



Topology optimization of on-chip integrated laser-driven particle accelerator

Yang-Fan He^{1,2} · Bin Sun^{1,2,3} · Ming-Jiang Ma⁴ · Wei Li^{1,2} · Qiang-You He^{1,2} · Zhi-Hao Cui^{1,2} · Shao-Yi Wang^{1,2} · Zong-Qing Zhao¹

Received: 8 April 2022 / Revised: 26 July 2022 / Accepted: 31 July 2022 / Published online: 22 September 2022

© The Author(s), under exclusive licence to China Science Publishing & Media Ltd. (Science Press), Shanghai Institute of Applied Physics, the Chinese Academy of Sciences, Chinese Nuclear Society 2022

Abstract Particle accelerators are indispensable tools in both science and industry. However, the size and cost of conventional RF accelerators limits the utility and scope of this technology. Recent research has shown that a dielectric laser accelerator (DLA) made of dielectric structures and driven at optical frequencies can generate particle beams with energies ranging from MeV to GeV at the tabletop level. To design DLA structures with a high acceleration gradient, we demonstrate topology optimization, which is a method used to optimize the material distribution in a specific area based on given load conditions, constraints, and performance indicators. To demonstrate the effectiveness of this approach, we propose two schemes and design several acceleration structures based on them. The optimization results demonstrate that the proposed method can

be applied to structure optimization for on-chip integrated laser accelerators, producing manufacturable structures with significantly improved performance compared with previous size or shape optimization methods. These results provide new physical approaches to explore ultrafast dynamics in matter, with important implications for future laser particle accelerators based on photonic chips.

Keywords Laser-driven particle acceleration · Dielectric grating accelerator · Inverse Smith–Purcell effect · Topology optimization

1 Introduction

A particle accelerator is a device that accelerates charged particles (such as protons and electrons). The input energy that induces acceleration is achieved by applying an electric field. The field provides kinetic energy to the particles such that they can reach speeds at a significant fraction of the speed of light. Originally invented for scientific research, particle accelerators now play an important role in improving health and prosperity. They are relevant to our daily lives in a variety of applications, including cancer treatment, material analysis, and removal of harmful microorganisms from food and water [1–5]. Radio frequency (RF)-powered devices are the traditional choice for acceleration elements [5–8]. However, their large size, high input power requirements, and expensive infrastructure limit their utility and accessibility to the wider scientific community. This has promoted the exploration of more compact and economical alternative technologies. In recent years, dielectric laser accelerators have attracted significant interest as a method for accelerating

This work was supported by the National Natural Science Foundation of China (Nos. 12004353, 11975214, 11991071, 11905202, 12174350), Key Laboratory Foundation of The Sciences and Technology on Plasma Physics Laboratory (No. 6142A04200103), and Independent scientific research (No. JCKYS2021212011).

✉ Yang-Fan He
y-f-he@foxmail.com

Zong-Qing Zhao
zhaozongqing99@caep.ac.cn

¹ Laser Fusion Research Center, CAEP,
P. O. Box 919 986, Mianyang 621900, China

² The Sciences and Technology on Plasma Physics Laboratory,
CAEP, Mianyang 621900, China

³ Department of Plasma Physics and Fusion Engineering, CAS
Key Laboratory of Geospace Environment, University of
Science and Technology of China, Hefei 230026, China

⁴ The Institute of Optics, University of Rochester, Rochester,
NY 14627, USA

charged particles driven by solid-state lasers [9–17]. The use of infrared lasers to power particle accelerators fabricated using optical-scale lithography is a developing area of research, in which evanescent acceleration modes within dielectric structures can synchronize charged particles. This enables a more compact accelerator design and a much higher acceleration gradient compared to RF accelerators, which provides the opportunity to achieve higher acceleration gradients and reduce cost and size simultaneously (see [18] and the references therein).

A dielectric laser accelerator (DLA) aims to reduce the dimensions associated with RF accelerators by using a near-infrared (NIR) ultrafast femtosecond laser to drive dielectric structures and to provide a higher acceleration gradient. Additionally, dielectric materials such as silicon and silicon dioxide can survive at energy levels one to two orders of magnitude higher than the metals used in RF accelerators [19, 20]. Theoretically, the DLA can provide acceleration gradients in excess of GV/m [21–24] compared to the 30–100 MV/m achieved by RF accelerators, which are limited because of the need to avoid material damage. In various experiments, an acceleration gradient of nearly 1 GV/m has been achieved (see [25] and the references therein). Compared with other advanced acceleration techniques, DLA currently provides the highest gradient among non-plasma accelerators. Because dielectric laser accelerators consist of dielectric structure devices and are driven at optical frequencies, cascade acceleration can be easily integrated on-chip [26, 27]. To make DLA competitive with larger devices, the Accelerator on a Chip International Program (ACHIP) [28], which is working toward building miniature particle accelerators on a chip using advanced laser and nanofabrication technology, has divided research topics in this field into several branches for detailed study; however, as stated in [29], the nanostructures of DLA are particularly important, and different nanostructures can be expected to have very different effects on particles driven by the same field distribution.

Most optimization studies and experiments in the field of DLA are based on grating structures [29]. Additionally, related theories have been gradually developed to explain the acceleration phenomenon. However, these relevant theories either do not consider how the structural parameters affect the particle dynamics or are limited to the grating structure and how its local parameters affect the particle dynamics [29]. In general, past research has focused on shape and size optimization, with examples shown in Fig. 1a and b. Obviously, the results obtained in this manner are limited by the initial architecture, and the results obtained are a subset of various possibilities depending on the structure used. A pioneering study on topology optimization (Fig. 1c) of DLA structures was performed using adjoint methods by Hughes et al. [30].

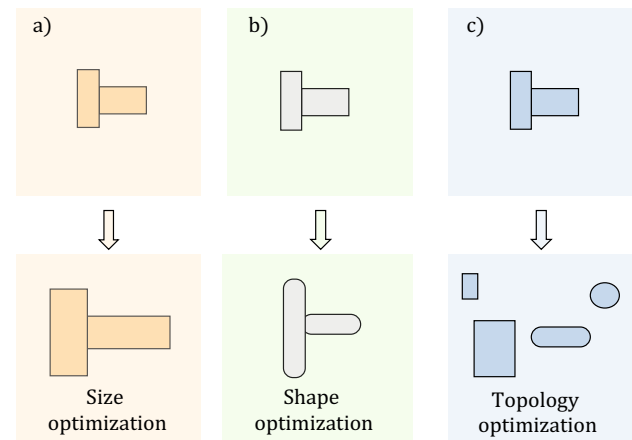


Fig. 1 (Color online) Three means of optimization: **a** size optimization, which takes the structural shape or the shape of the hole as the optimization object to adjust the size; **b** shape optimization, looking for new concave-convex distributions on the existing structure; **c** topology optimization, which can obtain the optimal distribution in the material design space

When using the adjoint method for topology optimization, the adjoint source must be given [31, 32]. In principle, the adjoint source is the first derivative of the objective function with respect to the design parameters. This means that there are situations in which this derivative cannot be solved for, so it is not possible to provide an accurate accompanying source for simulation purposes. In other words, it may be difficult to use adjoint methods to implement the inverse design of arbitrary objective functions. Additionally, to produce structures that can be practically fabricated, a mathematical method that can generate additional binary structure distributions is needed. Therefore, an effective method for designing an acceleration structure that can be manufactured is highly desirable.

