Fast-speed algorithm to compute tight focusing of laser beams: Effectiveness of circularly polarized vortex beam series as a mathematical basis

Stepan Boichenko ®*

*Irkutsk Branch of the Institute of Laser Physics of the Siberian Branch of the Russian Academy of Sciences,
Lermontov Street 130a, 664033 Irkutsk, Russian Federation*

We suggest a time-effective single-integral algorithm to calculate tight focusing of a collimated continuous-wave coherent laser beam with an arbitrary cross-section light vector distribution by a high-aperture microscope objective into a planar microcavity. This algorithm is based on the mathematical fact that any beam can be decomposed into a superposition (either finite or infinite) of circularly polarized vortex vector beams, which allows one to factorize the focal field into two parts, one of which depends only on distance coordinates ρ and z and the other one only on an azimuth φ in cylindrical coordinates. We compare the suggested algorithm with that based on the direct use of the double-integral Richards-Wolf method and demonstrate that the former is from several times to two orders faster for single-point computations and from two to five orders faster for typical focal-region computations. Some existing algorithms based on the discrete Fourier transform can exceed the single-integral algorithm in computational speed up to two orders of magnitude but their accuracy is significantly less and error level is poorly controllable, whereas the single-integral algorithm is able to reproduce field values obtained by the direct Richards-Wolf integration with unlimited accuracy and completely controllable error level.

I. INTRODUCTION

A variety of microscopy techniques exploiting tightly focused laser beams is widely used nowadays. These techniques are laser-scanning fluorescent microscopy [1–4], stimulated emission depletion microscopy (STED) [5–7], multiphoton excitation microscopy [8,9], and some others [10]. In many researches, it is necessary to evaluate the focal-region light field formed by tight focusing of a laser beam by a microscope objective and in some cases it is desirable to know precisely the focal-region light field distribution [2,11–13]. Therefore, one needs to have a tool to calculate theoretically such a distribution, knowing the entrance laser beam light field distribution and microscope objective parameters.

A beam can be focused either in free space—that means an inhomogeneity-free medium with continuous refractive index, not vacuum—or under refractive-index-discontinuity conditions, for example, into an optical microcavity [14–16], below or above a glass-air interface [1], near a microcrack [17], inside an optical nanoparticle [13], and near other complicated geometrical shapes. The task of laser beam focusing simulation in free space, near a planar interface or in a planar microcavity, is completely solvable analytically and field values can be calculated numerically with practically unlimited precision. Possible restrictions can take place only due to the uncertainties in the microscope and sample parameters. Although such uncertainties cannot be eliminated completely, they can be decreased in any experiment to minimize differences between simulations and experimental measurements.

On the contrary, precise simulation of laser field focusing near optical nanoparticles, microcracks, and other nano-objects with stochastic geometrical shapes is nearly impossible due to the lack of knowledge of these shapes. But even when the shapes can be defined, the complexity of the task of calculating a focal field near such a structure is normally orders of magnitude greater than that of near-planar-structure focusing and it is rarely solved in practice. So, the tight focusing of a laser beam into a planar microcavity can be considered as the most generalized analytically solvable task and we will explore it here. The tight focusing in free space and near a planar interface are particular cases of this task.

Richards-Wolf theory [18] is a key universal tool to calculate tight focusing of laser beams and, in principle, it allows one to calculate tight focusing of any beam. Although the original method developed by Richards and Wolf was designed for free-space focal field calculations, it can be adapted to calculate focusing near planar interfaces without significant changes in the approach [15]. The critical point is that the original method includes double integration, whereas in most cases it is possible to reduce the focal field calculation procedure down to single integration [10], which allows one to decrease calculation time drastically. Such a reduction can normally be done for any entrance beam and attempts to develop the respective generalized algorithm for some families of entrance beams have been made. So, tight focusing of nonvortex radially and azimuthally polarized laser beams into planar microcavities was considered by Meixner's group [15] and the suggested algorithm can be readily applied to this family of beams but it cannot be applied, for example, to a vortex beam. Free-space focusing of vortex laser beams was considered in Ref. [19] but the algorithm suggested there

---
*ste89@yandex.ru

cannot be applied to into-microcavity focusing because of
the mutual affection of radial and azimuthal components of
a vortex beam. In fact, a generalized single-integral algorithm
applicable to an arbitrary entrance beam had not been devel-
oped at present, to the best of our knowledge. In most cases,
one has to develop it for a given entrance beam individually.
Meanwhile, for some types of entrance beams it might be a
complicated time-consuming problem, on the one hand. On
the other hand, a generalized algorithm would save some time
anyway even for entrance beams, to which the Richards-Wolf
method can be applied easily. In the present investigation we
develop a mathematical basis for a single-integral algorithm,
which allows one to compute tight focusing of an arbitrary
collimated beam. Finally, such an algorithm is suggested.

Algorithms based on the fast Fourier transform (FFT)
allow one to reduce computational times by orders of mag-
nitude. Such an algorithm was suggested for the first time in
Ref. [20], developed further in Refs. [21,22], and then inte-
grated into enhanced algorithms developed to calculate tight
focusing of partially coherent beams [23,24]. FFT-based algo-
rithms are pretty fast and accurate, being applicable to many
types of focused beams. However, the FFT procedure is based
on the discrete Fourier transform (DFT) and, hence, calculates
numerical values of a set of double integrals directly without
exploiting the possibility of reducing a double integral down
to a single one. As computational cost of a double integral is
significantly higher than that of a single one, a single-integral
algorithm will be potentially advantageous compared to FFT
or any other DFT-based one.

The paper is organized as follows. In Sec. II we briefly de-
scribe the problem and introduce the Richards-Wolf method.
In Sec. III we explore mathematical properties of an entrance
beam and find a generalized mathematical representation of an
arbitrary beam; then we adopt the transfer-matrix method to
calculate a tightly focused light field inside a planar microcav-
ity. In the next step, a generalized single-integral algorithm is
described schematically and the problem of an entrance beam
expressed as an infinite Fourier series is explored. In Sec. IV
we test the suggested algorithm on some basic entrance beams
and a slowly convergent infinite-series beam as representative
examples. In Sec. V calculations of focal fields inside finite
spatial regions are investigated and the algorithm is adapted
to this task. In Sec. VI the results are summarized and the
suggested algorithm is analyzed and discussed.

II. FOCAL FIELD CALCULATION: RICHARDS-WOLF THEORY

Let us consider a laser beam focused by a microscope
objective into a planar microcavity as shown in Fig. 1 and we
want to find the light field in a given layer of interest inside the
microcavity (core layer). Here and below we will consider the
Richards-Wolf method as adapted by Novotny and Hecht and
described in Ref. [10]. The objective medium (left-hand-side
medium in the figure) refractive index is n_1, the microcavity
core layer index is n, and the right-hand-side medium index
is n_2. It is assumed that the microscope objective satisfies
the Abbe sine condition, which means that a focused laser
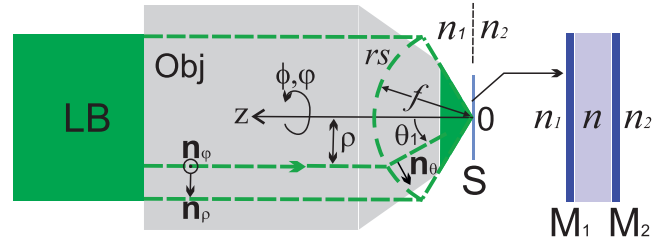beam with a plane wave front is converted into a convergent

FIG. 1. A focusing scheme. A laser beam (LB) enters a micro-
scope objective (Obj) to be focused onto a sample (S), which is
assumed to be a planar microcavity. The objective medium refractive
index is assumed to be equal to n_1, the microcavity core layer index
n, and the right-hand side medium index n_2. M_1 and M_2 are the
microcavity mirrors. The vectors $\mathbf{n}_i$ with $i = \rho, \varphi, \theta$ are unit vectors
of the respective curvilinear coordinate systems; $\rho, \theta, \phi, \varphi$, and z are
coordinates; rs is the reference sphere, and f is the objective focal
distance.

spherical wave and the entrance beam wave front is projected
onto a sphere of radius f (the reference sphere).

To calculate a focal-region light field, we have to define
a coordinate system. The present focusing model assumes
that (i) rays of the laser beam are refracted at the reference
sphere to be directed to the geometrical focus and (ii) the
light field in a focal-region point is formed by a superposition
of fields propagating from all points of the reference sphere.
Hence, one needs to integrate the reference-sphere field. A
microscope objective is a rotationally symmetric system with
respect to the longitudinal axis; the reference sphere is a part
of a spherically symmetric system. To this end, we have to
use a cylindrical coordinate system with inclusion of spherical
coordinates.

