LASER & PHOTONICS REVIEWS
www.lpr-journal.org

Transmission Efficiency Limit for Nonlocal Metalenses

*Shiyu Li and Chia Wei Hsu**

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. This study introduces 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 shows that the transmission efficiency must decrease with the numerical aperture for the commonly adopted designs with equal entrance and output aperture diameters. It also shows that reducing the size of the entrance aperture can raise the efficiency bound. This study reveals a fundamental limit on the transmission efficiency as well as provides guidance for the design of high-efficiency multi-channel optical systems.

1. Introduction

Over the past decade, nanophotonic design and fabrication became more and more advanced. Oftentimes, what limits the device's 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 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, such as metalenses[2–6] with a wide field of view (FOV). Their large angular diversity puts a bound on the performance and spatial footprint, such as bandwidth,[7] resolution,[8] and thickness.[9] Meanwhile, another major consideration of metalenses is the efficiency. For example, 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 across all angles compatible with the angle dependence in the desired response? What is the highest efficiency allowed by fundamental laws?

Here, we introduce a matrix-based formalism that sets a fundamental bound on the channel-averaged transmission efficiency of any passive linear multi-channel system given its functionality. Then we apply it to diffraction-limited lenses as an example. The bound applies to any design that realizes the target functionality, so it is applicable not only to metasurfaces but also to conventional refractive and diffractive optical systems.

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.[10,11] Inverse design,[10–18] grating averaging technique[19], and stitching separately designed sections together[20–22] 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.[23,24] 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 FOV.[9]

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.[9,25,26] Their strong nonlocality can be realized by the interaction among adjacent meta-atoms[27] or the excitation of guided resonances that travel across the device over large distances.[28,29] Nonlocal metalenses based on doublets[30,31] or aperture stops[32,33] can support focusing efficiencies higher than 50% over a wide FOV, but with NA lower than 0.5. Multilayer 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%.[34] 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

S. Li, C. W. Hsu
Ming Hsieh Department of Electrical and Computer Engineering
University of Southern California
Los Angeles, California 90089, USA
E-mail: cwhsu@usc.edu

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/lpor.202300201

DOI: 10.1002/lpor.202300201

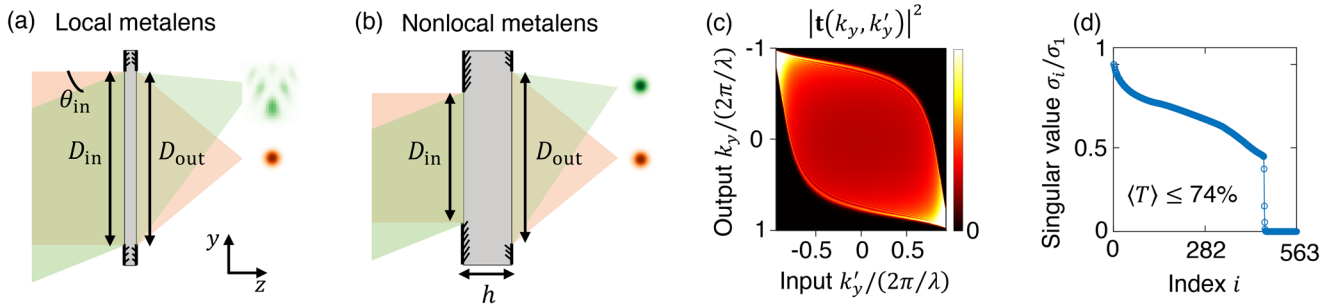

Figure 1. Nonlocal metalenses and their transmission matrix. a) A local metalens with subwavelength thickness, for which the input and output apertures have the same size, i.e., $D_{in} = D_{out}$. Diffraction-limited focusing can only be achieved within a narrow range of incident angles. b) A nonlocal 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 Equation (4) as $\langle T \rangle \leq 74\%$.

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.

