

# Fourier modal method for inverse design of metasurface-enhanced micro-LEDs

MARTIN F SCHUBERT<sup>1,\*</sup> AND ALEC M HAMMOND<sup>1</sup>

<sup>1</sup>*Meta, 1 Hacker Way, Menlo Park, CA 94025, USA*

<sup>\*</sup>*mfschubert@gmail.com*

**Abstract:** We present a simulation capability for micro-scale light-emitting diodes ( $\mu$ LEDs) that achieves comparable accuracy to CPU-based finite-difference time-domain simulation but is more than  $10^7$  times faster. Our approach is based on the Fourier modal method (FMM)—which, as we demonstrate, is well suited to modeling thousands of incoherent sources—with extensions that allow rapid convergence for  $\mu$ LED structures that are challenging to model with standard approaches. The speed of our method makes the inverse design of  $\mu$ LEDs tractable, which we demonstrate by designing a metasurface-enhanced  $\mu$ LED that doubles the light extraction efficiency of an unoptimized device.

## 1. Introduction

Micro-scale light-emitting diodes ( $\mu$ LEDs) with lateral dimensions near  $1\ \mu\text{m}$  are of interest for a variety of applications including augmented reality displays [1, 2]. In this context, high light extraction efficiency (LEE) is essential due to the requirement for high brightness (e.g. for outdoor usage) and the need to operate for extended periods on battery power. It would be desirable to apply methods such as inverse design to obtain  $\mu$ LED designs with higher performance; inverse design is a powerful technique that automatically discovers the topology and shapes of designs that minimize some objective function [3–5], and would be suitable to the creation of metasurfaces that enhance  $\mu$ LED LEE. However, light generated in the  $\mu$ LED is spatially incoherent; spatially incoherent sources (arising in e.g. spontaneous emission [6–8] and thermal emission [9, 10]) are extremely expensive to model by standard approaches (requiring many independent simulations), making inverse design of structures including incoherent sources computationally intractable. Consequently, previous works on modeling and optimization of  $\mu$ LEDs have focused on symmetric or low-dimensional cases, or used a widely-spaced dipoles to approximate a planar active region [11, 12], leaving the challenge of general 3D  $\mu$ LED inverse design unaddressed.

In this work, we present a new simulation capability based on the Fourier modal method (FMM) which overcomes these limitations. For the  $\mu$ LED, our method gives results that are in excellent agreement with finite-difference time-domain (FDTD) simulations while being more than  $10^7$  times faster (Sec. 4). This makes it practical for use in an inverse design setting, which we demonstrate by designing metasurface-enhanced  $\mu$ LEDs that significantly improve LEE (Sec. 6). The speed of our method also enables calculation of quantities such as high-resolution spatial maps of LEE, allowing new insights into the physics of  $\mu$ LED devices.

Related to our work, novel factorization methods for multi-channel inverse design [13, 14] and a trace formulation of photonic inverse design [15] have recently been introduced, which have lower computational cost compared to conventional inverse design formulations. Our work shows that—for problems such as the  $\mu$ LED—with an appropriate choice of the simulation algorithm, the standard formulation may be sufficient. In particular, the FMM has several characteristics that enable efficient modeling of many incoherent sources. The FMM treats fields in periodic stratified media in a truncated Fourier basis, which has the potential to represent fields accurately with relatively few terms. As in other methods, with the FMM one solves a linear system to obtain the fields generated by a source; this system has dimension given by the size of the basis, which for FMM is related only to the details of the structure in *two* dimensions, as the third

dimension is handled analytically. The resulting linear system can be relatively small, enabling direct solution methods to be employed. Finally, while constructing the linear system involves eigendecompositions and is expensive, the results of these operations can often be reused e.g. in an optimization setting where many layer profiles may be unchanged between iterations.

Despite these potential advantages, there are challenges that have discouraged the use of the FMM for problems such as the  $\mu$ LED. For example, the original and common FMM formulation exhibits poor convergence in structures containing metals [16]. In addition, FMM naturally treats periodic structures and sources; while  $\mu$ LEDs are generally formed in periodic arrays, treatment of a *source* as periodic gives rise to nonphysical interference effects [17]. We address these issues with a vector formulation of the FMM—including an improved method for automatically computing vector fields—and Brillouin zone integration [17–19], which dramatically improve convergence (Sec. 5) and allow modeling of localized dipoles in a periodic  $\mu$ LED array.

Our implementation of the FMM is `FMMAX`, which is based on *Jax* [20] and joins several recent codes that support automatic differentiation [21–24], enabling the gradient calculation needed for inverse design. Our code is distinguished by a flexible low-level programming interface which admits uncommon use cases such as the  $\mu$ LED simulations carried out here, and is freely available at [github.com/facebookresearch/fmmx](https://github.com/facebookresearch/fmmx).

## 2. Method

Throughout this section, we discuss extensions of the basic FMM needed to enable  $\mu$ LED simulation and inverse design. For a full description of the FMM, we refer the reader to [25] and [26], which our implementation closely follows.

### 2.1. Vector FMM formulations

It has long been recognized that the original FMM formulation (referred to as *FFT*, as in [25]) converges well only when electric fields are tangent to material interfaces [16]; this was improved by vector FMM formulations, which introduce a local coordinate system with unit vectors that are tangent and normal to the boundaries of features. Vector formulations allow orientation-dependent Fourier factorizations of permittivity to be used for tangent and normal components of the electric displacement field, which is the essential change of the vector FMM methods from the original formulation [27]. In an inverse design setting, the vector fields must be automatically generated from the evolving geometry. Various schemes have been developed based on interpolation [28] and functional minimization [25]; we adopt the latter due to its applicability to continuously varying permittivity, as is encountered during the course of optimization.

As in [25], the vector field  $\mathbf{t} = [t_x, t_y]^T$  for a permittivity array  $\varepsilon$  is obtained by minimizing a loss function  $\mathcal{L}$ . High-quality vector fields should be smoothly varying and normal to the permittivity gradient, i.e. tangent to the interfaces in the structure. This is achieved using

$$\mathcal{L}(\mathbf{t}, \varepsilon) = \sum |t_x \nabla_y \varepsilon - t_y \nabla_x \varepsilon| + |\nabla_{xy} t_x|^2 + |\nabla_{xy} t_y|^2, \quad (1)$$

where the sum is over all the elements in  $\mathbf{t}$ . The first term is minimized when  $\mathbf{t}$  is correctly oriented, while the remaining terms are minimized when  $\mathbf{t}$  varies slowly. The resulting optimization problem,

$$\begin{aligned} \mathbf{t} = \operatorname{argmin}_{\mathbf{t}^*} \mathcal{L}(\mathbf{t}^*, \varepsilon), \\ \text{s.t. } |\mathbf{t}^*| \leq 1 \end{aligned} \quad (2)$$

can be solved by standard gradient-based methods.

