Письма в ЖЭТФ, том 114, вып. 6, с. 360 – 364 

# Оптимизация многослойных фотонных структур с помощью искусственных нейронных сетей для получения заданного оптического отклика

*К. Р. Сафронов*[1]*, В. О. Бессонов, А. А. Федянин*

*Физический факультет, МГУ им. М. В. Ломоносова, 119991 Москва, Россия*

Поступила в редакцию 11 августа 2021 г.
После переработки 12 августа 2021 г.
Принята к публикации 12 августа 2021 г.

Предложен новый метод глубокого машинного обучения для задачи подбора параметров многослойной фотонной структуры по заданному оптическому спектру коэффициента отражения. Предложенный метод обучения основан на соединении искусственной нейронной сети для решения обратной задачи и аналитического метода матриц распространения. Такой подход позволяет добиться высокой точности работы сети. Показана возможность применения разработанного метода для дизайна структуры, выполняющей взятия производной по координате для падающего оптического сигнала.

DOI: 10.31857/S123456782118004X

Устройства нанофотоники обеспечивают сложные функциональные возможности за счет структурирования материалов на микро- и наноуровне [1–3]. Одной из наиболее простых, но при этом универсальных нанофотонных структур является массив чередующихся слоев с различной толщиной и показателем преломления. Оптический отклик такой многослойной структуры (МС) регулируется параметрами слоев и может подстраиваться под требования конкретной задачи. Благодаря такой гибкости, МС успешно применяются в таких областях, как топологическая фотоника [4], нелинейная оптика [5, 6], магнитооптика [7], оптические вычисления [8], поляритоника [9] и др. [10–12]. Однако для реализации всех этих возможностей необходима разработка методов подбора параметров (дизайна) МС, обладающей заданным оптическим откликом.

Дизайн МС может быть осуществлен перебором всевозможных конфигураций и расчетом оптического отклика для каждого набора параметров. Такой расчет может быть проведен одним из многих хорошо разработанных методов, включая метод конечных элементов, метод конечных разностей во временной области, метод матриц распространения (ММР) и др. Недостатком такого подхода являются его большие временные и вычислительные затраты. Другой подход основан на решении обратной задачи, т. е. определении параметров МС напрямую из заданного отклика [13]. Для решения обратной задачи разработано множество оптимизационных методов [14–17]. Однако они являются итеративными и, следовательно, дорогостоящими с точки зрения вычислений, что делает их неподходящими для крупномасштабных и сложных структур. Кроме того, даже незначительная модификация целевого оптического отклика заставляет начинать оптимизацию заново.

Чтобы преодолеть это препятствие, недавно было предложено использовать подход, основанный на искусственных нейронных сетях (ИНС) [18, 19]. ИНС можно использовать одним из двух способов. Первый вариант – предсказание оптического отклика структуры по геометрическим параметрам [19–21]. В этом случае ИНС заменяет дорогостоящее вычислительное моделирование и повышает скорость стандартных методов оптимизации. Второй вариант – обучение ИНС для предсказания структуры с требуемым откликом [22, 23], т. е. для решения обратной задачи. Это позволяет подбирать структуры с заданным оптическим откликом за доли секунды. Основной проблемой второго варианта использования ИНС является неединственность решения обратной задачи, что приводит к несогласованности данных и плохой сходимости ИНС [18]. Для решения этой проблемы было предложено несколько подходов. Одним из наиболее распространенных из них является двухэтапное обучение ИНС: сначала обучается одна сеть для предсказания оптического отклика по заданным параметрам, а затем эта предварительно обученная ИНС соединяется с другой сетью для решения обратной задачи, чтобы сформировать тандем