2. Results

2.1. Nonlocality

As schematically illustrated in **Figure 1**a, subwavelength-thick local metalenses (such as metalenses with the hyperbolic phase profile[35]) perform well only over a limited input angular range. Nonlocal metalenses can achieve diffraction-limited focusing over a much wider FOV but need a minimal thickness to provide the nonlocality.[9,36] 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 Figure S1 (Supporting Information), 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_{in} and output diameter D_{out} can be different [Figure 1b]. The different aperture sizes provide an additional degree of freedom compared to local metalenses with subwavelength thicknesses (for which $D_{in} = D_{out}$ intrinsically).

2.2. Transmission Matrix from the Target Response

While the formalism introduced in this work applies to different systems, for concreteness, we will consider metalenses here. Consider monochromatic, transverse magnetic waves of a 2D system invariant in the x dimension, where $\mathbf{E} = E_x(y,z)\hat{x}$ and $\mathbf{H} = H_y(y,z)\hat{y} + H_z(y,z)\hat{z}$, in s-polarization. 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 chief ray and a marginal ray from a point $\mathbf{r}_b = (y, z = h)$ on the back surface ($z = h$) of the metalens to the focus $\mathbf{r}_f(\theta_{in}) = (y = f\tan\theta_{in}, z = f + h)$ for incident angle θ_{in}. This yields an incident-angle-dependent ideal phase profile on the back surface,[9,32,37]

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

where λ is the wavelength, f and h are the focal length and lens thickness respectively, and $\psi(\theta_{in})$ is an angle-dependent but spatially invariant global phase with no effect on the focusing quality. To perfectly focus to the focal spot $\mathbf{r}_f(\theta_{in})$, the field profile on the back surface of the metalens should be proportional to the conjugated field radiated from a point source at $\mathbf{r}_f$. Therefore,

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

here, $A(\theta_{in})$ specifies the transmitted amplitude for incident angle θ_{in}. The factor $1/\sqrt{r}$ comes from the decay rate of the radiated field from a point source in 2D, with $r = |\mathbf{r}_b - \mathbf{r}_f|$. Note that here we consider transmitted field across the entire back aperture of the lens to be used for focusing, for maximal utility of the lens area. An alternative telecentric configuration that uses a subset of the back aperture is considered in Section S5 (Supporting Information).

Equations (1, 2) specify the target response, from which we can write down the transmission matrix. To do so, we project the incident and the transmitted wavefronts onto a finite number of channels. The incoming wavefront $E_x(y', z = 0)$ at the front surface of the system ($z = 0$) can be projected onto N_{in} propagating plane waves with input transverse momenta $k'_y = \{k_y^a\} = (2\pi/\lambda)\sin\theta_{in}$, yielding complex-valued amplitude α_a for the a-th plane-wave input. Each incident plane wave is truncated to the size of the input aperture, i.e., $|y'| < D_{in}/2$. The input momentum is restricted to the FOV of interest, $|k_y^a| < (2\pi/\lambda)\sin(\text{FOV}/2)$. We sample k_y^a within the FOV with $2\pi/D_{in}$ spacing at the Nyquist–Shannon sampling rate,[38] which is sufficient to reconstruct the incident wavefront since the wavefront is

blocked beyond the entrance size D_{in}. Doing so also makes incident channels flux-orthogonal (see Section S2, Supporting Information for more information). 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 $|y| < D_{\text{out}}/2$, 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 $2\pi/D_{\text{out}}$ at the Nyquist rate. The incoming and transmitted wavefronts are then parametrized as column vectors $\boldsymbol{\alpha} = [\alpha_1; \cdots; \alpha_{N_{\text{in}}}]$ and $\boldsymbol{\beta} = [\beta_1; \cdots; \beta_{N_{\text{out}}}]$. For any linear optical system, the input and the output vectors are related through the transmission matrix elements $t_{ba} = \mathbf{t}(k_y^b, k_y^a)$,

$$\beta_b = \sum_{a=1}^{N_{\text{in}}} t_{ba}\alpha_a \tag{3}$$

or $\boldsymbol{\beta} = \mathbf{t}\boldsymbol{\alpha}$ in matrix notation. The incident flux in z is proportional to $\sum_a |\alpha_a|^2$, and the transmitted flux is proportional to $\sum_b |\beta_b|^2$. The transmission efficiency for incident angle θ_{in}^a is $T_a = \sum_b |t_{ba}|^2$.