This study was inspired by the utilization of a self-guided derivation technique used in artificial intelligence to compute gradients for an arbitrary objective function, and applies a hyperparameter control function distribution to guide the generation of a more binary structure distribution to achieve more general topology optimization. In this paper, we outline in detail how topology optimization can be used to achieve the inverse design of DLA structures. The remainder of this paper is organized as follows. In Sect. 2, we first outline methods for simulating the physical phenomena associated with DLA, and then demonstrate the optimization principle for realizing a DLA structure using automatic derivation techniques. Finally, we present a control method that can realize a manufacturable structure. In Sect. 3, the validity of the proposed method is verified first, two optimization schemes that can be used in DLA are discussed, and design demonstrations of these two optimization schemes are provided.

2 Theoretical descriptions

All macroscopic electromagnetism, including the propagation of light in a photonic crystal, is governed by Maxwell's four macroscopic equations. In SI units, we seek vector fields $\mathbf{D}, \mathbf{E}, \mathbf{B}, \mathbf{H} : \Omega \rightarrow \mathbb{R}^3$ such that the Maxwell equations

$$\begin{aligned}\nabla \cdot \mathbf{B} &= 0, \\ \nabla \cdot \mathbf{D} &= \rho,\end{aligned}$$

hold for a given $\mathbf{J} : \Omega \rightarrow \mathbb{R}^3$ and $\rho : \Omega \rightarrow \mathbb{R}$. We call $\mathbf{E}$ the electric field, $\mathbf{D}$ electric displacement field (or electric flux density), $\mathbf{H}$ magnetic field intensity, $\mathbf{B}$ magnetic flux density, and $\mathbf{J}$ and ρ electric current intensity and electric charge density, respectively. Although the results obtained by analytically solving Maxwell's equations have theoretical significance, when combined with the research object of this study, it is very difficult to obtain their analytical solutions. Therefore, only numerical solutions of Maxwell's equations are considered in this study, which can be divided into two categories: time domain (TD) methods and frequency domain (FD) methods. The most famous time domain method is the finite difference time domain (FDTD) [33], which is usually suitable for solving for the transient change process under external excitation. The frequency domain method is based on time harmonic differential and integral equations to study the steady-state field distribution after infinite time under a time-harmonic excitation. Considering the small memory and low time consumption requirements, this study uses the finite-difference frequency-domain (FDFD) approach to solve for the field distribution of the structure at a given frequency (ω) for a DLA coupled with non-magnetic materials under normal conditions. For mathematical convenience, we employ the standard trick of using a complex-valued field and taking the real part to obtain the physical fields. This allows us to write a harmonic mode as a spatial pattern (or "mode profile") multiplied by a complex exponential as follows.

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{-i\omega t} \quad \text{and} \quad \mathbf{H}(\mathbf{r}, t) = \mathbf{H}(\mathbf{r})e^{-i\omega t}$$

We assume that all fields have the $e^{-i\omega t}$ dependence, with ω being the angular frequency. Through the curl operation, Maxwell's equations can be expressed as

$$\frac{1}{\mu} \nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \omega^2 \epsilon \mathbf{E}(\mathbf{r}) = i\omega \mathbf{J}(\mathbf{r}), \quad (1)$$

where $\mathbf{E}(\mathbf{r})$ is a complex vector representing the (unknown) electric field distribution in the domain. Generally, ignoring higher-order relationships, a dielectric medium with permeability $\mu = \mu_0 \mu_r$ and permittivity $\epsilon = \epsilon_0 \epsilon_r$ is considered, where μ_0 is the vacuum permeability, and ϵ_0 is the

vacuum permittivity. The working principle of FDFD [34, 35] is to represent the electric field, dielectric constant distribution, and source on the Yee grid [33], and then represent the $\nabla \times \nabla \times$ operator as a finite-difference differential matrix $\tilde{\mathbf{A}}$. In the discrete case, Eq. (1) can be written in matrix form as

$$[\tilde{\mathbf{A}} - \omega^2 \epsilon_0 \tilde{\epsilon}_r] \mathbf{E}(\mathbf{r}) = \tilde{\mathbf{S}}, \quad (2)$$

where $\tilde{\mathbf{A}}$ is a matrix that performs the finite difference derivative $\frac{1}{\mu} \nabla \times \nabla \times$, and does not depend on our permittivity distribution. $\tilde{\epsilon}_r$ is a diagonal complex-valued matrix and each element along the diagonal corresponds to the relative permittivity of a cell in our domain. $\tilde{\mathbf{S}}$ is a complex vector that represents the distribution of the driving current source. In practice, when applying the FDFD method, we first construct the matrix $\tilde{\Phi} = \frac{1}{\mu} \nabla \times \nabla \times - \omega^2 \epsilon_0 \tilde{\epsilon}_r$ and the source vector $\tilde{\mathbf{S}}$ and then

solve for the electric field $\mathbf{E} = \tilde{\Phi}^{-1} \tilde{\mathbf{S}}$ using an iterative or direct method. Because the fundamental features of DLA appear in the simplified 2D analysis, the full-field variation with z is not considered for the time being. This means that

$$\partial \mathbf{E} / \partial z = 0 \quad \text{and} \quad \partial \mathbf{H} / \partial z = 0. \quad (3)$$

In 2D problems, the electric and magnetic fields can be separated into two orthogonal modes: **TM** and **TE**, where the **TM** mode can be fully described by H_z and the **TE** mode by E_z , as shown in Fig. 2. It is worth mentioning that the simulation area of the DLA cannot be infinite, that is, the simulation boundary must be truncated, and the commonly used condition is the perfect matching layer and periodic boundary (see [36] and the references therein). The process will be discussed in more detail in Sect. 3 for specific examples.

To implement topology optimization, the DLA structure is divided into three parts: laser injection region (LIR), optimization design region (ODR), and particle injection channel region (PICR), as shown in Fig. 3a. To obtain the device structure under the optimal objective function $\xi(x_i)$ value, a straightforward idea is to mesh the design area such that the permittivity value in each grid is selected at random as 0 or 1, where 0 represents vacuum and 1 represents the permittivity constant of the material, and then change the permittivity value state of one grid, run the simulation with the permittivity value of the other grid unchanged, record the objective function $\xi(x)$ value and repeat the calculations until all the combined results are traversed, and finally select an optimal structure from all simulation results, as shown in Fig. 3b.

