npj computational materials

www.nature.com/npjcompumats

ARTICLE OPEN

Check for updates

Low-overhead distribution strategy for simulation and optimization of large-area metasurfaces

Jinhie Skarda[1,5], Rahul Trivedi[1,2,3,5]✉, Logan Su[1,5], Diego Ahmad-Stein[1], Hyounghan Kwon[1], Seunghoon Han[4], Shanhui Fan[1] and Jelena Vučković[1]✉

Fast and accurate electromagnetic simulation of large-area metasurfaces remains a major obstacle in automating their design. In this paper, we propose a metasurface simulation distribution strategy which achieves a linear reduction in the simulation time with the number of compute nodes. Combining this distribution strategy with a GPU-based implementation of the Transition-matrix method, we perform accurate simulations and adjoint sensitivity analysis of large-area metasurfaces. We demonstrate ability to perform a distributed simulation of large-area metasurfaces (over $600\lambda \times 600\lambda$), while accurately accounting for scatterer-scatterer interactions significantly beyond the locally periodic approximation.

npj Computational Materials (2022)8:78 ; https://doi.org/10.1038/s41524-022-00774-y

INTRODUCTION

Being able to achieve full phase control of optical fields is a central challenge in optical engineering, with diverse applications in imaging, sensing, augmented, and virtual reality systems[1,2]. The past decades have seen a rapid development of metasurface-based optical elements that exploit collective scattering properties of subwavelength structures for phase-shaping the incoming fields and are significantly more compact and integrable when compared to the conventional refractive optical elements[3–9]. The most commonly adopted metasurface-design strategy proceeds in two steps — first, a library of periodic meta-atoms with varying transmission amplitudes and phases is generated by varying a few geometric parameters specifying the meta-atom. *Next*, an aperiodic meta-surface is generated by laying out the periodic meta-atoms corresponding to the target spatially-varying phase profile[10–17]. This approach suffers from two major limitations — first, the resulting metasurface should be almost periodic, and thus this strategy cannot be used for reliably designing rapidly varying phase-profiles. Second, generating the metasurface library becomes increasingly difficult for multi-functional design problems. For instance, while it is usually not difficult to generate a library for designing a simple phase-mask operating at a few operating modes[15,18,19], it becomes increasingly difficult to scale up the number of modes since the same metasurface is required to simultaneously satisfy multiple design conditions corresponding to the different input modes.

Fully automating design of metasurfaces can provide a potential solution to this problem. Gradient-based optimization has been successful in designing integrated optical elements that are more compact, robust and high performing than their classical counterparts[20–27]. A key ingredient in these approaches is the ability to rapidly simulate the full electromagnetic structure. This presents a challenge for metasurface designs, since practical metasurfaces could be $\sim 10^2$–$10^3\lambda$ in the linear dimension, making it impractical to use general-purpose electromagnetic solvers such as Finite-Difference Time-Domain (FDTD)[28], Finite-Difference Frequency-Domain (FDFD)[29], or Finite Element Method (FEM)[30]. Inverse-design approaches that use discrete general-purpose electromagnetic solvers to simulate and design the full surface are limited to small design areas or a small number of optimization iterations[31,32], or restrict the parameter space through a specific symmetry that allows for fast simulations[33–35]. Consequently, nearly all the current methods for inverse-designing large-scale 3D metasurfaces rely on approximate electromagnetic simulations of the metasurface locally using either periodic or radiation boundary conditions[9,36–47], which do not accurately account for interactions between different meta-atoms. These approaches are thus fundamentally limited to designing metasurfaces with slow phase variations due to the implicit local approximation. A coupled-mode formalism can also be applied for metasurface simulation and optimization[48] but this approach is not guaranteed to yield exact fields, particularly for metasurfaces with multiple low quality-factor modes.

In this paper, we propose and demonstrate a numerically accurate simulation strategy that can be used to design and analyze large-area metasurfaces. Our strategy relies on a distribution of the simulation method where the simulation time scales linearly with the compute resources. This is achieved by a Nyquist-sampling decomposition of the fields incident on the metasurface, similar to that used recently to characterize the discrete impulse response of aperiodic metasurfaces[49]. Our distribution strategy, by ensuring minimal communication between compute nodes, allows for a linear reduction in the simulation time with the number of compute nodes, indicating that arbitrarily large metasurfaces can be simulated in reasonable time with sufficiently large number of compute nodes. On each compute node, we implement a GPU-based transition-matrix (T-matrix) simulation[50–52]. Though there are GPU-optimized FDTD implementations that allow fast simulation of unit-cells up to $100\lambda \times 100\lambda$[53], these approaches do not currently provide a low-overhead means of parallel simulation distribution. We demonstrate numerically accurate simulations of metasurfaces of size 1 mm × 1 mm at a wavelength of 1.55 µm (about $645\lambda \times 645\lambda$) on a cluster of 48 GPU nodes. Finally, we demonstrate the ability to efficiently compute the gradients with respect to both the geometry and the positions of the meta-atoms,

[1]E. L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA. [2]Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, 85748 Garching, Germany. [3]Department of electrical and computer engineering, University of Washington, Seattle 98195, USA. [4]Samsung Advanced Institute of Technology, Samsung Electronics, Suwon-si, Gyeonggi-do 443-803, Republic of Korea. [5]These authors contributed equally: Jinhie Skarda, Rahul Trivedi, Logan Su. ✉email: rtriv@uw.edu; jela@stanford.edu

