Applied Physics Letters

A perspective on the pathway toward full wave simulation of large area metalenses

Cite as: Appl. Phys. Lett. **119**, 150502 (2021); https://doi.org/10.1063/5.0071245
Submitted: 13 September 2021 • Accepted: 21 September 2021 • Published Online: 11 October 2021

Tyler W. Hughes, Momchil Minkov, Victor Liu, et al.

View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

A perspective on flexible sensors in developing diagnostic devices
Applied Physics Letters **119**, 150501 (2021); https://doi.org/10.1063/5.0057020

Probabilistic computing with p-bits
Applied Physics Letters **119**, 150503 (2021); https://doi.org/10.1063/5.0067927

Demonstration of a compact x-ray free-electron laser using the optical klystron effect
Applied Physics Letters **119**, 151102 (2021); https://doi.org/10.1063/5.0064934

© 2021 Author(s).

Applied Physics Letters PERSPECTIVE scitation.org/journal/apl

A perspective on the pathway toward full wave simulation of large area metalenses

Cite as: Appl. Phys. Lett. **119**, 150502 (2021); doi: 10.1063/5.0071245
Submitted: 13 September 2021 · Accepted: 21 September 2021 ·
Published Online: 11 October 2021

View Online Export Citation CrossMark

Tyler W. Hughes,[1] Momchil Minkov,[1] Victor Liu,[1] Zongfu Yu,[1,2] and Shanhui Fan[1,3,a)]

AFFILIATIONS

[1]Flexcompute, Inc., 130 Trapelo Rd., Belmont, Massachusetts 02478, USA
[2]Department of Electrical and Computer Engineering, University of Wisconsin—Madison, 1415 Engineering Dr., Madison, Wisconsin 53706, USA
[3]Department of Electrical Engineering, Stanford University, 350 Jane Stanford Way, Stanford, California 94305, USA

[a)]Author to whom correspondence should be addressed: shanhui@stanford.edu

ABSTRACT

Metalenses for optical beam manipulation have a significant impact in many exciting applications due their compact, planar geometry and ease of fabrication. However, the enormous physical size of metalenses relative to the optical wavelength provides a barrier to performing accurate simulations in a reasonable time frame. In principle, full-wave simulation techniques, such as the finite-difference time-domain (FDTD) method, are ideal for metalens modeling as they give an accurate picture of the device performance. However, when applied using traditional computing platforms, this approach is infeasible for large-diameter metalenses and requires hours and days to simulate even devices of modest size. To alleviate these issues, the standard approach has been to apply approximations, which typically employ simplified models of the metalens unit cells or ignore coupling between cells, leading to inaccurate predictions. In this Perspective, first, we summarize the current state of the art approaches in simulating large scale, three-dimensional metalenses. Then, we highlight that advances in computing hardware have now created a scenario where the full-wave simulation of large area metalenses is feasible within a reasonable time frame, providing significant opportunities to the field. As a demonstration, we show that a hardware-accelerated FDTD solver is capable of simulating a fully 3D, large area metalens of size $100\lambda \times 100\lambda$, including the focal length, in under 5 min. The application of hardware-accelerated, full-wave simulation tools to metalens simulation should have a significant impact in the metalens field and the greater photonics community.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0071245

Metasurfaces are planar photonic devices comprised of a collection of a large number of unit cells, which are selected and arranged in ways that give rise to a wide range of useful behaviors.[1] Metalenses are one common class of metasurfaces, in which the unit cells are designed to impart a specific, position-dependent phase shift to transmitted light, which gives rise to a global focusing effect.[2] As metalenses are typically fabricated as a wavelength-scale thick layer on a planar surface using standard nanofabrication techniques, their compactness makes them compelling alternatives to traditional, bulky optical components. This functionality presents many exciting opportunities for optical beam manipulation in microscopy,[3] imaging,[4] and augmented reality,[5] among other applications.