The choice of initial  $\mathbf{t}^*$  is consequential. In a novel method termed *Jones direct*, we compute an initial real-valued vector field (even for complex  $\varepsilon$ ) and convert to a complex Jones field using

the method of [29]. The final complex-valued  $\mathbf{t}$  is obtained by directly optimizing this Jones field, and advantageously lacks both discontinuities and zeros.

A real-valued  $\mathbf{t}$  can be obtained by avoiding the Jones conversion and directly optimizing a real  $\mathbf{t}^*$ ; this will lack discontinuities but include zeros, and is equivalent to the *Pol* method of [25]. The field can be normalized to have magnitude 1 everywhere (the *Normal* method of [25]) or it can be converted to a Jones field as a post-processing step (the *Jones* method of [25]). In general, the *Normal* method will yield fields having discontinuities, while the *Jones* method will yield fields that lack discontinuities and zeros, but vary more rapidly than those obtained by the *Jones direct* method. Thus, the *Jones direct* scheme has favorable characteristics for generation of a local coordinate system represented in a Fourier basis.

Example vector fields for these methods are in Appendix A. Consistent with the theoretical benefits discussed above, we observe best convergence for the FMM with the *Jones direct* formulation (Appendix B) and use the method throughout this work.

## 2.2. Brillouin zone integration for aperiodic sources

Sources in periodic structures can be modeled in the FMM by expanding the source spatial profile in terms of the layer eigenmodes [26]. However, in this approach both the sources and the structure are periodic, which is not the case in a real  $\mu$ LED and can give rise to unphysical interference effects.

In principle, larger supercells including multiple  $\mu$ LEDs could be modeled, but due to the unfavorable scaling of the FMM algorithm this may be prohibitive. Absorbing boundary conditions could also be introduced, as used previously in extensions of the FMM to accommodate aperiodic structures [30, 31]. A further alternative—the approach adopted here—is the Brillouin zone (BZ) integration method of [17–19], in which the electric field from an isolated source in a periodic structure is found by,

$$E(r) = \frac{1}{A_{BZ}} \int_{BZ} E_{k_B}(r) dk_B, \quad (3)$$

where  $E_{k_B}$  is the field with Bloch-periodic boundary conditions for wavevector  $k_B$ , and  $A_{BZ}$  is the BZ area. The calculation of each  $E_{k_B}$  requires a simulation of a single unit cell, and thus the problem of modeling a large supercell is transformed into one of several smaller simulations.

Throughout this work, we approximate integrals over the Brillouin zone by averaging over a regular  $k_B$  grid. When the grid has shape  $M \times M$ , we refer to the case as  $M \times M$  BZ integration; when no BZ integration is performed, we refer to this as the periodic dipole approximation (PDA).

## 2.3. Calculation of light extraction efficiency

For the  $\mu$ LED modeled here, we are primarily concerned with the total LEE and the power radiated by the dipole. For a given wavelength, the LEE is obtained by,

$$LEE = \sum_{r,p,k_B} w_p P_{r,p,k_B}^{extracted} \bigg/ \sum_{r,p,k_B} w_p P_{r,p,k_B}^{emitted} \quad (4)$$

Here,  $P_{r,p,k_B}^{emitted}$  and  $P_{r,p,k_B}^{extracted}$  are the power emitted by a dipole, and power emitted by the dipole which is extracted from the  $\mu$ LED. The powers are measured with planar monitors, obtained by summing over the powers in all the Fourier orders as described in [25], and are indexed by  $r$ ,  $p$ , and  $k_B$ , corresponding to the dipole position, dipole orientation, and BZ point, respectively. We use a dipole-orientation-dependent weight  $w_p$ , with values of 1 for the  $x$ - and  $y$ -oriented dipoles and 0.1 for the  $z$ -oriented dipoles. These generally have lower extraction efficiency and can be suppressed in material systems such as AlInGaP by appropriately straining the quantum wells [32].

### 3. $\mu$ LED structure and optimization problem

We now describe the  $\mu$ LED geometry and corresponding optimization problem used throughout the rest of our analysis. We consider an array of simplified cylindrical  $\mu$ LEDs with a  $1.4\ \mu\text{m}$  pitch, which is illustrated in Fig. 1. The semiconductor region has a diameter of  $1\ \mu\text{m}$ , the conducting oxide has a diameter of  $0.8\ \mu\text{m}$ , and the passivation sidewall has a thickness of  $0.1\ \mu\text{m}$ . Other layer thicknesses are optimizable parameters specified elsewhere. A spatial resolution of  $10\ \text{nm}$  is used for all layers. Refractive indices of the materials are as follows: semiconductor,  $3.0$ ; conducting oxide,  $1.9 + 0.005i$ ; passivation,  $1.5$ ; and metal,  $0.2 + 3.3i$ . Values are chosen to be representative of AlInGaP, indium-tin oxide, silicon oxide, and gold, respectively.

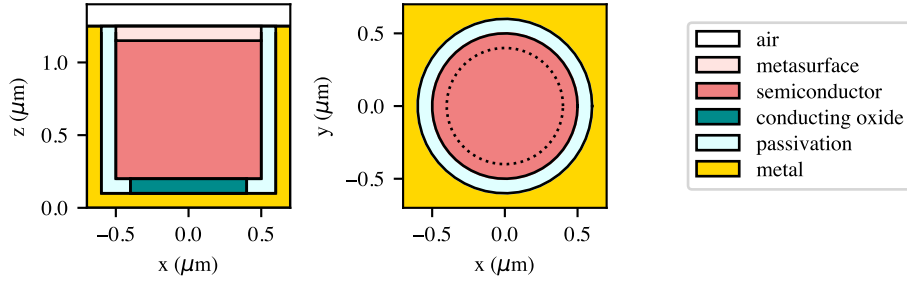


Fig. 1. Side and top view of the  $\mu$ LED unit cell analyzed throughout this work. A high index semiconductor ( $n = 3.0$ ) cylinder is embedded within a metallic substrate ( $n = 0.2 + 3.3i$ ) and separated by a thin passivation layer ( $n = 1.5$ ) and transparent conducting oxide ( $n = 1.9 + 0.005i$ ). The metasurface “design region” of the  $\mu$ LED unit cell resides on top of the semiconductor cavity and emits into air. The material of the metasurface is composed of a refractive index that is linearly interpolated between the index of the semiconductor and air. The metasurface itself is parameterized by a 2D grid of “pixels” [33] in  $x/y$  and projected into the  $z$ -direction using a thickness chosen by the optimizer. Dipoles within the  $\mu$ LED are located in a planar source region embedded within the semiconductor layer, which is free to move as the optimizer dictates.