Fourier transforming the output field of the ideal metalens in Equation (2) from y to k_y yields the transmission matrix $\mathbf{t}$; see Section S2 (Supporting Information) for details. 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 $D_{\text{in}} = D_{\text{out}} = 300\lambda$, NA $= \sin(\arctan(D_{\text{out}}/(2f))) = 0.8$ and FOV $= 140°$ is shown in Figure 1c.

For an ideal lens with no vignetting, here we choose the amplitude $A(\theta_{\text{in}}^a)$ in Equation (2) such that the transmission efficiency $T_a = \sum_b |t_{ba}|^2$ is the same for all incident angles θ_{in}^a within the FOV. However, an overall amplitude prefactor of the transmission matrix that determines this T_a is yet to be specified. Larger transmission efficiency is desired. But as shown below, setting $T_a = 1$ for all a will violate energy conservation because doing so leads to more transmitted flux than the incident flux for some non-plane-wave incident wavefronts, which is not allowed in a passive system. Next, we will derive a bound on the transmission efficiency for systems that yield the desired angle-dependent response while satisfying energy conservation.

The paragraphs above consider metalenses. One can similarly write down the desired transmission matrix for other multichannel optical systems of interest.

2.3. Transmission Efficiency Bound

For any system of interest, the transmission efficiency averaged over inputs within the prescribed FOV is $\langle T\rangle = \sum_a T_a/N_{\text{in}}$. To obtain an upper bound on $\langle T\rangle$ without specifying the overall amplitude prefactor of the transmission matrix, we consider the singular values $\{\sigma_i\}$ of the transmission matrix.[39] With the singular value decomposition, the transmission matrix is factorized into orthonormal inputs and outputs as $\mathbf{t} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger = \sum_i \sigma_i(u_i v_i^\dagger)$, 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. Per min-max theorem, $\sigma_{\max}^2$ and $\sigma_{\min}^2$ are the highest-possible and lowest-possible transmission efficiency among all possible input states (singular or not). Figure 1d shows the normalized singular values of the transmission matrix in Figure 1c. Since $\text{tr}(\mathbf{t}^\dagger\mathbf{t}) = \Sigma_a(\mathbf{t}^\dagger\mathbf{t})_{aa} = \sum_{b,a}|t_{ba}|^2$ equals the sum of the eigenvalues $\lambda_i = \sigma_i^2$ of matrix $\mathbf{t}^\dagger\mathbf{t}$, we have $\langle T\rangle = \sum_{b,a}|t_{ba}|^2/N_{\text{in}} = \sum_i \sigma_i^2/N_{\text{in}}$. The obstacle is that the overall amplitude prefactor of $\{\sigma_i\}$ is not known.

Energy conservation imposes that the transmitted energy cannot exceed the input energy in a passive system, so we must have $\sigma_i^2 \le 1$ for all i with the actual overall amplitude prefactor. One way to establish a bound on the average efficiency is to recognize that the highest possible transmission efficiency is unity, so $\langle T\rangle \le (\sum_i \sigma_i^2/N_{\text{in}})/\sigma_{\max}^2$; this expression does not depend on the overall amplitude prefactor. But since the response of a designed structure generally does not perfectly match the target response, the actual maximal singular value can also deviate from that of the target, and a bound that depends sensitively on one singular value will not be robust. To provide a robust bound that collectively depends on the system's response to all possible inputs, we quantify the effective number of high-transmission channels through an inverse participation ratio[40] as $N_{\text{eff}} = (\sum_{i=1}\sigma_i^2)^2/\sum_{i=1}\sigma_i^4$. This N_{eff} considers all singular values and is independent of the overall prefactor of the transmission matrix. If $\{\sigma_i\}$ were an array of zeros and ones, this N_{eff} would simply count the number of ones in this array. For a perfectly transmitting system, all of the singular values are ones, and $N_{\text{eff}} = N_{\text{in}}$. For a generic set of singular values, N_{eff} provides an effective count of high-transmission channels. To bound the average efficiency using N_{eff}, we recognize that $\sum_i \sigma_i^4 \le \sum_i \sigma_i^2$ since $\sigma_i^2 \le 1$, so that

$$\langle T\rangle \equiv \frac{\sum_{b,a}|t_{ba}|^2}{N_{\text{in}}} = \frac{N_{\text{eff}}}{N_{\text{in}}}\frac{\sum_i \sigma_i^4}{\sum_i \sigma_i^2} \le \frac{N_{\text{eff}}}{N_{\text{in}}} \tag{4}$$

In Figure 1d, $\langle T\rangle \le N_{\text{eff}}/N_{\text{in}} \approx 74\%$.

