoting the normalized sphere radius as \( \eta_0 = \frac{kR}{kr_0} \), and defining \( x = kr_0 \) (focal radius relative to wavelength), we extend up to quartic order in \( \eta_0 \) or \( x \) (the asymptotic bound turns out to be quartic):

\[
S_{\text{max}} = 1 - \frac{2 \left[ (1 - x^2/6 + x^4/120) - \frac{R^2}{4k} \eta_0(x/3 - x^3/30) \right]^2}{2(1 - x^2/6 + x^4/120)^2 - \frac{R^2}{4k} \eta_0(1 - x^2/6 + x^4/120)(x/3 - x^3/30) + \frac{R^2}{4k^2} \eta_0^2(x/3 - x^3/30)^2}. \tag{82}
\]

Although not obvious from the above expression, straightforward but tedious algebra and repeated Taylor expansion show that the leading order term is quartic just in \( \eta_0 \) and that other quadratic terms and a quartic term in \( x \) all cancel out:

\[
S_{\text{max}} = \frac{\eta_0^2}{432}, \quad \eta_0 \ll 1, \tag{83}
\]

As Eq. (83) shows, the maximum Strehl ratio for a focal sphere also scales quartically with the normalized focal radius \( \eta_0 \).

### IV. 2D FAR-ZONE BOUNDS

Mirroring the derivation presented in Sec. II B, we can bound the maximum intensity for the 2D problem where we have translational invariance along one direction (chosen to be the \( y \)-axis in what follows). Here we do not constrain only the zeroth-order mode as in the previous two sections, but solve for the zero-field constraint exactly, as the zero-field contour involves just two evaluation points.

Working in the \( xz \)-coordinate, we appropriately modify the Green’s function taking the symmetry into account (electric Green’s function explicitly shown in Eq. (39) of Ref. [9], albeit with \( z \)-axis chosen as symmetry direction). Using the asymptotic form for the Hankel function \( H_{1/2} \sim \sqrt{\frac{2}{\pi u}} e^{(u - \frac{1}{4})} \) for large \( u \) and neglecting the transverse \( x \)-coordinates relative to the aperture-focus distance \( z \),

\[
\psi_0 = \frac{k e^{-i(kz - \frac{\pi}{4})}}{4c} \sqrt{\frac{2}{k\pi z}} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \mu = \frac{A e^{-ikz}}{\sqrt{z}} \hat{\mu} \tag{84}
\]
where \( A = \frac{\sqrt{2k/\pi}}{4c} \) is fixed for a given wavelength and \( \tilde{\mu} \equiv K \mu \) encoding the directions of \( \psi_0 \) and \( \psi_1 \) (same direction as \( \psi_0 \) in far zone).

Adopting the same notation as in Sec. II B (except that now we do not consider the \( y \)-coordinate), we can write down the expression for \( \psi \), which now involves just two points \( x = [\rho_0, -\rho_0] \) instead of all points around a spot-size ring:

\[
\psi_1 = \frac{A}{\sqrt{z}} \tilde{\mu} \left[ e^{-ikz\sqrt{1-\frac{2\rho_0^2}{kz^2} + \frac{k^2}{4z^2}}} + e^{-ikz\sqrt{1+\frac{2\rho_0^2}{kz^2} + \frac{k^2}{4z^2}}} \right].
\]  

(85)

Dropping the quadratic terms in the exponent in the far zone and Taylor expanding the remaining term, we obtain

\[
\psi_1 = \frac{Ae^{-ikz}}{\sqrt{z}} \tilde{\mu} \left[ e^{ikz}\eta_0 + e^{-ikz}\eta_0 \right] = \frac{2Ae^{-ikz}}{\sqrt{z}} \cos \left( \frac{kx'\rho_0}{z} \right) \tilde{\mu}.
\]  

(86)

Using the fields \( \psi_0 \) and \( \psi_1 \) obtained above, we can calculate the maximum intensity (\( R \) is the half-length of the aperture):

\[
I^{2D} \leq \frac{2R|A|^2}{z} - 2\frac{R|A|^2}{z} \left| \int_0^1 \cos(kR\rho_0 x'/z) \, dx' \right|^2 = \frac{kR}{4\pi z} - \frac{1}{\pi \rho_0} \sin^2(kR\rho_0/z) + \frac{1}{2\rho_0} \sin^2(2kR\rho_0/z)
\]  

(87)

Note that the unit here differs from that in Eq. (73) by a length dimension, as the aperture considered here is essentially a one-dimensional array of effective currents (extended infinitely along the \( y \)-axis) in contrast to a closed surface for the 3D case.