As shown in Fig. 1, the coordinate origin is placed at the
geometrical focus of the objective and the longitudinal axis z
is directed toward the objective pupil. We will use cylindrical
coordinates (ρ, φ, z), where ρ is the distance from the longi-
tudinal axis z and φ is the azimuth. To define the azimuth, we
need to introduce Cartesian coordinates x and y in the cross
section. We can use any pair of mutually orthogonal axes that
form a right-hand coordinate system in conjunction with the
shown z axis as x and y. Positive values of the azimuth φ
correspond to the counterclockwise direction as shown in the
figure.

We need to define two spatial points: the observation point
in the focal region where we calculate the field and the field
source point on the reference sphere (in the entrance beam
cross section). Below, focal-region points will be described
by coordinates ρ, φ, and z, reference-sphere points by a polar
angle θ_1 and azimuthal angle ϕ, and beam cross-section points
by ρ, ϕ, and z. We use the same names for the longitudinal and
distance coordinates in the beam cross section and in the focal
region but these groups of coordinates will not appear in the
same equations. The following three polar angles will be used:
θ_1 in the objective space (refractive index n_1), θ inside the
microcavity core layer (refractive index n), and θ_2 in the right-
hand-side semispace (refractive index n_2). In this notation, the
reference-sphere polar angle is θ_1. The entrance beam cross-
section distance coordinate ρ and reference-sphere polar angle
θ_1 are linked by the relation $\rho = f \sin \theta_1$. Vectors $\mathbf{n}_\rho$ and $\mathbf{n}_\varphi$

are unit vectors of the distance and azimuth coordinates in a cylindrical coordinate system, and $\mathbf{n}_\theta$ is the unit vector of the polar coordinate in spherical coordinates. The conversion of the entrance beam into a spherical wave does not affect $\mathbf{n}_\varphi$ as it is orthogonal to the plane in which a ray is refracted: $\mathbf{n}_\varphi$ remains the same on the reference sphere as it is in the cross section; $\mathbf{n}_\rho$ is converted into $\mathbf{n}_\theta$.

We assume that the entrance beam light field is a plane-wave-like monochromatic wave propagating along the negative z direction,

$$\mathbf{E}_{eb}(\rho, \phi, z, t) = \mathbf{E}_0(\rho, \phi)e^{ik_1 z}e^{i\omega t}, \tag{1}$$

with t being time, ω the angular frequency, $k_1 = n_1\omega/c$ the wave number, c the vacuum speed of light, and $\mathbf{E}_0(\rho, \phi)$ the vector amplitude lying in the beam cross-section plane. The conversion of the beam into a converging spherical wave transforms the vector amplitude into

$$\mathbf{E}_{rs}(\theta_1, \phi) = \sqrt{\cos(\theta_1)}\{[\mathbf{E}_0(\theta_1, \phi) \cdot \mathbf{n}_\rho]\mathbf{n}_\theta \\ + [\mathbf{E}_0(\theta_1, \phi) \cdot \mathbf{n}_\varphi]\mathbf{n}_\varphi\}. \tag{2}$$

To obtain a light field value in a free-space focal-region point, we have to calculate a superposition of partial secondary plane waves propagating from the reference sphere to the focal region:

$$\mathbf{E}_{foc}(\mathbf{r}) = -\frac{ik_1 f e^{ik_1 f}}{2\pi} \int_0^{\theta_{\max}} \int_0^{2\pi} \mathbf{E}_{rs}(\theta_1, \phi) \\ \times e^{i\mathbf{k}(\theta_1, \phi)\cdot\mathbf{r}} \sin\theta_1 \, d\phi \, d\theta_1, \tag{3}$$

with $\mathbf{r} = (x, y, z)$ being the Cartesian radius vector of a focal-region point, $\theta_{\max}$ being the objective angular aperture, and

$$\mathbf{k}(\theta_1, \phi) = k_1 \begin{pmatrix} \sin\theta_1 \cos\phi \\ \sin\theta_1 \sin\phi \\ \cos\theta_1 \end{pmatrix} \tag{4}$$

a wave vector. This expression is an angular plane-wave spectrum representation of the focal field composed by waves with the vector amplitudes $\mathbf{E}_{rs}(\theta_1, \phi)\sin\theta_1$ and wave vectors $\mathbf{k}(\theta_1, \phi)$. To obtain the focal field inside a microcavity, we need to take into account action of such a cavity on the field (3). As this field is represented by a superposition of mutually independent plane waves, we can readily explore refraction and reflection of each partial wave separately and then integrate cavity-converted waves to obtain the intracavity focal field. Doing this, we can write the intracavity focused field as

$$\mathbf{E}_{foc}(\mathbf{r}) = -\frac{ik_1 f e^{ik_1 f}}{2\pi} \int_0^{\theta_{\max}} \int_0^{2\pi} \widehat{MC}[\mathbf{E}_{rs}(\theta_1, \phi) \\ \times e^{i\mathbf{k}(\theta_1, \phi)\cdot\mathbf{r}} \sin\theta_1] \, d\phi \, d\theta_1. \tag{5}$$

The operator $\widehat{MC}$ represents the action of the microcavity at a partial plane wave. Details of its action will be considered in Sec. III B. Equation (5) in conjunction with Eqs. (1) and (2) is the starting point in any collimated laser beam tight focusing simulation.

III. GENERALIZED SINGLE-INTEGRAL ALGORITHM

In this section, we will explore the problem of a generalized single-integral algorithm. The section is organized as follows. In Sec. III A we solve the problem of an appropriate generalized mathematical representation of an entrance field vector amplitude $\mathbf{E}_0(\theta_1, \phi)$ appearing in Eq. (5). Finally, we derive that a circularly polarized vortex vector beam series turns out to be an optimal representation. In Sec. III B single-integral focusing equations for an arbitrary beam are derived. In Sec. III C we describe a generalized scheme of the suggested algorithm. In Sec. III D problems of an entrance beam represented by an infinite series are investigated.

A. Generalized mathematical representation of vector light fields

Let us consider the vectorial representation of an entrance field. The curvilinear unit vectors in Eq. (2) are expressed in terms of the Cartesian unit vectors $\mathbf{n}_x$ and $\mathbf{n}_y$ as

$$\mathbf{n}_\rho = \cos\varphi\, \mathbf{n}_x + \sin\varphi\, \mathbf{n}_y, \tag{6a}$$

$$\mathbf{n}_\varphi = -\sin\varphi\, \mathbf{n}_x + \cos\varphi\, \mathbf{n}_y, \tag{6b}$$

$$\mathbf{n}_\theta = \cos\theta\, \mathbf{n}_\rho - \sin\theta\, \mathbf{n}_z. \tag{6c}$$

The use of these vectors was imposed by the cylindrical symmetry of the microscope objective system: it is very convenient to treat the refraction at the reference sphere in a cylindrical vector basis. Focusing of a radially polarized entrance beam with an azimuth-independent amplitude gives radially and longitudinally polarized focal-region components; an azimuthally polarized entrance beam with an azimuth-independent amplitude gives an azimuthal focal-region component [10,15,25]. It makes the cylindrical vector basis even more convenient but in general the field amplitudes are azimuth dependent. Because of this dependence, a radially polarized entrance component becomes able to produce an azimuthally polarized focal-region one, and vice versa. This mutual influence makes the cylindrical vector basis inconvenient and leads to the problem of an optimal polarization basis: the reference sphere refraction requires a cylindrical vector basis but the integration procedure in Eq. (5) makes a stationary (not curvilinear) basis preferable. But regardless of this problem, Eq. (5) will include the ϕ integration of functions $E_0^i(\theta_1, \phi)\cos\phi$ and $E_0^i(\theta_1, \phi)\sin\phi$ ($i = \rho, \varphi$). Such an integration requires an adequate series representation of entrance beam scalar amplitudes. Below, we will first consider the problem of generalized representation of scalar amplitudes and then study the problem of an optimal polarization basis.

