arXiv:2303.00726v1 [physics.optics] 1 Mar 2023

Transmission efficiency limit for nonlocal metalenses

Shiyu Li and Chia Wei Hsu*

Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA

The rapidly advancing capabilities in nanophotonic design are enabling complex functionalities limited mainly by physical bounds. The efficiency of transmission is a major consideration, but its ultimate limit remains unknown for most systems. Here, we introduce a matrix formalism that puts a fundamental bound on the channel-averaged transmission efficiency of any passive multi-channel optical system based only on energy conservation and the desired functionality, independent of the interior structure and material composition. Applying this formalism to diffraction-limited nonlocal metalenses with a wide field of view, we show that the transmission efficiency must decrease with the numerical aperture for the commonly adopted designs with equal entrance and output aperture diameters. We also show that reducing the size of the entrance aperture can raise the efficiency bound. This work reveals a fundamental limit on the transmission efficiency as well as providing guidance for the design of high-efficiency multi-channel optical systems.

I. INTRODUCTION

Over the past decade, nanophotonic design and fabrication became more and more advanced. Oftentimes, what limits the device performance is no longer fabrication constraints or the cleverness of the design, but fundamental physical bounds. Furthermore, the design process typically requires time consuming development, simulation, and optimization. It is invaluable to know beforehand what the fundamental bounds are and how they are related to the design choices [1]. Such a knowledge can significantly reduce the time spent in blind explorations and also point to better design choices that are not otherwise obvious. Of particular interest are multi-channel optical systems. For example, with a metalens [2-6], one would like the incident wave from each angle to be focused to the corresponding focal spot with unity efficiency. But is a uniformly perfect efficiency compatible with the desired angle-dependent response? What is the highest efficiency allowed by fundamental laws?

Here, we introduce a matrix-based formalism that sets a fundamental bound on the efficiency of any linear multichannel system given its functionality, and apply it to metalenses.

Metalenses are compact lenses made with metasurfaces, which show great potential for thinner and lighter imaging systems with performances comparable to or exceeding conventional lenses [2–6]. Metalenses designed from a library of unit cells have limited focusing efficiency, which can be overcome by more flexible designs [7, 8]. Inverse design [7–15], grating averaging technique [16], and stitching separately designed sections together [17–19] are effective approaches. However, achieving high focusing efficiency at large numerical aperture (NA) remains difficult, as all such "local" metasurfaces have limited deflection efficiency at large angles [20, 21]. Since local metasurfaces have a spatial impulse response close to a delta function, they provide the same response for different incident angles, so they are also limited in their angular field of view (FOV) [22].

Nonlocal metalenses with tailored interactions between adjacent building blocks (*i.e.*, the spatial impulse response is extended beyond a delta function) can overcome the limited angular diversity of local metalenses to enable diffraction-limited focusing over a large FOV [22– 24]. Nonlocal metalenses based on doublets [25, 26] or aperture stops [27, 28] can support focusing efficiencies higher than 50% over a wide FOV, but with NA lower than 0.5. Multi-layer structures obtained from inverse design have achieved diffraction-limited focusing with NA = 0.7 over FOV = 80°, but the averaged focusing efficiency is only about 25% [29]. However, there was no guidance on the efficiency bound of these nonlocal metalenses.

From the desired response of a multi-channel optical system, we can write down its transmission matrix that relates the input to the output. Here, we rigorously bound the channel-averaged transmission efficiency using the singular values of the transmission matrix and the fact that the transmitted energy must not exceed the input energy. For commonly adopted designs with equal entrance and output apertures, we find the transmission efficiency bound of a nonlocal metalens to drop with the NA. We also find that reducing the entrance aperture size can raise the efficiency bound to close to unity. This approach is general and can guide the design of not only metalenses but also other multi-channel optical systems.

II. RESULTS

A. Nonlocality

As schematically illustrated in Fig. 1(a), subwavelength-thick local metalenses (such as metalenses with the hyperbolic phase profile [30]) perform well only over a limited input angular range. Nonlocal

^{*} cwhsu@usc.edu



FIG. 1. Nonlocal metalenses and their transmission matrix. (a) A local metalens with subwavelength thickness, for which the input and output apertures have the same size, *i.e.*, $D_{in} = D_{out}$. Diffraction-limited focusing can only be achieved within a narrow range of incident angles. (b) A nonlocal metalens, whose larger thickness allows nonlocal coupling, under which diffraction-limited focusing may be achieved over a wide angular range. Here, D_{in} and D_{out} can be different. (c,d) The squared amplitude of the transmission matrix $|\mathbf{t}(k_y, k'_y)|^2$ (c) and its normalized singular values σ_i/σ_1 (d) for an ideal nonlocal metalens with diameter $D_{in} = D_{out} = 300\lambda$, NA = 0.8, FOV = 140°. Here, the averaged transmission efficiency is bounded by (3) as $\langle T \rangle \leq 74\%$.

metalenses can achieve diffraction-limited focusing over a much wider FOV but need a minimal thickness to provide the nonlocality [22, 31]. Since a large FOV requires an angle-dependent response (i.e., angular)dispersion), a spatially localized incident wave must spread as it propagates through the metalens under space-angle Fourier transform. Thus more angular diversity necessitates more nonlocality. As shown in Supplementary Fig. S1, nonlocal effects become important when the FOV is larger than a threshold. For FOV above such a threshold, a sufficient thickness is needed for light to spread and create nonlocality, so the input diameter $D_{\rm in}$ and output diameter $D_{\rm out}$ can be different [Fig. 1(b)]. The different aperture sizes provide an additional degree of freedom compared to local metalenses with subwavelength thicknesses (for which $D_{\rm in} = D_{\rm out}$ intrinsically).

B. Transmission matrix and target response

While the transmission-matrix formalism introduced in this work is general, for concreteness we will consider metalenses. For ideal (aberration-free) lenses under plane-wave incidence, we can obtain the transmitted phase profile on the output surface of the metalens by matching the optical path lengths between the marginal rays and the chief ray at the focus $\mathbf{r}(\theta_{\rm in}) =$ $(y = f \tan \theta_{\rm in}, z = f + h)$. In 2D, we get [22, 27, 32]

$$\phi_{\text{out}}^{\text{ideal}}(y, z = h, \theta_{\text{in}}) = \psi(\theta_{\text{in}}) - \frac{2\pi}{\lambda} \sqrt{f^2 + (y - f \tan \theta_{\text{in}})^2},$$
(1)

where θ_{in} is the incident angle; f and h are the focal length and lens thickness respectively; $\psi(\theta_{in})$ is an angledependent but spatially invariant global phase with no effect on the focusing quality.