Denoting the normalized spot size width as \( \eta_0 = \frac{kR\rho_0}{z} \), the maximum Strehl ratio is as follows:

\[
S_{\text{max}}^{2D} = 1 - \frac{\int_0^1 \cos(\eta_0 x) \, dx}{\int_0^1 \cos^2(\eta_0 x) \, dx} = 1 - \frac{4}{\eta_0} \frac{\sin^2(\eta_0)}{\sin(2\eta_0) + 2\eta_0}.
\]  

(88)

For small spot sizes (\( \eta_0 \ll 1 \)), an asymptotic bound of \( S_{\text{max}}^{2D} \) again emerges as quartic in \( \eta_0 \):

\[
S_{\text{max}}^{2D} = 1 - \frac{4}{\eta_0} \left( \frac{(1 - \eta_0^3/3!) + \eta_0^5/(5!)}{1 - (2\eta_0)^3/3! + (2\eta_0)^5/(5!) + 2\eta_0} \right) = \frac{\eta_0^4}{45}, \ \eta_0 \ll 1.
\]  

(89)

We can also derive the optimal effective current (\( x \) denotes the aperture coordinate):

\[
\zeta_{\text{opt}}^{2D}(x) = \frac{Ae^{-ikz}}{\sqrt{z}} \left[ 1 - \frac{4}{\eta_0} \frac{\sin(\eta_0)}{\sin(2\eta_0) + 2\eta_0} \cos(\eta_0 x) \right] \tilde{\mu}.
\]  

(90)

Defining the normalized coordinate on the focal plane as \( \eta = \frac{kR\rho}{z} \) (where \( \rho \) is the focal-plane coordinate), the optimal field profile

\[
|\psi^{2D}(\eta)| = \psi_0 \left[ \frac{\sin(\eta)}{\eta} - \frac{4}{\sin(2\eta_0) + 2\eta_0} \frac{\eta_0 \sin(\eta_0) \cos(\eta) - \eta \cos(\eta_0) \sin(\eta)}{\eta_0^2 - \eta^2} \right]
\]  

(0 if \( \eta = \eta_0 \))

(91)

where \( \psi_0 \) is a normalization factor, resembles an Airy-like pattern (sinc function) near the diffraction limit, with increasing sidelobe relative to focal intensity for smaller spot sizes.

V. STREHL RATIO RECONSTRUCTION OF EXPERIMENTAL RESULTS IN FIG. 4B

To obtain the simulated results in Fig. 4b based on Ref. [10–16], we have reconstructed the various lens designs by discretizing an aperture with the same physical dimensions. The incident monochromatic waves passing through these lenses were then simulated by appropriately choosing the effective currents on the aperture (with \( 800 \times 800 = 640000 \) equally spaced grid points, each grid allowing for transverse electric and magnetic current—longitudinal currents do not appear in the equivalence principle [17]). To calculate the Strehl ratio, we normalize the intensities from each lens by the unconstrained (no zero-field condition) maximum intensity from a circular aperture of the same radius. As
FIG. 1: Reconstructed intensity profile for the lens design in Ref. [14], using effective currents to simulate the beam incident on the lens. The physical dimensions of the lens and focal distance are also fixed as in [ref], and our figure compares favorably with the experimental intensity profile in Fig. 5d of Ref. [14]. The intensity is normalized such that the peak intensity gives the Strehl ratio.

As an example, Fig. 1 shows the reconstructed intensity profile for the lens in [14] (scaled such that the peak intensity coincides with the Strehl ratio). There is close agreement in the spot sizes and peak widths (Fig. 1 has full width at half maximum (FWHM) of 0.4λ) compared to the experimental result (Fig. 5d of Ref. [14], FWHM of 0.42λ), the difference arising from discretization and experimental errors (and more pronounced further away from the peak, whose details are not important).