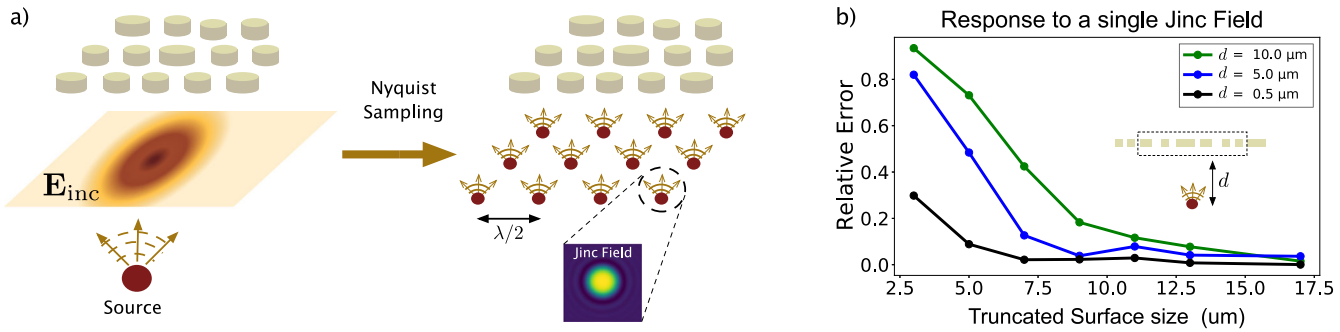
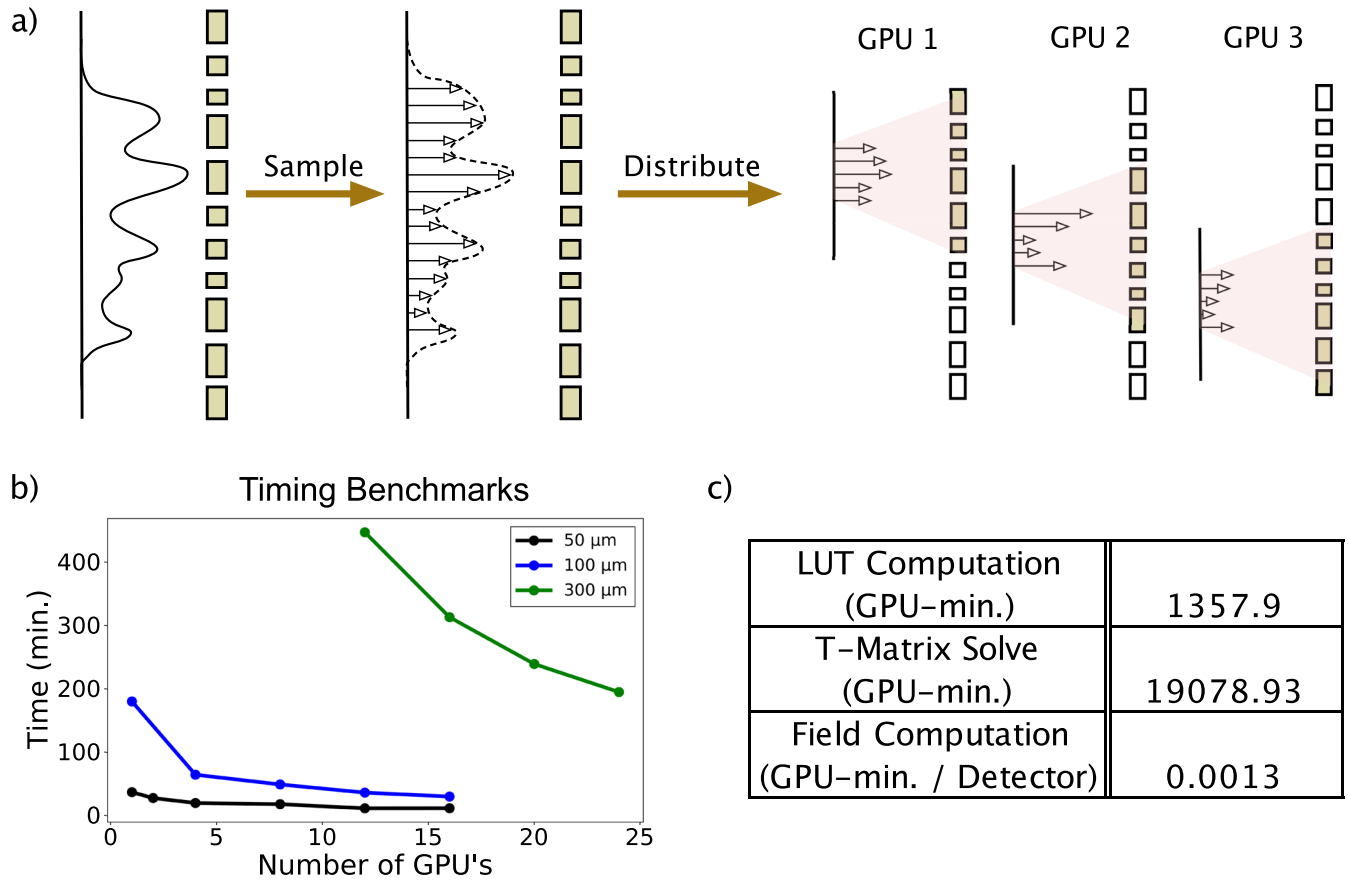