The ability to accurately simulate metalenses in a reasonable time frame is vital both to understanding their fundamental properties and to engineering and optimizing their performance. Ideally, one would use a general electromagnetic simulation tool, such as a finite-difference time-domain (FDTD)[6,7] package, to model the device. However, as the diameter of most metalenses operating in the visible wavelength spectrum can range from several hundred micrometers to centimeters,[2] the run-time and memory requirements render traditional simulation processes infeasible. As such, most full-wave metalens modeling studies are restricted to structures exhibiting either translational[8,9] or axial[10,11] symmetry, restricting the simulation to two dimensions and limiting the number of degrees of freedom available in the design as a result. While recent works have shown the simulation and optimization of fully 3D metalenses with diameters just over 50 λ,[12,13] these simulations each required over an hour of run-time on CPU platforms containing over eight cores. As such, an alternative approach is required to handle the typical case of metalenses that are significantly larger and, therefore, require far more computational resources.

© Author(s) 2021

The standard approach used to model large scale, three-dimensional metalenses is to *approximate* the device response by breaking the calculation into simpler components, typically at the scale of the individual unit cell. A common strategy is to perform individual simulations of each unit cell within a set of designs, which gives a library of phase and amplitude responses that can be used in models that estimate the far-field response of a full device.[14] To improve the accuracy of this approach, hybrid models have been developed, which retain the full-wave field patterns in individual simulations through either locally periodic approximation approaches[8,15] or unit cell simulations with overlapping domains.[16] While these techniques may improve the accuracy of the simulations, they require substantial computational resources and can be challenging to implement. Moreover, the errors introduced by these techniques remain uncontrolled. Not only is the error difficult to estimate, but it is also challenging to reduce systematically through these methods. While these types of approaches can provide some important understanding of the metalens response, they ignore much of the full-wave information and neglect cell-to-cell coupling effects, motivating the need for methods with greater accuracy.[17]

A straightforward strategy for scaling full-wave simulations to larger devices is to employ massive parallelization of CPU-based computing platforms.[18] However, this approach quickly becomes infeasible due to memory bandwidth limitations of CPUs, meaning that continued parallelization can only provide benefit up to a limit that is quickly reached before this approach is useful for large scale metalens problems. As such, a fundamentally different approach to computing hardware is worth considering for metalens simulation.

In recent years, specialized computing hardware, such as graphics processing units (GPUs) and tensor processing units (TPUs), have revolutionized computation in industries such as computer graphics[19] and machine learning[20] by providing significant acceleration beyond the capability of traditional CPUs. In the area of scientific simulation, the application of specialized hardware is emerging, and impressive speed up of calculations by orders of magnitude when compared to conventional CPU-based modeling[21] has been demonstrated. This approach has also been applied to electromagnetic simulation[22–24] with promising results.

In this Perspective, we argue that a similar application of hardware-accelerated electromagnetic simulation will enable the simulation of large area metalenses with full-wave accuracy. As a demonstration that highlights the potential of this approach, we employ the use of a hardware-accelerated FDTD solver developed within our group, titled "Tidy3D,"[25] for the simulation of a three-dimensional metalens of record size. We show that the solver can perform simulations of large area metalenses with a turnaround time on the minutes scale. Then, we make use of this capability to achieve significant performance improvement over a conventional metalens design through optimization.

As illustrated in Fig. 1(a), the structure of interest is a fully 3D metalens structure with a diameter of 100λ, and the simulation region includes the space containing the focal plane. The design of the metalens itself is based on a commonly cited recent experimental work[3] and is described as follows. The metalens is defined on top of a square SiO_2 substrate with dimensions of $100\lambda \times 100\lambda$ in the xy plane. The metalens has a desired numerical aperture (NA) of 0.8, giving a focal length of 37λ in the z direction. To accommodate the fields along the focal axis, the simulation extends for a total of 46λ in the z direction. Periodic boundary conditions are employed in the x and y boundaries, and a convolutional perfectly matched layer[26] is employed in the z boundary. A resolution of $\lambda/18$ is used along with subpixel smoothing of the permittivity distribution. This choice was verified through a resolution scan and is consistent with other works.[17] The total simulation size is $100\lambda \times 100\lambda \times 46\lambda$ ($66 \times 66 \times 30.36\,\mu m^2$), which requires 2.68×10^6 grid cells at this resolution.