Fourier series is the most attractive solution to the problem of generalized representation of scalar amplitudes because, first, it is closely connected to the functions $\cos\phi$ and $\sin\phi$, and second, the field components are inherently ϕ-periodic functions with the period equal to 2π. Some examples from practice may call into question this periodicity and Fig. 2 explains it in detail. Figure 2(a) shows a laser beam cross section. The axes x and y are Cartesian coordinate axes, ϕ is the azimuth angle, and ρ the distance coordinate in a cylindrical coordinate system. We will consider a function $F(\rho, \phi)$ as a common field component. Such a function is shown schematically in grayscale tones. The angle ϕ is
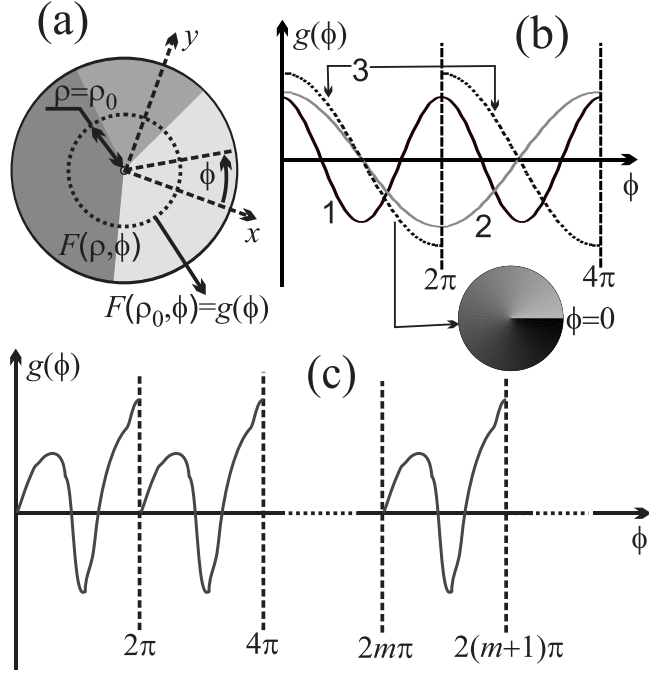
FIG. 2. An explanation of the light field azimuthal periodicity. (a) A laser beam cross-section scheme. (b) Examples of periodic functions. (c) An example of a generalized periodic function.

varied from zero (it corresponds to the x direction) to 2π and a 2π term added to an azimuth corresponds to a full circle around the cylinder axis, which guarantees the periodicity condition $F(\rho, \phi + 2\pi) = F(\rho, \phi)$. However, there exist some light beams described by aperiodic functions, for example, fractional vortex beams [26,27]. An example of such a beam is $\mathbf{E}_0(\rho, \phi) = A(\rho)e^{i\phi/2}\mathbf{n}_x$ with $A(\rho)$ being a scalar amplitude and the azimuth dependence is described by a function with a period of 4π instead of 2π. Figure 2(b) displays some examples of periodic functions to settle this point. We denote a function $F(\rho, \phi)$ on a circle with the radius ρ_0 as $g(\phi)$: $F(\rho = \rho_0, \phi) = g(\phi)$. Next, a function $g(\phi)$ is displayed in rectangular coordinates in Fig. 2(b). The azimuth range is spread from $[0, \dots, 2\pi]$ to $[0, \dots, 4\pi]$ to explore the property of periodicity. Three functions are displayed. Curve 1 corresponds to the function $g(\phi) = \cos\phi$ and its period is equal to 2π. Curve 2 corresponds to the function $g(\phi) = \cos(\phi/2)$ and its period is equal to 4π, which leads to the problem that the function takes on different values in the ranges $[0, \dots, 2\pi]$ and $[2\pi, \dots, 4\pi]$, while $(\rho_0, \phi + 2\pi)$ corresponds to the same spatial point as (ρ_0, ϕ) and light field components are single-valued functions of spatial coordinates. The solution to this problem is demonstrated by curve 3 and the respective inset. Curves 2 and 3 correspond to the same function (with different amplitudes for the sake of visibility) in the range $[0, \dots, 2\pi]$, but at $\phi = 2\pi$ curve 3 has a jump. So, if a light field component is described by a ϕ-aperiodic function, the aperiodicity indicates not the component's ambiguity but its discontinuity at $\phi = 0$ (or 2π). It means that we can (mathematically) spread the azimuth range from $[0, \dots, 2\pi]$ to $[0, \dots, \infty]$ and a g function will be periodic (with or without jumps) as shown in Fig. 2(c).

This conclusion is crucially important: it guarantees that an arbitrary light field can be expressed mathematically as a Fourier series in the range $\phi = [0, \dots, 2\pi]$.

Next, we will derive an optimal vector basis matching the Fourier series representation. A Fourier series can be expressed either in terms of real functions $\sin(n\phi)$ and $\cos(n\phi)$ or in terms of complex functions $\exp(in\phi)$, where n is an integer. If we use the former approach and represent an entrance field in a cylindrical vector basis, the need to integrate terms $\sin(n\phi)\sin\phi$, $\sin(n\phi)\cos\phi$, $\cos(n\phi)\sin\phi$, and $\cos(n\phi)\cos\phi$ in Eq. (5) will appear. The latter approach will allow us to treat products of complex exponents $\exp(im\varphi)\exp(in\phi)$ instead of products of sines and cosines if the cylindrical basic vectors can be expressed through the complex exponents. Obviously, the exponents are significantly more convenient than the sines and cosines and it is reasonable to try the latter approach. Substituting the expressions $\cos\phi = 0.5[\exp(i\phi) + \exp(-i\phi)]$ and $\sin\phi = -0.5i[\exp(i\phi) - \exp(-i\phi)]$ into Eqs. (6), we obtain

$$\mathbf{n}_\rho = 2^{-0.5}(e^{-i\varphi}\mathbf{n}_+ + e^{i\varphi}\mathbf{n}_-), \tag{7a}$$

$$\mathbf{n}_\varphi = -2^{-0.5}i(e^{-i\varphi}\mathbf{n}_+ - e^{i\varphi}\mathbf{n}_-), \tag{7b}$$

with

$$\mathbf{n}_\pm = 2^{-0.5}(\mathbf{n}_x \pm i\mathbf{n}_y) \tag{8}$$

being unit vectors of a circular polarization basis. The reverse transform reads as

$$\mathbf{n}_+ = 2^{-0.5}e^{i\varphi}(\mathbf{n}_\rho + i\mathbf{n}_\varphi), \tag{9a}$$

$$\mathbf{n}_- = 2^{-0.5}e^{-i\varphi}(\mathbf{n}_\rho - i\mathbf{n}_\varphi). \tag{9b}$$

So, the cylindrical vector basis and circular polarization basis are expressed in terms of each other with coefficients proportional to $\exp(\pm i\phi)$, which makes a circular polarization basis maximally convenient. Therefore, we will explore light field Fourier series representation in a circular polarization vector basis.

Any light field with a total angular momentum defined by a topological charge m, propagating along the negative z direction, can be expressed as a complex superposition of the following two fields:

$$\mathbf{E}_+(\rho, \phi, z) = \exp[i(m-1)\phi]A_+(\rho, z)\mathbf{n}_+$$
$$+ \exp(im\phi)A_z^+(\rho, z)\mathbf{n}_z, \tag{10a}$$

$$\mathbf{E}_-(\rho, \phi, z) = \exp[i(m+1)\phi]A_-(\rho, z)\mathbf{n}_-$$
$$+ \exp(im\phi)A_z^-(\rho, z)\mathbf{n}_z, \tag{10b}$$

with the amplitudes

$$A_\pm(\rho, z) = \int_0^\infty Q_\pm(q)J_{m\mp1}(q\rho)\exp[ik(q)z]dq, \tag{11a}$$

$$A_z^\pm(\rho, z) = \mp i\int_0^\infty [q/k(q)]Q_\pm(q)J_m(q\rho)\exp[ik(q)z]dq, \tag{11b}$$

where $k(q) = \sqrt{(2\pi n/\lambda)^2 - q^2}$ is a longitudinal wave number, q a transverse extinguish parameter, $Q_\pm(q)$ functions of the parameter q, and m an integer [28]. The parameter q

varies up to infinity in the Bessel beam field representation in Eqs. (11), but for a laser beam to be paraxial (as we are considering such a beam), the functions $Q(q)$ must be nonvanishing only in a narrow range of q values and these values should be of the order of $1/f$. Although Bessel beams with high q values are nondiffractive like those with low q values, their plane-wave angular spectrum representation is formed by waves with large transverse components of the wave vector [29] and they can be neither tightly focused nor considered paraxial. Therefore, we have the approximation $q/k(q) \propto \lambda/f \propto 10^{-3}$ as light wavelength λ is of the order of hundreds of nanometers and the focal length f is of the order of millimeters. Hence, the longitudinal component in Eq. (10) can be neglected and the integral in Eq. (11a) can be approximated as

$$\int_0^\infty Q_\pm(q) J_{m\mp 1}(q\rho) \exp[ik(q)z] dq \approx G_m^\pm(\rho) \exp(ikz), \quad (12)$$

where the factor G, in general, weakly depends on z but this dependence is negligible and we assume G to be dependent on the ρ coordinate only. Thus, a light field of a paraxial laser beam with a topological charge m can be expressed as

$$\mathbf{E}_m(\rho, \phi, z) = e^{im\phi}[e^{-i\phi}G_m^+(\rho)\mathbf{n}_+ + e^{i\phi}G_m^-(\rho)\mathbf{n}_-]e^{ikz}. \quad (13)$$