Fig. 1 Nyquist sampling of bandlimited incident field. a Schematic of Nyquist sampling of the incident electric field, which is bandlimited because it is propagating. **b** Percent error in scattered field power versus spatial-extent of metasurface included in the simulation for a single jinc source placed 10 μm (green), 5 μm (blue), and 0.5 μm (black) from the metasurface. The full metasurface is a 25 μm × 25 μm metasurface with focal length of 10 μm, and the surface size on the x-axis of this convergence plot refers to the spatial-extent around the center of this metasurface that is included in the simulation. The y-axis relative error is computed assuming the simulation including the full metasurface is the converged result.



LUT Computation (GPU-min.)	1357.9
T-Matrix Solve (GPU-min.)	19078.93
Field Computation (GPU-min. / Detector)	0.0013

Fig. 2 Low-overhead parallelization scheme for simulation of arbitrarily large metasurfaces. a Schematic of the simulation distribution scheme — the incident field is first sampled and represented as a superposition of jinc sources, and then smaller groups of jinc sources and the locally surrounding metasurface regions are simulated on independent GPUs. **b** Total simulation time versus number of V100 GPU's used for simulation for a 50 μm (black), 100 μm (blue), and 300 μm (green) metasurface. All metasurfaces have focal length of 25 μm and are designed from a library of silicon cylinders with height 940 nm, radii range of 50–250 nm, lattice period of 1070 nm, air background, and source wavelength of 1550 nm (based on scatterer library from Arbabi et al.[61]). **c** Computation time for the key stages of the large-area 1 mm × 1 mm metasurface simulation (metalens with focal length 0.4 mm designed with the same scatterer library used in (b)): top row – computing the Look-Up Tables (LUT) used to efficiently perform T-matrix simulation (Supplementary Note 1); middle row – computing the T-matrices (Supplementary Note 1.b) and solving the resulting linear system of equations for the scattered field coefficients (Supplementary Note 1.c, Supplementary Eq. 23); bottom row – computing the E and H fields from the scattered field coefficients for each desired detector point (Supplementary Note 1.c, Supplementary Eq. 24).

thus enabling the application of optimization-based design to large-scale metasurfaces.

RESULTS

Low-overhead multi-GPU simulation strategy

To simulate millimeter-scale metasurfaces, it is essential to parallelize the simulation method across multiple compute nodes. In order to be scalable, however, this parallelization scheme should introduce only a modest communication overhead in the simulation as this communication overhead can potentially offset any time savings achieved due to the parallelization ([54–56]).

For metasurface simulations, however, by utilizing the property that the incident fields generated by far-field sources will be within the light-cone in the $\mathbf{k}-$space, a parallelization strategy can be devised that requires minimal communication between the compute nodes. The fundamental principle behind this parallelization is to represent the bandlimited incident field by its samples using the Nyquist-sampling theorem[57]. More precisely, consider an incident field propagating along the $z-$direction — the transverse polarization of this field, $\mathbf{E}^T_{inc}(x,y,z)$, at any z can be expressed as

$$\mathbf{E}^{\mathrm{T}}_{\mathrm{inc}}(x,y,z)=\sum_{i,j}\mathbf{E}^{\mathrm{T}}_{\mathrm{inc}}(x_i,y_j,z)f_{i,j}(x,y), \tag{1}$$

where $x_i, y_j = i\lambda/2, j\lambda/2$ with λ being the wavelength in the background medium, and $f_{i,j}(x,y)$ is a *jinc* function[58] centered at (x_i, y_j). Each term in the Nyquist decomposition can be considered to be an independent source, which falls off to zero with distance (Fig. 1a), and the response of a metasurface to these individual sources can be obtained by considering only a spatially-truncated portion of the metasurface in the simulation. This is numerically demonstrated in Fig. 1b, in which we consider the scattered power obtained on exciting a metasurface with a single jinc source as a function of the size of the metasurface included in the simulation. As the size of the metasurface is increased, the scattered power converges, indicating that a local simulation is sufficient to capture the metasurface response. The portion of the metasurface required to achieve a particular accuracy in the simulation is governed by diffraction of the jinc source as it propagates to the metasurface.

To parallelize the simulation, we can then divide up the jinc sources that compose the incident electric fields into smaller groups, and simulate the local response of the metasurface for each source group by performing an independent solve on a single compute node (Fig. 2a). This parallelization strategy only requires communication between the compute nodes at the start and at the end of the simulation — once to distribute the simulation data corresponding to the local subregions, and then to consolidate the electric field data computed per subregion. On each compute node, we implement a GPU-parallelized transition-matrix (T-matrix) electromagnetic solver[59,60] (See Supplementary Note 1 for details of the T-matrix method and our implementation of it). In order to accurately account for the diffraction of the jinc source while computing the local response of the metasurface, we include a padding region around the group of sources for each compute node. While in principle, we should ensure that the performance metric being analyzed (e.g. metasurface efficiency) converges with respect to the padding, in practice and for typical metasurfaces, the thickness of padding required for accurate simulations can be estimated simply by studying the response of a local patch of the metasurface to one source (similar to the study performed in Fig. 1b). After having performed all the simulations, the electric fields obtained can be added together to compute the total electric field. Furthermore, because each compute node performs roughly the same amount of compute, the total simulation time scales as $1/N_{nodes}$ (Fig. 2b). Details of our jinc source computation for the T-matrix method single-node simulation can be found in Supplementary Note 2.