Given any angle-dependent response of interest, one can write down the corresponding transmission matrix (without specifying the overall amplitude prefactor), evaluate its N_{in} singular values $\{\sigma_i\}$ and the associated N_{eff}, and the efficiency bound follows as $\langle T\rangle \le N_{\text{eff}}/N_{\text{in}}$. Equation (4) is rigorously derived, with the only assumption being passivity, $\sigma_i^2 \le 1$. Therefore, this bound is fundamental and applies to any passive linear optical system for which the desired response is specified. Unlike the thickness bound of References [9, 36], Equation (4) does not rely on any empirical relation or any heuristic argument.

A bound is useful only when it is sufficiently tight. Figure S2 (Supporting Information) shows from full-wave numerical simulations[41] that Equation (4) indeed provides a reasonably tight upper bound for the averaged transmission efficiency $\langle T\rangle$, considering hyperbolic and quadratic metalenses as examples.

Figure 2a shows the transmission efficiency bound $N_{\text{eff}}/N_{\text{in}}$ as a function of NA for aberration-free nonlocal metalenses with FOV larger than the threshold shown in Figure S1 (Supporting Information). In general, both the FOV and the output diameter D_{out} can affect the transmission efficiency. For the metalens example considered here, Figures S3 and S4 (Supporting Information) show that they happen to have a small influence on $N_{\text{eff}}/N_{\text{in}}$, so in Figure 2a we map out how the efficiency bound depends on

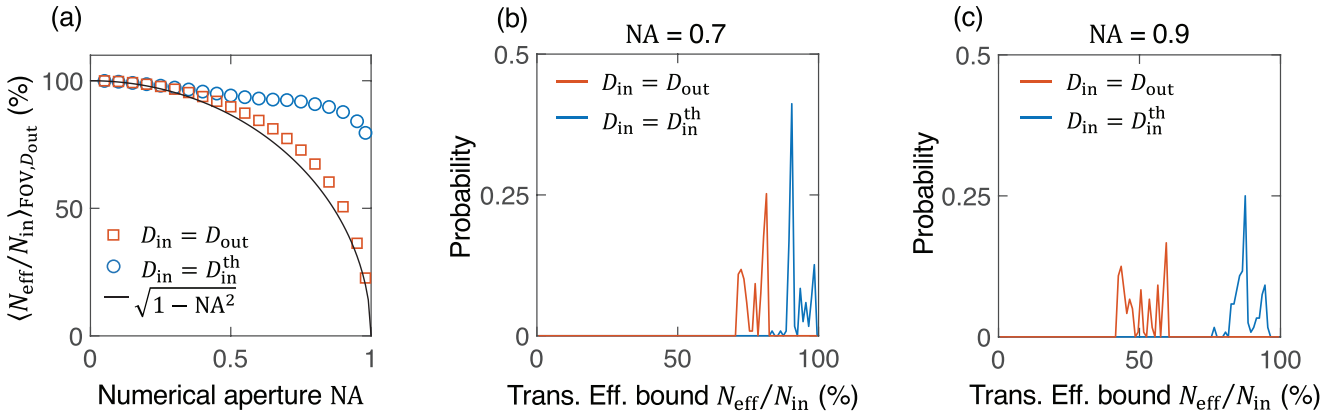

Figure 2. Transmission efficiency bound $N_{\rm eff}/N_{\rm in}$ of ideal nonlocal metalenses when $D_{\rm in} = D_{\rm out}$ and when using $D_{\rm in} = D_{\rm in}^{\rm th}$ from Equation (5). a) $N_{\rm eff}/N_{\rm in}$ averaged over FOV and output diameter $D_{\rm out}$ as a function of NA. The black line is $\sqrt{1-{\rm NA}^2}$. 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.