We model dipoles in a plane within the semiconductor material, and consider  $x$ -,  $y$ -, and  $z$ -oriented dipoles on a  $20\ \text{nm}$  grid (5847 total dipoles). The spatial profile of the dipoles is Gaussian with a  $60\ \text{nm}$  full-width at half-maximum. Power emitted by each dipole is measured using planar monitors positioned  $50\ \text{nm}$  from the dipole plane. We model a single plane, corresponding e.g. to a single quantum well active region, but note that it would be possible to model multi-quantum-well active region with no additional eigendecompositions, and therefore relatively cheaply.

We turn now to the  $\mu$ LED optimization problem. In this paper, we exercise unconstrained optimization, as our primary aim is to demonstrate the functioning of the simulator and inverse design pipeline, rather than to obtain designs suitable for manufacture, or to form estimates of the potential performance of metasurface-enhanced  $\mu$ LEDs. We use a monochromatic objective  $\mathcal{L}(p) = 1 - LEE(p)$  for a  $630\ \text{nm}$  emission wavelength. Here,  $p$  represents the optimizable parameters, including layer thicknesses and the metasurface density  $\rho$ . The density is a two-dimensional array with values in the range  $[0, 1]$  from which the permittivity at each point is computed by interpolation between the permittivity of air and of semiconductor using the method of [34].

Although one could directly optimize the elements of the density array  $\rho$ , this can produce

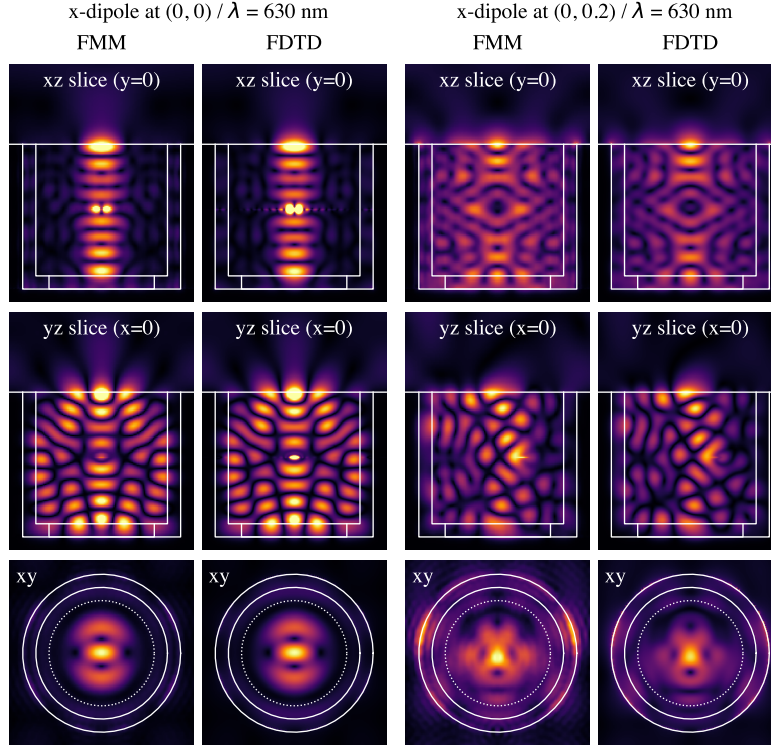


Fig. 2. Cross sections of the steady-state electric field magnitude for dipoles at  $(x, y) = (0, 0)$  and  $(0, 0.2)$  computed by FMM and FDTD. The  $xz$  and  $yz$  are taken through the center of the  $\mu$ LED, while the  $xy$  section is taken at the semiconductor/air interface. The methods show strong agreement at all slices and for both dipole positions, as demonstrated by the consistent representation of the complicated modal patterns.

designs with very fine features that would not be practical to manufacture. Therefore, we choose to instead optimize a *latent density*  $\tilde{\rho}$  also in the range  $[0, 1]$ , from which  $\rho$  is obtained by,

$$\rho = \frac{1}{2} (1 + \tanh(\beta(2\tilde{\rho} - 1) \otimes k)) \quad (5)$$

where  $k$  is a Gaussian kernel with a full-width at half-maximum of 40 nm and  $\beta = 2$ . This expression is similar to the transforms commonly used in density-based topology optimization [35] and help ensure that  $\rho$  lacks features which are too small by encouraging an implicit lengthscale. Furthermore, this approach helps *regularize* the optimization problem, which is solved using the L-BFGS-B algorithm [36].

#### 4. Field calculation and validation

To establish the validity of our method, we first make a direct comparison of the steady-state fields computed by *Meep* [37], a mature FDTD solver, to those computed by FMM. We consider a simple  $\mu$ LED lacking a metasurface, with semiconductor and conducting oxide thicknesses of 1.0  $\mu\text{m}$  and 0.1  $\mu\text{m}$ , respectively. Dipoles emitting at 630 nm are centered vertically in the semiconductor and positioned (with units of  $\mu\text{m}$ ) at  $(x, y) = (0, 0)$  and  $(0, 0.2)$ , both oriented in the  $x$ -direction.

Figure 2 shows cross sections of the electric field magnitude computed by the two methods. The field structure for both dipole locations computed by FMM and FDTD are in excellent agreement; in particular, the nodal lines found by the two methods are nearly indistinguishable.

Our FMM calculation uses the PDA with the *Jones direct* formulation and  $N = 2000$ .

The FDTD calculation uses a spatial resolution of 200 voxels /  $\mu\text{m}$ , periodic boundary conditions in the  $x$ - and  $y$ -directions, and perfectly electric conductor (PEC) boundary conditions in the  $z$ -direction. To mitigate backreflections from the upper  $z$ -boundary, we used a 2  $\mu\text{m}$  thick, adiabatic, conductive absorber layer above the  $\mu\text{LED}$ . Using a nonlinear optimization algorithm [38], we fit a Lorentz-drude material model to the desired complex refractive index values at  $\lambda = 630\text{ nm}$  to ensure simulation stability [39]. The FDTD simulation using a broadband Gaussian pulse centered at 630 nm and terminated the simulation after the fields were sufficiently decayed. Just like the FMM solver, the spatial profile of the FDTD source was a normalized Gaussian with a FWHM of 60 nm. The steady-state fields for the 630 nm wavelength were recorded using rolling discrete-time Fourier transform monitors at the specified cross sections.