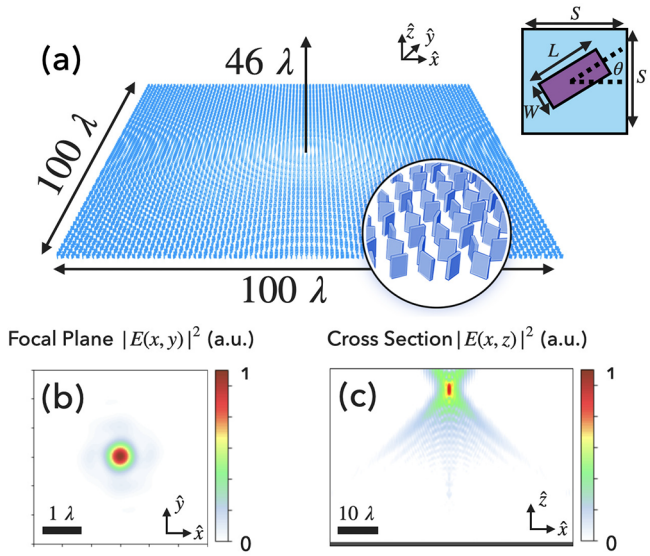

FIG. 1. (a) 3D rendering of metalens being simulated, including the focal length. (Inset) Description of the metalens unit cell. $S = 660$ nm, $L = 410$ nm, and $W = 80$ nm. (b) Electric field intensity at a limited section of the x-y focal plane. (c) Electric field intensity at a limited section of the x-z cross-sectional plane. Note the difference in the scale bars between (b) and (c).

The metalens unit cell is defined by a rectangular TiO_2 structure sitting on the substrate. The dimension of the cell is chosen such that it operates as a half waveplate, converting a fraction of the right-hand circularly polarized incident light to left-hand circular polarization and creating constructive interference at the focal spot. To obtain the desired phase shift for focusing, each TiO_2 structure centered at position (x, y) in the plane is oriented at an angle θ from the x-axis, given by

$$\theta(x, y) = \frac{\pi}{\lambda}\left(f - \sqrt{f^2 + x^2 + y^2}\right). \tag{1}$$

This formula can be derived by a simple model of constructive interference assuming each metacell imparts a phase shift of 2θ, the details of which are described in other works.[3] The entire metalens contains 10 000 such unit cells, which are arranged in a rectangular grid as shown in Fig. 1(a).

An approach commonly employed in metalens simulation is to record only the transmitted electric and magnetic near-field distribution and use these fields to extrapolate far-field information using the free space Green's function,[8] beam propagation method,[27] or near-field to far-field transformation.[6] Such a technique enables one to simulate a thin slice in the z direction instead of a large volume, which reduces run-time and memory requirements. While these techniques

© Author(s) 2021

are compatible with our simulation, the speed of our solver gives us the benefit of not relying on such an approach to extract simulation results. As such, we choose to employ a simpler, full domain modeling that retains the full accuracy solution of Maxwell's equations in the entire domain.

A right circularly polarized source was injected below the substrate with Gaussian time dependence. The spectral bandwidth of the pulse is 0.1 times the central wavelength at $\lambda = 660$ nm. The simulation was run for 6235 time steps, and the steady state fields at the central wavelength were recorded along several output planes using a running discrete Fourier transform method.[6] Although only the fields at the central frequency were measured, full spectral information may be obtained from the same simulation by including additional frequency probes with only marginal degradation to the performance. The electric intensity patterns at the focal plane and the x-z cross section are shown in Figs. 1(b) and 1(c), respectively.

The simulation was run on an elastic network of NVIDIA GPUs. The total simulation took 288 seconds, consisting of 99 s of setup and 189 s of time stepping through the FDTD algorithm.

From the resulting field patterns, the focal spot exhibited a FWHM of 440 nm and a Strehl ratio of 0.83, as defined in Ref. 3. The focusing efficiency, as defined in our case by the ratio of power passing through a square with side length six times the FWHM of the diffraction limited spot (420 nm) at the focal plane to the incident power, was measured at 57.7%. These metrics are close to the experimental measurements performed on the similar design from Ref. 3.