To obtain a generalized decomposition of a light field of an arbitrary paraxial beam, we should summarize fields (13) over all the topological charges: $m = [-\infty, \dots, \infty]$. Summarizing them and substituting $G_{m+1}^+(\rho)$ by $G_m^+(\rho)$ and $G_{m-1}^-(\rho)$ by $G_m^-(\rho)$, we obtain the generalized expression for an entrance field $\mathbf{E}(\rho, \phi)$ in Eq. (10):

$$\mathbf{E}_0(\rho, \phi) = E_+(\rho, \phi)\mathbf{n}_+ + E_-(\rho, \phi)\mathbf{n}_- \quad (14)$$

with

$$E_+(\rho, \phi) = \sum_{m=-\infty}^{\infty} G_m^+(\rho)\exp(im\phi), \quad (15a)$$

$$E_-(\rho, \phi) = \sum_{m=-\infty}^{\infty} G_m^-(\rho)\exp(im\phi). \quad (15b)$$

This equation is a complex-function Fourier series representation of an entrance field in a circular polarization vector basis. It can be called a circularly polarized vortex beam series representation, as each term of the series $G_m^\pm(\rho)\exp(im\phi)\mathbf{n}_\pm$ is physically a circularly polarized vortex vector beam (CPVVB).

As (i) the set of complex exponential functions $\{\exp(im\phi)\}$ is a complete basis [30] and (ii) an entrance field complex vector amplitude is an inherently periodic function of the variable ϕ as was explained above (see Fig. 2), we can guarantee that amplitudes $E_+(\rho, \phi)$ and $E_-(\rho, \phi)$ in Eq. (14) are Fourier-series-expandable functions and any collimated z-propagating beam complex field amplitude can be expressed in the form given by Eqs. (14) and (15). Hence, one can decompose any entrance laser beam into a linear superposition of CPVVBs and treat it in such a form.

B. Circularly polarized vortex vector beam: The focusing process

To calculate focusing of an entrance field represented by a series (14), we need to calculate focusing of each partial CPVVB of the series

$$\mathbf{E}_m^\pm(\rho, \phi) = G_m^\pm(\rho)\exp(im\phi)\mathbf{n}_\pm \quad (16)$$

and then summarize the partial focal fields. So, focusing of a basic CPVVB [Eq. (16)] is the first task that we need to solve. Exploiting Eqs. (9) and (6c), one obtains the respective reference-sphere field:

$$\mathbf{E}_{rs}(\theta_1, \phi) = 2^{-0.5}\sqrt{\cos\theta_1}F_m^\pm(\theta_1)\exp[i(m \pm 1)\phi]$$
$$\times [\cos\theta_1\, \mathbf{n}_\rho \pm i\, \mathbf{n}_\varphi - \sin\theta_1\, \mathbf{n}_z]. \quad (17)$$

The index m was omitted at the reference-sphere field as it is an intermediate entity and a new function $F_m^\pm(\theta_1) = G_m^\pm(f \sin\theta_1)$ was introduced for the sake of brevity. Substituting this field into Eq. (3), we obtain for a partial plane wave of the angular spectrum representation

$$\mathbf{E}_{pw}(\mathbf{r}; \theta_1, \phi) = \mathbf{E}_{pw}^{(s)}(\mathbf{r}; \theta_1, \phi) + \mathbf{E}_{pw}^{(p)}(\mathbf{r}; \theta_1, \phi) \quad (18)$$

with

$$\mathbf{E}_{pw}^{(s)}(\mathbf{r}; \theta_1, \phi) = \pm iA(\theta_1, \phi)\, \mathbf{n}_\varphi e^{i\mathbf{k}(\theta_1, \phi)\cdot\mathbf{r}}, \quad (19a)$$

$$\mathbf{E}_{pw}^{(p)}(\mathbf{r}; \theta_1, \phi) = A(\theta_1, \phi)(\cos\theta_1\, \mathbf{n}_\rho - \sin\theta_1\, \mathbf{n}_z)e^{i\mathbf{k}(\theta_1, \phi)\cdot\mathbf{r}}, \quad (19b)$$

where indices s and p denote s-polarized (or transverse-electric) and p-polarized (or transverse-magnetic) waves, respectively; $A(\theta_1, \phi) = 2^{-0.5}\sqrt{\cos\theta_1}\sin\theta_1 F_m^\pm(\theta_1)\exp[i(m \pm 1)\phi]$. An azimuthal polarization corresponds to s-polarized waves and a radial polarization to p waves in the angular spectrum representation. Exploiting Eq. (18), we can explore
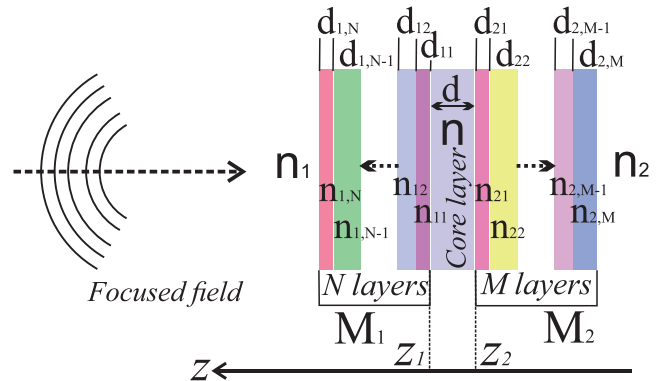
FIG. 3. The detailed structure of a microcavity shown in Fig. 1. M_1 and M_2 are planar layered mirrors, the objective medium semispace refractive index is n_1, and the right-hand-side medium refractive index is n_2. The focused laser field is calculated in a core layer and left and right bounds of the core layer are described by z_1 and z_2 coordinates, respectively. Mirror M_1 consists of N layers and mirror M_2 of M layers; the jth layer of an ith mirror is characterized by a refractive index $n_{i,j}$ and thickness $d_{i,j}$.

microcavity action on partial plane waves. Denoting the microcavity operator as $\widehat{MC}$ and applying it to Eqs. (19), one obtains

$$\widehat{MC}\big[\mathbf{E}_{pw}^{(s)}(\mathbf{r};\theta_1,\phi)\big] = \pm i\, T_\varphi(\theta_1,z)A(\theta_1,\phi)\,\mathbf{n}_\varphi \mathrm{e}^{i\mathbf{k}_\perp(\theta_1,\phi)\cdot\mathbf{r}_\perp}, \tag{20a}$$

$$\widehat{MC}\big[\mathbf{E}_{pw}^{(p)}(\mathbf{r};\theta_1,\phi)\big] = A(\theta_1,\phi)[T_\rho(\theta_1,z)\cos\theta\,\mathbf{n}_\rho \\ - T_z(\theta_1,z)\sin\theta\,\mathbf{n}_z]\mathrm{e}^{i\mathbf{k}_\perp(\theta_1,\phi)\cdot\mathbf{r}_\perp}. \tag{20b}$$

The symbol $\perp$ denotes transverse vector components (orthogonal to the z direction). As the transverse wave number $k_\perp$ conserves according to Snell's law, the dot product takes the form

$$\mathbf{k}_\perp(\theta_1,\phi)\cdot\mathbf{r}_\perp = k_1\rho\sin\theta_1\cos(\varphi-\phi). \tag{21}$$

The angle θ depends on the angle θ_1 as $n\sin\theta = n_1\sin\theta_1$; here and below such dependence is implied, although not displayed apparently. $T_i(\theta_1,z)$ ($i=\rho,\varphi,z$) are intracavity interference coefficients defined by the cavity parameters.