Consider monochromatic, transverse magnetic waves of 2D systems at wavelength λ , where $\mathbf{E} = E_x(y, z)\hat{x}$. The incoming wavefront $E_x(y', z = 0)$ can be projected onto $N_{\rm in}$ propagating plane waves with input transverse momenta $k'_y = \{k^a_y\} = (2\pi/\lambda)\sin\theta_{\rm in}$ truncated to the size of the input aperture *i.e.*, $|y'| < D_{in}/2$, yielding complex-valued amplitude α_a of the a-th plane-wave input, where z = 0 is the front surface of the system, $|k_{u}^{a}| <$ $(2\pi/\lambda)\sin(\text{FOV}/2)$ is restricted to the FOV of interest with k_u^a discretized with $2\pi/D_{\rm in}$ spacing following the Nyquist-Shannon theorem [33]. Similarly, the transmitted wavefront $E_x(y, z = h)$ can be written as a superposition of N_{out} plane waves with $k_y = \{k_y^b\} = (2\pi/\lambda) \sin \theta_{\text{out}}$ truncated to $|y| < D_{out}/2$, and the complex-valued amplitude of the *b*-th plane-wave output is β_b , where θ_{out} is the output angle and $|k_y^b| < 2\pi/\lambda$ is sampled with momentum spacing $2\pi/D_{out}$. For any linear optical system, these amplitudes are related through the transmission matrix t_{ba} ,

$$\beta_b = \sum_{a=1}^{N_{\rm in}} t_{ba} \alpha_a. \tag{2}$$

We can parameterize the incoming and transmitted wavefronts as column vectors $\boldsymbol{\alpha} = [\alpha_1; \cdots; \alpha_{N_{\text{in}}}]$ and $\boldsymbol{\beta} = [\beta_1; \cdots; \beta_{N_{\text{out}}}]$. Then, $\boldsymbol{\beta} = \mathbf{t}\boldsymbol{\alpha}$, where $\mathbf{t} = \mathbf{t}(k_y, k_y') = t(k_y^b, k_y^a) = t_{ba}$ is the transmission matrix in the angular basis.

Fourier transforming the output field of the ideal metalens of Eq. (1) from y to k_y yields the transmission matrix \mathbf{t} . Such transmission matrix applies to any lens with diffraction-limited focusing over this FOV. The squared amplitude of the transmission matrix $|\mathbf{t}(k_y, k'_y)|^2$ for a metalens with $D_{\rm in} = D_{\rm out} = 300\lambda$, NA = $\sin(\arctan(D_{\rm out}/(2f))) = 0.8$ and FOV = 140° is shown in Fig. 1(c). More details on the transmission matrix can be found in Supplementary Sec. 2. Notably, Eq. (1) does not specify the overall amplitude prefactor of the transmission matrix, so we cannot directly assess the transmission efficiency based on this target response.



FIG. 2. Transmission efficiency bound $N_{\rm eff}/N_{\rm in}$ of ideal nonlocal metalenses when $D_{\rm in} = D_{\rm out}$ and when using $D_{\rm in} = D_{\rm th}^{\rm th}$ from (4). (a) $N_{\rm eff}/N_{\rm in}$ averaged over FOV and output diameter $D_{\rm out}$ as a function of NA. The black line is $\sqrt{1 - \mathrm{NA}^2}$. (b,c) The distribution of $N_{\rm eff}/N_{\rm in}$ among different FOV and $D_{\rm out}$ when (b) NA = 0.7 and (c) NA = 0.9.

C. Transmission efficiency bound

When the transmission matrix includes the overall amplitude prefactor and is flux-normalized, the total transmission efficiency (summed over all transmitted flux) averaged over inputs within the prescribed FOV is $\langle T \rangle =$ $\sum_{a} T_a/N_{\rm in}$, where $T_a = \sum_{b} |t_{ba}|^2$ is the total transmission efficiency for incident angle $\theta_{\rm in}^a$ and $N_{\rm in}$ is the number of input channels (*i.e.*, the number of columns in the transmission matrix). To obtain an upper bound on $\langle T \rangle$ without knowledge of the overall amplitude prefactor of the transmission matrix, we consider the singular values $\{\sigma_i\}$ of the transmission matrix. With the singular value decomposition, the transmission matrix is factorized into $\mathbf{t}(k_y, k'_y) = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\dagger} = \sum_i \sigma_i(u_i v_i^{\dagger})$, where each right-singular column vector v_i is a normalized incident wavefront being a linear superposition of the input channels. The corresponding transmitted wavefront is $\sigma_i u_i$, with a transmission efficiency of σ_i^2 and with the normalized transmitted wavefront being column vector u_i . Energy conservation imposes that the transmitted energy cannot exceed the input energy, so $0 \leq \sigma_i^2 \leq 1$ for all *i*. Since $\operatorname{tr}(\mathbf{t}^{\dagger}\mathbf{t}) = \sum_{a} (\mathbf{t}^{\dagger}\mathbf{t})_{aa} = \sum_{b,a} |t_{ba}|^{2}$ equals the sum of the eigenvalues $\lambda_{i} = \sigma_{i}^{2}$ of matrix $\mathbf{t}^{\dagger}\mathbf{t}$, we have $\langle T \rangle = \sum_{b,a} |t_{ba}|^{2} / N_{\text{in}} = \sum_{i} \sigma_{i}^{2} / N_{\text{in}}$. In addition, $\sum_{i} \sigma_{i}^{4} \leq \sum_{i} \sigma_{i}^{2}$ follows from energy conservation. Therefore, we have a rigorous inequality

$$\langle T \rangle \equiv \frac{\sum_{b,a} |t_{ba}|^2}{N_{\rm in}} = \frac{N_{\rm eff}}{N_{\rm in}} \frac{\sum_i \sigma_i^4}{\sum_i \sigma_i^2} \le \frac{N_{\rm eff}}{N_{\rm in}}.$$
 (3)

Here, $N_{\text{eff}} = (\sum_{i=1} \sigma_i^2)^2 / \sum_{i=1} \sigma_i^4$ is the effective number of high-transmission channels characterized through an inverse participation ratio [34], and it is independent of the overall prefactor of the transmission matrix. Figure 1(d) shows the normalized singular values of the transmission matrix $\mathbf{t}(k_y, k'_y)$ in Fig. 1(c), for which $\langle T \rangle \leq N_{\text{eff}} / N_{\text{in}} \approx 74\%$.