To demonstrate the usefulness of this simulation capability for improving device performance, here we demonstrate a brute force optimization of the previous design. For this demonstration, we take a smaller version of the metalens with a $21\lambda \times 21\lambda$ array of unit cells. Starting from the design outlined in Eq. (1), we seek to adjust each of the 441 angles sequentially to maximize the focusing efficiency (η) of the lens as defined in the previous paragraph.

The optimization routine is described as follows. The starting design exhibits an η of 57.5%, which we take as our baseline. We visit each unit cell in the array row by row and perform a line search of the unit cell angle (θ) to locally maximize η. In this line search, θ is first perturbed in the counterclockwise direction and the change in η is measured. If the change is positive, then θ is further perturbed counterclockwise until η no longer increases. On the other hand, if the initial change is negative, then η is perturbed in the clockwise direction until η no longer increases. All perturbations are done in increments of 5°. After perturbing the angle of the unit cell, we leave the angle at its most optimal value for maximizing η and continue to the next unit cell. After visiting each unit cell in the metalens, the routine is repeated twice over the entire structure in the same order, leading to more improvement each pass.

The results of the optimization are shown in Fig. 2. As shown in Fig. 2(a), at the end of the optimization routine, η is improved from 57.5% to 63.5%, which is a relative change of 10.4%. The individual simulations completed in 4 s with 3000 total simulations required for a total optimization time of 3.5 hours. The final device exhibits minor shifts in angle at most unit cells with the change in angle visualized in Fig. 2(b).

While the speed of our solver enables one to handle large scale, brute force optimizations of this kind, it may be equally applied to other, more sophisticated algorithms for metalens design.[28,29] One

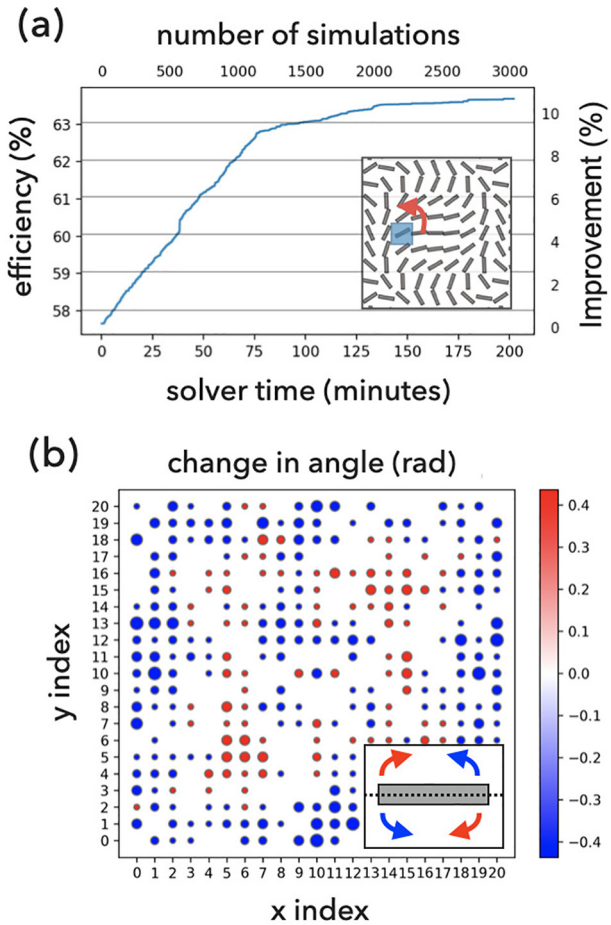

FIG. 2. (a) Progress of optimization in focusing efficiency (left) and relative improvement (right) vs number of simulations (top) and total solver time (bottom). (b) Change in angle of each cell resulting in optimization. Red circles indicate a clockwise change, and blue cells indicate a counterclockwise change. The circle radius is proportional to the magnitude of the angle change, and the maximum circle size corresponds to a magnitude of 0.436 rad.

such example is the adjoint method,[12,30] wherein one may compute the gradient of η with respect to all unit cell angles using just two simulations, providing the ability to do parallel updates of all design parameters and thereby reduce the total number of simulations required. While advanced optimization algorithms are useful for improving the device performance, sequential routines involving numerous simulations, such as parameter scans, are still commonplace, and having a high speed simulation capability alone will greatly reduce the effort and wait time required by such studies.