## 5. Solver convergence and performance

Next, we study the convergence and performance of the FMM in the context of our  $\mu\text{LED}$  simulations and compare the results to FDTD. We note that the numerical accuracy of each algorithm is fundamentally dictated by the underlying discretization resolution; for FDTD, this corresponds to the actual spatial resolution of the grid, or the total number of simulation voxels, whereas with FMM this refers to the spatial-frequency resolution, or the total number of Fourier orders. To quantify the relative convergence characteristics of both algorithms, we simulated the  $\mu\text{LED}$  structure and dipole configuration of Sec. 4 with multiple resolutions.

First, we examine the importance of the FMM formulation itself. Fig. 3 shows the LEE for the  $x$ -oriented dipoles at  $(x, y) = (0, 0)$  and  $(0, 0.2)$  as a function of the Fourier orders  $N$  and for the *Jones direct* and *FFT* formulations. For the *Jones direct* formulation, the LEE for both dipoles quickly converges; the values with  $N = 4800$  are 36.7% and 30.4% respectively, but with only  $N = 400$  the values are 37.4% and 29.4%, in excellent agreement with the converged results. By contrast, the *FFT* formulation exhibits very poor convergence; as  $N$  increases the values trend in the direction of the *Jones direct* results, but they exhibit substantial fluctuation and remain far from the converged result even with  $N = 4800$ .

We now directly compare the convergence of FMM and FDTD. Fig. 3 also shows the convergence of LEE for both dipole positions calculated by FDTD with varying spatial resolution. With a resolution of 200 voxels /  $\mu\text{m}$  ( $\Delta x = 5\text{ nm}$ ), we obtain values of 39.2% and 31.7%, in good agreement with the converged *Jones direct* result. Such a high spatial resolution for FDTD is expected due to the high-contrast metals within the  $\mu\text{LED}$  model, although the application of subpixel-smoothing algorithms or other conformal meshing approaches compatible with dispersive media should accelerate the convergence [40]. In combination with the field profiles in Fig. 2, this shows that the vector FMM can be used for modeling and LEE calculation of  $\mu\text{LEDs}$ , with results comparable to high-resolution FDTD.

Despite the similarity in results between FMM and FDTD, there are major differences in the performance and computational burden of the two algorithms. Fig. 4 shows the elapsed time for the FMM and FDTD simulations in Fig. 3 as a function of Fourier order and resolution respectively. All calculations were performed using 10 cores of a 2.20GHz Intel® Xeon® CPU E5-2698 v4 with 512 GB of RAM and two NUMA nodes. Although *FMMAX* supports GPU acceleration, this was not used to allow a direct comparison. The FMM results are shown for the case where three incoherent dipoles are modeled, and where all  $\sim 6000$  dipoles are modeled; the results show the negligible incremental cost of simulating additional dipoles.

The FDTD elapsed times are shown for centered and offset dipoles; the centered-dipole simulation used odd  $x$ -symmetry and even  $y$ -symmetry to reduce the computational burden,

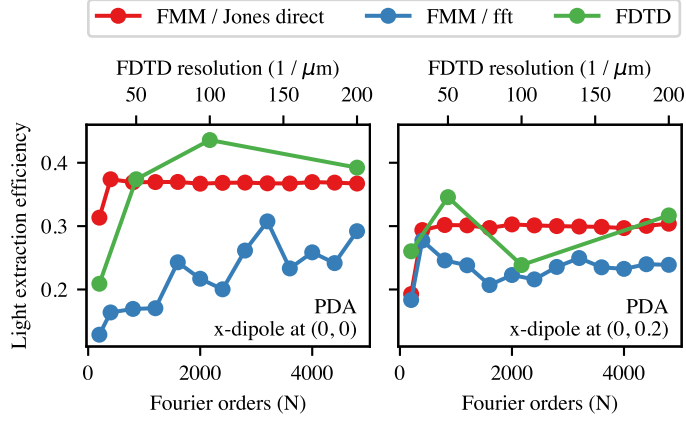


Fig. 3. Convergence of the light extraction efficiency for FMM and FDTD as a function of Fourier order and spatial resolution respectively. Results are given for both the centered and offset  $x$ -oriented dipole. While the *Jones direct* FMM formulation of FMM converges for  $N \geq 400$ , the *FFT* formulation fails to converge; the calculated efficiency trends in the direction of the *Jones direct* result, but does not reach agreement even for  $N = 4800$ . FDTD requires a resolution of 200 voxels /  $\mu\text{m}$  ( $\Delta x = 5\text{ nm}$ ) to reach suitable convergence, largely due to the high-contrast metals within the  $\mu\text{LED}$  model.

while the offset dipole considered the full simulation volume and is representative of the general case. Considering this offset dipole, there is approximately a  $3 \times 10^3$  speedup when comparing the 200 voxels /  $\mu\text{m}$  FDTD results (the coarsest resolution that yields converged results) to the  $N = 800$  Fourier-term results (the lowest  $N$  we used in the inverse design setting).

Practically speaking, the speedup is even more significant; if FDTD simulations were carried out for the  $\sim 6000$  dipoles considered in this work, we project that CPU time would be  $\sim 2 \times 10^7$  times greater than for the equivalent calculation by FMM. If simulations considering multiple wavelengths were carried out, the advantage would be reduced; FDTD automatically handles multiple wavelengths within a single simulation, while with the FMM each wavelength requires an independent simulation. However, for the  $\mu\text{LED}$  we have found that  $O(10)$  wavelengths are sufficient, and in any case these simulations can easily be distributed across many machines (the task is embarrassingly parallel). This makes FMM a strong choice for problems such as the  $\mu\text{LED}$ .

We note a few other important observations regarding the convergence trends of each method. For low  $N$ , for example, the vector field calculation contributes significantly to the total time of the FMM method, causing the non-polynomial scaling seen in Fig. 4. The FDTD elapsed times also exhibit some deviation from polynomial behavior and a crossover at  $100 / \mu\text{m}$  resolution. *Meep* requires the user to properly choose a hardware configuration (e.g. the proper number of processes to launch) to maximize the timestepping rate [41]. Choosing these parameters correctly is highly dependent on the underlying simulation resolution, and most likely accounts for the aforementioned crossover point. The  $\mu\text{LED}$  structure may also be more or less resonant for dipoles at different positions in a manner that depends on resolution; for the FDTD algorithm, this can also cause simulation times to vary.