Figure 3 displays the detailed structure of a microcavity under consideration. Mirrors can be considered planar stratified media and described by the matrices

$$M_1 = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1,N} \\ n_{11} & n_{12} & \cdots & n_{1,N} \end{pmatrix}, \tag{22a}$$

$$M_2 = \begin{pmatrix} d_{21} & d_{22} & \cdots & d_{2,M} \\ n_{21} & n_{22} & \cdots & n_{2,M} \end{pmatrix}, \tag{22b}$$

where d_{ij} and n_{ij} are the thickness and refractive index of a jth layer of the ith mirror; $i=1,2$; $j=1,\ldots,N$ at $i=1$, $j=1,\ldots,M$ at $i=2$ (mirror M_1 consists of N layers and mirror M_2 of M layers). The focused field is passed through mirror M_1 from the objective medium with the refractive index n_1 into the core layer of a thickness d and with the refractive index n, then multiply reflected from the two mirrors to form the final focal field inside the core layer. A partial plane wave undergoes the following three processes: (i) transmission from medium 1 into the core layer through mirror M_1, (ii) reflection from mirror M_2, and (iii) reflection from mirror M_1. We will denote the transmission coefficients as $t_1^\alpha(\theta_1)$ and the reflection coefficients as $r_i^\alpha(\theta_1)$ with $\alpha = s, p$ being the wave polarization index and $i=1,2$ the mirror index. These coefficients can be calculated by means of the transfer-matrix method described, for example, in Ref. [31]. The intracavity interference coefficients take the form

$$T_m(\theta_1,z) = t_1^{\alpha_m}(\theta_1)\frac{\exp[-ik_z(\theta_1)z] + s_m r_2^{\alpha_m}(\theta_1)\exp[ik_z(\theta_1)(z-2z_2)]}{1 - r_2^{\alpha_m}(\theta_1)r_1^{\alpha_m}(\theta_1)\exp[2ik_z(\theta_1)d]} \tag{23}$$

with $m = \rho,\varphi,z$; $k_z(\theta_1) = k_0 n\cos[\theta(\theta_1)]$, $s_\rho = -1$, $s_\varphi = 1$, $s_z = 1$; $\alpha_\rho,\alpha_z = p$, $\alpha_\varphi = s$.

A partial plane wave (18) transformed by a microcavity takes the form

$$\widehat{MC}[\mathbf{E}_{pw}(\mathbf{r};\theta_1,\phi)] = \big[E_{rs}^{+,mc}(\theta_1,\phi;z)\mathbf{n}_+ + E_{rs}^{-,mc}(\theta_1,\phi;z)\mathbf{n}_- + E_{rs}^{z,mc}(\theta_1,\phi;z)\mathbf{n}_z\big]\mathrm{e}^{i\mathbf{k}_\perp(\theta_1\phi)\cdot\mathbf{r}_\perp} \tag{24}$$

with the amplitudes

$$E_{rs}^{+,mc}(\theta_1,\phi;z) = 0.5\sqrt{\cos\theta_1}F_m^\pm(\theta)[T_\rho(\theta,z)\cos\theta \pm T_\varphi(\theta,z)]\exp[i(m\pm 1-1)\phi], \tag{25a}$$

$$E_{rs}^{-,mc}(\theta_1,\phi;z) = 0.5\sqrt{\cos\theta_1}F_m^\pm(\theta)[T_\rho(\theta,z)\cos\theta \mp T_\varphi(\theta,z)]\exp[i(m\pm 1+1)\phi], \tag{25b}$$

$$E_{rs}^{z,mc}(\theta_1,\phi;z) = -(1/\sqrt{2})\sqrt{\cos\theta_1}F_m^\pm(\theta)T_z(\theta,z)\sin\theta\exp[i(m\pm 1)\phi]. \tag{25c}$$

To integrate this field over the azimuthal angle, we will exploit the relationship

$$\int_0^{2\pi}\exp(il\phi)\exp[ik_1\rho\sin\theta_1\cos(\varphi-\phi)]d\phi = 2\pi J_l(k_1\rho\sin\theta_1)\exp[i(\varphi+\pi/2)l], \tag{26}$$

valid for an integer l with $J_l(k\rho\sin\theta)$ being the lth-order Bessel function of the first kind [10]. Substituting Eq. (21) into Eq. (24) and integrating over ϕ, one obtains the focal field of a partial CPVVB:

$$\mathbf{E}_{m,\pm}^f(\rho,\varphi,z) = 0.5ik_1 f\mathrm{e}^{ik_1f}\exp[i(\varphi+\pi/2)(m\pm 1)]\big[A_{+,\pm}^m(\rho,z)\mathrm{e}^{-i(\varphi+\pi/2)}\mathbf{n}_+ + A_{-,\pm}^m(\rho,z)\mathrm{e}^{i(\varphi+\pi/2)}\mathbf{n}_- + A_{z,\pm}^m(\rho,z)\mathbf{n}_z\big] \tag{27}$$

with

$$A_{+,\pm}^m(\rho,z) = \int_0^{\theta_{\max}}\sqrt{\cos\theta_1}\sin\theta_1 F_m^\pm(\theta_1)J_{m\pm 1-1}(k_1\rho\sin\theta_1)[T_\rho(\theta_1,z)\cos\theta \pm T_\varphi(\theta_1,z)]d\theta_1, \tag{28a}$$

$$A_{-,\pm}^m(\rho,z) = \int_0^{\theta_{\max}}\sqrt{\cos\theta_1}\sin\theta_1 F_m^\pm(\theta_1)J_{m\pm 1+1}(k_1\rho\sin\theta_1)[T_\rho(\theta_1,z)\cos\theta \mp T_\varphi(\theta_1,z)]d\theta_1, \tag{28b}$$

$$A_{z,\pm}^m(\rho,z) = -\sqrt{2}\int_0^{\theta_{\max}}\sqrt{\cos\theta_1}\sin\theta_1\sin\theta F_m^\pm(\theta_1)J_{m\pm 1}(k_1\rho\sin\theta_1)T_z(\theta_1,z)d\theta_1. \tag{28c}$$

C. Generalized single-integral algorithm: Formalized scheme

Now, the problem can be formulated as follows. We have an entrance laser beam with a given light field vector amplitude $\mathbf{E}_0(\rho, \phi)$. In the first step, we need to represent the amplitude in a circular polarization basis as in Eq. (14). Then the amplitudes $E_+(\rho, \phi)$ and $E_-(\rho, \phi)$ should be decomposed into complex-function Fourier series (15). For azimuth-continued simple fields this series reduces to a finite sum, but for complicated fields the series is infinite. Finally, we should calculate focusing of all the CPVVBs representing $\mathbf{E}_0(\rho, \phi)$ and summarize their focal fields.

Mathematically, we can associate with amplitudes $E_+(\rho, \phi)$ and $E_-(\rho, \phi)$ a set of integer-valued numbers M,

$$M = \{m_1, m_2, \ldots, m_p\}, \tag{29}$$

which consists of p elements, and light field radial amplitudes can be expressed as vectors

$$\mathbf{G}_+(\rho) = \left(G^+_{m_1}(\rho), G^+_{m_2}(\rho), \ldots, G^+_{m_p}(\rho)\right), \tag{30a}$$

$$\mathbf{G}_-(\rho) = \left(G^-_{m_1}(\rho), G^-_{m_2}(\rho), \ldots, G^-_{m_p}(\rho)\right). \tag{30b}$$

Introducing a vector of exponents

$$\mathbf{e}(\phi) = (\exp(im_1\phi), \exp(im_2\phi), \ldots, \exp(im_p\phi)), \tag{31}$$

one can express amplitudes (15) as scalar products

$$E_+(\rho, \phi) = \mathbf{G}_+(\rho) \cdot \mathbf{e}(\phi), \tag{32a}$$

$$E_-(\rho, \phi) = \mathbf{G}_-(\rho) \cdot \mathbf{e}(\phi). \tag{32b}$$

Calculation of a focal field [Eq. (27)] induced by a partial entrance CPVVB [Eq. (16)] at known microcavity mirror parameters and other technical settings is a finely programmable procedure. Thus, to define an entrance field for a computer program, a user needs to set vectors $\mathbf{G}_+(\rho)$ and $\mathbf{G}_-(\rho)$ or (which would be better in practice) vectors $\mathbf{F}_+(\theta_1)$ and $\mathbf{F}_+(\theta_1)$ obtained from the vectors $\mathbf{G}_\pm(\rho)$ through the substitution $\rho = f \sin \theta_1$:

$$\mathbf{F}_+(\theta_1) = \left(F^+_{m_1}(\theta_1), F^+_{m_2}(\theta_1), \ldots, F^+_{m_p}(\theta_1)\right), \tag{33a}$$

$$\mathbf{F}_-(\theta_1) = \left(F^-_{m_1}(\theta_1), F^-_{m_2}(\theta_1), \ldots, F^-_{m_p}(\theta_1)\right). \tag{33b}$$

Below, we will call an entrance field description with such vectors followed by a respective integer-valued set M a single-integral treatable representation.

Finally, one can develop a program script to compute partial focal fields and calculate focusing of any entrance collimated beam with use of such a script, performing the following simple steps.