This N_{eff} comes directly from the desired angledependent response encoded in the transmission matrix, regardless of what structural design and material composition are used to realize such response. Together with energy conservation, N_{eff} imposes a fundamental bound on the average transmission of the system through (3). This formalism applies to any linear optical system for which the desired response is known.

A bound is useful only when it is sufficiently tight. Supplementary Fig. S2 shows from full-wave numerical simulations [35] that Eq. (3) indeed provides a reasonably tight upper bound for the averaged transmission efficiency $\langle T \rangle$, considering hyperbolic and quadratic metalenses as examples.

Figure 2(a) shows the transmission efficiency bound $N_{\rm eff}/N_{\rm in}$ as a function of NA for aberration-free nonlocal metalenses with FOV larger than the threshold shown in Supplementary Fig. S1. Since FOV and the output diameter $D_{\rm out}$ have a very small influence on $N_{\rm eff}/N_{\rm in}$ (Supplementary Figs. S3,S4), in Fig. 2(a) we map out how the efficiency bound depends on the NA while averaging over FOV and $D_{\rm out}$. We see that using equal entrance and output diameters, *i.e.*, $D_{\rm in} = D_{\rm out}$, results in an efficiency bound that drops approximately as $\sqrt{1 - \mathrm{NA}^2}$. This bound applies regardless of how complicated or optimized the design is and regardless of what materials are used.

In Fig. 2(b,c), we show the distribution (among different FOV and output diameters) of the transmission efficiency bound with NA = 0.7 and 0.9. The bound is consistent with the inverse design results of Ref. [29], where the achieved average absolute focusing efficiency (considering only the transmitted power within three fullwidths at half-maximum around the focal peak) is 25% for a nonlocal metalens with NA = 0.7 and FOV = 80°.

D. Optimal aperture size

In addition to establishing a transmission efficiency limit, it would be even more useful to know what strategy one may adopt to raise such an efficiency limit. To this end, we examine the singular values of the transmission matrix. In Fig. 1(d) where $D_{\rm in} = D_{\rm out} = 300\lambda$ and



FIG. 3. The transmission efficiency bound $N_{\rm eff}/N_{\rm in}$ and the condition number κ of the ideal transmission matrix as a function of the entrance diameter $D_{\rm in}$. Black and gray vertical dotted lines indicate $D_{\rm in}^{\rm th} = 185\lambda$ and $D_{\rm out} = 300\lambda$. Lens parameters: output diameter $D_{\rm out} = 300\lambda$, NA = 0.8, FOV = 140°.

NA = 0.8, we see there are many zero singular values. These zero singular values lower $N_{\rm eff}$ and the transmission efficiency. Removing these zero singular values (*i.e.*, eliminating non-transmitting wavefronts) can raise the transmission efficiency bound $N_{\rm eff}/N_{\rm in}$. A zero singular value corresponds to a superposition of the columns of the transmission matrix that yields a zero vector, meaning those columns are linearly dependent. Therefore, we can eliminate zero singular values by reducing the number of columns in the transmission matrix. Because the input wave vectors $|k'_y| < (2\pi/\lambda) \sin (\text{FOV}/2)$ are sampled with momentum spacing $2\pi/D_{\rm in}$ at the Nyquist rate, we expect that reducing $D_{\rm in}$ can lower the number of input columns in the transmission matrix to raise the transmission efficiency bound. Figure 3 shows this strategy indeed works: reducing the input aperture size $D_{\rm in}$ increases the efficiency bound $N_{\rm eff}/N_{\rm in}$.

The next question is: what would be an optimal input diameter D_{in} to use? While reducing D_{in} raises the transmission efficiency bound, doing so also reduces the amount of light that can enter the metalens, which is not desirable. To find a balance, we examine the condition number κ , defined as the ratio between the maximal and the minimal singular values. When zero σ_i exist, the condition number κ diverges to infinity (subject to numerical precision). The right-axis of Fig. 3 shows that κ abruptly shoots up by many orders of magnitude when the input diameter $D_{\rm in}$ raises above a threshold that we label as $D_{\rm in}^{\rm th}$ (black dotted line). When $D_{\rm in} < D_{\rm in}^{\rm th},\,\kappa$ is of order unity, all singular values are comparable with no zerotransmission wavefronts, so we have $N_{\rm eff} \approx N_{\rm in}$, and the transmission efficiency bound is close to unity. When $D_{\rm in} > D_{\rm in}^{\rm th}$, near-zero-transmission wavefronts start to appear, which results in a fast reduction of the transmission efficiency bound. Therefore, this threshold value $D_{\rm in}^{\rm th}$ is an optimal input diameter to use, providing maximal entrance flux while keeping a near-unity transmission efficiency bound.

To automate the determination of D_{in}^{th} , we examine

the slope $\partial \kappa / \partial D_{\rm in}$, which transitions from near zero to a very large number at $D_{\rm in}^{\rm th}$. Supplementary Tab. S1 shows that different threshold values for $\partial \kappa / \partial D_{\rm in}$ yield almost identical $D_{\rm in}^{\rm th}$ and $N_{\rm eff}/N_{\rm in}$. Note that the choice of the global phase $\psi(\theta_{\rm in})$ in (1) does not influence $D_{\rm in}^{\rm th}$ and the transmission efficiency bound, as shown in Supplementary Fig. S5.

To guide future designs, it is desirable to know how $D_{\rm in}^{\rm th}$ depends on the various lens parameters. Figure 4 plots $D_{\rm in}^{\rm th}$ as a function of the FOV, output diameter $D_{\rm out}$, and NA. As described in Sec. 2II A, lenses with a very small FOV do not require nonlocality; for such local metalenses, we find $D_{\rm in}^{\rm th} \approx D_{\rm out}$ as expected from the schematic in Fig. 1(a). As the FOV increases and nonlocality emerges, we observe in Fig. 4(a) that $D_{\rm in}^{\rm th}$ drops below $D_{\rm out}$ (as expected from the preceding discussions) and reaches a constant value that depends on NA but not on the FOV. This $D_{\rm in}^{\rm th}$ for nonlocal metalenses is proportional to $D_{\rm out}$ [Fig. 4(b)]. Figure 4(a,b) fix $D_{\rm out} = 300\lambda$ and FOV = 140° respectively; other lens parameters share similar dependencies (Figs. S6-S7 of Supplement 1). In Fig. 4(c), we find empirically that the NA dependence is well described by

$$D_{\rm in}^{\rm th} = D_{\rm out} \sqrt{1 - \rm NA^2}.$$
 (4)

This result provides a recipe for choosing the input and output aperture sizes for nonlocal high-NA metalenses with high transmission.