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 $\sqrt{1-{\rm NA}^2}$. This bound applies regardless of how complicated or optimized the design is, what materials are used, and how thick the device is.

In Figure 2b,c, we show the distribution (among different FOV and output diameters) of the transmission efficiency bound with NA = 0.7 and 0.9, considering 120 combinations of different FOV and $D_{\rm out}$ for each NA. The bound is consistent with the inverse design results of Reference [34], where the achieved average absolute focusing efficiency (considering only the transmitted power within three full-widths at half-maximum around the focal peak) is 25% for a nonlocal metalens with NA = 0.7 and FOV = 80°.

In general, the transmission efficiency limit can depend on the details of the target response. In Figure S5 (Supporting Information), we show the efficiency bound for telecentric lenses where only positions within an effective output diameter $D_{\rm out}^{\rm eff} = D_{\rm in}$ are used for focusing, which is comparable to the non-telecentric bound when the FOV is small. In Figure S6 (Supporting Information), we consider non-telecentric systems described by Equation (2) and show that the choice of the angle-dependent amplitude $A(\theta_{\rm in}^a)$ has little effect on the efficiency bound.

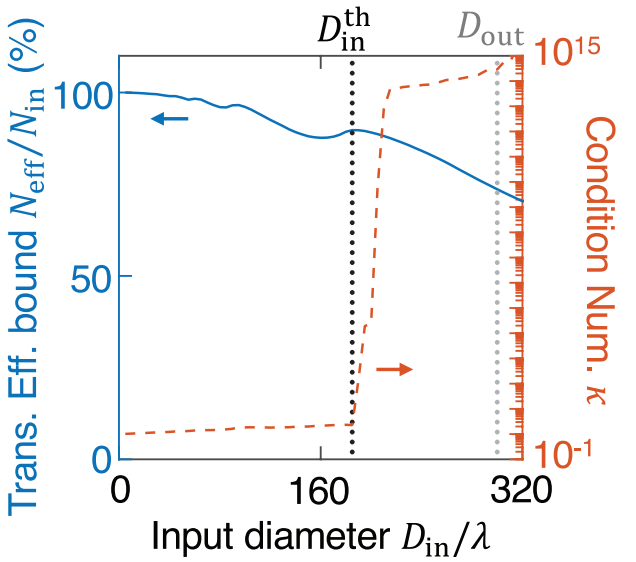

Figure 3. The transmission efficiency bound $N_{\rm eff}/N_{\rm in}$ and the condition number κ of the ideal transmission matrix as a function of the entrance diameter $D_{\rm in}$. Black and gray vertical dotted lines indicate $D_{\rm in}^{\rm th} = 185\lambda$ and $D_{\rm out} = 300\lambda$. Lens parameters: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

2.4. 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 Figure 1d where $D_{\rm in} = D_{\rm out} = 300\lambda$ and 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 $|k_y'| < (2\pi/\lambda)\sin({\rm FOV}/2)$ 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 $D_{\rm in}$ to use? While reducing $D_{\rm in}$ 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 Figure 3 shows that κ abruptly shoots up by many orders of magnitude when the input