Thus, given a sufficiently large number of compute nodes, we expect the simulation strategy to be able to handle large problems – on a compute cluster with 48 V100 GPU nodes, we were able to simulate a metasurface of size about $645\lambda \times 645\lambda$ in ~10 hours. This total time is broken down into the compute times for the key simulation parts in Fig. 2c. The simulated surface is a 1 mm × 1 mm metalens with focal length 0.4 mm (NA = 0.78) designed from a library of silicon cylinders with height 940 nm, radii range of

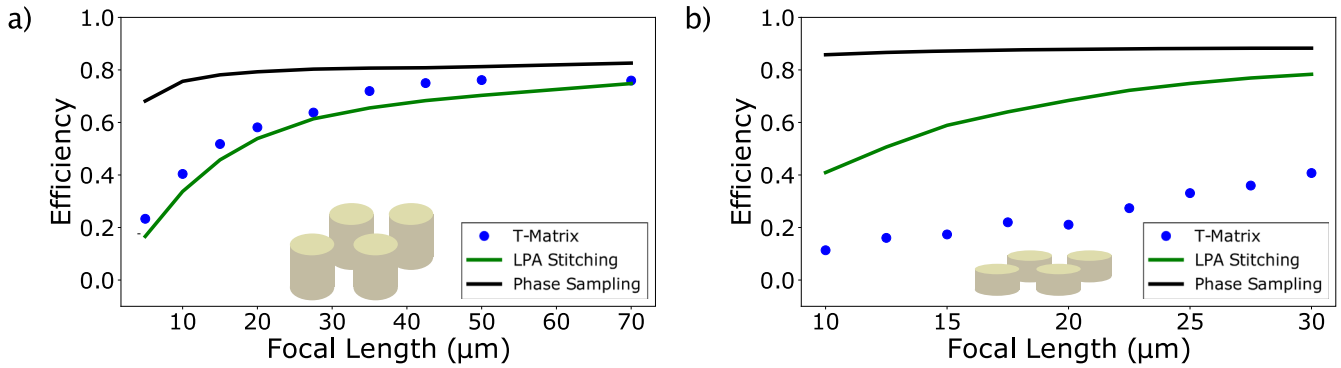


Fig. 3 Comparison of T-matrix method simulations with locally periodic assumption (LPA) simulations. a Efficiency versus focal length for 25 μm × 25 μm metasurfaces designed from a library of high aspect-ratio scatterers with a large period (silicon cylinders with height 940 nm, radii range of 50–250 nm, lattice period of 1070 nm, and air background; source wavelength of 1550 nm – transmission and phase response shown in Supplementary Fig. 4a, based on scatterer library from Arbabi et al.[61]) — efficiencies are computed using the T-matrix approach (blue dots), the commonly-used LPA transmission mask phase sampling approach (black curve), and the LPA field-stitching method (green curve). The metalens efficiency is defined as the ratio of the power within a circle of radius 3 × FWHM in the focal plane to the power incident on the metasurface. The T-matrix and LPA-stitching methods agree fairly well here because the scatterers are high aspect-ratio and the lattice constant is large, hence the interactions between neighboring scatterers is negligible. **b** Efficiency versus focal length for 15 μm × 15 μm metasurfaces designed from a library of low aspect-ratio scatterers with a small period (silicon cylinders with height 220 nm, radii range of 175–280 nm, lattice period of 666 nm, and background refractive index 1.66; source wavelength of 1340 nm – using scatterer library from Gigli et al.[62], scatterer transmission and phase response shown in Supplementary Fig. 4b) — efficiencies are computed using the T-matrix approach (blue dots), the commonly-used LPA transmission mask phase sampling approach (black curve), and the LPA field-stitching method (green curve). The metalens efficiency is defined as the ratio of the power within a circle of radius 3 × FWHM in the focal plane to the power incident on the metasurface. The T-matrix and LPA-stitching methods do not agree here because the scatterers are low aspect-ratio and the lattice constant is small, hence the interaction between neighboring scatterers is significant.