To demonstrate the increased transmission efficiency bound, we also show in Fig. 2(a–c) the transmission efficiency bound when the input aperture size $D_{\rm in}$ is set to the optimal $D_{\rm in}^{\rm th}$ in (4). We observe a large transmission efficiency bound $N_{\rm eff}/N_{\rm in}$ that overcomes the $\sqrt{1 - NA^2}$ limit when equal entrance and output apertures are used.

III. DISCUSSION

While this work was originally motivated by the efficiency of high-NA nonlocal metalenses, the formalism we introduce is very general. Given the desired response of any multi-channel optical system, one may write down its transmission matrix and apply (3) to establish a bound on its transmission efficiency. Like the thickness bound introduced in Refs. [22, 31], the transmission efficiency bound here is functionality-driven and designindependent.

Nonlocal metasurfaces open up a wide range of applications and tailored angle-dependent responses that are impossible for traditional local metasurfaces [24]. The efficiency bound in this work provides valuable guidance for this rapidly evolving field.

ACKNOWLEDGMENTS

We thank X. Gao for helpful discussions. This work is supported by the National Science Foundation CA-



FIG. 4. Dependence of the threshold input diameter D_{in}^{th} on the lens parameters. (a) D_{in}^{th} as a function of the FOV when $D_{out} = 300\lambda$. (b) D_{in}^{th} as a function of the output diameter D_{out} when FOV = 140°. Symbols are D_{in}^{th} computed from the ideal transmission matrices, and solid lines are linear fits. (c) D_{in}^{th}/D_{out} averaged over different FOV for nonlocal metalenses as a function of NA. Black solid line is $\sqrt{1 - NA^2}$.

REER award (ECCS-2146021) and the Sony Research Award Program. **Disclosures:** The authors declare no conflicts of interest. **Data availability:** All data needed to evaluate the conclusions in this study are presented in the paper and in the supplemental document.

- P. Chao, B. Strekha, R. Kuate Defo, S. Molesky, and A. W. Rodriguez, Physical limits in electromagnetism, Nature Reviews Physics 4, 543 (2022).
- [2] P. Lalanne and P. Chavel, Metalenses at visible wavelengths: past, present, perspectives, Laser Photonics Reviews 11, 1600295 (2017).
- [3] M. Khorasaninejad and F. Capasso, Metalenses: versatile multifunctional photonic components, Science 358, eaam8100 (2017).
- [4] H. Liang, A. Martins, B.-H. V. Borges, J. Zhou, E. R. Martins, J. Li, and T. F. Krauss, High performance metalenses: numerical aperture, aberrations, chromaticity, and trade-offs, Optica 6, 1461 (2019).
- [5] M. Pan, Y. Fu, M. Zheng, H. Chen, Y. Zang, H. Duan, Q. Li, M. Qiu, and Y. Hu, Dielectric metalens for miniaturized imaging systems: progress and challenges, Light. Sci. Appl. 11, 195 (2022).
- [6] A. Arbabi and A. Faraon, Advances in optical metalenses, Nat. Photonics 17, 16 (2023).
- [7] H. Chung and O. D. Miller, High-NA achromatic metalenses by inverse design, Optics Express 28, 6945 (2020).
- [8] K. Schab, L. Jelinek, M. Capek, and M. Gustafsson, Upper bounds on focusing efficiency, Opt. Express 30, 45705 (2022).
- [9] D. Sell, J. Yang, S. Doshay, and J. A. Fan, Periodic dielectric metasurfaces with high-efficiency, multiwavelength functionalities, Advanced Optical Materials 5, 1700645 (2017).
- [10] R. Pestourie, C. Pérez-Arancibia, Z. Lin, W. Shin, F. Capasso, and S. G. Johnson, Inverse design of large-area metasurfaces, Optics Express 26, 33732 (2018).
- [11] H. Liang, Q. Lin, X. Xie, Q. Sun, Y. Wang, L. Zhou, L. Liu, X. Yu, J. Zhou, T. F. Krauss, et al., Ultrahigh

numerical aperture metalens at visible wavelengths, Nano Lett. **18**, 4460 (2018).

- [12] M. Mansouree and A. Arbabi, Large-scale metasurface design using the adjoint sensitivity technique, in 2018 Conference on Lasers and Electro-Optics (CLEO) (2018) p. FF1F.7.
- [13] T. Phan, D. Sell, E. W. Wang, S. Doshay, K. Edee, J. Yang, and J. A. Fan, High-efficiency, large-area, topology-optimized metasurfaces, Light Science Applications 8, 48 (2019).
- [14] K.-F. Lin, C.-C. Hsieh, S.-C. Hsin, and W.-F. Hsieh, Achieving high numerical aperture near-infrared imaging based on an ultrathin cylinder dielectric metalens, Applied Optics 58, 8914 (2019).
- [15] Z.-P. Zhuang, R. Chen, Z.-B. Fan, X.-N. Pang, and J.-W. Dong, High focusing efficiency in subdiffraction focusing metalens, Nanophotonics 8, 1279 (2019).
- [16] A. Arbabi, E. Arbabi, M. Mansouree, S. Han, S. M. Kamali, Y. Horie, and A. Faraon, Increasing efficiency of high numerical aperture metasurfaces using the grating averaging technique, Sci. Rep. 10, 7124 (2020).
- [17] S. J. Byrnes, A. Lenef, F. Aieta, and F. Capasso, Designing large, high-efficiency, high-numerical-aperture, transmissive meta-lenses for visible light, Optics Express 24, 5110 (2016).
- [18] R. Paniagua-Dominguez, Y. F. Yu, E. Khaidarov, S. Choi, V. Leong, R. M. Bakker, X. Liang, Y. H. Fu, V. Valuckas, L. A. Krivitsky, *et al.*, A metalens with a near-unity numerical aperture, Nano Lett. **18**, 2124 (2018).
- [19] M. Kang, Y. Ra'di, D. Farfan, and A. Alù, Efficient focusing with large numerical aperture using a hybrid metalens, Physical Review Applied 13, 044016 (2020).