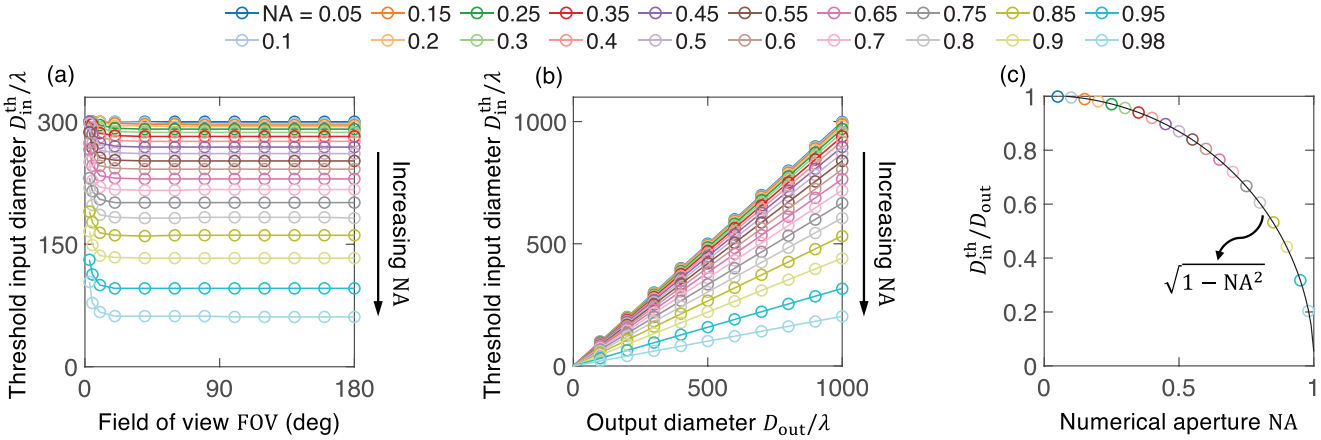

Figure 4. Dependence of the threshold input diameter D_{in}^{th} on the lens parameters. a) D_{in}^{th} as a function of the FOV when $D_{out} = 300\lambda$. b) D_{in}^{th} as a function of the output diameter D_{out} when FOV = 140°. Symbols are D_{in}^{th} computed from the ideal transmission matrices, 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 $\sqrt{1-NA^2}$.

diameter D_{in} raises above a threshold that we label as D_{in}^{th} (black dotted line). When $D_{in} < D_{in}^{th}$, κ is of order unity, all singular values are comparable with no zero-transmission wavefronts, so we have $N_{eff} \approx N_{in}$, and the transmission efficiency bound is close to unity. When $D_{in} > D_{in}^{th}$, near-zero-transmission wavefronts start to appear, which results in a fast reduction of the transmission efficiency bound. Therefore, this threshold value D_{in}^{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 $\partial\kappa/\partial D_{in}$, which transitions from near zero to a very large number at D_{in}^{th}. Table S1 (Supporting Information) shows that different threshold values for $\partial\kappa/\partial D_{in}$ yield almost identical D_{in}^{th} and N_{eff}/N_{in}. Note that the choice of the global phase $\psi(\theta_{in})$ in Equation (1) does not influence D_{in}^{th} and the transmission efficiency bound, as shown in Figure S7 (Supporting Information).

To guide future designs, it is desirable to know how D_{in}^{th} depends on the various lens parameters. **Figure 4** plots D_{in}^{th} as a function of the FOV, output diameter D_{out}, and NA. As described in Section 2.1, lenses with a very small FOV do not require nonlocality; for such local metalenses, we find $D_{in}^{th} \approx D_{out}$ as expected from the schematic in Figure 1a. As the FOV increases and nonlocality emerges, we observe in Figure 4a that D_{in}^{th} drops below D_{out} (as expected from the preceding discussions) and reaches a constant value that depends on NA but not on the FOV. This D_{in}^{th} for nonlocal metalenses is proportional to D_{out} [Figure 4b]. Figure 4a,b fix $D_{out} = 300\lambda$ and FOV = 140° respectively; other lens parameters share similar dependencies (Figures S8 and S9, Supporting Information). In Figure 4c, we find empirically that the NA dependence is well described by

$$D_{in}^{th} = D_{out}\sqrt{1-NA^2} \tag{5}$$

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 Figure 2a–c the transmission efficiency bound when the input aperture size D_{in} is set to the optimal D_{in}^{th} in Equation (5). We observe a large transmission efficiency bound N_{eff}/N_{in} that overcomes the $\sqrt{1-NA^2}$ limit when equal entrance and output apertures are used.

3. Discussion

While we use the efficiency of high-NA nonlocal metalenses as the example here, the formalism we introduce is very general and applies to any passive linear optical system. Once the target response is specified, this formalism can be used to put a bound on the channel-averaged transmission efficiency. The bound holds for any structural design and material composition, including not only metasurfaces but also volumetric nanophotonic structures and conventional optical systems.