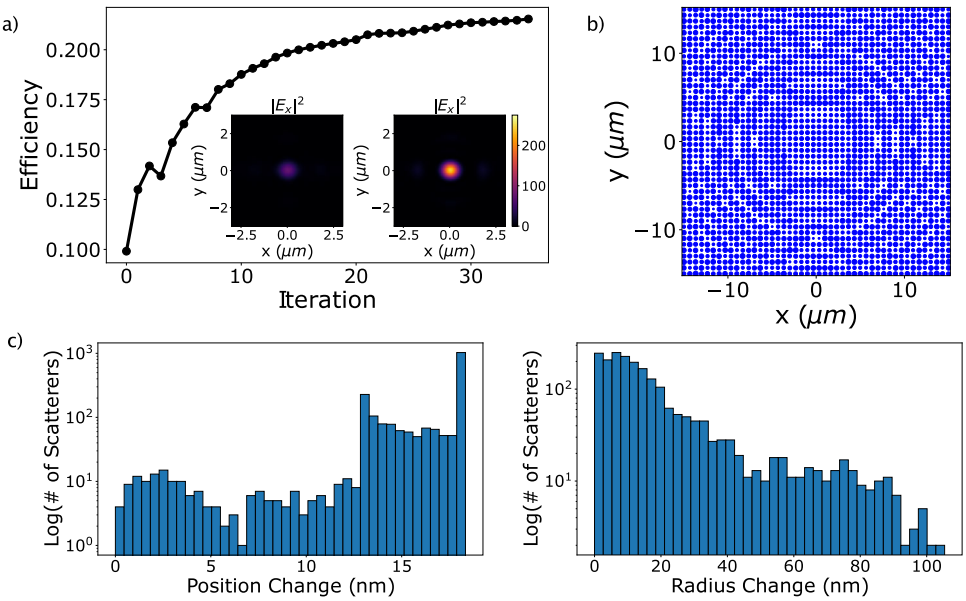


Fig. 4 Distributed Gradient-based optimization improvement of metalens design. a Lens efficiency versus optimization iteration, where lens efficiency is defined as the ratio of the power within a circle of radius 3 × FWHM in the focal plane to the power incident on the metasurface. The initial metasurface is a 30 μm × 30 μm metalens with focal-length 20 μm designed from the low aspect-ratio scatterer library used in Fig. 3b using the traditional periodic-approximation metasurface design approach. The metalens is designed and optimized for x-polarized light only. In 35 optimization iterations, the metalens efficiency is almost doubled. The inset shows the intensity of the X-component of the electric field in the focal plane before optimization (left) and after optimization (right). **b** Schematic of the cylindrical metasurface scatterers after optimization. **c** Histograms of the distance between the final scatterer positions and the initial scatterer positions (left) and the absolute radius difference between the final scatterer cylinders and the initial scatterer cylinders (right). As can be seen in these histograms, both the scatterer positions and radii change as a result of the optimization.

50–250 nm, lattice period of 1070 nm, air background, and source wavelength of 1550 nm (based on scatterer library from Arbabi et al.[61]). The simulation is performed on 48 V100 GPUs and is distributed between these compute nodes using a subregion size of 20 μm × 20 μm and a padding of 6.5 μm, resulting in 2601 subregion simulations.

Comparison with locally periodic approximation

Approximate simulations of large-area metasurfaces often rely on the locally periodic approximation (LPA)[19,37,61], wherein the local electromagnetic response of the metasurface is approximated with that of a periodic metasurface. To demonstrate that the full metasurface simulation approach captures meta-atom interactions beyond LPA, we compare the T-matrix simulation method with two commonly-used LPA approaches in Fig. 3. First is a simple transmission mask approximation, where we assume that the metasurface imparts a smooth position-dependent amplitude and phase to the incident field as determined by the periodic simulation[61], and ignore the variation of the fields within a single unit cell. Second, we consider a more exact field stitching method[9], where the field near the metasurface within each unit cell is approximated with the fields from the periodic simulation and then this stitched field is propagated. For high aspectratio scatterers, we find that while the transmission mask method significantly deviates from the T-matrix method, the field stitching method does not. However, for small aspect-ratio scatterers, which are expected to have larger inter meta-atom interactions, both the transmission mask and the field stitching LPA approximation methods significantly deviate from the T-matrix method[62]. These results are a strong indication of the ability of the T-matrix method to capture meta-atom interactions and accurately simulate the metasurface response.

Distributed optimization-based design

An essential ingredient for optimization-based design of metasurfaces is an efficient evaluation of the gradient of the figure of merit with respect to the design parameters. A particularly useful method to evaluate gradients is based on adjoint-sensitivity analysis[63,64] which analytically differentiates through Maxwell's equations and computes the gradients with respect to all the design parameters with a cost proportional to only two electromagnetic simulations. The distributed T-matrix simulation method is also amenable to distributed adjoint sensitivity analysis and can allow for scalable evaluation of the gradient of a performance metric defined on the electric fields scattered from the metasurface with respect to both the meta-atom shape and positions (see Supplementary Note 4 for details). Figure 4 demonstrates a distributed gradient-based optimization with respect to the positions and radii of the cylindrical meta-atoms of a cost function evaluating the amount of power within a spot at the focal plane for a 30 μm × 30 μm metalens with focal-length 20 μm initially designed using the traditional periodic-approximation approach with the same scatterer library used in Fig. 3b. The distributed optimization was performed on 9 T4 GPUs with the metalens divided into 9 subregions (subregion size of 10 μm × 10 μm, and padding size of 6 μm). The forward simulations performed took an average of about 120 GPU-min and the gradient computations with respect to radius and position took an average of 150 GPU-min). The metalens has a very high NA over 0.996 and the optimization improves the efficiency of the metalens by ~2×, giving a final efficiency of about 24%. Although thin low aspect-ratio metasurfaces (Huygens metasurfaces) are of interest because they are more amenable to large-scale fabrication, they have not found widespread adoption due to their very limited efficiencies and angular responses[62]. Our ability to accurately model the scatterer-scatterer effects in our metasurface inverse-design may allow discovery of more practical Huygens metasurfaces[65,66]. Combining this multi-GPU gradient computation with the

multi-GPU forward simulation, we have opened the door to gradient-based optimization over the many degrees of freedom afforded by arbitrarily large metasurfaces. In particular, our method allows optimizing both the shape and position of the scatterers composing the large-area metasurface — optimizing the scatterer positions is very difficult for any inverse-design approach that relies on a periodic assumption.