However, this enumeration method will become difficult to implement in practice because even when enumerating only 10 grids in the optimization design region, it must be

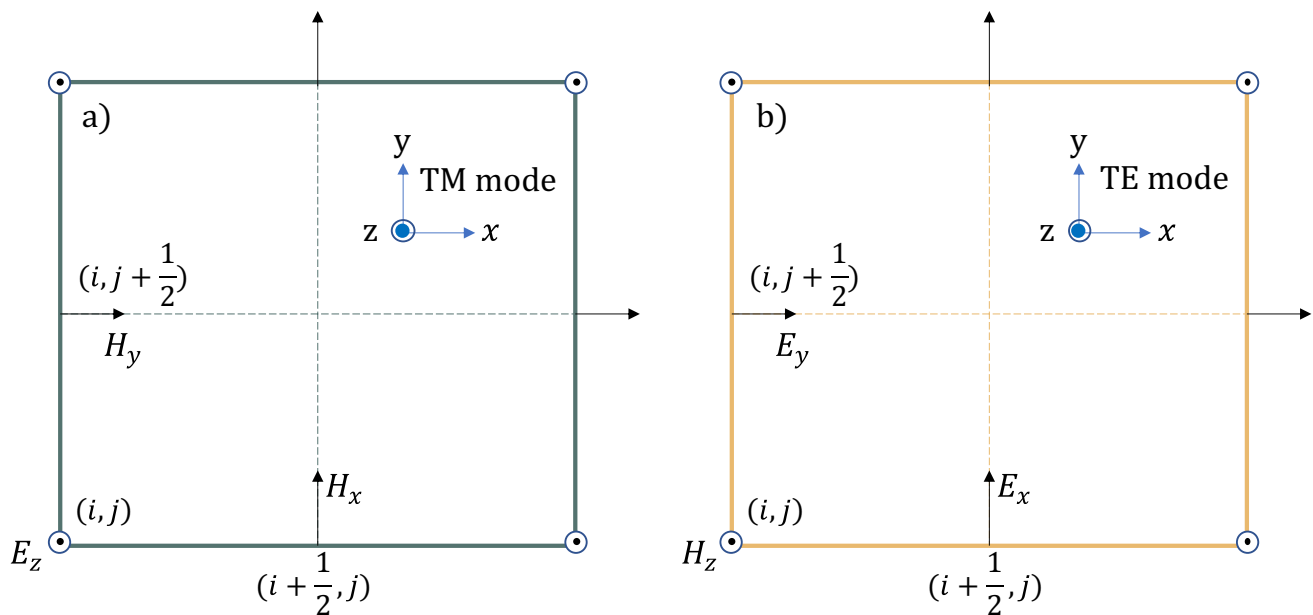


Fig. 2 (Color online) Fundamental features of DLA used in the simplified 2D analysis, in which the full-field variation with z is not considered. In 2D problems, the electric and magnetic fields can be separated into orthogonal **TM** and **TE** modes, where the **TM** mode

can be fully described by E_z , and the **TE** mode can be fully described by H_z . The field distribution of the **TM** (**TE**) mode is discretized onto a square Yee grid, as shown in **a** and **b**

run 2^{10} times. This also means that with further refinement of the grid in the optimization design region, the number of calculations will increase exponentially. In other words, the idea is straightforward, but not practical. Therefore, we use the gradient of the objective function to update and optimize the parameters to minimize the objective function $\xi(x_i)$. To make the computation of the derivative of the objective function $\xi(x_i)$ with respect to the design variables more flexible and extensive, we employ an automatic derivation technique. Automatic derivation is a method that lies between symbolic derivation and numerical derivation (it is a numerical calculation method that calculates an approximation of the derivative). In general, automatic derivation is the application of a symbolic derivation to the most basic operations, such as constant, power function, exponential function, logarithmic functions, trigonometric function, and other basic function. The value of the independent variable is substituted to obtain its derivative value, which is retained as an intermediate result. Then, the derivative value of the entire function is calculated according to the derivation results of these basic operation units. The difference between automatic derivation and symbolic derivation is that instead of computing the analytical solution, the chain rule is used to decompose the complex function into individual units, differentiate these

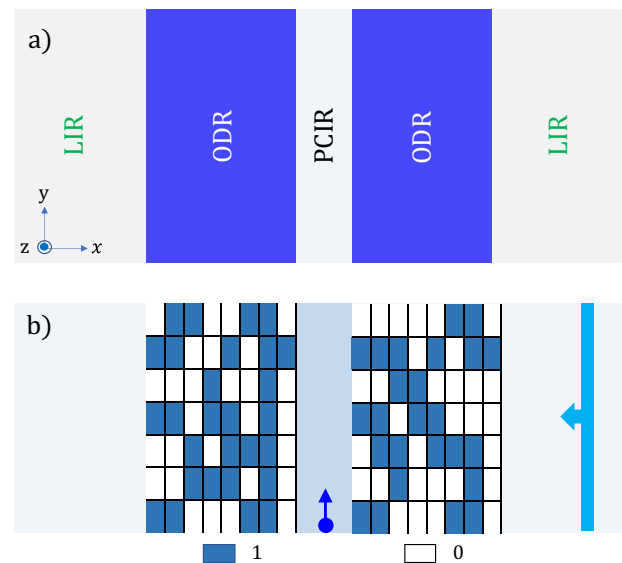


Fig. 3 (Color online) In our topology optimization method, the DLA structure is divided into three parts: laser injection region (LIR), optimization design region (ODR) and particle injection channel region (PICR), as shown in **a**. The permittivity function of the design area is randomly distributed as 0 (white block) or 1 (dark block), where 0 represents vacuum, and 1 represents the permittivity function of the material, as shown in **b**

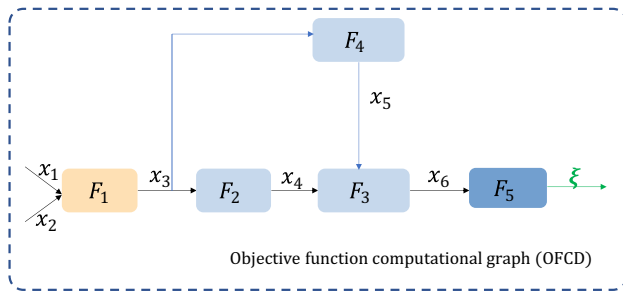


Fig. 4 (Color online) An objective function is divided into a calculation graph consisting of multiple calculation units obtained according to the calculation order. This directed acyclic graph charts the flow of data

small units, and finally combine them to obtain the solution to the full derivative.