While not clear from the data in Fig. 4, the FMM solver also offers advantages when used within an optimization pipeline where many solves are performed sequentially. Specifically, the eigendecomposition for layers whose cross-sections do not change can be cached and reused. With *Jax*, this optimization is easily done by *jit*-ing (just-in-time compiling) the simulation and



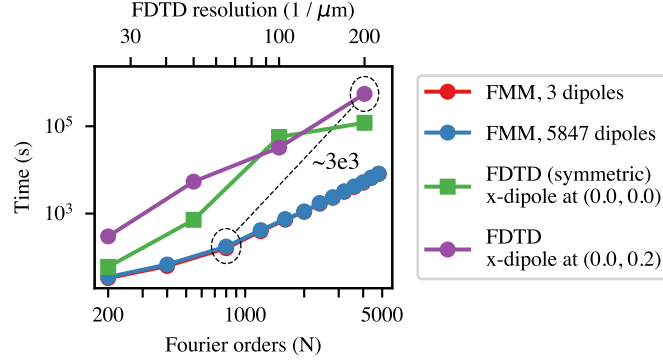


Fig. 4. Comparison of elapsed time for FMM and FDTD LEE calculations. The results indicate an approximate  $\sim 3 \times 10^3$  times speedup when comparing the 200 voxels /  $\mu\text{m}$  FDTD simulation to the  $N = 800$  Fourier term FMM simulation (resolutions where both algorithms are converged). Notably, the FDTD results correspond to a single dipole simulation. The expected computational cost for additional dipoles scales linearly with the number of dipoles. This is in direct contrast with the FMM method shown here, where the marginal cost of an additional  $\sim 6000$  dipoles is negligible. Therefore, the projected speedup for a full  $\mu\text{LED}$  simulation is  $\sim 2 \times 10^7$ .

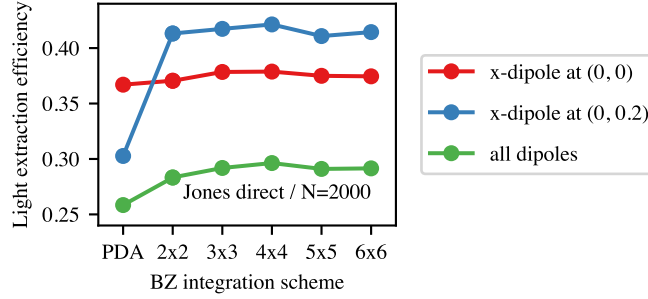


Fig. 5. Light extraction efficiency for various BZ integration schemes computed with the *Jones direct* FMM formulation and  $N = 2000$ . Values are for individual  $x$ -oriented dipoles and a full calculation including all  $x$ -,  $y$ -, and  $z$ -oriented dipoles in the basic  $\mu\text{LED}$ . Each “scheme” is defined the regular grid of wavevectors in the first BZ used to approximate a BZ integral, such that  $3 \times 3$  corresponds to nine evenly distributed points (three samples along  $k_x$  and three samples along  $k_y$ ). Importantly, for the  $\mu\text{LED}$  problem, the LEE appears to converge with a relatively small number of points.

specifying that eigendecompositions be carried out at compile time. For the  $\mu\text{LED}$  problem, the savings are substantial—while eigendecomposition for three patterned layers is needed at the first step, at subsequent steps only one eigendecomposition (for the metasurface) is needed. For the  $N = 800$  case on our Xeon<sup>®</sup> system, this results in a step time of 49 seconds for the second and subsequent steps, compared to 175 seconds for the first step.

Finally, to study the effect of BZ integration, we calculated LEE using the PDA and BZ integration schemes up to  $6 \times 6$ , using the *Jones direct* formulation and  $N = 2000$ . The results are in Fig. 5; values are shown for the two  $x$ -oriented dipoles and the total LEE considering all  $\sim 6000$  dipoles in the active region. The quality of the PDA apparently depends on the dipole position: for the dipole at  $(0, 0)$ , going to  $2 \times 2$  BZ integration and beyond only slightly increases



	unoptimized	thicknesses	thicknesses and metasurface
metasurface	–	–	0.102 $\mu\text{m}$
semiconductor height	1.000 $\mu\text{m}$	0.945 $\mu\text{m}$	0.966 $\mu\text{m}$
source offset	0.500 $\mu\text{m}$	0.531 $\mu\text{m}$	0.507 $\mu\text{m}$
conducting oxide	0.100 $\mu\text{m}$	0.054 $\mu\text{m}$	0.102 $\mu\text{m}$

Table 1. Thicknesses for the unoptimized and optimized  $\mu\text{LEDs}$ .

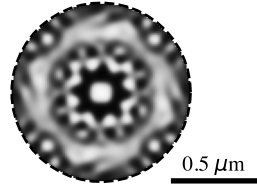


Fig. 6. The optimized metasurface density  $\rho$ ; light and dark colors correspond to  $\rho = 0$  (air) and  $\rho = 1$  (semiconductor), respectively. The dashed line indicates the passivation/semiconductor interface.

the computed LEE value. By contrast, the dipole at  $(0, 0.2)$  has a dramatic increase in the LEE when BZ integration is used, and a sizeable increase is also seen in the all-dipole case. This indicates that avoiding the PDA is critical in the calculation of the LEE for a  $\mu\text{LED}$ .

Motivated by these results, for the remainder of this work we make use of two simulation configurations. In the inverse design setting, to balance accuracy against speed, we use the *Jones direct* method with  $N = 800$  and  $2 \times 2$  BZ integration. For validation, we use  $N = 2000$ , also with *Jones direct* and  $2 \times 2$  BZ integration.

## 6. Optimization results

To exercise our inverse design pipeline and examine the potential of metasurface-enhanced  $\mu\text{LEDs}$ , we consider three cases: the first is the initial, unoptimized  $\mu\text{LED}$  with structure discussed earlier, i.e. with no metasurface, conducting oxide thickness of 0.1  $\mu\text{m}$ , semiconductor thickness of 1  $\mu\text{m}$ , and dipoles are located 0.5  $\mu\text{m}$  above the oxide. In the second, the film thicknesses are optimized, using the given values as the initial solution. In the third case, both the film thicknesses and the metasurface density are optimizable. Here, the initial semiconductor thickness is reduced to 0.95  $\mu\text{m}$  and the initial metasurface thickness is 0.1  $\mu\text{m}$ , resulting in a structure whose optical thickness approximately matches that of the unoptimized  $\mu\text{LED}$ . The latent density is uniformly initialized with a value of 0.5.

The thickness-optimized  $\mu\text{LED}$  optimization proceeded for 17 iterations before the optimizer converged, reaching an LEE of 40.1% at 630 nm, compared to 28.4% for the initial  $\mu\text{LED}$ . The optimized film thicknesses are given in Table 1. Changes to thicknesses from the initial values are relatively small, consistent with a loss landscape having many local minima, as frequently encountered in thin film optimization problems. The  $\mu\text{LED}$  with optimized film thicknesses and metasurface proceeded for 69 iterations, ultimately reaching an LEE of 56.2%. The film thicknesses are given in Table 1, and the metasurface density is shown in Fig. 6.