We note that even though the transmission efficiency limit here depends only on the target functionality and not on the device thickness, this does not mean that thin and thick structures will perform similarly. Thin structures have a limited degree of nonlocality, which may not be sufficient for realizing the target angle-dependent response of interest[9]—not even with a low transmission efficiency. The efficiency bound here and the thickness bound of Reference [9] are independent constraints.

In establishing a target-driven thickness bound, Reference [9] had to use an intuition-based and empirically established relation between lateral spreading and thickness, and Reference [36] had to use a heuristic relation between channel number and thickness and the estimation of a maximal internal angle that is not clearly defined in the presence of wave scattering. In contrast, the efficiency bound here is rigorously derived with no heuristics or empirical assumption and is clearly defined.

When the input diameter is above its threshold value D_{in}^{th}, the efficiency loss may be associated with vignetting but is not necessarily so. We also note that even an ideal system with 100% transmission efficiency for all incident angles will still experience

a natural falloff of the illumination flux density governed by the $\cos^4\theta$ law due to projection,[42] which is distinct from the transmission efficiency limit here.

Nonlocal metasurfaces open up a wide range of applications and tailored angle-dependent responses that are impossible for traditional local metasurfaces.[26] The efficiency bound in this work provides valuable guidance for this rapidly evolving field. For example, this work suggests that inverse-designed multilayer wide-FOV metalenses such as in Reference [34] may benefit from expanding the size of each successive layer.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors thank X. Gao for helpful discussions. This work was supported by the National Science Foundation CAREER award (ECCS-2146021) and the Sony Research Award Program.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

All data needed to evaluate the conclusions in this study are presented in the paper and in the supplemental document.