The objective function $\xi(x_i)$ is divided into multiple computing units obtained according to the order of computation, which we call a computation graph, which is a directed acyclic graph that charts the flow of data, as shown in Fig. 4. The purpose is to find the derivative $\nabla \xi(x_i) = [\partial \xi / \partial x_1, \partial \xi / \partial x_2]$, so we apply the chain rule to the first term, $\partial \xi / \partial x_1$, which yields

$$\frac{\partial \xi}{\partial x_1} = \frac{\partial \xi}{\partial x_6} \left(\frac{\partial x_6}{\partial x_5} \left(\frac{\partial x_5}{\partial x_3} \frac{\partial x_3}{\partial x_1} \right) + \frac{\partial x_6}{\partial x_4} \left(\frac{\partial x_4}{\partial x_3} \frac{\partial x_3}{\partial x_1} \right) \right), \quad (4)$$

or

$$\frac{\partial \xi}{\partial x_1} = \left(\left(\frac{\partial \xi}{\partial x_6} \frac{\partial x_6}{\partial x_5} \right) \frac{\partial x_5}{\partial x_3} + \left(\frac{\partial \xi}{\partial x_6} \frac{\partial x_6}{\partial x_4} \right) \frac{\partial x_4}{\partial x_3} \right) \frac{\partial x_3}{\partial x_1}. \quad (5)$$

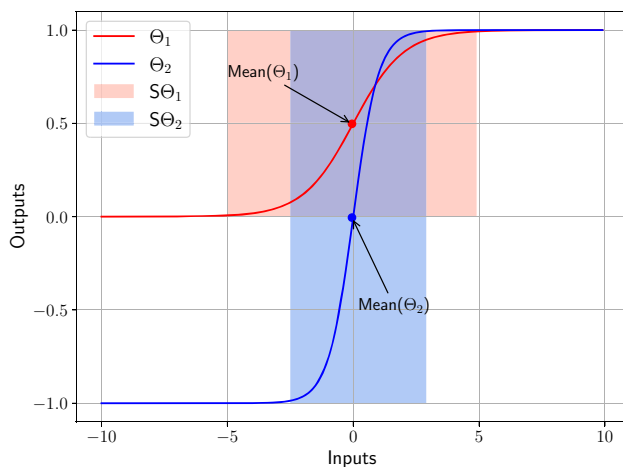


Fig. 5 (Color online) We show two specific modification methods using Θ_1 and Θ_2 , where Θ_1 is a “Sigmoid function,” and Θ_2 is a “Tanh function.” S_{Θ_1} and S_{Θ_2} are the regions corresponding to the non-extreme points of the function, and the size of this region can be controlled by introducing one or more hyperparameters. That is, the function can be made more binary by introducing certain hyperparameters a

We require only one forward sweep to obtain the result. The calculation graph from left to right corresponds to Eq. (4) from inside to outside. If we use the chain rule in another way, such as using Eq. (5), the method needs to first calculate the rightmost derivative and then calculate the derivative to its left in turn. Therefore, not only a forward propagation, but a reverse sweep is required. In automatic derivation, the first method is called the forward mode and the second is called the reverse mode. An automatic differentiation library allows the user to compute the exact Jacobian dx_i/dx_j for any i, j using the rules of differentiation and some knowledge of the partial derivatives of each operation (see [37, 38] and the references therein). When the objective function is convex, the solution of the gradient descent method is globally optimal. However, in general, the objective function is often not a convex function; therefore, it cannot be guaranteed that its solution is the global optimal solution. To prevent the solution from falling into a local optimum, adaptive moment estimation [39] is used in our study.

To obtain a device that can actually be manufactured or meaningful, we need to modify the equipment parameterization to encourage the use of more binary features in the optimization, because if the optimized equipment is simply represented as a set of data, the eigenvalues during optimization will change constantly, and the final output may not produce a device that can be manufactured. In general, we can modify the parameters into the following form:

$$f(\mu_r) = \Theta H(\mu_1, \mu_2), \quad (6)$$

$$\begin{cases} H(\mu_1, \mu_2) = \mu_1 \Theta^{-1} + (\mu_2 - \mu_1) \Theta, \Theta = \Theta_1 = \frac{1}{1 + e^{-ax}}, \\ H(\mu_1, \mu_2) = \left(\frac{\mu_1 + \mu_2}{2} \right) \Theta^{-1} + \left(\frac{\mu_2 - \mu_1}{2} \right) \Theta, \Theta = \Theta_2 = \frac{1 - e^{-ax}}{1 + e^{-ax}}. \end{cases} \quad (7)$$

where Θ is a common s-shaped function, as shown in Fig. 4. We demonstrate two specific modification methods involving Θ_1 and Θ_2 , respectively, where Θ_1 is a “Sigmoid-like function”, and Θ_2 is a “Tanh-like function”, and a represents the hyperparameters. S_{Θ_1} and S_{Θ_2} are the regions corresponding to the non-extremum points of the function, and the size of either region can be controlled by introducing hyperparameters, that is, the function can be made more binary by introducing certain hyperparameters a . Alternatively, Heaviside’s step function (κ , Eq. (8)) with projection significance can be used.

$$\kappa = \frac{\tanh(\tau\sigma) + \tanh(\tau(\mu(r) - \sigma))}{\tanh(\tau\sigma) + \tanh(\tau(1 - \sigma))} \quad (8)$$

Here, σ is the projection threshold (when the value is less than σ , and the Heaviside function (κ) projects the value of

the unit in the 0 direction. In contrast, when the density of the unit is greater than σ , the Heaviside function projects the value of the unit in the direction of 1), where τ is a set of hyperparameters that controls the steepness around the threshold parameter σ . In the actual process, convolution filtering is required to smooth the edges of the design structure [32].

3 Results and discussion

The two important parameters (ϵ and μ in Eq. (1)) that affect the electromagnetic distribution can be written in the following form:

$$\epsilon_r = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{pmatrix}, \quad \mu_r = \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \\ \mu_{31} & \mu_{32} & \mu_{33} \end{pmatrix}. \quad (9)$$

Now, consider a material that consists of infinite columns in the x_3 -direction modeled by a permittivity such that $(x_1, x_2) \mapsto \epsilon(x_1, x_2)$. We assume an isotropic linear dielectric medium with permeability $\mu = \mu_0 \mu_r$ and permittivity $\epsilon = \epsilon_0 \epsilon_r$. However, only isotropic non-magnetic materials are considered in this study, which are the preferred materials for chip accelerators. Their properties can be described by

$$\epsilon_{ij} = \begin{cases} \epsilon_r, & i = j \\ 0, & i \neq j \end{cases} \quad \text{and} \quad \mu_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \quad (10)$$

The coordinate system used for DLA optimization is shown in the figure; therefore, the TE mode should be used to realize particle acceleration. At this time, the electromagnetic distribution of the system can be calculated using Eq. (11).

$$\begin{aligned} (\nabla_x^E \epsilon_{22}^{-1} \nabla_x^H + \nabla_y^E \epsilon_{11}^{-1} \nabla_y^H + \omega^2/c^2) H_z &= i\omega J(z) \\ \nabla_y^H H_z &= i\omega \epsilon_0 \epsilon_{11} E_x, \quad \nabla_x^H H_z = i\omega \epsilon_0 \epsilon_{22} E_y \end{aligned} \quad (11)$$