[1]e-mail: safronov@nanolab.phys.msu.ru

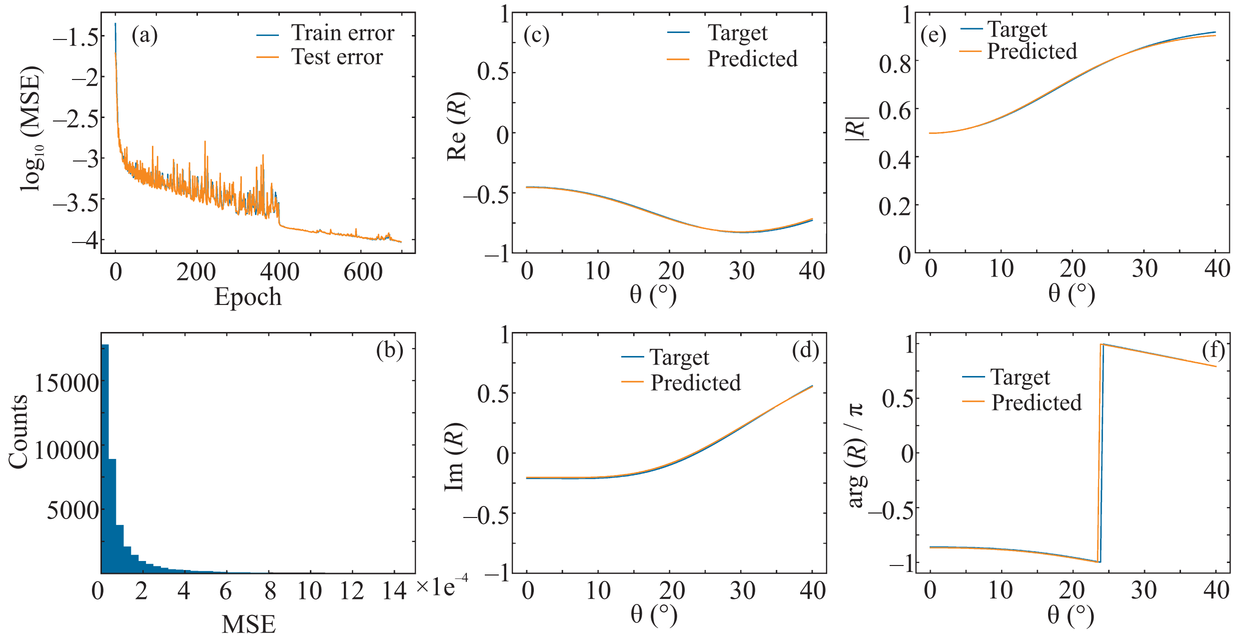


Рис. 1. (Цветной онлайн) (a) – Зависимость десятичного логарифма СКО (MSE) для тренировочного (оранжевая кривая) и тестового (синяя кривая) наборов данных от номера эпохи. (b) – Гистограмма распределения СКО по тестовому набору данных. (c)–(f) – Зависимости действительной части $\mathrm{Re}(R)$, мнимой части $\mathrm{Im}(R)$, модуля $|R|$ и фазы $\arg(R)$, нормированной на $\pi$, коэффициента отражения от угла падения. Целевой коэффициент отражения показан синей кривой, коэффициент отражения МС, предсказанной ИНС, показан оранжевой кривой

сетей [18, 23]. В этом случае от ИНС не требуется какого-то конкретного решения обратной задачи и сеть выбирает какое-то одно из множества решений, которое приводит к лучшей сходимости. Другой примечательный подход основан на генеративных состязательных сетях. В этом случае две ИНС (генератор и дискриминатор) обучаются одновременно [24]. Третий подход использует уменьшение размерности пространства параметров с помощью автоэнкодера [25]. При этом предполагается, что обратная задача имеет единственное решение в пространстве меньшей размерности. Одной общей проблемой во всех предлагаемых методах является необходимость обучения нескольких ИНС, что неизбежно приводит к дополнительным ошибкам.

В данной работе предложен альтернативный подход, использующий только одну ИНС для решения обратной задачи дизайна МС, что позволяет значительно повысить качество работы сети. Показано, что обученная ИНС успешно обходит проблему неединственности решения обратной задачи. Данный подход может использоваться для определения параметров различных нанофотонных устройств.

В работе рассматривается МС, состоящая из 10 чередующихся слоев $SiO_2$ ($n = 1.45$)/$Ta_2O_5$ ($n = 2.08$) толщиной от 0 до $d_{\max} = 350$ нм, нанесенных на стеклянную подложку ($n = 1.52$). Цель состоит в определении толщин слоев $d_i$ ($i = 1...10$) МС с заданным оптическим откликом. Входными данными для ИНС является угловой спектр амплитудного коэффициента отражения для TE-поляризованного излучения при $\lambda = 800$ нм в диапазоне углов падения $\theta$ от 0° до 40°. Спектр дискретизован на 100 точек. На вход ИНС поступает массив, состоящий из действительной $r_i'$ и мнимой $r_i''$ части коэффициента отражения для каждого угла падения, т.е. массив имеет длину 200. Действительная и мнимая части спектра отражения были выбраны вместо фазы и амплитуды, поскольку в фазе могут наблюдаться разрывы, ограничивающие точность ИНС [26]. Выходные данные ИНС представляют собой массив $D$ из 10 нормированных толщин $d_i/d_{\max}$ МС. Материал каждого слоя считается известным и фиксированным. Слой на границе с воздухом выполнен из $SiO_2$.