Step 1. Describe the entrance beam light field in terms of amplitude vectors $\mathbf{F}_+(\theta_1)$ and $\mathbf{F}_-(\theta_1)$ and input these vectors and a set M. A component $F^s_{m_j}(\theta_1)$ of a given entrance field $\mathbf{E}_0(\theta_1, \phi)$ can be calculated in general as

$$F^s_{m_j}(\theta_1) = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{E}_0(\theta_1, \phi) \cdot \mathbf{n}_s \exp(-im_j\phi) d\phi, \tag{34}$$

although normally it is obtained in simpler ways as will be shown in the next section. Note that in calculating a dot product of complex vectors, $\mathbf{E}_0 \cdot \mathbf{n}_s$, one should multiply $\mathbf{E}_0$ by the complex conjugate of the vector $\mathbf{n}_s$, according to complex

vector dot product rules. For example, if we have a vector $\mathbf{E}_0 = (5 + 2i, 1 + i)$ (in Cartesian coordinates) and need to find the dot product $\mathbf{E}_0 \cdot \mathbf{n}_+$, it will be calculated as $\mathbf{E}_0 \cdot \mathbf{n}_+ = 2^{-0.5} \times [(5 + 2i) \times \bar{1} + (1 + i) \times \bar{i}] = 2^{-0.5} \times (6 + i)$.

Step 2. Input microcavity mirror matrices M_1 and M_2 as in Eq. (22) and the following other numeric parameters: the objective angular aperture $\theta_{\max}$, excitation wavelength λ_{ex}, core layer thickness d, core layer left bound z_1, and refractive indices n_1, n, and n_2.

Step 3. Obtain the resultant focal field as

$$\mathbf{E}^f(\rho, \varphi, z) = 0.5ik_1 f e^{ik_1 f} [E_+(\rho, \varphi, z)e^{-i\varphi}\mathbf{n}_+$$
$$+ E_-(\rho, \varphi, z)e^{i\varphi}\mathbf{n}_- + E_z(\rho, \varphi, z)\mathbf{n}_z] \tag{35}$$

with

$$E_j(\rho, \varphi, z) = \left[\mathbf{A}_{j,+}(\rho, z)e^{i(\varphi+\pi/2)} + \mathbf{A}_{j,-}(\rho, z)e^{-i(\varphi+\pi/2)}\right]$$
$$\cdot \mathbf{e}(\varphi + \pi/2), \tag{36}$$

where

$$\mathbf{A}_{j,\pm}(\rho, z) = \left(A^{m_1}_{j,\pm}(\rho, z), A^{m_2}_{j,\pm}(\rho, z), \ldots, A^{m_p}_{j,\pm}(\rho, z)\right) \tag{37}$$

with $A^{m_l}_{j,s}(\rho, z)$ being calculated programmatically according to Eqs. (28), where $j = +, -, z$, $s = +, -$, and $l = [1, \ldots, p]$. Vectors $\mathbf{A}_{j,s}(\rho, z)$ are induced by vectors $\mathbf{F}_s(\theta_1)$ and both of them are p dimensional.

D. The infinite-series field problem

1. Field negligibility criterion

If an entrance beam is represented by a finite superposition of vortex beams, there is no practical issue to define set M in Eq. (29) and entrance vectors $\mathbf{F}_\pm(\theta_1)$ in Eq. (33). If it is represented by an infinite series, one needs to solve the problem of rejection of negligible terms to reduce the series down to a finite sum. Below we derive a simple practical criterion that allows one to reject superfluous terms.

Let us consider a field component E_j expressed by Eq. (36) and its calculation error. We will denote its precise value as E^0_j and practically computed one as E^c_j; the absolute error is equal to $\Delta E_j = |E^c_j - E^0_j|$. In practice, we want to provide a given acceptable error level. It can be defined by a relative error, or by an absolute error, or by an average error, and so on. The major problem is that we cannot find the precise value of the error because the precise value of the field E^0_j is not known *a priori*. The concept of a relative error relies on a comparison of the absolute error and the field value: $\Delta E_j \ll |E_j|$. The concept of an absolute error relies on a comparison of the absolute error with a reference constant field value E_{ref}: $\Delta E_j \ll E_{\mathrm{ref}}$. The first of these concepts poorly works for small fields ($|E_j| \to 0$). In the second one, we encounter the impossibility to set a criterion of the absolute error negligibility because the magnitude of a calculated field cannot be known in advance. Hence, neither of these concepts can be used. The problem might be solved by exploiting the concept of a proper comparison of the calculated focal field value with the entrance field as the former is inherently proportional to the latter. Below we will develop this concept.

We will use the root-mean-square value of an entrance beam light field,

$$A_{en} = \frac{\sqrt{\int_0^{\theta_{\max}} \int_0^{2\pi} |\mathbf{E}_0(\theta_1, \phi)|^2 \sin \theta_1 \cos \theta_1 d\phi d\theta_1}}{\sqrt{\pi} \sin \theta_{\max}}, \quad (38)$$

as a reference field magnitude. To compare the focal field with the reference entrance amplitude A_{en}, we will exploit the energy conservation concept. The cycle-averaged electric field energy density of a field with a frequency ω_0 at a location $\mathbf{r}$ is expressed as

$$w_E(\mathbf{r}, \omega_0) = \frac{\varepsilon_0}{4} \frac{d}{d\omega} \{\omega \text{Re}[\varepsilon(\omega, \mathbf{r})]\} \bigg|_{\omega=\omega_0} |\mathbf{E}(\mathbf{r})|^2, \quad (39)$$

where $\varepsilon(\omega, \mathbf{r})$ is the permittivity [10]. As the field under consideration propagates along the negative z direction, its energy inside an infinitesimal layer of a thickness dz is a constant. In general, the derivative factor in Eq. (39) is different in an objective medium and the core layer, but this difference can barely exceed one or two orders in magnitude and might be temporally neglected. From this, the assessment

$$A_f = \sqrt{\frac{S_{en}}{S_f}} A_{en} \quad (40)$$

can be derived, where A_f is a root mean square of the focal field, and S_{en} and S_f are the entrance beam cross-section area and focal-region target plane effective area, respectively. The former can be expressed as $S_{en} = \pi f^2 \sin^2 \theta_{\max}$ and the latter can be assessed as $S_f = \pi (2\pi\alpha/k)^2$ with α being a real number of the order of magnitude from 1 to about 10. Conceptually, the effective area is concerned with a circle of radius $d = \alpha\lambda$, inside which the most part of the field energy is concentrated and it can be finely applied to tightly focused fields. For example, for focused linearly polarized Gaussian beams the circle diameter normally is about from 0.5λ to 3λ depending on the filling factor [10], for cylindrical vector beams to about λ [32], and for vortex vector beams from 2λ up to 6λ [19,33,34]. In general, an entrance beam may be poorly suitable for tight focusing and the effective radius may be arbitrarily large, but the concept of tight focusing requires the effective circle to be pretty small. So, we can assign the coefficient α to be $\alpha \leqslant 10$. Substituting S_{en} and S_f into Eq. (40), we obtain

$$A_f = kf \frac{\sin \theta_{\max}}{2\pi\alpha} A_{en} \quad (41)$$

and the assessment for α above gives

$$A_f \gtrsim 0.01 kf A_{en}. \quad (42)$$

Additionally, the following aspects should be taken into account. First, we neglected the derivative factor in Eq. (39) and now assume that it is able to decrease focal field up to one order in magnitude. Second, energy losses can occur in layered structures. We will assume that they are able to decrease the field up to two orders in magnitude. Third, one is normally interested not only in maximal and average field values, but also in small values located in dark areas of the focal region. So, we will consider field values being two orders of magnitude smaller than an average one to be

small and smaller values to be negligible. Finally, for a field value to be considered as negligible, it must be $1 + 2 + 2 = 5$ orders smaller than an average one given by Eq. (42). Hence, for a field normalized to $kf A_{en}$ the criterion of negligibility reads as

$$\frac{A_f}{kf A_{en}} < 10^{-7}. \quad (43)$$

For abnormal conditions—such as significant field energy losses, locations far beyond the effective circle, extremely large refractive index of the target layer, and similar—this criterion can be not suitable and should be properly corrected. For normal conditions we recommend to use the threshold of 10^{-7}. This criterion can be readily applied to the field in Eq. (35): because of the factor $k_1 f$ on the right-hand side, its components $E_j(\rho, \varphi, z)$ $(j = +, -, z)$ are already normalized to this factor and the negligibility criterion for them should be reformulated as

$$\frac{\left|E_j^{\text{rest}}(\rho, \varphi, z)\right|}{A_{en}} < 10^{-7}, \quad (44)$$

where a field component is divided into two parts as

$$E_j(\rho, \varphi, z) = E_j^0(\rho, \varphi, z) + E_j^{\text{rest}}(\rho, \varphi, z) \quad (45)$$

with $E_j^0(\rho, \varphi, z)$ being the calculated field value and $E_j^{\text{rest}}(\rho, \varphi, z)$ negligible part.