Keywords

fundamental limits, nonlocal metasurfaces, transmission efficiency

Received: March 2, 2023
Revised: May 23, 2023
Published online: June 17, 2023

[1] P. Chao, B. Strekha, R. Kuate Defo, S. Molesky, A. W. Rodriguez, *Nat. Rev. Phys.* **2022**, *4*, 543.
[2] P. Lalanne, P. Chavel, *Laser Photonics Rev.* **2017**, *11*, 1600295.
[3] M. Khorasaninejad, F. Capasso, *Science* **2017**, *358*, eaam8100.
[4] H. Liang, A. Martins, B.-H. V. Borges, J. Zhou, E. R. Martins, J. Li, T. F. Krauss, *Optica* **2019**, *6*, 1461.
[5] M. Pan, Y. Fu, M. Zheng, H. Chen, Y. Zang, H. Duan, Q. Li, M. Qiu, Y. Hu, *Light Sci. Appl.* **2022**, *11*, 195.
[6] A. Arbabi, A. Faraon, *Nat. Photonics* **2023**, *17*, 16.
[7] K. Shastri, F. Monticone, *EPJ Appl. Metamat.* **2022**, *9*, 16.
[8] A. Martins, J. Li, B.-H. V. Borges, T. F. Krauss, E. R. Martins, *Nanophotonics* **2022**, *11*, 1187.
[9] S. Li, C. W. Hsu, *Light Sci. Appl.* **2022**, *11*, 338.
[10] H. Chung, O. D. Miller, *Opt. Express* **2020**, *28*, 6945.
[11] K. Schab, L. Jelinek, M. Capek, M. Gustafsson, *Opt. Express* **2022**, *30*, 45705.
[12] D. Sell, J. Yang, S. Doshay, J. A. Fan, *Adv. Opt. Mater.* **2017**, *5*, 1700645.
[13] R. Pestourie, C. Pérez-Arancibia, Z. Lin, W. Shin, F. Capasso, S. G. Johnson, *Opt. Express* **2018**, *26*, 33732.
[14] H. Liang, Q. Lin, X. Xie, Q. Sun, Y. Wang, L. Zhou, L. Liu, X. Yu, J. Zhou, T. F. Krauss, J. Li, *Nano Lett.* **2018**, *18*, 4460.
[15] M. Mansouree, A. Arbabi, in *2018 Conference on Lasers and Electro-Optics (CLEO)*, **2018**, https://doi.org/10.1364/CLEO_QELS.2018.FF1F.7.
[16] T. Phan, D. Sell, E. W. Wang, S. Doshay, K. Edee, J. Yang, J. A. Fan, *Light Sci. Appl.* **2019**, *8*, 48.
[17] K.-F. Lin, C.-C. Hsieh, S.-C. Hsin, W.-F. Hsieh, *Appl. Opt.* **2019**, *58*, 8914.
[18] Z.-P. Zhuang, R. Chen, Z.-B. Fan, X.-N. Pang, J.-W. Dong, *Nanophotonics* **2019**, *8*, 1279.
[19] A. Arbabi, E. Arbabi, M. Mansouree, S. Han, S. M. Kamali, Y. Horie, A. Faraon, *Sci. Rep.* **2020**, *10*, 7124.
[20] S. J. Byrnes, A. Lenef, F. Aieta, F. Capasso, *Opt. Express* **2016**, *24*, 5110.
[21] 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, A. I. Kuznetsov, *Nano Lett.* **2018**, *18*, 2124.
[22] M. Kang, Y. Ra'di, D. Farfan, A. Alù, *Phys. Rev. Appl.* **2020**, *13*, 044016.
[23] N. M. Estakhri, A. Alù, *Phys. Rev. X* **2016**, *6*, 041008.
[24] V. S. Asadchy, M. Albooyeh, S. N. Tcvetkova, A. Díaz-Rubio, Y. Ra'di, S. Tretyakov, *Phys. Rev. B* **2016**, *94*, 075142.
[25] A. Overvig, A. Alù, *Laser Photonics Rev.* **2022**, *16*, 2100633.
[26] K. Shastri, F. Monticone, *Nat. Photonics* **2023**, *17*, 36.
[27] M. G. Silveirinha, *Phys. Rev. Lett.* **2009**, *102*, 193903.
[28] S. Fan, J. D. Joannopoulos, *Phys. Rev. B* **2002**, *65*, 235112.
[29] X. Gao, C. W. Hsu, B. Zhen, X. Lin, J. D. Joannopoulos, M. Soljačić, H. Chen, *Sci. Rep.* **2016**, *6*, 31908.
[30] A. Arbabi, E. Arbabi, S. M. Kamali, Y. Horie, S. Han, A. Faraon, *Nature Commun.* **2016**, *7*, 13682.
[31] B. Groever, W. T. Chen, F. Capasso, *Nano Lett.* **2017**, *17*, 4902.
[32] M. Y. Shalaginov, S. An, F. Yang, P. Su, D. Lyzwa, A. M. Agarwal, H. Zhang, J. Hu, T. Gu, *Nano Lett.* **2020**, *20*, 7429.
[33] C.-Y. Fan, C.-P. Lin, G.-D. J. Su, *Sci. Rep.* **2020**, *10*, 15677.
[34] Z. Lin, C. Roques-Carmes, R. E. Christiansen, M. Soljačić, S. G. Johnson, *Appl. Phys. Lett.* **2021**, *118*, 041104.
[35] F. Aieta, P. Genevet, M. A. Kats, N. Yu, R. Blanchard, Z. Gaburro, F. Capasso, *Nano Lett.* **2012**, *12*, 4932.
[36] D. A. Miller, *Science* **2023**, *379*, 41.
[37] A. Kalvach, Z. Szabó, *J. Opt. Soc. Am. B* **2016**, *33*, A66.
[38] H. Landau, *Proc. IEEE* **1967**, *55*, 1701.
[39] D. A. Miller, *Adv. Opt. Photonics* **2019**, *11*, 679.
[40] C. W. Hsu, S. F. Liew, A. Goetschy, H. Cao, A. Douglas Stone, *Nat. Phys.* **2017**, *13*, 497.
[41] H.-C. Lin, Z. Wang, C. W. Hsu, *Nat. Comput. Sci.* **2022**, *2*, 815.
[42] S. Ray, in, *Applied Photographic Optics*, 3 ed., Routledge, England, UK**2002**.