DISCUSSION

We have demonstrated a scalable distribution method to accurately simulate arbitrarily large-area metasurfaces. Our method uses the Nyquist sampling theorem to allow parallel distribution of compute across multiple GPU nodes, on which a T-matrix method formulation is used to efficiently simulate the subregion. We show a roughly $\frac{1}{N_{GPU}}$ scaling of the total simulation time and demonstrate that our method accurately accounts for all scatterer interactions. Finally, we demonstrate our ability to apply our distribution method to the computation of the gradient with respect to all design parameters. Our distributed simulation method provides a solution to the long-standing problem of simulating large-area metasurfaces and opens the door to gradient-based optimization of the full metasurface, taking advantage of all the design degrees of freedom.

METHODS

Our low-overhead distribution strategy works by using the Nyquist sampling of the incident field to split the simulation into subregion simulations, each of which can be performed in parallel. We use RabbitMQ (https://www.rabbitmq.com) to create a queue of the metasurface subregion simulations and manage the distribution of these simulations to the available GPU compute nodes in a fault-tolerant manner. On each GPU compute node, we run our T-matrix method code implemented in C++ using CUDA for the single-node GPU parallelization of incident field, matrix-vector product, and electric and magnetic field computations. For our GPU compute nodes, we used Google Cloud V100 GPUs for the timing benchmarks and 1 mm × 1 mm metasurface simulation in Fig. 2, and T4 GPUs for the distributed inverse-design in Fig. 4. We interface our distributed metasurface solver with the photonic optimization framework software SPINS to perform the inverse-design.

DATA AVAILABILITY

The data that support the plots within this paper and other findings of this study are available from the corresponding authors upon reasonable request.

Received: 9 October 2021; Accepted: 26 March 2022;
Published online: 21 April 2022