The terms ∇_x^E and ∇_y^E are banded matrices that calculate first order spatial derivatives of the electric fields across the grid. As a quick example of these matrices, they were computed for a small two-dimensional grid composed of 3×3 cells. Using periodic boundary conditions, the matrix derivative operators ∇_x^E and ∇_y^E for this simple case are

$$\nabla_x^E = \frac{1}{\Delta y} \begin{bmatrix} -1 & 1 & & & & & & & \\ & -1 & 1 & & & & & & \\ U_x & & -1 & 0 & & & & & \\ & & -1 & 1 & & & & & \\ & & & -1 & -1 & 1 & & & \\ & & & & U_x & & -1 & 0 & \\ & & & & & & -1 & 1 & \\ & & & & & & & -1 & 1 \\ & & & & & & & & U_x & -1 \end{bmatrix}$$

$$\nabla_y^E = \frac{1}{\Delta x} \begin{bmatrix} -1 & & & & 1 & & & & \\ & -1 & & & & 1 & & & \\ & & -1 & & & & 1 & & \\ & & & -1 & & & & 1 & \\ & & & & -1 & & & & 1 \\ & & & & & -1 & & & \\ V_y & & & & & & -1 & & \\ & V_y & & & & & & -1 & \\ & & V_y & & & & & & -1 \end{bmatrix}$$

where $\Delta x(y)$ denotes the FDFD Yee grid size along the $x(y)$ direction, $U_x = \exp(ik_x^{\text{bloch}})$, $V_y = \exp(ik_y^{\text{bloch}})$, and $k_{x(y)}^{\text{bloch}}$ denotes the phase across the periodic boundary in the $x(y)$ direction. Similarly, the terms ∇_x^H and ∇_y^H calculate first order spatial derivatives of the magnetic fields across the grid. The derivative operators for the electric and magnetic fields are different because of the staggered nature of the Yee grid (Fig. 2). However, they are related through

$$\nabla_x^H = -(\nabla_x^E)^{\text{HT}}, \quad \nabla_y^H = -(\nabla_y^E)^{\text{HT}}.$$

The “HT” superscript indicates a Hermitian transpose operation. For the TE mode, averaging is needed for ϵ_{11} and ϵ_{22} . All involved matrices are sparse; therefore, we can apply sparse matrix techniques to save computation time and memory.

The acceleration gradient (or the particle energy gain per unit length) is an important figure of merit. For a laser operating in normal incidence mode (laser propagating in the y direction), phase velocity matching between the particle and the electromagnetic fields is established by introducing a spatial periodicity in our structure having a period of $\beta\lambda$ along the x direction, where λ is the laser wavelength. The acceleration gradient function $\Xi_{\text{gain}}(E)$ can be calculated using a periodic structure as follows.

$$\Xi_{\text{gain}}(E_y) = \frac{1}{\lambda} \int_{\lambda/2}^{-\lambda/2} E_y \left(x, y; t = \frac{y}{\beta c} \right) dy$$

Here, we focus only on the relationship between the permittivity distribution (ϵ_r) and the energy gain ($\Xi_{\text{gain}}(\mathbf{E})$). In other words, this is exactly the same case for nonmagnetic materials mentioned earlier. As described in Sect. 2, we must compute $\Xi_{\text{gain}}(\mathbf{E})$ for ϵ_r as a derivative of

$$\frac{\partial \Xi_{\text{gain}}(\mathbf{E})}{\partial \epsilon_r} = \text{Real} \left\{ 2 \frac{\partial \Xi_{\text{gain}}(\mathbf{E})}{\partial \mathbf{E}} \frac{\partial \mathbf{E}}{\partial \epsilon_r} \right\}, \quad (12)$$

where $\Xi_{\text{gain}}(\mathbf{E})$ is equivalent to the objective function ($\xi(x_i)$) mentioned previously. $\partial \Xi_{\text{gain}}(\mathbf{E}) / \partial \epsilon_r$ is an important parameter that indicates how to change ϵ_r to increase $\Xi_{\text{gain}}(\mathbf{E})$. Two optimization schemes for maximizing the acceleration gradient function $\Xi_{\text{gain}}(E_y)$ are presented below.

Case 1: Considering the energy conversion rate, the acceleration gradient should be higher than the peak value of the incident electric field as much as possible. This means that the “conversion factor” (Γ) given by dividing the acceleration gradient by the peak amplitude of the incident electric field needs to be optimized. This quantity will reveal the maximum conversion rate that the structure can achieve. Using mathematical language, we express this principle as

$$\text{maximum } \Gamma = \frac{\Xi_{\text{gain}}(\mathbf{E})}{E_0},$$

where E_0 is the initial injection electric field, which should be less than the material damage threshold ($E_{\text{material damage}}$).

Case 2: Specifically considering the protection or reuse of the structure, the optimal energy gain should be selected so that the acceleration gradient is as high as possible without causing damage to the structure. This implies that the value of the maximum electric field in the structure must be determined so that it cannot exceed the local peak laser fluence at the damage threshold. Therefore, another relevant quantity to maximize is the “acceleration factor” (χ) given by the acceleration gradient divided by the maximum electric field magnitude in the system. This amount ultimately limits the amount of acceleration gradient that can be achieved so that the damage threshold is not exceeded. Using mathematical language, we express this principle as

$$\begin{aligned} \text{maximum } \chi &= \frac{\Xi_{\text{gain}}(\mathbf{E})}{E_{\text{max}}}, \\ \text{subject to } E_{\text{max}} &< E_{\text{material damage}}. \end{aligned} \quad (13)$$

Here, E_{max} is the maximum value of the electric field magnitude vector in the optimization design region (ODR) and $E_{\text{material damage}}$ is the material damage threshold expressed in terms of the electric field. The damage threshold of the DLA structure can be improved by the choice of material.

The results of “Case 1” are shown in Figs. 6 and 7. The parameter settings in Fig 6 are the same as those in [30], in which the purpose is to test the solver and optimization scheme involved in this study. To accomplish this, a plane wave (E_0 is the initial injected electric field) is introduced into the DLA chip from the left side using the periodic condition. The laser wavelength (λ) used in the simulations is 2000 nm. The normalized velocity (β) of the injected electrons is 0.5, and the square grid size ($\Delta x(y)$) used in the simulation is 20 nm. The material used in the optimization process is fused silica (SiO_2), the index is $n = 1.45$, that is, $\epsilon_{11} = \epsilon_{22} = 2.1$. The cross-sectional area utilized for the optimization design region is $5\beta\lambda \times \lambda$, and the particle injection channel region is $5\beta\lambda \times \frac{1}{5}\lambda$. An acceleration gradient function ($\Xi_{\text{gain}}(\mathbf{E})$) of $0.32E_0$ ($\Gamma = 0.32$) can be achieved by optimizing the maximum conversion factor (Γ). The results of the gradient calculation and structure optimization are consistent with those in Ref.