2. Bessel-function integral series properties

Here, we will explore the behavior of series corresponding to the vectors $\mathbf{A}_{j,\pm}(\rho, z)$ expressed by Eq. (37) to apply the suggested criterion of negligibility to them and reveal their components that can be considered negligible and rejected.

First, series describing entrance field amplitudes (32) must converge and amplitudes $F_{m_j}^s(\theta_1)$ in Eq. (33) must decrease on average, starting from a number m_0^{en} in the positive direction of numbers (from $-\infty$ to ∞) and similarly in the negative direction. In general, the decreasing behavior can be non-monotonic with respect to m_j and dependent on a spatial point defined by the polar angle θ_1. Here, we assume the behavior to be decreasing and later will explore this problem of an entrance coefficient. Below, we will investigate only a positive direction as the series behavior is similar in both positive and negative directions. We will represent focal amplitudes (28) in a generalized form

$$A_{j,\pm}^m(\rho, z) = \int_0^{\theta_{\max}} \sin \theta_1 \cos \theta_1 F_m^{\pm}(\theta_1)$$

$$\times J_{l(m)}(k_1 \rho \sin \theta_1) K_j(\theta_1, z) d\theta_1 \quad (46)$$

with $j = +, -, z$, $K_+(\theta_1, z) = [T_\rho(\theta_1, z) \cos \theta \pm T_\varphi(\theta_1, z)]$ $/\sqrt{\cos \theta_1}$, $K_-(\theta_1, z) = [T_\rho(\theta_1, z) \cos \theta \mp T_\varphi(\theta_1, z)]/\sqrt{\cos \theta_1}$, $K_z(\theta_1, z) = -\sqrt{2} T_z(\theta_1, z)/\sqrt{\cos \theta_1}$, and $l(m) = m \pm 1 - 1$ for a positive focal component, $l(m) = m \pm 1 + 1$ for a negative one, and $l(m) = m \pm 1$ for a longitudinal one. The behavior of such a function depends on an entrance function $F_m^{\pm}(\theta_1)$ and Bessel function $J_{l(m)}(k_1 \rho \sin \theta_1)$. From the definition of the index m_0^{en} given above, the entrance field terms $F_m^{\pm}(\theta_1)$ do not decrease in the range $m = [0, \dots, m_0^{en}]$. Hence, we cannot guarantee their negligibility and reject this

range. But at $m > m_0^{en}$ the entrance function does not prevent focal series convergence and there exists a probability that terms starting from $m_0 > m_0^{en}$ can be omitted. It is defined by properties and behavior of the Bessel functions and we need to explore them.

The Bessel function of the first kind, $J_n(x)$, of an integer order n obeys Bessel's differential equation

$$x^2 \frac{d^2 J_n(x)}{dx^2} + x \frac{dJ_n(x)}{dx} + (x^2 - n^2) J_n(x) = 0 \qquad (47)$$

and, from its definition, it exhibits the following two properties:

$$J_{-n}(x) = (-1)^n J_n(x) \qquad (48)$$

and

$$J_n(x) > 0, \frac{dJ_n(x)}{dx} > 0, \frac{d^2 J_n(x)}{dx^2} > 0 \bigg|_{x \to 0+}. \qquad (49)$$

The former guarantees that the absolute value of a function behaves similarly at positive and negative n's, which allows us to extrapolate results for a positive-valued series branch to a negative-valued one; the latter states that, at small positive values of an argument x, the Bessel function and its first and second derivatives are positive. Next, we will explore the function's behavior in the range $x = [0, \ldots, n]$ at a positive n. All over this range, coefficients x^2 and x in Eq. (47) are positive and the coefficient $n^2 - x^2$ is negative. Next, it follows from Eq. (49) that, at small positive x values, the Bessel function and its first derivative are not only positive, but also increasing functions. Hence, the first and second terms in Eq. (47) are positive and the third one is negative. The function $J_n(x)$ can reach a local maximum and start decreasing if and only if its first derivative becomes zero and then negative. Hence, the first derivative must start decreasing preliminarily, which is possible if and only if the second derivative becomes negative. So, at $x = 0$, $J_n(x)$ and both the derivatives are equal to zero and increasing. At any $x = \epsilon > 0$ with ϵ being an infinitesimal value, the function is positive and increasing and, at least, the first derivative is positive and increasing too. Thus, the second derivative will start decreasing and take on negative values first. At a negative second derivative and positive first one and function itself, the first and the third terms of Eq. (47) will be negative and the second one positive. It makes the zero balance possible in principle and allows us to permit the second derivative to reach a zero value at any $x = x_0$ in the range $x = [0, \ldots, n]$ and then take on negative values. After that, the first derivative can reach its zero value and go to the negative range, but (i) $J_n(x)$ cannot become negative at the same x value as its first derivative and (ii) the second derivative cannot become positive instantly after the first one reaches its zero and goes to the negative range. It means that if the first derivative becomes negative, a range $x = [x_1, \ldots, x_2]$ where $J_n(x)$ is positive and both its derivatives are negative must exist inside the considered range $x = [0, \ldots, n]$. In such a range all three terms in Eq. (47) will be negative, which makes the zero balance impossible and brings us to the conclusion that in the range $x = [0, \ldots, n]$ the Bessel function $J_n(x)$ and its

first derivative cannot take on negative values:

$$J_n(x) > 0, \frac{dJ_n(x)}{dx} > 0 \bigg|_{x=[0,\ldots,n]}. \qquad (50)$$

Next, Bessel functions obey the following universal recursive relation:

$$J_n(x) = \frac{n+1}{x} J_{n+1}(x) + \frac{dJ_{n+1}(x)}{dx}. \qquad (51)$$

From this and Eq. (50), it directly follows that in the range $x = [0, \ldots, n]$ an inequality

$$J_{n+p+1}(x) < \frac{n}{n+p+1} J_{n+p}(x) \qquad (52)$$

for $p = [0, 1, 2, \ldots, \infty]$ is valid. Applying it to Eq. (46), we conclude that the Bessel function decreases with an increasing number at $m > m_0^{bf}$, where m_0^{bf} is defined from the equation

$$l(m_0^{bf}) = [k\rho \sin \theta_{max}] + 1 \qquad (53)$$

and $[\alpha]$ denotes the integer part of a real number α. So, starting from $m_0 = \sup\{m_0^{bf}, m_0^{en}\}$ (m_0^{en} was defined above as a number from which the entrance coefficient starts decrease), both functions $F_m^{\pm}(\theta_1)$ and $J_{l(m)}(k\rho \sin \theta_1)$ decrease with an increasing number.

A residual part of a partial amplitude in Eq. (36) takes the form

$$E_{j,\text{rest}}^{\pm}(\rho, \varphi, z) = e^{\pm(\varphi + \pi/2)} \sum_{k=m_0+p+1}^{\infty} A_{j,\pm}^k(\rho, z) e^{ik(\varphi + \pi/2)} \qquad (54)$$

and we need to satisfy the condition of its negligibility,

$$|E_{j,\text{rest}}^{\pm}(\rho, \varphi, z)| < \varepsilon, \qquad (55)$$

where ε is an acceptable error level. An assessment of the residual field gives

$$|E_{j,\text{rest}}^{\pm}(\rho, \varphi, z)| \leqslant \sum_{k=m_0+p+1}^{\infty} |A_{j,\pm}^k(\rho, z)| \qquad (56)$$

and now terms $|A_{j,\pm}^k(\rho, z)|$ should be explored. Exploiting Eq. (46), we derive the assessment

$$|A_{j,\pm}^k(\rho, z)| \leqslant \int_0^{\theta_{max}} \sin \theta_1 \cos \theta_1 |F_k^{\pm}(\theta_1)|$$
$$\times J_{l(k)}(k_1 \rho \sin \theta_1) |K_j(\theta_1, z)| d\theta_1. \qquad (57)$$