To understand the mechanism for LEE improvement in the  $\mu\text{LEDs}$  obtained by optimization, we computed high-resolution maps of the LEE and emitted power for the three  $\mu\text{LED}$  structures.

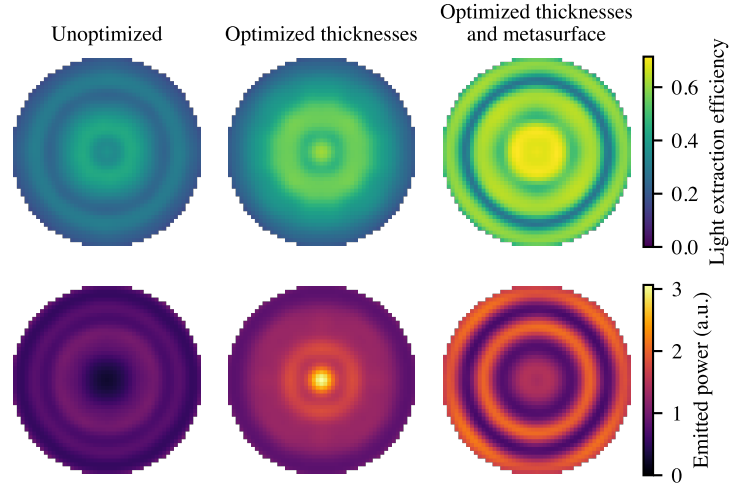


Fig. 7. Spatial maps of light extraction efficiency and emitted power for the initial, thickness-optimized, and metasurface-enhanced  $\mu$ LED. Values are for 630 nm wavelength, averaged over dipole orientations, and for  $N=2000$  with  $2\times 2$  BZ integration.

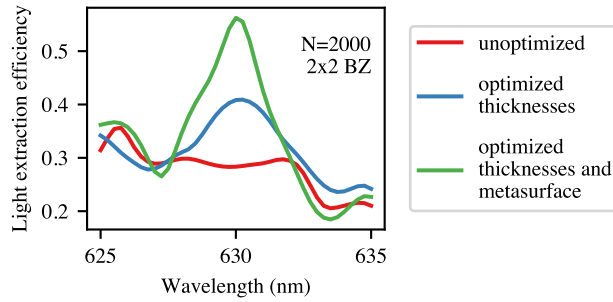


Fig. 8. Light extraction efficiency vs wavelength for the initial, unoptimized  $\mu$ LED; the  $\mu$ LED with optimized film thicknesses; and the  $\mu$ LED with optimized thicknesses and a metasurface.

These are shown in Fig. 7. As noted, our calculation considers dipoles at  $\sim 2000$  locations in a plane; each pixel in a map corresponds to a single physical dipole location, with value found by appropriately combining those for the three dipole orientations at the location. It is evident that the large increase in LEE for the optimized  $\mu$ LEDs is not simply due to a uniform increase in the LEE, but rather a nonuniform increase together with an enhancement in the dipole power in regions where the LEE is high. We can also observe sharp features in the LEE and emitted power. This suggests that simulations with few dipoles on a coarse grid may yield poor estimates of the total LEE.

Finally, Fig. 8 shows the wavelength-dependent LEE for the three cases, and illustrates the strong enhancement of LEE at the 630 nm target wavelength. Away from the target, both optimized  $\mu$ LEDs exhibit a significant drop in LEE, reaching values comparable to the unoptimized LED with less than 5 nm of wavelength shift. This suggests that a broadband objective will be required in order to obtain  $\mu$ LED designs which do not have strong wavelength dependence to LEE.

We emphasize that the  $\mu$ LEDs discussed above are merely examples of designs that can be

found through optimization, and we have seen many other designs with different metasurface densities and associated maps of LEE and emitted power. In practice, an objective should be crafted which is minimized for devices having the actual characteristics needed for  $\mu$ LED applications.

## 7. Conclusion

We have shown that the FMM can be used to model  $\mu$ LED structures with accuracy comparable to high-resolution FDTD, while achieving a speedup greater than  $10^7$  compared to CPU-based FDTD in situations where a large number of incoherent dipoles is considered. This result is enabled by the inherent characteristics of the FMM, vector FMM formulations that dramatically improve convergence compared to the basic FMM scheme, and BZ integration to model localized sources. These features are available in FMMAX, a *Jax*-based implementation of the FMM which we are open-sourcing alongside this work.

We used FMMAX to optimize several  $\mu$ LED designs, and showed that a metasurface-enhanced  $\mu$ LED can increase LEE by 98% over an unoptimized design, and by 38% over a  $\mu$ LED with optimized film thicknesses but no metasurface. This result indicates that metasurfaces are a promising technology to improve  $\mu$ LED performance as required for applications such as AR displays. Next steps could include consideration of  $\mu$ LED structures that are more realistic, use of optimization objectives that better target the actual requirements of  $\mu$ LEDs in AR applications (e.g. which maximize directional emission), the incorporation of manufacturability constraints in the inverse design scheme, and alternate initialization for  $\mu$ LED film thicknesses to obtain better local optima. It would also be interesting to study the new *Jones direct* FMM formulation across a wider of structures.

Finally, our method could be applied to a range of problems where spatially incoherent emission is considered, to problems where localized sources interact with a periodic structure, or to problems where FMM is more typically employed, such as modeling of gratings or photonic crystals.

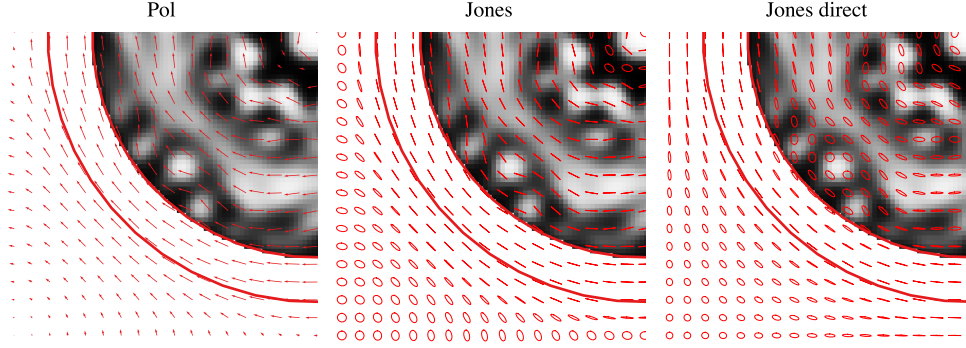


Fig. 9. Automatically-generated vector fields for the metasurface layer in Fig. 6. *Pol* yields a real-valued vector field tangent to material interfaces. *Jones* yields a complex Jones vector field that is linear and tangent at interfaces, becoming circular away from interfaces. *Jones direct* yields a field that is similar to *Jones*—but critically, does not become circular between parallel interfaces, and is therefore smoother. The depicted vectors are sampled from an underlying, higher-resolution vector field.