As these findings demonstrate, the application of hardware-acceleration to electromagnetic simulation provides a promising path toward the full-wave modeling and optimization of large scale metalenses within a reasonable time frame. Such techniques put simulation of metalenses of the millimeter and centimeter scale within reach, which is of significant interest for both accurate modeling and performance improvement of industrial-scale metalens designs. Furthermore, the use of full-wave simulation methods, such as FDTD,

© Author(s) 2021

enables opportunities in simulating more complex physics in metalens systems such as re-programmable[31] and nonlinear[32] devices.

Access to simulation tools with this capability remains a challenge as the most commonly used electromagnetic simulation tools do not currently support hardware-acceleration. Simulation tools that do support hardware-acceleration are largely proof-of-principle and, therefore, lack the features required for complex photonic device modeling. Additionally, there are challenges associated with deploying FDTD and other numerical methods at scales beyond those that have been tested before. For example, numerical dispersion can become a significant problem when modeling wave propagation over large length scales, and numerical instabilities associated with dispersive material models often cause divergences in simulations running for a large number of time steps. Despite these challenges, the application of hardware-accelerated simulation will inevitably have wide ranging implications, both to metalenses and to the broader field of photonics. Some exciting possibilities include the simulation of entire photonic integrated circuits, photo-lithography systems, and photonic crystal devices without approximation and at a scale previously thought to be infeasible.

The authors wish to acknowledge the help of Lei Zheng for technical assistance.

All authors have a financial interest in Flexcompute, Inc., which developed the Tidy3D solver used in this work.

DATA AVAILABILITY

The data that support the findings of this study are openly available in https://github.com/flexcompute/metalens.