- [20] N. M. Estakhri and A. Alu, Wave-front transformation with gradient metasurfaces, Physical Review X 6, 041008 (2016).
- [21] V. S. Asadchy, M. Albooyeh, S. N. Tcvetkova, A. Díaz-Rubio, Y. Ra'di, and S. Tretyakov, Perfect control of reflection and refraction using spatially dispersive metasurfaces, Physical Review B 94, 075142 (2016).
- [22] S. Li and C. W. Hsu, Thickness bound for nonlocal widefield-of-view metalenses, Light Sci. Appl. 11, 338 (2022).
- [23] A. Overvig and A. Alù, Diffractive nonlocal metasurfaces, Laser Photonics Reviews 16, 2100633 (2022).
- [24] K. Shastri and F. Monticone, Nonlocal flat optics, Nat. Photonics 17, 36 (2023).
- [25] A. Arbabi, E. Arbabi, S. M. Kamali, Y. Horie, S. Han, and A. Faraon, Miniature optical planar camera based on a wide-angle metasurface doublet corrected for monochromatic aberrations, Nature Commun. 7, 13682 (2016).
- [26] B. Groever, W. T. Chen, and F. Capasso, Meta-lens doublet in the visible region, Nano Lett. 17, 4902 (2017).
- [27] M. Y. Shalaginov, S. An, F. Yang, P. Su, D. Lyzwa, A. M. Agarwal, H. Zhang, J. Hu, and T. Gu, Single-element diffraction-limited fisheye metalens, Nano Lett. 20, 7429 (2020).
- [28] C.-Y. Fan, C.-P. Lin, and G.-D. J. Su, Ultrawide-angle and high-efficiency metalens in hexagonal arrangement,

Sci. Rep. 10, 15677 (2020).

- [29] Z. Lin, C. Roques-Carmes, R. E. Christiansen, M. Soljačić, and S. G. Johnson, Computational inverse design for ultra-compact single-piece metalenses free of chromatic and angular aberration, Applied Physics Letters 118, 041104 (2021).
- [30] F. Aieta, P. Genevet, M. A. Kats, N. Yu, R. Blanchard, Z. Gaburro, and F. Capasso, Aberration-free ultrathin flat lenses and axicons at telecom wavelengths based on plasmonic metasurfaces, Nano Lett. 12, 4932 (2012).
- [31] D. A. Miller, Why optics needs thickness, Science 379, 41 (2023).
- [32] A. Kalvach and Z. Szabó, Aberration-free flat lens design for a wide range of incident angles, J. Opt. Soc. Am. B 33, A66 (2016).
- [33] H. Landau, Sampling, data transmission, and the Nyquist rate, Proceedings of the IEEE 55, 1701 (1967).
- [34] C. W. Hsu, S. F. Liew, A. Goetschy, H. Cao, and A. Douglas Stone, Correlation-enhanced control of wave focusing in disordered media, Nat. Phys. 13, 497 (2017).
- [35] H.-C. Lin, Z. Wang, and C. W. Hsu, Fast multi-source nanophotonic simulations using augmented partial factorization, Nat. Comput. Sci. 2, 815 (2022).

Transmission efficiency limit for nonlocal metalenses: supplemental document

This document provides supplemental information to "Transmission efficiency limit for nonlocal metalenses". It consists of seven sections. Section 1 characterizes whether a metalens is local or nonlocal. Section 2 details how to construct the transmission matrix of an ideal metalens. Section 3 compares the efficiency bound $N_{\text{eff}}/N_{\text{in}}$ with the transmission efficiency of actual metalenses from full-wave simulations. Sec. 4 provides comprehensive data of $N_{\text{eff}}/N_{\text{in}}$ for different lens parameters. Section 5 shows that the slope of the condition number, $\partial \kappa / \partial D_{\text{in}}$, is a robust parameter for determining the threshold input diameter $D_{\text{in}}^{\text{th}}$. In Sec. 6, we show that $D_{\text{in}}^{\text{th}}$ and transmission efficiency bound $N_{\text{eff}}/N_{\text{in}}$ are independent of the global phase function $\psi(\theta_{\text{in}})$. In Sec. 7, we provide complete data plots for $D_{\text{in}}^{\text{th}}$ as a function of the lens parameters.

CONTENTS

1	Nonlocality of metalenses	1
2	Transmission matrix of an ideal metalens	2
3	Transmission efficiency of hyperbolic and quadratic metalenses	3
4	Dependence of transmission efficiency bound on lens parameters	4
5	Threshold input diameter from $\partial \kappa / \partial D_{in}$	4
6	$\psi(heta_{ ext{in}})$ dependence	4
7	Comprehensive data on the threshold input diameter	6

1. NONLOCALITY OF METALENSES

When a metalens achieves diffraction-limited focusing across a wide range of incident angles, its response must be angledependent, requiring nonlocality. Conversely, if the angular range is small, the response does not need to vary with angle, and a local metalens would suffice. Thus we expect a threshold field of view (FOV) that separates local and nonlocal metalenses.

Assume free space on the incident and transmitted sides of the metalens. An ideal lens needs to match the optical path lengths of the marginal rays and the chief ray from the focal spot position ($y = f \tan \theta_{in}, z = h + f$) to the lens surface (y, z = h) for diffraction-limited focusing, so the phase distribution on the back surface of an ideal metalens is

$$\phi_{\text{out}}^{\text{ideal}}(y, z = h, \theta_{\text{in}}) = \psi(\theta_{\text{in}}) - \frac{2\pi}{\lambda} \sqrt{f^2 + (y - f \tan \theta_{\text{in}})^2},$$
(S1)

where h, θ_{in} and f are the lens thickness, incident angle and focal length, respectively. The phase shift provided by the metalens is then

$$\Delta\phi_{\text{ideal}}(y,\theta_{\text{in}}) = \phi_{\text{out}}^{\text{ideal}}(y,\theta_{\text{in}}) - \phi_{\text{in}}(y,\theta_{\text{in}}), \tag{S2}$$

where $\phi_{in}(y, \theta_{in}) = k_y^{in} y = (2\pi/\lambda) \sin \theta_{in} y$ is the phase profile of the incident light.