REFERENCES

1. Lee, Y.-H., Zhan, T. & Wu, S.-T. Prospects and challenges in augmented reality displays. *Virtual Real. Intell. Hardw.* **1**, 10–20 (2019).
2. Berkovic, G. & Shafir, E. Optical methods for distance and displacement measurements. *Adv. Opt. Photon.* **4**, 441–471 (2012).
3. Chen, W. T., Zhu, A. Y. & Capasso, F. Flat optics with dispersion-engineered metasurfaces. *Nat. Rev. Mater.* **5**, 604–620 (2020).
4. Zhou, Y., Zheng, H., Kravchenko, I. I. & Valentine, J. Flat optics for image differentiation. *Nat. Photon.* **14**, 316–323 (2020).
5. Colburn, S., Zhan, A. & Majumdar, A. Metasurface optics for full-color computational imaging. *Sci. Adv.* **4**, eaar2114 (2018).
6. Horie, Y., Arbabi, A., Arbabi, E., Kamali, S. M. & Faraon, A. Wide bandwidth and high resolution planar filter array based on dbr-metasurface-dbr structures. *Opt. Express* **24**, 11677–11682 (2016).
7. Kwon, H., Arbabi, E., Kamali, S. M., Faraji-Dana, M. & Faraon, A. Single-shot quantitative phase gradient microscopy using a system of multifunctional metasurfaces. *Nat. Photon.* **14**, 109–114 (2020).
8. Lee, G.-Y. et al. Metasurface eyepiece for augmented reality. *Nat. Commun.* **9**, 1–10 (2018).
9. Li, Z. et al. Inverse design enables large-scale high-performance meta-optics reshaping virtual reality. *Preprint at* https://arxiv.org/abs/2104.09702 (2021).
10. McClung, A., Mansouree, M. & Arbabi, A. At-will chromatic dispersion by prescribing light trajectories with cascaded metasurfaces. *Light Sci. Appl.* **9**, 1–9 (2020).
11. Arbabi, A. et al. Increasing efficiency of high numerical aperture metasurfaces using the grating averaging technique. *Sci. Rep.* **10**, 1–10 (2020).
12. Pestourie, R., Mroueh, Y., Nguyen, T. V., Das, P. & Johnson, S. G. Active learning of deep surrogates for pdes: Application to metasurface design. *Npj Comput. Mater.* **6**, 1–7 (2020).
13. Zhan, A. et al. Low-contrast dielectric metasurface optics. *ACS Photon.* **3**, 209–214 (2016).
14. Aieta, F. et al. Aberration-free ultrathin flat lenses and axicons at telecom wavelengths based on plasmonic metasurfaces. *Nano Lett.* **12**, 4932–4936 (2012).
15. Arbabi, E., Arbabi, A., Kamali, S. M., Horie, Y. & Faraon, A. Multiwavelength polarization-insensitive lenses based on dielectric metasurfaces with meta-molecules. *Optica* **3**, 628–633 (2016).
16. Devlin, R. C., Khorasaninejad, M., Chen, W. T., Oh, J. & Capasso, F. Broadband high-efficiency dielectric metasurfaces for the visible spectrum. *Proc. Natl. Acad. Sci. USA* **113**, 10473–10478 (2016).
17. Fan, Z.-B. et al. Silicon nitride metalenses for close-to-one numerical aperture and wide-angle visible imaging. *Phys. Rev. Appl.* **10**, 014005 (2018).
18. Shi, Z. et al. Single-layer metasurface with controllable multiwavelength functions. *Nano Lett.* **18**, 2420–2427 (2018).
19. Khorasaninejad, M. et al. Polarization-insensitive metalenses at visible wavelengths. *Nano Lett.* **16**, 7229–7234 (2016).
20. Molesky, S., Piggott, A. Y., Weiliang, J., Vučković, J. & Rodriguez, A. W. Inverse design in nanophotonics. *Nat. Photon.* **9**, 659 (2018).
21. Yang, K. Y. et al. Inverse-designed non-reciprocal pulse router for chip-based lidar. *Nat. Photon.* **14**, 369–374 (2020).
22. Wang, F., Jensen, J. S. & Sigmund, O. Robust topology optimization of photonic crystal waveguides with tailored dispersion properties. *J. Opt. Soc. Am. B* **28**, 387–397 (2011).
23. Hughes, T. W., Minkov, M., Williamson, I. A. & Fan, S. Adjoint method and inverse design for nonlinear nanophotonic devices. *ACS Photon.* **5**, 4781–4787 (2018).
24. Sapra, N. V. et al. On-chip integrated laser-driven particle accelerator. *Science* **367**, 79–83 (2020).
25. Dory, C. et al. Inverse-designed diamond photonics. *Nat. Commun.* **10**, 1–7 (2019).
26. Piggott, A. Y. et al. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nat. Photon.* **9**, 374–377 (2015).
27. Su, L. et al. Nanophotonic inverse design with spins: Software architecture and practical considerations. *Appl. Phys. Rev.* **7**, 011407 (2020).
28. Taflove, A. & Hagness, S. C. *Computational electrodynamics*, vol. 28 (Artech house publishers Norwood, MA, 2000).
29. Rumpf, R. C. Simple implementation of arbitrarily shaped total-field/scattered-field regions in finite-difference frequency-domain. *Prog. Electromagn. Res.* **36**, 221–248 (2012).
30. Reddy, J. N. *Introduction To The Finite Element Method* (McGraw-Hill Education, 2019).
31. Camayd-Muñoz, P., Ballew, C., Roberts, G. & Faraon, A. Multifunctional volumetric meta-optics for color and polarization image sensors. *Optica* **7**, 280–283 (2020).
32. Mansouree, M. et al. Multifunctional 2.5 d metastructures enabled by adjoint optimization. *Optica* **7**, 77–84 (2020).
33. Christiansen, R. E. et al. Fullwave maxwell inverse design of axisymmetric, tunable, and multi-scale multi-wavelength metalenses. *Opt. Express* **28**, 33854–33868 (2020).
34. Lin, Z., Roques-Carmes, C., Christiansen, R. E., Soljačić, M. & Johnson, S. G. Computational inverse design for ultra-compact single-piece metalenses free of chromatic and angular aberration. *Appl. Phys. Lett.* **118**, 041104 (2021).
35. Chung, H. & Miller, O. D. High-na achromatic metalenses by inverse design. *Opt. Express* **28**, 6945–6965 (2020).
36. Lin, Z., Liu, V., Pestourie, R. & Johnson, S. G. Topology optimization of freeform large-area metasurfaces. *Opt. Express* **27**, 15765–15775 (2019).
37. Pestourie, R. et al. Inverse design of large-area metasurfaces. *Opt. Express* **26**, 33732–33747 (2018).
38. Chung, H. & Miller, O. D. Tunable metasurface inverse design for 80% switching efficiencies and 144 angular deflection. *ACS Photon.* **7**, 2236–2243 (2020).
39. Sell, D., Yang, J., Doshay, S. & Fan, J. A. Periodic dielectric metasurfaces with high-efficiency, multiwavelength functionalities. *Adv. Opt. Mater.* **5**, 1700645 (2017).
40. Phan, T. et al. High-efficiency, large-area, topology-optimized metasurfaces. *Light Sci. Appl.* **8**, 1–9 (2019).
41. Lin, Z. & Johnson, S. G. Overlapping domains for topology optimization of large-area metasurfaces. *Opt. Express* **27**, 32445–32453 (2019).