REFERENCES

[1] D. Lin, P. Fan, E. Hasman, and M. L. Brongersma, "Dielectric gradient metasurface optical elements," Science **345**, 298–302 (2014).
[2] M. Khorasaninejad and F. Capasso, "Metalenses: Versatile multifunctional photonic components," Science **358**, eaam8100 (2017).
[3] M. Khorasaninejad, W. T. Chen, R. C. Devlin, J. Oh, A. Y. Zhu, and F. Capasso, "Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging," Science **352**, 1190–1194 (2016).
[4] S. Wang, P. C. Wu, V.-C. Su, Y.-C. Lai, M.-K. Chen, H. Y. Kuo, B. H. Chen, Y. H. Chen, T.-T. Huang, J.-H. Wang *et al.*, "A broadband achromatic metalens in the visible," Nat. Nanotechnol. **13**, 227–232 (2018).
[5] G.-Y. Lee, J.-Y. Hong, S. Hwang, S. Moon, H. Kang, S. Jeon, H. Kim, J.-H. Jeong, and B. Lee, "Metasurface eyepiece for augmented reality," Nat. Commun. **9**, 4562 (2018).
[6] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, 2005).
[7] A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson, "Meep: A flexible free-software package for electromagnetic simulations by the fdtd method," Comput. Phys. Commun. **181**, 687–702 (2010).
[8] R. Pestourie, C. Pérez-Arancibia, Z. Lin, W. Shin, F. Capasso, and S. G. Johnson, "Inverse design of large-area metasurfaces," Opt. Express **26**, 33732–33747 (2018).
[9] E. Bayati, R. Pestourie, S. Colburn, Z. Lin, S. G. Johnson, and A. Majumdar, "Inverse designed metalenses with extended depth of focus," ACS Photonics **7**, 873–878 (2020).
[10] R. E. Christiansen, Z. Lin, C. Roques-Carmes, Y. Salamin, S. E. Kooi, J. D. Joannopoulos, M. Soljačić, and S. G. Johnson, "Fullwave Maxwell inverse design of axisymmetric, tunable, and multi-scale multi-wavelength metalenses," Opt. Express **28**, 33854–33868 (2020).
[11] 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," Appl. Phys. Lett. **118**, 041104 (2021).
[12] M. Mansouree, H. Kwon, E. Arbabi, A. McClung, A. Faraon, and A. Arbabi, "Multifunctional 2.5D metastructures enabled by adjoint optimization," Optica **7**, 77–84 (2020).
[13] M. Mansouree, A. McClung, S. Samudrala, and A. Arbabi, "Large-scale parametrized metasurface design using adjoint optimization," ACS Photonics **8**, 455–463 (2021).
[14] S. Shrestha, A. C. Overvig, M. Lu, A. Stein, and N. Yu, "Broadband achromatic dielectric metalenses," Light: Sci. Appl. **7**, 85 (2018).
[15] Z. Lin, V. Liu, R. Pestourie, and S. G. Johnson, "Topology optimization of free-form large-area metasurfaces," Opt. Express **27**, 15765–15775 (2019).
[16] Z. Lin and S. G. Johnson, "Overlapping domains for topology optimization of large-area metasurfaces," Opt. Express **27**, 32445–32453 (2019).
[17] H. Chung and O. D. Miller, "High-NA achromatic metalenses by inverse design," Opt. Express **28**, 6945–6965 (2020).
[18] A. Smirnov, A. Semenov, and A. Pozdneev, "Performance analysis of parallel fdtd algorithm on IBM bluegene supercomputer series," in PIERS Proceedings (2014).
[19] A. R. Brodtkorb, T. R. Hagen, and M. L. Sætra, "Graphics processing unit (GPU) programming strategies and trends in GPU computing," J. Parallel Distributed Comput. **73**, 4–13 (2013).
[20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," arXiv:1912.01703 (2019).
[21] D. Michéa and D. Komatitsch, "Accelerating a three-dimensional finite-difference wave propagation code using gpu graphics cards," Geophys. J. Int. **182**, 389–402 (2010).
[22] C. Warren, A. Giannopoulos, A. Gray, I. Giannakis, A. Patterson, L. Wetter, and A. Hamrah, "A CUDA-based GPU engine for gprMax: Open source FDTD electromagnetic simulation software," Comput. Phys. Commun. **237**, 208–218 (2019).
[23] P. Micikevicius, "3D finite difference computation on GPUs using CUDA," in Proceedings of 2nd *Workshop* on General Purpose Processing on Graphics Processing Units (2009) pp. 79–84.
[24] J. Skarda, R. Trivedi, L. Su, D. Ahmad-Stein, H. Kwon, S. Han, S. Fan, and J. Vučković, "Simulation of large-area metasurfaces with a distributed transition matrix method," arXiv:2107.09879 (2021).
[25] See https://www.flexcompute.com for "Tidy3d."
[26] A. Giannopoulos, "An improved new implementation of complex frequency shifted PML for the FDTD method," IEEE Trans. Antennas Propag. **56**, 2995–3000 (2008).
[27] J. Van Roey, J. Van der Donk, and P. Lagasse, "Beam-propagation method: Analysis and assessment," J. Opt. Soc. Am. **71**, 803–810 (1981).
[28] 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: Sci. Appl. **8**, 48 (2019).
[29] S. J. Byrnes, A. Lenef, F. Aieta, and F. Capasso, "Designing large, high-efficiency, high-numerical-aperture, transmissive meta-lenses for visible light," Opt. Express **24**, 5110–5124 (2016).
[30] T. W. Hughes, M. Minkov, I. A. Williamson, and S. Fan, "Adjoint method and inverse design for nonlinear nanophotonic devices," ACS Photonics **5**, 4781–4787 (2018).
[31] L. Li, H. Ruan, C. Liu, Y. Li, Y. Shuang, A. Alù, C.-W. Qiu, and T. J. Cui, "Machine-learning reprogrammable metasurface imager," Nat. Commun. **10**, 1082 (2019).
[32] A. Mekawy, D. L. Sounas, and A. Alù, "Free-space nonreciprocal transmission based on nonlinear coupled fano metasurfaces," Photonics **8**, 139 (2021).

© Author(s) 2021