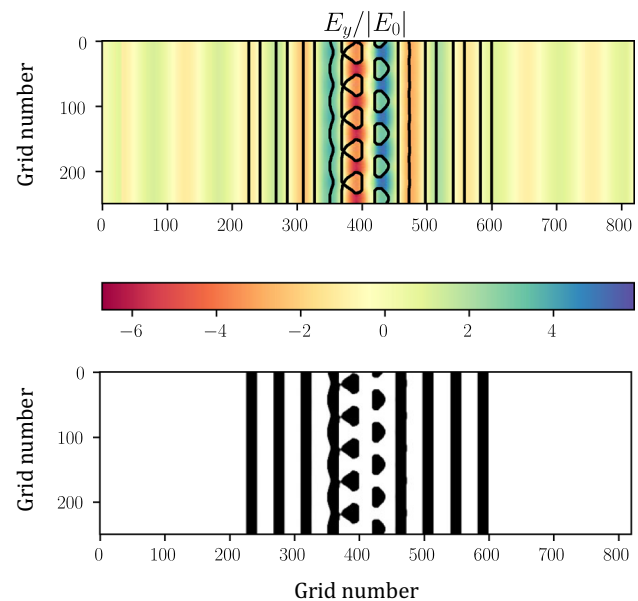


Fig. 6 (Color online) The parameter settings are the same as those used in Ref. [30]. The material used in the optimization process is fused silica (SiO_2). A plane wave (E_0 is the initial injected electric field) is introduced into the chip from the left side using the periodic condition. The laser wavelength (λ) is 2000 nm, The normalized velocity (β) of the injected electrons is 0.5, $\epsilon_{11} = \epsilon_{22} = 2.1$, the square grid size ($\Delta x(y)$) used in the simulation is 20 nm, and the achieved gradient is $0.32E_0$ ($\Gamma = 0.32$)

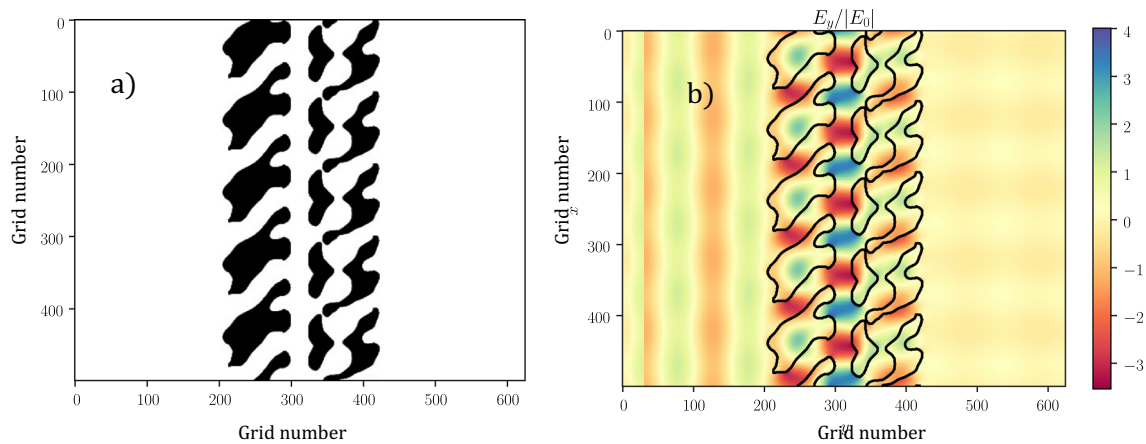


Fig. 7 (Color online). Topology optimization results of DLA structure using “Case 1” scheme. The parameters used are the same as those used in [29]. A plane wave (E_0 is the initial injected electric field) is introduced into the chip from the left side using the periodic

condition. The laser wavelength (λ) is 800 nm, the normalized velocity (β) of the injected electrons is 0.99, and the square grid size ($\Delta x(y)$) used in the simulation is 8 nm [30] (the achieved gradient is $0.31E_0$). As calculated based on [30], Fig. 6 corresponds to an acceleration gradient of 1.37 GV/m. It can be observed from Fig. 6 that the structural design given by the optimization results tends to include mirrors on both sides of the structure. By observing the normalized electric field diagram in the y direction ($E_y/|E_0|$), it can be observed that for higher acceleration gradients, it is useful to use dielectric mirror resonance to enhance the field in the gap.

The parameters in Fig. 7 are the same as those used in Ref. [29], and the purpose is to highlight the advantage of using the topology optimization method. A plane wave (E_0 is the initial injected electric field) is introduced into the DLA chip from the left side using the periodic condition. The material used in the optimization process is crystal

quartz, the index is $n = 1.55$, i.e., $\epsilon_{11} = \epsilon_{22} = 2.4$. The cross-sectional area utilized for the optimization design region is $4\beta\lambda \times \lambda$, and the particle injection channel region is $4\beta\lambda \times \frac{1}{4}\lambda$. The acceleration gradient function ($\Xi_{\text{gain}}(E)$) of $3.1E_0$ ($\Gamma = 3.1$) is achieved by optimizing the maximum conversion factor (Γ). In Ref. [29], the gradient achieved was $0.49E_0$ ($\Gamma = 0.49$). Therefore, the topology-optimized structures achieve an approximately six-fold improvement in the acceleration gradient. It is worth noting from Fig. 7a that the optimized structure at this time deviates significantly from the intuitive physical structure. Although the structure differs greatly from the intuitive structure, it can be observed from the normalized electric field ($E_y/|E_0|$) in the y-direction, as shown in Fig. 7b, that the topologically optimized DLA structure has an obvious standing wave

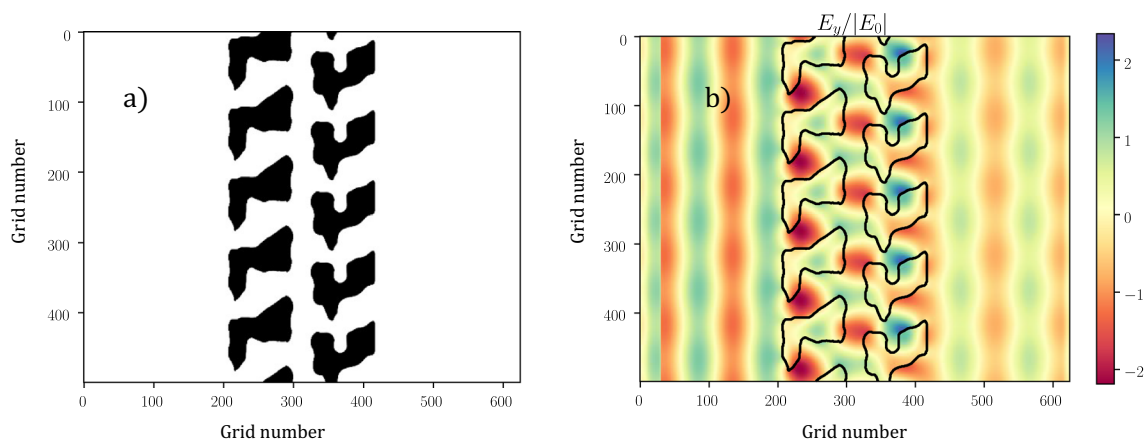


Fig. 8 (Color online) Topology optimization results of DLA structure using “Case 2” scheme. The parameter settings used are the same as those used in [29]

field distribution in the particle injection channel region (PICR), which benefits the acceleration of charged particles significantly.