42. Sell, D., Yang, J., Doshay, S., Yang, R. & Fan, J. A. Large-angle, multifunctional metagratings based on freeform multimode geometries. *Nano Lett.* **17**, 3752–3757 (2017).
43. Bayati, E. et al. Inverse designed extended depth of focus meta-optics for broadband imaging in the visible. Nanophotonics https://doi.org/10.1515/nanoph-2021-0431 (2021).
44. Zhelyeznyakov, M. V., Brunton, S. & Majumdar, A. Deep learning to accelerate scatterer-to-field mapping for inverse design of dielectric metasurfaces. *ACS Photon.* **8**, 481–488 (2021).
45. Jiang, J. & Fan, J. A. Global optimization of dielectric metasurfaces using a physics-driven neural network. *Nano Lett.* **19**, 5366–5372 (2019).
46. Jiang, J. et al. Free-form diffractive metagrating design based on generative adversarial networks. *ACS Nano* **13**, 8872–8878 (2019).
47. Byrnes, S. J., Lenef, A., Aieta, F. & Capasso, F. Designing large, high-efficiency, high-numerical-aperture, transmissive meta-lenses for visible light. *Opt. Express* **24**, 5110–5124 (2016).
48. Zhou, M. et al. Inverse design of metasurfaces based on coupled-mode theory and adjoint optimization. *ACS Photon.* **8**, 2265–2273 (2021).
49. Torfeh, M. & Arbabi, A. Modeling metasurfaces using discrete-space impulse response technique. *ACS Photon.* **7**, 941–950 (2020).
50. Zhan, A. et al. Controlling three-dimensional optical fields via inverse mie scattering. *Sci. Adv.* **5**, eaax4769 (2019).
51. Zhan, A., Fryett, T. K., Colburn, S. & Majumdar, A. Inverse design of optical elements based on arrays of dielectric spheres. *Appl. Opt.* **57**, 1437–1446 (2018).
52. Zhelyeznyakov, M. V., Zhan, A. & Majumdar, A. Design and optimization of ellipsoid scatterer-based metasurfaces via the inverse t-matrix method. *OSA Contin.* **3**, 89–103 (2020).
53. Hughes, T. W., Minkov, M., Liu, V., Yu, Z. & Fan, S. Full wave simulation and optimization of large area metalens. OSA Opt. Design and Fab. *OSA Optical Design and Fabrication 2021. Congress 2021* (2021).
54. Tang, J., Zheng, Y., Yang, C., Wang, W. & Luo, Y. Parallelized implementation of the finite particle method for explicit dynamics in gpu. *Comput. Model Eng. Sci.* **122**, 5–31 (2020).
55. Hermann, E., Raffin, B., Faure, F., Gautier, T. & Allard, J. *European Conference on Parallel Processing*, p. 235–246 (Springer, 2010).
56. Dziekonski, A., Sypek, P., Lamecki, A. & Mrozowski, M. Communication and load balancing optimization for finite element electromagnetic simulations using multi-gpu workstation. *IEEE Trans. Microw. Theory Tech.* **65**, 2661–2671 (2017).
57. Landau, H. Sampling, data transmission, and the nyquist rate. *Proc. IEEE* **55**, 1701–1706 (1967).
58. Goodman, J. W. *Introduction to Fourier Optics* (Roberts and Compnay Publishers, 2005).
59. Egel, A. et al. Extending the applicability of the t-matrix method to light scattering by flat particles on a substrate via truncation of sommerfeld integrals. *J. Quant. Spectrosc. Radiat. Transf.* **202**, 279–285 (2017).
60. Doicu, A., Wriedt, T. & Eremin, Y. A. *Light Scattering By Systems Of Particles: Null-field Method With Discrete Sources: Theory And Programs*. Vol. 124 (Springer, 2006).
61. Arbabi, A., Horie, Y., Ball, A. J., Bagheri, M. & Faraon, A. Subwavelength-thick lenses with high numerical apertures and large efficiency based on high-contrast transmitarrays. *Nat. Commun.* **6**, 1–6 (2015).
62. Gigli, C. et al. Fundamental limitations of huygens' metasurfaces for optical beam shaping. *Laser Photon. Rev.* **15**, 2000448 (2021).
63. Lalau-Keraly, C. M., Bhargava, S., Miller, O. D. & Yablonovitch, E. Adjoint shape optimization applied to electromagnetic design. *Opt. Express* **21**, 21693–21701 (2013).
64. Piggott, A. Y., Petykiewicz, J., Su, L. & Vučković, J. Fabrication-constrained nanophotonic inverse design. *Sci. Rep.* **7**, 1–7 (2017).
65. Ollanik, A. J., Smith, J. A., Belue, M. J. & Escarra, M. D. High-efficiency all-dielectric huygens metasurfaces from the ultraviolet to the infrared. *ACS Photon.* **5**, 1351–1358 (2018).
66. Cai, H. et al. Inverse design of metasurfaces with non-local interactions. *Npj Comput. Mater.* **6**, 1–8 (2020).

ACKNOWLEDGEMENTS

This work was supported by the Samsung GRO program. J.S. acknowledges support from the National Science Foundation Graduate Research Fellowship (grant no. DGE-1656518) and Cisco Systems Stanford Graduate Fellowship (SGF). R.T acknowledges support from Max Planck Harvard research center for Quantum Optics (MPHQ) fellowship, and Sarah and Kailath Stanford Graduate Fellowship (SGF).

AUTHOR CONTRIBUTIONS

R.T. and L.S. conceived the idea. J.S., R.T., and L.S. designed and implemented the software. J.S., R.T., L.S., D.A-S., and H.K. ran benchmarks and conducted numerical experiments. J.V., S.H., and S.F. supervised the project. All authors assisted with data analysis and manuscript preparation.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-022-00774-y.

Correspondence and requests for materials should be addressed to Rahul Trivedi or Jelena Vučković.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.