ММР был использован для создания набора данных, содержащих спектры коэффициента отражения случайно сгенерированных 10-слойных структур. Это гарантирует, что целевые спектры могут быть реализованы с помощью рассматриваемой в работе одномерной МС. Затем набор данных делится на тренировочную часть (используется для обучения ИНС) и тестовую (для проверки качества работы ИНС). Наш метод обучения ИНС основан на соединении сети с ММР [27]. В процессе обучения целевой угловой спектр $R_{\text{target}}(\theta)$ используется в качестве входа ИНС, а массив $D$ нормированных толщин является выходом ИНС. Затем $D$ используется в ММР для вычисления углового спектра $R_{\text{predicted}}(\theta)$ МС, сгенерированной ИНС. Затем вычисляется среднеквадратичная ошибка (СКО) между спектрами $R_{\text{predicted}}$ и $R_{\text{target}}$. После этого веса и смещения слоев ИНС обновляются с помощью обратного распространения ошибки, чтобы минимизировать СКО. Для того, чтобы ММР не становился узким местом в обучении ИНС, был разработан оптимизированный код, в котором ММР реализован путем перемножения многомерных матриц (тензоров). ММР естественным образом вписывается в процесс обучения ИНС, поскольку он основан на умножении матриц – основной операции глубокого обучения. Кроме того, ММР определяет аналитическую зависимость между толщиной слоя и спектром отражения и может быть использован для обратного распространения ошибки, т.е. расчета градиента СКО относительно весов и смещений ИНС. В описанном процессе обучения параметры МС из тренировочного набора данных никак не используются. В этом случае ИНС выучивает закономерности, заложенные в ММР, а не просто интерполирует данные из тренировочного набора.

Метод обучения ИНС был протестирован на примере сети, содержащей 4 скрытых слоя по 500 нейронов. ИНС обучалась на 100 000 образцов из тренировочного набора данных в течение 700 эпох. СКО резко падает через 10 эпох, а затем постепенно уменьшается до $10^{-4}$ (рис. 1a). Распределение СКО в тестовом наборе данных локализовано около нуля, что доказывает хорошее качество обучения (рис. 1b). Рисунки 1c, d демонстрируют работу нашей сети на случайном примере из тестового набора данных. Наша сеть успешно справляется с обратной задачей. Рисунки 1e, f демонстрируют амплитуду $|R|$ и фазу $\arg(R)$ коэффициента отражения, которые вычисляются из действительной $r_i'$ и мнимой $r_i''$ части коэффициента отражения по формулам:

$$|R_i| = \sqrt{(r_i')^2 + (r_i'')^2}, \tag{1}$$

$$\arg(R_i) = \arctan(r_i''/r_i'), \tag{2}$$

где $i = 1...100$. Так как $\arg(R)$ лежит в пределах от $-\pi$ до $\pi$, то в спектре коэффициента отражения может наблюдаться скачок на $2\pi$ (рис. 1f).

Важный вопрос заключается в том, как разработанная ИНС справляется с проблемой неединственности решения обратной задачи. На рисунках 2a, b

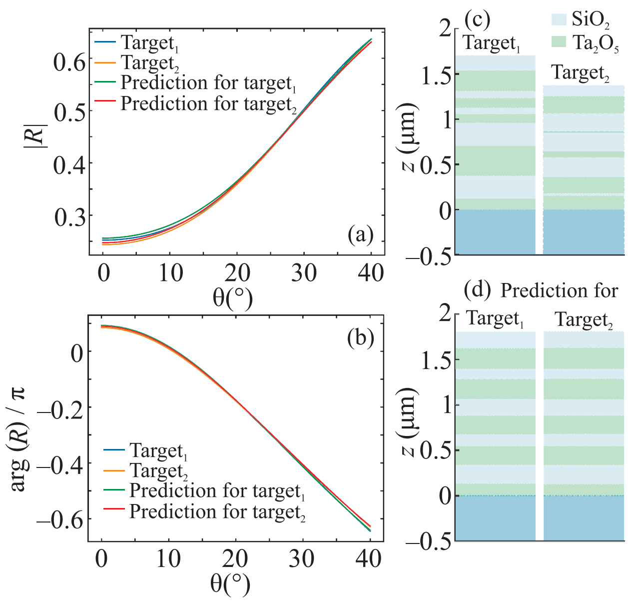


Рис. 2. (Цветной онлайн) (a), (b) – Зависимости модуля $|R|$ и фазы $\arg(R)$, нормированной на $\pi$, коэффициента отражения от угла падения для двух близких по спектру примеров из тестового набора данных (синяя и оранжевая кривые) и спектры МС, предсказанных ИНС (зеленая и красная кривые). (c), (d) – МС с близкими спектрами из тестового набора данных и МС, предсказанные ИНС