The $\psi(\theta_{in})$ in Eq. (S1) is an angle-dependent but spatially-invariant global phase, with no influence on the focusing performance. One sensible choice for the global phase $\psi(\theta_{in})$, applied in all calculations in this work, is

$$\psi(\theta_{\rm in}) = \frac{2\pi}{\lambda} \left\langle \sqrt{f^2 + (y - f \tan \theta_{\rm in})^2} + y \sin \theta_{\rm in} \right\rangle_y \equiv \psi_0(\theta_{\rm in}), \tag{S3}$$

where $\langle \cdots \rangle_y$ denotes averaging over *y* within the output aperture *i.e.*, $|y| < D_{out}/2$. With this $\psi_0(\theta_{in})$, the *y* average of the phase shift, $\langle \Delta \phi_{ideal}(y, \theta_{in}) \rangle_y$, is the same for different incident angles, which minimizes the required thickness of the metalens [1]. We will show in Sec. 6 that the transmission efficiency bound studied in this work is independent of the choice of $\psi(\theta_{in})$.

A local hyperbolic metalens achieves diffraction-limited focusing at the normal incidence $\theta_{in} = 0^\circ$, with an angle-independent phase-shift profile [2, 3]

$$\Delta\phi_{\rm hyp}(y) = \Delta\phi_{\rm ideal}(y,\theta_{\rm in}=0^\circ) = \psi_0(\theta_{\rm in}=0^\circ) - \frac{2\pi}{\lambda}\sqrt{f^2 + y^2}.$$
 (S4)

Away from the normal incident angle, $\Delta \phi_{hyp}(y) \neq \Delta \phi_{ideal}(y, \theta_{in})$, so a hyperbolic metalens no longer achieves ideal focusing.

Given the lens parameters (FOV, NA, and *D*_{out}), we want to determine whether it can be realized with a local metalens or whether nonlocality is required. To do so, we evaluate

$$\max_{y,\theta_{\rm in}} |\Delta \phi_{\rm ideal}(y,\theta_{\rm in}) - \Delta \phi_{\rm hyp}(y)| \tag{S5}$$

across all positions $|y| < D_{out}/2$ and all incident angles within $|\theta_{in}| < FOV/2$. If Eq. (S5) $< \pi$, we consider a local hyperbolic metalens to be sufficient, and the metalens is classified as local. Otherwise, it is classified as nonlocal.

Figure S1(a) shows whether an ideal metalens is local or nonlocal over FOV \in (0°, 180°) and numerical aperture NA \in (0, 1) when $D_{out} = 300\lambda$. When the FOV is very small, all lenses are local regardless of NA; this is the reason that in Fig. 4(a) of the main text, $D_{in}^{th} \approx D_{out}$ when the FOV is very small. Above a threshold FOV, the lens becomes nonlocal, for which $D_{in}^{th} < D_{out}$. The threshold FOV and D_{in}^{th} both depend on the NA. Figure S1(b) further shows the threshold FOV for different D_{out} and NA.



Fig. S1. (a) Designation of whether a metalens is local or nonlocal; $D_{out} = 300\lambda$. (b) The threshold FOV that separates local and nonlocal metalenses for different output diameter D_{out} and NA.

2. TRANSMISSION MATRIX OF AN IDEAL METALENS

Following Eq. (2) of the main text, the transmission matrix relates the incoming wavefront to the outgoing wavefront as $\beta_b = \sum_a t_{ba} \alpha_a$.

We express the input wavefront in a flux-orthogonal basis of truncated plane waves $\{f_a\}$,

$$f_a(y,z) = \begin{cases} \frac{1}{\sqrt{D_{\text{in}}}} \frac{1}{\sqrt{k_z^a}} e^{i[k_y^a y + k_z^a z)]} & \text{for } |y| < \frac{D_{\text{in}}}{2} \\ 0 & \text{otherwise} \end{cases}$$
(S6)

with $\{k_y^a\} = \left\{a(2\pi/D_{\text{in}}) \text{ such that } a \in \mathbb{Z} \text{ and } |k_y^a| < (2\pi/\lambda) \sin(\text{FOV}/2)\right\}$ and $(k_y^a)^2 + (k_z^a)^2 = (2\pi/\lambda)^2$.

Consider incident plane wave from a fixed angle θ_{in}^a within the FOV. Here, $E_x^a(y', z = 0) = f_a(y', z = 0)$, corresponding to one column of the transmission matrix, so $\beta_b = t_{ba}$. To perfectly focus to the focal spot, the field profile on the output surface (z = h) of the metalens should be proportional to the conjugated field radiated from a point source at the focal spot. Therefore,

$$E_{x}^{a}(y, z = h) = \begin{cases} A(\theta_{in}^{a}) \frac{e^{i\phi_{out}^{ideal}(y, \theta_{in}^{a})}}{[f^{2} + (y - f \tan \theta_{in}^{a})^{2}]^{1/4}} & \text{for } |y| < \frac{D_{out}}{2} \\ 0 & \text{otherwise} \end{cases}.$$
(S7)

The distance between a point (y, z = h) on the back surface of the metalens and the focal spot $(y = f \tan \theta_{in}, z = h + f)$ is $r = \sqrt{f^2 + (y - f \tan \theta_{in})^2}$. The amplitude factor $1/\sqrt{r}$ comes from the decay rate of the radiated field from a point source in 2D. We expand this ideal output in a basis of flux-orthogonal truncated plane waves,

$$g_b(y,z) = \begin{cases} \frac{1}{\sqrt{D_{\text{out}}}} \frac{1}{\sqrt{k_z^b}} e^{i[k_y^b y + k_z^b(z-h)]} & \text{for } |y| < \frac{D_{\text{out}}}{2} \\ 0 & \text{otherwise} \end{cases}$$
(S8)

with $\{k_y^b\} = \{b(2\pi/D_{\text{out}}) \text{ such that } b \in \mathbb{Z} \text{ and } |k_y^b| < 2\pi/\lambda \}$ and $(k_y^b)^2 + (k_z^b)^2 = (2\pi/\lambda)^2$. Projecting onto this basis, we obtain $E_x(y, z = h) = \sum_b \beta_b g_b(y, z = h)$ with

$$t_{ba} = \beta_b = \sqrt{\frac{k_z^b}{D_{\text{out}}}} \int_{-\frac{D_{\text{out}}}{2}}^{\frac{D_{\text{out}}}{2}} E_x^a(y, z = h) e^{-ik_y^b y} dy.$$
(S9)