The results of “**Case 2**” are shown in Fig. 8. The parameters used are the same as those used in Ref. [29]. The purpose is to point out that the method using automatic derivation can optimize the objective function more flexibly than the adjoint method, because in the “**Case 2**” scheme, the max function ($E_{\max}$) is not differentiable, which is a case when it is impossible to directly provide the accurate adjoint source. The cross-sectional area utilized for the optimization design region is $4\beta\lambda \times \lambda$, and the particle injection channel region is $4\beta\lambda \times \frac{1}{4}\lambda$. An acceleration gradient function ($\Xi_{\text{gain}}(E)$) of $1.4E_0$ ($\Gamma = 1.4$) is achieved by optimizing the maximum conversion factor (Γ). In Ref. [29], the gradient achieved is $0.49E_0$ ($\Gamma = 0.49$). Therefore, the topology-optimized structures show an approximately three-fold improvement in the acceleration gradient compared to previously established structures. As in Fig. 8 a, it can also be clearly observed that the optimized structure is very different from the intuitive physical structure. According to the normalized electric field ($E_y/|E_0|$), Fig. 8b indicates that the electric field intensity of the topologically optimized DLA chip is weak. For comparison, we calculated the normalized maximum electric field ($E_{\max}/|E_0|$) shown in Figs. 7a and 8a, and the results indicate that under the same parameters, the maximum electric field strength in the “**Case 2**” optimized structure is only one-fifth of the maximum electric field strength in the “**Case 1**” optimized structure. Therefore, it can be shown that when the “**Case 2**” scheme is implemented, the electric field intensity in the structure is reduced, which means that the structure is less likely to suffer damage.

The optimization results show that the optimized structures obtained by the two schemes (“**Case 1**”, and “**Case 2**”) are significantly different from the intuitive structures, and it is difficult to explain these results using physical intuition. It is worth noting that in the comparison between “**Case 1**” and “**Case 2**,” the structures obtained by the two optimization schemes are completely different. If χ is used as the evaluation index, the result for the acceleration gradient function in Fig. 7 is three times that in Fig. 8. If Γ is used as the evaluation index, the result for the acceleration gradient function in Fig. 8 is twice that of Fig. 7. The acceleration gradients of the two schemes are different, but they are significantly improved when compared to the grating structure [29]. A phenomenological explanation is that in the y-direction of the topologically optimized DLA structures (Figs. 7b and Fig. 8b), the electric field ($E_y/|E_0|$) exhibits a distinct standing wave field pattern in the particle injection channel region (PICR),

which is very favorable for the acceleration of charged particles. This also indicates that the DLA optimized by either of these two schemes can pass more particle beams, which undoubtedly can increase particle excitation (by appropriately increasing the width of the particle injection channel region). Although increasing the width of the particle injection channel region (PICR) reduces the acceleration gradient of the DLA, the optimized structures obtained by these two schemes provide an intrinsic acceleration gradient that is several times greater than that achieved by the grating-structured accelerator [29].

4 Summary

The topology optimization algorithm presented in this study can realize the optimization of any objective function, and the obtained structure also has a high-contrast binary structure. The advantage of this method is that the structure can be designed according to the expected goal, so that the optimal geometry can be found more intelligently without tedious adjustment of the associated parameters. By changing the material parameterization function, we can transition to shape optimization or size optimization. We study the acceleration structure of an accelerator based on the high-gradient and compact characteristics of the DLA. The two design schemes presented ignore the influence of E_x that the particle feels. However, the effect of E_x will lead to the deflection of accelerated particles over long distances; therefore, it may require that the design include other structures embedded in the junction of the accelerating structure to focus the beam in order to prevent the effect of E_x in the material.

In future work, our goal is to design a complete dielectric laser accelerator that supports larger sizes and increased focus effects. To achieve simple manufacturing, we may need to add more constraints to the current parameterization.

This technique is also applicable to the design of other dielectric-based accelerator structures. This is based on the inverse Cherenkov Effect of acceleration. As such, our work provides an opportunity to significantly increase the energy gain of DLAs, which is crucial for the practical application of these exciting micro accelerators.

Acknowledgements The authors thank Dr. Fu-Long Liu of the China Institute of Atomic Energy, Dr. Wen-Bo Mo of Tsinghua University, Dr. Dan Wang of Sichuan University, and Dr. Han-Sheng Ye of Tsinghua University for the illuminating discussions. We thank the reviewers for their constructive comments.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yang-Fan He and Bin Sun. The first draft of the