демонстрируются два похожих спектра от совершенно разных МС из тестового набора данных (рис. 2c). Среднее отклонение толщин слоев, определяемое формулой:

$$\Delta d = \sum_{i=1}^{N} \frac{d_i^{\text{target}_1} - d_i^{\text{target}_2}}{N},$$

где $N = 10$, было использовано как мера сходства МС. Для выбранных образцов $\Delta d = 96$ нм. Как видно из рис. 2a, b, ИНС успешно справляется с дизайном МС с выбранными спектрами. Однако средняя разница в толщине слоев $\Delta d$ для сгенерированных МС составляет всего 3 нм (рис. 2d). Далее была оценена корреляция между разницей спектров $\Delta R$ и разницей толщин слоев $\Delta d$ для образцов из тестового набора данных и для МС, сгенерированных ИНС. В качестве меры корреляции был использован коэффициент Пирсона $\rho_{(\Delta R,\Delta d)}$. Для тестового набора данных $\rho^{\text{test}}_{(\Delta R,\Delta d)} = -0.05$, а для МС, подобранных ИНС, $\rho^{NN}_{(\Delta R,\Delta d)} = 0.48$. Видно, что в тестовом наборе данных нет корреляции между разницей спектров и разницей толщин, т. е. похожий оптический отклик могут демонстрировать МС с абсолютно разными параметрами. Однако существует довольно сильная по-