In practice, we can approximate the continuous integration over *y* in Eq. (S9) by a discrete summation. A sampling spacing of $\Delta y = \lambda/2$ is chosen following the Nyquist-Shannon sampling theorem [4]. The discretized Eq. (S9) can be evaluated efficiently using fast Fourier transform:

$$t_{ba} \approx \Delta y \sqrt{\frac{k_z^b}{D_{\text{out}}}} e^{-ik_y^b (-\frac{D_{\text{out}}}{2} + \frac{\Delta y}{2})} \sum_{n=0}^{N-1} E_x^a (y_n, z = h) e^{-i\frac{2\pi}{N}bn},$$
(S10)

where $y \in [-D_{\text{out}}/2, D_{\text{out}}/2]$ is discretized to $\{y_n \equiv -\frac{D_{\text{out}}}{2} + \frac{\Delta y}{2} + n\Delta y\}$, and $N \equiv D_{\text{out}}/\Delta y \in \mathbb{Z}$ is the length of $\{y_n\}$.

Eq. (S7) includes an amplitude factor $A(\theta_{in}^a)$ that can depend on the incident angle. This means that each column of the transmission matrix has an undetermined amplitude prefactor. While the average transmission efficiency bound $N_{\text{eff}}/N_{\text{in}}$ in this work does not depend on a global amplitude prefactor, we do need to specify the relative amplitude between the columns. Here, we choose $A(\theta_{in}^a)$ such that $\sum_b |t_{ba}|^2$ is the same for all incident angles within the FOV. The reason is twofold. First, it is typically desirable that the focal power is independent of the incident angle so there is no vignetting in the image, and the focal power is proportional to the transmitted flux $\sum_b |t_{ba}|^2$ times the Strehl ratio (which is unity for an ideal lens). Second, in this work we are interested in the upper bound on the transmission efficiency, so we want to find the largest possible N_{eff} . A large N_{eff} corresponds to all possible inputs having similar total transmission, so making $\sum_b |t_{ba}|^2$ independent of the incident angle can increase N_{eff} .

3. TRANSMISSION EFFICIENCY OF HYPERBOLIC AND QUADRATIC METALENSES



Fig. S2. Transmission efficiency for 2D metalenses with (a) hyperbolic phase-shift profile and (b) quadratic phase-shift profile. Lens parameters: diameter $D = 100 \text{ }\mu\text{m}$, $\lambda = 532 \text{ }n\text{m}$.

As a verification, here we compare the transmission efficiency bound $N_{\rm eff}/N_{\rm in}$ with the transmission efficiency of actual metalenses.

We design 2D metalenses with hyperbolic [Eq. (S4)] and quadratic [5-7] phase-shift profiles

$$\Delta\phi_{\rm qua}(y) = -\frac{2\pi}{\lambda} \frac{y^2}{2f'},\tag{S11}$$

operating at wavelength $\lambda = 532$ nm, composed of ridges with thickness $h = 0.6 \mu$ m and varying widths between 45 nm and 200 nm. Each unit cell has a titanium dioxide ridge with refractive index $n_{ridge} = 2.43$ sitting on a silica substrate with $n_{sub} = 1.46$. The unit cell size is fixed at 240 nm. We calculate their transmission matrices by full-wave simulations using an open-source software MESTI [8, 9]. The simulation domain is discretized with 40 pixels per wavelength.

The transmission efficiency T_a at incident angle θ_{in}^a is

$$T_a = \sum_b |t_{ba}|^2 \tag{S12}$$

for a flux-normalized transmission matrix t_{ba} . Here we consider transmission averaged over incident angles with $|\theta_{in}^a| < 45^\circ$ (*ie.* $\langle T \rangle = \sum_{b,a} |t_{ba}|^2 / N_{in}$) and compare it to the efficiency bound N_{eff} / N_{in} . Figure S2 plots N_{eff} / N_{in} and $\langle T \rangle$ as a function of NA with the lens diameter fixed at $D = 100 \,\mu\text{m}$, for hyperbolic and quadratic

Figure S2 plots $N_{\text{eff}}/N_{\text{in}}$ and $\langle T \rangle$ as a function of NA with the lens diameter fixed at $D = 100 \,\mu\text{m}$, for hyperbolic and quadratic metalenses. Indeed, $\langle T \rangle$ is always below $N_{\text{eff}}/N_{\text{in}}$. Note that the range of NA is restricted to < 0.71 for the quadratic metalense because its largest possible effective NA is $1/\sqrt{2} = 0.71$ [7].

4. DEPENDENCE OF TRANSMISSION EFFICIENCY BOUND ON LENS PARAMETERS

Figure S3 plots the transmission efficiency bound $N_{\text{eff}}/N_{\text{in}}$ as a function of the FOV and D_{out} when $D_{\text{in}} = D_{\text{in}}^{\text{th}}$ and $D_{\text{in}} = D_{\text{out}}$, respectively, averaging over the other lens parameters. We see that FOV and D_{out} have very minor effects on the efficiency bound.

Some representative results of the transmission efficiency bound are plotted in Fig. S4.



Fig. S3. Transmission efficiency bound $N_{\text{eff}}/N_{\text{in}}$ of nonlocal metalenses (a) averaged over NA and output diameter D_{out} as a function of the FOV, and (b) averaged over NA and FOV as a function of the output diameter D_{out} , when $D_{\text{in}} = D_{\text{in}}^{\text{th}}$ and $D_{\text{in}} = D_{\text{out}}$.



Fig. S4. Transmission efficiency bound $N_{\text{eff}}/N_{\text{in}}$ of nonlocal metalenses (a) as a function of the FOV when $D_{\text{out}} = 300\lambda$, (b) as a function of the output diameter D_{out} when FOV = 80°, and (c) as a function of the NA when $D_{\text{out}} = 300\lambda$. Open circles and filled triangles represent the bound with $D_{\text{in}} = D_{\text{in}}^{\text{th}}$ and $D_{\text{in}} = D_{\text{out}}$, respectively.