manuscript was written by Yang-Fan He and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. B.W.J. McNeil, N.R. Thompson, X-ray free-electron lasers. *Nat. Photon* **4**, 814–821 (2010). <https://doi.org/10.1038/nphoton.2010.239>
2. Y. Zhang, W.C. Fang, X.X. Huang et al., Design, fabrication, and cold test of an S-band high-gradient accelerating structure for compact proton therapy facility. *Nucl. Sci. Tech.* **32**, 38 (2021). <https://doi.org/10.1007/s41365-021-00869-z>
3. S.V. Bulanov, V.S. Khoroshkov, Feasibility of using laser ion accelerators in proton therapy. *Plasma Phys. Rep.* **28**, 453–456 (2002). <https://doi.org/10.1134/1.1478534>
4. W. Nan, B. Guo, C.J. Lin et al., First proof-of-principle experiment with the post-accelerated isotope separator on-line beam at BRIF: measurement of the angular distribution of $^{23}\text{Na} + ^{40}\text{Ca}$ elastic scattering. *Nucl. Sci. Tech.* **32**, 53 (2021). <https://doi.org/10.1007/s41365-021-00889-9>
5. E.L. Ginzton, W.W. Hansen, W.R. Kennedy, A Linear Electron Accelerator. *Rev. Sci. Instrum.* **19**, 89–108 (1948). <https://doi.org/10.1063/1.1741225>
6. Z.-Y. Ma, S.-J. Zhao, X.-M. Liu et al., High RF power tests of the first 13 GHz fundamental power coupler prototypes for the SHINE project. *Nucl. Sci. Tech.* **33**, 10 (2022). <https://doi.org/10.1007/s41365-022-00984-5>
7. C. Wang, J.H. Tan, X.X. Huang et al., Design optimization and cold RF test of a 26-cell cryogenic RF gun. *Nucl. Sci. Tech.* **32**, 97 (2021). <https://doi.org/10.1007/s41365-021-00925-8>
8. G.-R. Li, S.-X. Zheng, H.-J. Zeng et al., Design and test of an RF acceleration system loaded with magnetic alloy for the proton synchrotron of the Xi'an Proton Application Facility. *Nucl. Sci. Tech.* **29**, 94 (2018). <https://doi.org/10.1007/s41365-018-0434-9>
9. R.J. England, R.J. Noble, K. Bane et al., Dielectric laser accelerators. *Rev. Mod. Phys.* **86**, 1337 (2014). <https://doi.org/10.1103/RevModPhys.86.1337>
10. J. Breuer, P. Hommelhoff, Laser-based acceleration of nonrelativistic electrons at a dielectric structure. *Phys. Rev. Lett.* **111**, 134803 (2013). <https://doi.org/10.1103/PhysRevLett.111.134803>
11. E.A. Peralta, K. Soong, R.J. England et al., Demonstration of electron acceleration in a laser-driven dielectric microstructure. *Nature* **503**, 91–94 (2013). <https://doi.org/10.1038/nature12664>
12. K.J. Leedle, R.F. Pease, R.L. Byer et al., Laser acceleration and deflection of 96.3 keV electrons with a silicon dielectric structure. *Optica* **2**, 158–161 (2015). <https://doi.org/10.1364/OPTICA.2.000158>
13. K.J. Leedle, R.F. Pease, R.L. Byer et al., Dielectric laser acceleration of sub-100 keV electrons with silicon dual-pillar grating structures. *Optics Lett.* **40**, 4344–4347 (2015). <https://doi.org/10.1364/OL.40.004344>
14. U. Niedermayer, T. Egenolf, O. Boine-Frankenheim et al., Alternating phase focusing for dielectric laser acceleration. *Phys. Rev. Lett.* **121**, 214801 (2018). <https://doi.org/10.1103/PhysRevLett.121.214801>
15. D.S. Black, K.J. Leedle, Y. Miao et al., Laser-driven electron lensing in silicon microstructures. *Phys. Rev. Lett.* **122**, 104801 (2019). <https://doi.org/10.1103/PhysRevLett.122.104801>
16. N. Schönenberger, A. Mittelbach, P. Yousefi et al., Generation and characterization of attosecond microbunched electron pulse trains via dielectric laser acceleration. *Phys. Rev. Lett.* **123**, 264803 (2019). <https://doi.org/10.1103/PhysRevLett.123.264803>
17. D.S. Black, U. Niedermayer, Y. Miao et al., Net acceleration and direct measurement of attosecond electron pulses in a silicon dielectric laser accelerator. *Phys. Rev. Lett.* **123**, 264802 (2019). <https://doi.org/10.1103/PhysRevLett.123.264802>
18. K.P. Wootton, J. McNeur, K.J. Leedle, Dielectric laser accelerators: Designs, experiments, and applications. *Rev. Accelerat. Sci. Technol.* **09**, 105–126 (2016). <https://doi.org/10.1142/S179362681630005X>
19. K. Soong, R.L. Byer, E.R. Colby et al., Laser damage threshold measurements of optical materials for direct laser accelerators. *AIP Conf. Proceed.* **1507**, 511 (2012). <https://doi.org/10.1063/1.4773749>
20. R. Agustsson, E. Arab, A. Murokh et al., Measuring single-shot, picosecond optical damage threshold in Ge, Si, and sapphire with a 5.1- μm laser. *Optical Mater. Expr.* **5**, 2835–2842 (2015). <https://doi.org/10.1364/OME.5.002835>
21. D. Cesar, J. Maxson, P. Musumeci et al., Optical design for increased interaction length in a high gradient dielectric laser accelerator. *Nucl. Instrum. Meth. A* **909**, 252–256 (2018). <https://doi.org/10.1016/j.nima.2018.01.012>
22. J. Breuer, J. McNeur, P. Hommelhoff, Dielectric laser acceleration of electrons in the vicinity of single and double grating structures-theory and simulations. *J. Phys. B. At. Mol. Opt. Phys.* **47**, 234004 (2014). <https://doi.org/10.1088/0953-4075/47/23/234004>
23. T. Plettner, R.L. Byer, B. Montazeri, Electromagnetic forces in the vacuum region of laser-driven layered grating structures. *J. Mod. Opt.* **58**, 1518 (2011). <https://doi.org/10.1080/09500340.2011.611914>
24. D.S. Black, Z.X. Zhao, K.J. Leedle et al., Operating modes of dual-grating dielectric laser accelerators. *Phys. Rev. Accel. Beams* **23**, 114001 (2020). <https://doi.org/10.1103/PhysRevAccelBeams.23.114001>
25. D. Cesar, S. Custodio, J. Maxson et al., High-field nonlinear optical response and phase control in a dielectric laser accelerator. *Commun. Phys.* **1**, 46 (2018). <https://doi.org/10.1038/s42005-018-0047-y>
26. T.W. Hughes, S. Tan, Z.X. Zhao et al., On-chip laser-power delivery system for dielectric laser accelerators. *Phys. Rev. Appl.* **9**, 054017 (2018). <https://doi.org/10.1103/PhysRevApplied.9.054017>
27. N.V. Sapra, K.Y. Yang, D. Vercruysse et al., On-chip integrated laser-driven particle accelerator. *Science* **367**, 79–83 (2020). <https://doi.org/10.1126/science.aay5734>
28. <https://www.laserphysics.nat.fau.eu/research/achip/>
29. T. Plettner, P.P. Lu, R.L. Byer, Proposed few-optical cycle laser-driven particle accelerator structure. *Phys. Rev. ST Accel. Beams* **9**, 111301 (2006). <https://doi.org/10.1103/PhysRevSTAB.9.111301>
30. T. Hughes, G. Veronis, K.P. Wootton et al., Method for computationally efficient design of dielectric laser accelerator structures. *Opt. Expr.* **25**, 15414–15427 (2017). <https://doi.org/10.1364/OE.25.015414>
31. C.M. Lalau-Keraly, S. Bhargava, O.D. Miller et al., Adjoint shape optimization applied to electromagnetic design. *Opt. Expr.* **21**, 21693–21701 (2013). <https://doi.org/10.1364/OE.21.021693>
32. A. Michaels, E. Yablonovitch, Leveraging continuous material averaging for inverse electromagnetic design. *Opt. Expr.* **26**, 31717–31737 (2018). <https://doi.org/10.1364/OE.26.031717>
33. K. Yee, Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE T. Antenn. Propagat.* **14**, 302–307 (1966). <https://doi.org/10.1109/TAP.1966.1138693>
34. R.C. Rumpf, Simple implementation of arbitrarily shaped total-field/scattered-field regions in finite-difference frequency-

- domain. *Prog. Electromag. Resear. B* **36**, 221–248 (2012). <https://doi.org/10.2528/PIERB11092006>
35. T.W. Hughes, I.A.D. Williamson, M. Minkov et al., Forward-mode differentiation of Maxwell's equations. *ACS Photon.* **6**, 3010–3016 (2019). <https://doi.org/10.1021/acsp Photonics.9b01238>
36. S.G. Johnson, Notes on perfectly matched layers (PMLs). [arxiv: 2108.05348](https://arxiv.org/abs/2108.05348)
37. M. Minkov, I.A.D. Williamson, L.C. Andreani et al., Inverse design of photonic crystals through automatic differentiation. *ACS Photon.* **7**, 1729–1741 (2020). <https://doi.org/10.1021/acsp Photonics.0c00327>
38. Autograd: efficiently computes derivatives of numpy code. <https://github.com/HIPS/autograd>
39. D.P. Kingma, J. Ba, Adam: a method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.