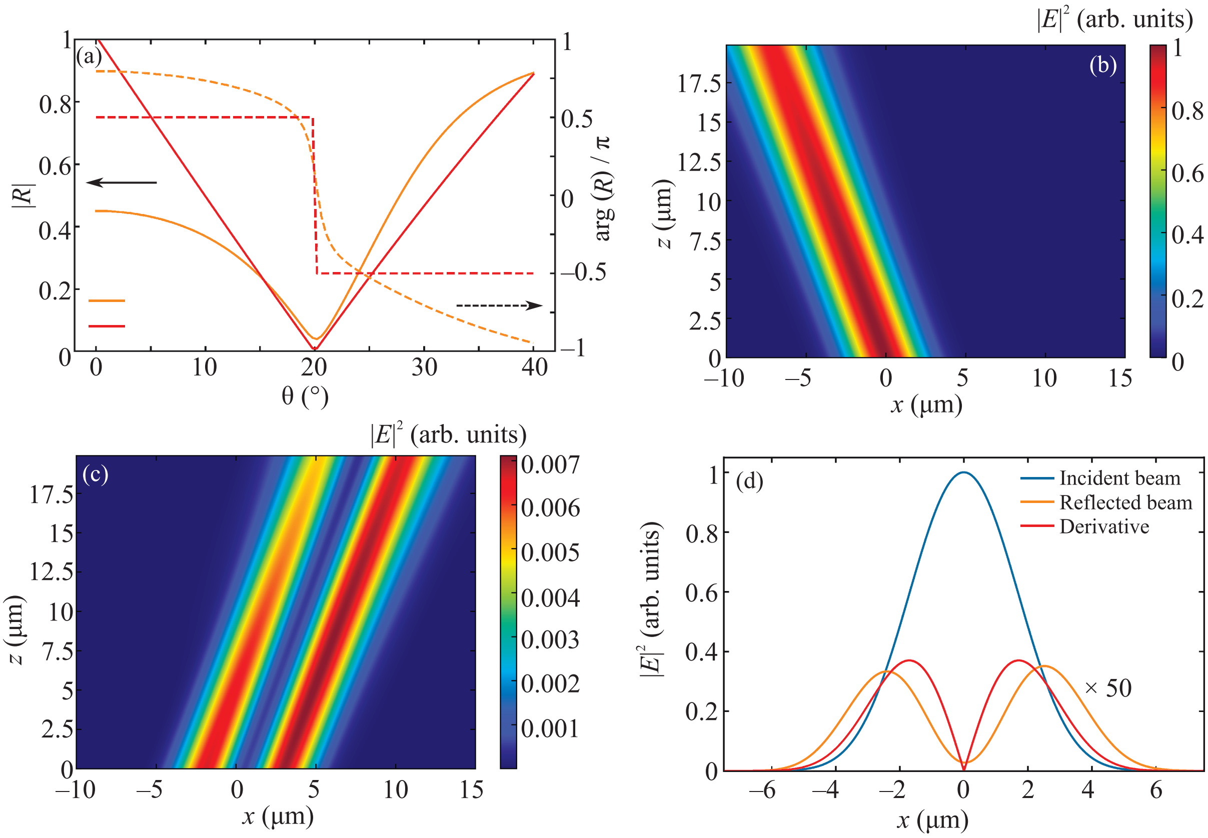


Рис. 3. (Цветной онлайн) (a) – Зависимости модуля $|R|$ (левая ось ординат) и фазы $\arg(R)$ (правая ось ординат), нормированной на $\pi$, коэффициента отражения от угла падения для идеального дифференциатора (красные кривые) и МС, предсказанной ИНС (оранжевые кривые). (b), (c) – Зависимость интенсивности падающего и отраженного от МС пучков от координаты. (d) – Сечение интенсивности падающего пучка (синяя кривая), отраженного (оранжевая) и аналитически вычисленная производная интенсивности падающего пучка (красная)

ложительная корреляция между разницей спектров и разницей толщин для образцов, подобранных сетью. Это означает, что для этих МС близкие спектры создаются похожими структурами. Таким образом, при обучении ИНС сглаживает пространство параметров и из множества возможных решений обратной задачи сходится к какому-то одному.

В качестве примера применения нашей сети были подобраны параметры МС, являющейся оптическим пространственным дифференциатором. Как известно, для реализации пространственной производной первого порядка такое устройство должно иметь передаточную функцию (в нашем случае коэффициент отражения) $R \approx ik_x$ [28], где $k_x = k_0 \sin\theta$ – тангенциальная компонента волнового вектора, $\theta$ – угол падения. Таким образом, амплитуда коэффициента отражения дифференциатора должна иметь форму, аналогичную модулю $x$, а фаза должна испытывать скачок на $\pi$. Для демонстрации работы сети была спроектирована МС, выполняющая дифференцирование при угле падения 20°. МС, подобранная ИНС, обладает спектром, похожим на отклик идеального дифференциатора (рис. 3a). Работа разработанного устройства была исследована с помощью собственного метода матриц распространения, который позволяет вычислять отражение не только для плоских волн, но и для сфокусированных гауссовых пучков. Это достигается путем разложения гауссова пучка на набор плоских волн [29]. Качество дифференцирования было исследовано для сфокусированного на поверхность МС 3-мкм гауссова пучка. Поскольку отражение в районе 20° близко к 0, то падающий (рис. 3b) и отраженный лучи (рис. 3c) визуализированы отдельно для наглядности. Рисунок 3d показывает поперечное сечение падающего и отраженного лучей, а также аналитически рассчитанную производную первого порядка. Действительно, наша ИНС подобрала МС с откликом, похожим на идеальный дифференциатор, но не идентичную ему. Увеличение количества слоев может обеспечить лучшее качество взятия производной.

Таким образом, в данной работе продемонстрирован метод обучения одной единственной искусственной нейронной сети для решения обратной задачи дизайна многослойной фотонной структуры по заданному угловому спектру коэффициента отражения. Предложенный метод позволяет достичь среднеквадратической ошибки меньше $10^{-4}$, и тем самым проектировать МС, обладающие целевым оптическим откликом с высокой степенью достоверности. Обученная ИНС успешно решает проблему неединственности решения обратной задачи путем сглаживания пространства параметров. Кроме того, ИНС подходит для проектирования устройств нанофотоники, например, оптического дифференциатора. Предложенный метод обучения ИНС может быть перенесен на частотный спектр коэффициента отражения, что откроет еще больше возможностей для создания нанофотонных устройств с различной функциональностью.

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации, грант # 14.W03.008.31 (разработка оптического дифференциатора) и грант # 075-15-2020-801 (разработка нейронной сети), Российского фонда фундаментальных исследований в рамках научного проекта # 19-32-90225 (разработка оптимизированного метода матриц распространения). Часть исследований выполнена при поддержке Центра квантовых технологий МГУ и некоммерческого Фонда развития науки и образования "Интеллект".