As a technical aside, most of these experiments are not carried out in the far zone and so the paraxial approximation [18] is not necessarily valid for these setups. Although the longitudinal field does not contribute at the peak (which renders the definition of Strehl ratio unambiguous), they can alter the overall intensity distribution across the focal plane. To facilitate comparisons with the far-zone case, for which the longitudinal components vanish, we plot the transverse (in-plane polarization) intensity in Fig. 1. Also, the normalized spot-size radius is modified to take the numerical aperture (NA) into account: \( \eta_0 = kNAP_0 \), which reduces to the definition given in the main text \( \eta_0 = kR\rho_0/z \) in the far zone. However, numerical experiments suggest that our bound also holds in the near zone with an error of at most a few percent even for NA \( \sim 0.5 \).

VI. SMALL-BUT-NONZERO FIELD CONSTRAINTS

In this section we aim to understand how the bounds change when the field at a specified spot size does not necessarily have to go to zero, but rather is small by some metric.

As discussed in the main text, the rank-two versions of all of the bounds that we derive can be written in the form

\[
\begin{align*}
\text{maximize} \quad & \psi^\dagger \psi_0 \psi_0^\dagger \psi \\
\text{subject to} \quad & \psi_1 \psi_1^\dagger \psi = 0 \\
& \psi^\dagger \psi = 1,
\end{align*}
\]

where the first-line term is the intensity at the origin at the second-line left-hand-side term is the total intensity on the spot-size contour. We showed in the main text that the optimum to this problem is attained for \( \psi = v \), where

\[
v = \psi_0 - \frac{\psi_1^\dagger \psi_0}{|\psi_1|^2} \psi_1.
\]
The vector \( v \) is orthogonal to \( \psi_1 \) (it is the projection of \( \psi \) onto \( \psi_0 \) in the subspace orthogonal to \( \psi_1 \)), and the maximum intensity for the zero-field constraint, which we denote here as \( I_0 \), is given by

\[
I_0 = v^\dagger v = \psi_0^\dagger \psi_0 - \frac{\vert \psi_0^\dagger \psi_1 \vert^2}{\vert \psi_1 \vert^2}.
\]  

(93)

In this case of a zero-field constraint, we did not have to be specific about the meaning of \( \psi_1^\dagger \psi_1 \psi \); it could be the total intensity integrated along the zero-field contour, it could have been an average, it could have been a discrete sum, etc.

Now that we want to allow for nonzero intensities, however, we must be specific about the precise value, and we will take \( \psi_1^\dagger \psi_1 \psi \) to represent the average intensity around the zero-field contour, which will allow for direct comparison to the intensity at the origin. (Note that this is a rescaled version of the \( \psi_1 \) that we compute in Eq. (17).) For a physically intuition figure of merit, we consider a scenario in which we allow the average intensity along the zero-field contour to be some fraction \( f \) of the originally optimal focusing intensity \( I_0 \), which would be represented by the constraint

\[
\psi_1^\dagger \psi_1 \psi \leq f I_0.
\]

(94)

We want to answer the following question: given this loosening of the constraint, how much can the focal intensity increase? We will see that the increase cannot be substantial, and thus that our zero-field results give realistic constraints even for nonzero constraints.

The new optimization problem is

\[
\begin{align*}
\text{maximize} & \quad \psi^\dagger \psi \psi_0^\dagger \psi_0 \\
\text{subject to} & \quad \psi_1^\dagger \psi_1 \psi \leq f I_0 \\
& \quad \psi^\dagger \psi = 1,
\end{align*}
\]

(95)

We know that \( \psi_0 \) and \( \psi_1 \) are not orthogonal, and in fact are very nearly parallel. And that the optimal field with the original zero-field constraint, \( v \), was orthogonal to \( \psi_1 \). Thus we can write

\[
\psi_0 = a \psi_1 + v,
\]