5. THRESHOLD INPUT DIAMETER FROM $\partial \kappa / \partial D_{IN}$

Table S1 shows the threshold input diameter D_{in}^{th} for high transmission determined by $(\partial \kappa / \partial D_{in})|_{D_{in}=D_{in}^{th}} = 0.04/\lambda$, $0.4/\lambda$ or $4/\lambda$, and the corresponding transmission efficiency bound N_{eff}/N_{in} at $D_{in} = D_{in}^{th}$ when $D_{out} = 300\lambda$. We can see that the results are insensitive to the threshold value used, indicating that $\partial \kappa / \partial D_{in}$ is a robust indicator for the threshold. We use $(\partial \kappa / \partial D_{in})|_{D_{in}=D_{in}^{th}} = 0.4/\lambda$ in the main text, though other values will yield similar results.

6. $\psi(\theta_{IN})$ DEPENDENCE

The incident-angle-dependent but spatially-invariant global phase $\psi(\theta_{in})$ in Eq. (S1) has no influence on the focusing quality, so it can be chosen at will. The $\psi(\theta_{in}) = \psi_0(\theta_{in})$ in Eq. (S3) was chosen to minimize the required thickness of the metalens [1]. Another common choice is

$$\psi(\theta_{\rm in}) = \frac{2\pi}{\lambda} \sqrt{f^2 + (f \tan \theta_{\rm in})^2} \equiv \psi_c(\theta_{\rm in}), \tag{S13}$$

	$\partial \kappa / \partial D_{\rm in} = 0.04 / \lambda$		<i>∂κ/</i> ∂Ι	$\partial \kappa / \partial D_{\rm in} = 0.4 / \lambda$		$\partial \kappa / \partial D_{\rm in} = 4/\lambda$	
	$D_{\rm in}^{\rm th}$	$N_{\rm eff}/N_{\rm in}$	$D_{\rm in}^{\rm th}$	$N_{\rm eff}/N_{\rm in}$	$D_{\rm in}^{\rm th}$	$N_{\rm eff}/N_{\rm in}$	
NA = 0.2, FOV = 10°	294λ	100%	300λ	98.7%	300λ	98.7%	
NA = 0.5, FOV = 10°	260λ	99.8%	265λ	98.8%	274λ	95.9%	
NA = 0.9, FOV = 10°	133λ	96.0%	136λ	95.0%	146λ	89.3%	
NA = 0.2, FOV = 80°	293λ	98.4%	294λ	98.7%	296λ	99.0%	
NA = 0.5, FOV = 80°	260λ	92.8%	261λ	93.0%	264λ	93.4%	
NA = 0.9, FOV = 80°	133λ	84.2%	133λ	84.2%	136λ	84.6%	
NA = 0.2, FOV = 160°	292λ	97.3%	294λ	97.9%	296λ	98.3%	
NA = 0.5, FOV = 160°	260λ	90.6%	261λ	90.8%	264λ	91.1%	
NA = 0.9, FOV = 160°	132λ	87.4%	133λ	87.7%	136λ	88.0%	

Table S1. Threshold input diameter $D_{\rm in}^{\rm th}$ and transmission efficiency bound $N_{\rm eff}/N_{\rm in}$ at $D_{\rm out}=300\lambda$



Fig. S5. Threshold input diameter D_{in}^{th} and transmission efficiency bound N_{eff}/N_{in} at $D_{in} = D_{in}^{th}$ as a function of the output diameter D_{out} when different $\psi(\theta_{in})$ is applied. Lens parameters: NA = 0.7, FOV = 80°.

which makes $\Delta \phi_{ideal}(y = 0, \theta_{in}) = \phi_{out}^{ideal}(y = 0, \theta_{in}) - \phi_{in}(y = 0, \theta_{in}) = 0.$ As shown in Fig. S5, the threshold input diameter D_{in}^{th} and transmission efficiency bound N_{eff}/N_{in} at $D_{in} = D_{in}^{th}$ are the same for both choices of the global phase $\psi(\theta_{in})$.

7. COMPREHENSIVE DATA ON THE THRESHOLD INPUT DIAMETER

Figure 4(a,b) of the main text plots the threshold input diameter D_{in}^{th} for output diameter $D_{out} = 300\lambda$ and FOV = 140° respectively, with varying NA. Figures S6–S7 plot D_{in}^{th} for other D_{out} and other FOV.



Fig. S6. Threshold input diameter D_{in}^{th} as a function of the FOV for different lens diameter D_{out} and NA.



Fig. S7. Threshold input diameter D_{in}^{th} as a function of the output diameter D_{out} for different FOV and NA.

REFERENCES

S. Li and C. W. Hsu, "Thickness bound for nonlocal wide-field-of-view metalenses," Light. Sci. Appl. 11, 338 (2022). 1.

- 2. E. Hecht, "Chapter 5.2: Lenses," in *Optics*, 5ed, (Pearson Education Limited, 2017).
- 3. F. Aieta, P. Genevet, M. A. Kats, N. Yu, R. Blanchard, Z. Gaburro, and F. Capasso, "Aberration-free ultrathin flat lenses and axicons at telecom wavelengths based on plasmonic metasurfaces," Nano Lett. **12**, 4932–4936 (2012).
- 4. H. Landau, "Sampling, data transmission, and the Nyquist rate," Proc. IEEE 55, 1701–1706 (1967).
- 5. M. Pu, X. Li, Y. Guo, X. Ma, and X. Luo, "Nanoapertures with ordered rotations: symmetry transformation and wide-angle flat lensing," Opt. Express 25, 31471–31477 (2017).
- 6. A. Martins, K. Li, J. Li, H. Liang, D. Conteduca, B.-H. V. Borges, T. F. Krauss, and E. R. Martins, "On metalenses with arbitrarily wide field of view," ACS Photonics 7, 2073–2079 (2020).
- E. Lassalle, T. W. Mass, D. Eschimese, A. V. Baranikov, E. Khaidarov, S. Li, R. Paniagua-Dominguez, and A. I. Kuznetsov, "Imaging properties of large field-of-view quadratic metalenses and their applications to fingerprint detection," ACS Photonics 8, 1457–1468 (2021).
- 8. H.-C. Lin, Z. Wang, and C. W. Hsu, "MESTI," https://github.com/complexphoton/MESTI.m.
- 9. H.-C. Lin, Z. Wang, and C. W. Hsu, "Fast multi-source nanophotonic simulations using augmented partial factorization," Nat. Comput. Sci. 2, 815–822 (2022).