## Appendix A: Vector field examples

Vector fields obtained for the metasurface layer in Fig. 6 and the *Pol*, *Jones*, and *Jones direct* methods are shown in Fig. 9. The metasurface exhibits approximate four-fold rotational symmetry, and so we show one quadrant to make the details of vector fields more visible.

The *Pol* field is obtained by solving Eq. (2) starting with a real-valued  $\mathbf{t}^*$ ; the result is a real-valued vector field that at high-contrast interfaces is large and tangent to the interfaces. Zeros are located between parallel interfaces, e.g. at the top left, bottom left, and bottom right of the quadrant.

The *Jones* field is obtained by converting the *Pol* field into a complex Jones vector field as in [25]. Where the *Pol* field has magnitude 1, the *Jones* field corresponds to a linear polarization aligned with material interfaces. At zeros of the *Pol* field, the *Jones* field corresponds to a circular polarization.

The *Jones direct* field is obtained by solving Eq. (2) starting with a complex-valued  $\mathbf{t}^*$ . To obtain the complex-valued  $\mathbf{t}^*$ , we simply take the real-valued initial field and convert to a complex Jones vector field by the aforementioned procedure. The resulting Jones vector field again corresponds to a tangent linear polarization at high-contrast interfaces, and circular at certain locations (e.g. the bottom left of the quadrant), but which remains linear at some locations where the *Jones* method produces a circular polarization (e.g. top left or bottom right of the quadrant). Thus, the *Jones direct* field is (in some sense) smoother, and may be represented with fewer terms in a Fourier expansion—leading to improved convergence.

There are a few other differences between the *Jones* and *Jones direct* fields, e.g. locations where the *Jones* field is fully linear and *Jones direct* is merely elliptical. The field obtained by each of these methods can in principle be fine-tuned by adding hyperparameters to Eq. (1) which scale the smoothness and alignment terms. In future work, it would be interesting to study convergence in the context of this hyperparameter space and with a wider range of structures.

## Appendix B: Comparison of FMM formulations

FMMAX implements several FMM formulations. These exhibit virtually identical convergence for the basic  $\mu$ LED structure of Fig. 1. However, for more complex  $\mu$ LEDs that include metasurfaces

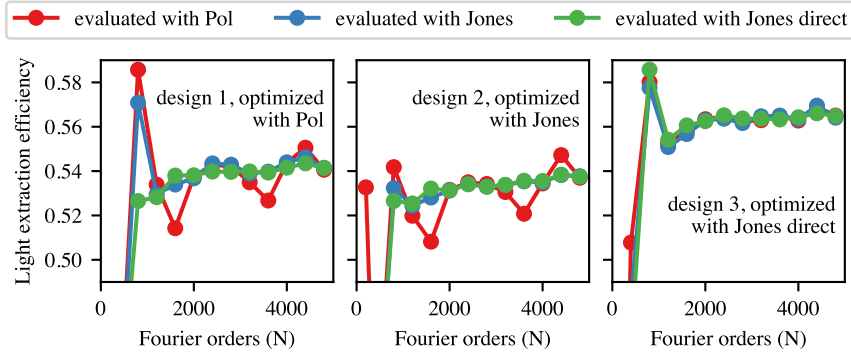


Fig. 10. Light extraction efficiency as a function of Fourier orders  $N$  for three separate metasurface-enhanced  $\mu$ LED designs. Each design was obtained by solving the inverse design problem of Sec. 3 using the *Pol*, *Jones*, and *Jones direct* formulations during the course of optimization. We then evaluated each design with various  $N$  and the three formulations. Overfitting to  $N = 800$  is seen in all cases, but the results converge as  $N$  increases. *Jones direct* exhibits the best convergence of the three formulations.

this may not be expected. Further, in an inverse design setting one may encounter “overfitting” where an optimized design performs best when evaluated with simulation settings (i.e. number of Fourier orders, FMM formulation, and BZ integration scheme) used during the optimization procedure. This makes it challenging to identify a preferred formulation.

Thus, to select the formulation best suited for  $\mu$ LED design, we solved the inverse design problem from Sec. 3 with the *Pol*, *Jones*, and *Jones direct* methods (with  $N = 800$ ), obtaining three separate designs. We then evaluated the total LEE of each design using the three formulations and Fourier orders up to  $N = 4800$ . The results are shown in Fig. 10.

The designs achieve similar performance and all show some degree of overfitting, where the simulation settings used for optimization yield a LEE higher than the converged result. However, increasing from  $N = 800$  to  $N = 2000$  (our validation setting, discussed in Sec. 5) is sufficient to obtain converged results. Of the three formulations, *Pol* converges least well, and exhibits some oscillation in the total LEE with an amplitude of  $\sim 5\%$ . *Jones* and *Jones direct* are quite similar, with *Jones direct* being slightly better on all three designs and avoiding the overfitting on the *Pol*-optimized design. The results support the selection of *Jones direct* and  $N = 2000$  for validation, although a more thorough study of convergence for a wider range of structures would be valuable.

**Disclosures.** The authors declare no conflicts of interest.

## References

1. J. Xiong, E.-L. Hsiang, Z. He, T. Zhan, and S.-T. Wu, “Augmented reality and virtual reality displays: emerging technologies and future perspectives,” *Light. Sci. & Appl.* **10**, 216 (2021).
2. J. Park, J. H. Choi, K. Kong, J. H. Han, J. H. Park, N. Kim, E. Lee, D. Kim, J. Kim, D. Chung, S. Jun, M. Kim, E. Yoon, J. Shin, and S. Hwang, “Electrically driven mid-submicrometre pixelation of InGaN micro-light-emitting diode displays for augmented-reality glasses,” *Nat. Photonics* **15**, 449–455 (2021).
3. S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, “Inverse design in nanophotonics,” *Nat. Photonics* **12**, 659–670 (2018).
4. A. M. Hammond, A. Oskooi, M. Chen, Z. Lin, S. G. Johnson, and S. E. Ralph, “High-performance hybrid time/frequency-domain topology optimization for large-scale photonics inverse design,” *Opt. Express* **30**, 4467–4491 (2022).
5. A. M. Hammond, A. Oskooi, S. G. Johnson, and S. E. Ralph, “Photonic topology optimization with semiconductor-foundry design-rule constraints,” *Opt. Express* **29**, 23916–23938 (2021).