(96)

where the first term indicates the portion of \( \psi_0 \) that aligns with \( \psi_1 \) while the second term is the portion of \( \psi_0 \) orthogonal to \( \psi_1 \). If we assume that solving the unconstrained problem yields nearly equal intensities whether at the origin or averaged along the zero-field contour (very reasonable due to the homogeneity of free space), then

\[
\psi_0^\dagger \psi_0 = \psi_1^\dagger \psi_1,
\]

which implies that \( |a| \lesssim 1 \). We can bound the metric of Eq. (95) by

\[
\begin{align*}
\psi^\dagger \psi_0 \psi_0^\dagger \psi & = \psi^\dagger (a \psi_1 + v) (a \psi_1 + v)^\dagger \psi \\
& = |a|^2 \psi^\dagger \psi_1 \psi_1^\dagger \psi + 2 \text{Re} \left[ a \psi_1^\dagger \psi \psi_1^\dagger \psi \right] + \psi^\dagger vv^\dagger \psi \\
& \leq |a|^2 f I_0 + 2 |a| \sqrt{\psi^\dagger \psi_1 \psi_1^\dagger \psi} \sqrt{v^\dagger vv^\dagger \psi} + I_0 \\
& \leq \left( |a|^2 f + 2 |a| \sqrt{f} + 1 \right) I_0.
\end{align*}
\]

We see that the focusing intensity increases only linearly with the spot-size intensity. Thus, dramatic increases in focusing intensity are not possible simply by allowing small, nonzero spot-size fields. (For example, getting even a factor three intensity enhancement at the origin requires a spot-size field that is 50% of the original focusing intensity.)

1. D. Slepian, Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty — I, Bell System Technical Journal 40, 43 (1961).
2. D. Slepian, Prolate spheroidal wave functions, fourier analysis, and uncertainty—v: The discrete case, Bell System Technical Journal 57, 1371 (1978).
3. P. J. Ferreira and A. Kempf, Superscillations: Faster than the Nyquist rate, IEEE Transactions on Signal Processing 54, 3732 (2006).
4. R. Courant and D. Hilbert, Methods of Mathematical Physics: Partial Differential Equations (Interscience, 1953).
[5] L. Levi, Fitting a bandlimited signal to given points, IEEE Transactions on Information Theory 11, 372 (1965).
[6] K. Kreutz-Delgado, The Complex Gradient Operator and the CR-Calculus, arXiv:0906.4835 (2009).
[7] W. C. Chew, Waves and fields in inhomogeneous media (IEEE press, 1995).
[8] M. Abramowitz and I. A. Stegun, Handbook of mathematical functions with formulas, graphs, and mathematical tables (Dover, 1974).
[9] K. Sarabandi, Dyadic Green’s Function (2009), URL: http://www.eecs.umich.edu/courses/eecs730/lect/DyadicGF_W09_port.pdf.
[10] D. M. de Juana, J. E. Oti, V. F. Canales, and M. P. Cagigal, Design of superresolving continuous phase filters, Optics Letters 28, 607 (2003).
[11] P. N. Gundu, E. Hack, and P. Rastogi, ‘Apodized superresolution’ - Concept and simulations, Optics Communications 249, 101 (2005).
[12] S. Kosmeier, M. Mazilu, J. Baumgartl, and K. Dholakia, Enhanced two-point resolution using optical eigenmode optimized pupil functions, Journal of Optics 13, 105707 (2011).
[13] E. T. Rogers, J. Lindberg, T. Roy, S. Savo, J. E. Chad, M. R. Dennis, and N. I. Zheludev, A super-oscillatory lens optical microscope for subwavelength imaging, Nature Materials 11, 432 (2012).
[14] E. T. Rogers, S. Savo, J. Lindberg, T. Roy, M. R. Dennis, and N. I. Zheludev, Super-oscillatory optical needle, Applied Physics Letters 102, 031108 (2013).
[15] F. Qin, K. Huang, J. Wu, J. Teng, C. W. Qiu, and M. Hong, A Supercritical Lens Optical Label-Free Microscopy: Sub-Diffraction Resolution and Ultra-Long Working Distance, Advanced Materials 29, 1602721 (2017).
[16] X. H. Dong, A. M. H. Wong, M. Kim, and G. V. Eleftheriades, Superresolution far-field imaging of complex objects using reduced superoscillating ripples, Optica 4, 1126 (2017).
[17] J. A. Kong, Theory of electromagnetic waves (Wiley-Interscience, New York, 1975).
[18] L. Novotny and B. Hecht, Principles of nano-optics (Cambridge university press, 2012).