1. A. F. Koenderink, A. Alu, and A. Polman, Science **348**, 516 (2015).
2. А. Д. Гартман, М. К. Кройчук, А. С. Шорохов, А. А. Федянин, Письма в ЖЭТФ **112**, 730 (2020).
3. A. M. Chernyak, M. G. Barsukova, A. S. Shorokhov, A. I. Musorin, and A. A. Fedyanin, JETP Lett. **111**, 46 (2020).
4. P. S. Pankin, B.-R. Wu, J.-H. Yang, K.-P. Chen, I. V. Timofeev, and A. F. Sadreev, Commun. Phys. **3**, 1 (2020).
5. B. I. Afinogenov, A. A. Popkova, V. O. Bessonov, B. Lukyanchuk, and A. A. Fedyanin, Phys. Rev. B **97**, 115438 (2018).
6. B. I. Afinogenov, V. O. Bessonov, I. V. Soboleva, and A. A. Fedyanin, ACS Photonics **6**, 844 (2019).
7. M. N. Romodina, I. V. Soboleva, A. I. Musorin, Y. Nakamura, M. Inoue, and A. A. Fedyanin, Phys. Rev. B **96**, 081401 (2017).
8. W. Wu, W. Jiang, J. Yang, S. Gong, and Y. Ma, Opt. Lett. **42**, 5270 (2017).
9. F. Barachati, A. Fieramosca, S. Hafezian, J. Gu, B. Chakraborty, D. Ballarini, L. Martinu, V. Menon, D. Sanvitto, and S. Kéna-Cohen, Nature Nanotech. **13**, 906 (2018).
10. I. S. Kriukova, V. A. Krivenkov, P. S. Samokhvalov, and I. R. Nabiev, JETP Lett. **112**, 537 (2020).
11. A. E. Schegolev, A. M. Popov, A. V. Bogatskaya, P. M. Nikiforova, M. V. Tereshonok, and N. V. Klenov, JETP Lett. **111**, 371 (2020).
12. K. R. Safronov, D. N. Gulkin, I. M. Antropov, K. A. Abrashitova, V. O. Bessonov, and A. A. Fedyanin, ACS Nano **14**, 10428 (2020).
13. S. D. Campbell, D. Sell, R. P. Jenkins, E. B. Whiting, J. A. Fan, and D. H. Werner, Opt. Mater. Express **9**, 1842 (2019).
14. R. L. Haupt and D. H. Werner, *Genetic algorithms in electromagnetics*, John Wiley & Sons, New Jersey (2007).
15. J. Robinson and Y. Rahmat-Samii, IEEE Trans. Antennas Propag. **52**, 397 (2004).
16. J. S. Jensen and O. Sigmund, Laser Photonics Rev. **5**, 308 (2011).
17. S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, Nat. Photonics **12**, 659 (2018).
18. D. Liu, Y. Tan, E. Khoram, and Z. Yu, ACS Photonics **5**, 1365 (2018).
19. J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljačić, Sci. Adv. **4**, eaar4206 (2018).
20. C. C. Nadell, B. Huang, J. Malof, and W. J. Padilla, Opt. Express **27**, 27523 (2019).
21. I. Sajedian, J. Kim, and J. Rho, Microsyst. Nanoeng. **5**, 1 (2019).
22. J. He, C. He, C. Zheng, Q. Wang, and J. Ye, Nanoscale **11**, 17444 (2019).
23. S. So, J. Mun, and J. Rho, ACS Appl. Mater. Interfaces **11**, 24264 (2019).
24. Z. Liu, D. Zhu, S. P. Rodrigues, K. T. Lee, and W. Cai, Nano Lett. **18**, 6570 (2018).
25. Y. Kiarashinejad, S. Abdollahramezani, and A. Adibi, Npj Comput. Mater. **6**, 1 (2020).
26. S. An, C. Fowler, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, K. A. Richardson, T. Gu, J. Hu, and H. Zhang, ACS Photonics **6**, 3196 (2019).
27. C. C. Katsidis and D. I. Siapkas, Appl. Opt. **41**, 3978 (2002).
28. A. Pors, M. G. Nielsen, and S. I. Bozhevolnyi, Nano Lett. **15**, 791 (2015).
29. F. I. Baida and M.-P. Bernal, Commun. Phys. **3**, 1 (2020).