6. P. Milonni, "Semiclassical and quantum-electrodynamical approaches in nonrelativistic radiation theory," *Phys. Reports* **25**, 1–81 (1976).
7. K.-J. Kim, "An analysis of self-amplified spontaneous emission," *Nucl. Instruments Methods Phys. Res. Sect. A: Accel. Spectrometers, Detect. Assoc. Equip.* **250**, 396–403 (1986).
8. A. G. Polimeridis, M. T. H. Reid, W. Jin, S. G. Johnson, J. K. White, and A. W. Rodriguez, "Fluctuating volume-current formulation of electromagnetic fluctuations in inhomogeneous media: Incandescence and luminescence in arbitrary geometries," *Phys. Rev. B* **92**, 134202 (2015).
9. V. P. Carey, G. Chen, C. Grigoropoulos, M. Kaviani, and A. Majumdar, "A review of heat transfer physics," *Nanoscale Microscale Thermophys. Eng.* **12**, 1–60 (2008).
10. S. Basu, Z. M. Zhang, and C. J. Fu, "Review of near-field thermal radiation and its application to energy conversion," *Int. J. Energy Res.* **33**, 1203–1232 (2009).
11. H.-Y. Ryu, J. Pyo, and H. Y. Ryu, "Light extraction efficiency of GaN-based micro-scale light-emitting diodes investigated using finite-difference time-domain simulation," *IEEE Photonics J.* **12**, 1–10 (2020).
12. H. Chung, "Computational upper-limit of directional light emission in nano-led via inverse design," *Opt. Express* **30**, 9008–9020 (2022).
13. H.-C. Lin, Z. Wang, and C. W. Hsu, "Fast multi-source nanophotonic simulations using augmented partial factorization," *Nat. Comput. Sci.* **2**, 815–822 (2022).
14. S. Li, H.-C. Lin, and C. W. Hsu, "Fast multi-channel inverse design through augmented partial factorization," (2023).
15. W. Yao, F. Verdugo, R. E. Christiansen, and S. G. Johnson, "Trace formulation for photonic inverse design with incoherent sources," *Struct. Multidiscip. Optim.* **65**, 336 (2022).
16. L. Li and C. W. Haggans, "Convergence of the coupled-wave method for metallic lamellar diffraction gratings," *J. Opt. Soc. Am. A* **10**, 1184–1189 (1993).
17. E. López-Fraguas, F. Binkowski, S. Burger, S. Hagedorn, B. García-Cámara, R. Vergaz, C. Becker, and P. Manley, "Tripling the light extraction efficiency of a deep ultraviolet led using a nanostructured p-contact," *Sci. Reports* **12**, 11480 (2022).
18. L. Zschiedrich, H. J. Greiner, S. Burger, and F. Schmidt, "Numerical analysis of nanostructures for enhanced light extraction from OLEDs," (2013). Presented at Light-Emitting Diodes: Materials, Devices, and Applications for Solid State Lighting XVII.
19. L. Zschiedrich, L. Blome, and H. J. Greiner, "Simulation of advanced OLED light extraction structures with novel FEM methods," (2014). Presented at Organic Photonics VI.
20. J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: Composable transformations of Python+NumPy programs," <http://github.com/google/jax> (2018).
21. C. Kim and B. Lee, "TORCWA: GPU-accelerated Fourier modal method and gradient-based optimization for metasurface design," *Comput. Phys. Commun.* **282**, 108552 (2023).
22. S. Colburn and A. Majumdar, "Inverse design and flexible parameterization of meta-optics using algorithmic differentiation," *Commun. Phys.* **4**, 65 (2021).
23. W. Jin, W. Li, M. Orenstein, and S. Fan, "Inverse design of lightweight broadband reflector for relativistic lightsail propulsion," *ACS Photonics* **7**, 2350–2355 (2020).
24. B. Vial, "nannos: Fourier modal method for multilayer metamaterials," <http://github.com/benvial/nannos> (2021).
25. V. Liu and S. Fan, "S4 : A free electromagnetic solver for layered periodic structures," *Comput. Phys. Commun.* **183**, 2233–2244 (2012).
26. D. M. Whittaker and I. S. Culshaw, "Scattering-matrix treatment of patterned multilayer photonic structures," *Phys. Rev. B* **60**, 2610 (1999).
27. L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A* **13**, 1870–1876 (1996).
28. P. Götz, T. Schuster, K. Frenner, S. Rafler, and W. Osten, "Normal vector method for the RCWA with automated vector field generation," *Opt. Express* **16**, 17295–17301 (2008).
29. R. Antos, "Fourier factorization with complex polarization bases in modeling optics of discontinuous bi-periodic structures," *Opt. Express* **17**, 7269–7274 (2009).
30. P. Lalanne and E. Silberstein, "Fourier-modal methods applied to waveguide computational problems," *Opt. Lett.* **25**, 1092–1094 (2000).
31. E. Silberstein, P. Lalanne, J.-P. Hugonin, and Q. Cao, "Use of grating theories in integrated optics," *J. Opt. Soc. Am. A* **18**, 2865–2875 (2001).
32. A. Moritz and A. Hangleiter, "Optical gain in ordered GaInP/AlGaInP quantum wells," *Appl. Phys. Lett.* **66**, 3340–3342 (1995).
33. O. Sigmund and K. Maute, "Topology optimization approaches," *Struct. Multidiscip. Optim.* **48**, 1031–1055 (2013).
34. R. E. Christiansen, J. Vester-Petersen, S. P. Madsen, and O. Sigmund, "A non-linear material interpolation for design of metallic nano-particles using topology optimization," *Comput. Methods Appl. Mech. Eng.* **343**, 23–39 (2019).
35. K. Svanberg and H. Svärd, "Density filters for topology optimization based on the pythagorean means," *Struct. Multidiscip. Optim.* **48**, 859–875 (2013).
36. R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM*

- J. on Sci. Comput. **16**, 1190–1208 (1995).
37. 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).
  38. S. G. Johnson, “The NLOpt nonlinear-optimization package,” <http://github.com/stevengj/nlopt>.
  39. A. Taflov, S. C. Hagness, and M. Picket-May, “Computational electromagnetics: the finite-difference time-domain method,” *The Electr. Eng. Handb.* **3**, 629–670 (2005).
  40. A. Deinega and I. Valuev, “Subpixel smoothing for conductive and dispersive media in the finite-difference time-domain method,” *Opt. letters* **32**, 3429–3431 (2007).
  41. A. Oskooi, C. Hogan, A. M. Hammond, M. Reid, and S. G. Johnson, “Factorized machine learning for performance modeling of massively parallel heterogeneous physical simulations,” *arXiv preprint arXiv:2003.04287* (2020).