In the case of a monotonically decreasing entrance coefficient with increasing index k, the coefficient's maximal value is readily assessed as

$$|F_k^{\pm}(\theta_1)| \leqslant |F_{m_0+p}^{\pm}(\theta_1)| \qquad (58)$$

and this value can be used in Eq. (57) as an upper bound. If the decrease is nonmonotonic or θ_1 dependent, such an

assessment is no longer valid and we have to revise it. To do this, one should first break the integral in Eq. (57) into two multipliers as

$$
\left| A_{j,\pm}^k(\rho, z) \right| \leqslant \int_0^{\theta_{\max}} \sin\theta_1 \cos\theta_1 |F_k^{\pm}(\theta_1)| d\theta_1
$$
$$
\times \int_0^{\theta_{\max}} J_{l(k)}(k_1 \rho \sin\theta_1)|K_j(\theta_1, z)| d\theta_1, \quad (59)
$$

and then find the upper bound of the first multiplier. To this end, one has to analyze the initial entrance field $\mathbf{E}_0(\theta_1, \phi)$ and exploit energy conservation. It cannot directly provide an assessment of a function $|F_k^{\pm}(\theta_1)|$ but is able to give some information about $|F_k^{\pm}(\theta_1)|^2$. So, we should first express the first multiplier in Eq. (59) in terms of the integral of $|F_k^{\pm}(\theta_1)|^2$. Such a relation can be obtained from the equation

$$
\int_0^{\theta_{\max}} \sin\theta_1 \cos\theta_1 [|F_k^{\pm}(\theta_1)| - \alpha]^2 d\theta_1 \geqslant 0 \quad (60)
$$

with α being a positive real number. Expressing the integral of $|F_k^{\pm}(\theta_1)|$ in terms of the integral of $|F_k^{\pm}(\theta_1)|^2$, α, and $\theta_{\max}$ and then minimizing the upper bound with respect to α, one obtains

$$
\int_0^{\theta_{\max}} \sin\theta_1 \cos\theta_1 |F_k^{\pm}(\theta_1)| d\theta_1
$$
$$
\leqslant 2^{-0.5} \sin\theta_{\max} \sqrt{\int_0^{\theta_{\max}} \sin\theta_1 \cos\theta_1 |F_k^{\pm}(\theta_1)|^2 d\theta_1}. \quad (61)
$$

Exploiting entrance field series decomposition, we directly obtain from Eq. (38)

$$
\sum_{m=-\infty}^{\infty} \int_0^{\theta_{\max}} |F_m^+(\theta_1)|^2 \sin\theta_1 \cos\theta_1 d\theta_1
$$
$$
+ \sum_{m=-\infty}^{\infty} \int_0^{\theta_{\max}} |F_m^-(\theta_1)|^2 \sin\theta_1 \cos\theta_1 d\theta_1
$$
$$
= 0.5 A_{en}^2 \sin^2 \theta_{\max}. \quad (62)
$$

Introducing a quantity

$$
\sum_{m=(m_0+p)_{neg}}^{(m_0+p)_{pos}} \int_0^{\theta_{\max}} |F_m^+(\theta_1)|^2 \sin\theta_1 \cos\theta_1 d\theta_1
$$
$$
+ \sum_{m=(m_0+p)_{neg}}^{(m_0+p)_{pos}} \int_0^{\theta_{\max}} |F_m^-(\theta_1)|^2 \sin\theta_1 \cos\theta_1 d\theta_1 = F_{en}^2,
$$
$$
\quad (63)
$$

where all (presumably) non-negligible field terms are summarized in both negative and positive directions and indices $(m_0 + p)_{pos}$ and $(m_0 + p)_{neg}$ are defined as in Eq. (54), one derives

$$
\int_0^{\theta_{\max}} \sin\theta_1 \cos\theta_1 |F_k^{\pm}(\theta_1)|^2 d\theta_1 \leqslant 0.5 A_{en}^2 \sin^2 \theta_{\max} - F_{en}^2
$$
$$
\quad (64)
$$

for $k > m_0 + p$ and the synthesis of Eq. (64) with Eq. (61) gives

$$
\int_0^{\theta_{\max}} \sin\theta_1 \cos\theta_1 |F_k^{\pm}(\theta_1)| d\theta_1
$$
$$
\leqslant \frac{\sin\theta_{\max}}{2} \sqrt{A_{en}^2 \sin^2 \theta_{\max} - 2F_{en}^2}. \quad (65)
$$

Now, combining Eq. (65) with Eq. (58) and introducing them into Eq. (57), we can write

$$
|A_{j,\pm}^k(\rho, z)| \leqslant \int_0^{\theta_{\max}} U_{\pm}(\theta_1) J_{l(k)}(k_1 \rho \sin\theta_1)|K_j(\theta_1, z)| d\theta_1,
$$
$$
\quad (66)
$$

where $U_{\pm}(\theta_1) = \sin\theta_1 \cos\theta_1 |F_{m_0+p}^{\pm}(\theta_1)|$ for monotonically decreasing entrance coefficients and $U_{\pm}(\theta_1) = 0.5 \sin\theta_{\max} [A_{en}^2 \sin^2 \theta_{\max} - 2F_{en}^2]^{0.5}$ for coefficients with nonstable behavior. Obviously, the second form of the U function is suitable for monotonically decreasing coefficients too but in that case it leads to overestimation of the upper bound and is not preferable. Exploiting Eq. (66) and the Bessel function properties derived above, we can further investigate the residual field absolute value given by Eq. (56). From inequality (52), we have

$$
\sum_{k=l(m_0)+p+1}^{\infty} J_k(x) < \left(1 + \frac{l(m_0)}{p+1}\right) J_{l(m_0)+p}(x) \quad (67)
$$

and

$$
|E_{j,\text{rest}}^{\pm}(\rho, \varphi, z)| < \left(1 + \frac{l(m_0)}{p+1}\right) \int_0^{\theta_{\max}} |K_j(\theta_1, z)| U_{\pm}(\theta_1)
$$
$$
\times J_{l(m_0)+p}(k_1 \rho \sin\theta_1) d\theta_1. \quad (68)
$$

Thus, to define index p and set entrance vectors $\mathbf{E}_{\pm}$ given by Eq. (33), we should assign p the least value at which the condition

$$
\left[1 + \frac{l(m_0)}{p+1}\right] \int_0^{\theta_{\max}} |K_j(\theta_1, z)| U_{\pm}(\theta_1)
$$
$$
\times J_{l(m_0)+p}(k_1 \rho \sin\theta_1) d\theta_1 < \varepsilon \quad (69)
$$

is satisfied. Applying criterion (44), we obtain the final equation for index p,

$$
\left[1 + \frac{l(m_0)}{p+1}\right] \int_0^{\theta_{\max}} \frac{|K_j(\theta_1, z)| U_{\pm}(\theta_1)}{A_{en}}
$$
$$
\times J_{l(m_0)+p}(k\rho \sin\theta) d\theta_1 < 10^{-7}, \quad (70)
$$

and the positive branch of the series will be described by a set $M_{pos} = \{0, 1, 2, \ldots, m_0 + p\}$. The negative branch is obtained in a similar way.

IV. PRACTICAL TESTING OF THE SINGLE-INTEGRAL ALGORITHM

In this section, we will consider some examples of practical implementation of the suggested algorithm. First, several

widely used azimuth-continued entrance laser fields will be considered and represented in the single-integral treatable form. Second, a fractional vortex beam will be considered and analyzed numerically as an example of an azimuth-discontinued beam.

A. Finite-series fields: Examples of single-integral treatable representations

Let us consider a linearly polarized beam with a vector amplitude $\mathbf{E}_0(\theta_1, \phi) = P(\theta_1)\mathbf{n}_x$, where $P(\theta_1)$ is a scalar amplitude. Depending on the scalar amplitude, it can be a true plane wave, Gaussian beam, Hermite-Gaussian mode, Laguerre-Gaussian mode, or any other. For example, for a Gaussian paraxial beam, we have the amplitude $P(\theta_1) = C \exp[-\sin^2\theta_1/(f_0 \sin^2\theta_{\max})]$, where C is a constant, f_0 the filling factor, and $\theta_{\max}$ the objective angular aperture [10]. To represent such a field in a single-integral treatable form, we need to express the vector $\mathbf{n}_x$ in terms of circular polarization vectors. From Eq. (8), one obtains $\mathbf{n}_x = 2^{-0.5}(\mathbf{n}_+ + \mathbf{n}_-)$ and we can write the respective single-integral treatable representation as $M = \{0\}$, $\mathbf{F}_+(\theta_1) = (2^{-0.5}P(\theta_1))$, and $\mathbf{F}_-(\theta_1) = (2^{-0.5}P(\theta_1))$. Here, M is a single-component set and $\mathbf{F}_\pm(\theta_1)$ are single-component vectors.

The next entrance beam that we will analyze is an elliptically polarized cylindrical vector Bessel beam [25,35]. Mathematically it is expressed as $\mathbf{E}_0(\theta_1, \phi) = J_1(\beta \sin\theta_1)(e^{i\Phi}\cos\alpha\,\mathbf{n}_\rho + e^{-i\Phi}\sin\alpha\,\mathbf{n}_\varphi)$ with β being a constant, α and Φ ellipticity parameters, and $J_1(x)$ the Bessel function of the first kind of a variable x. Again, we need to represent it in terms of circular polarization vectors. Exploiting Eqs. (7), we obtain $M = \{-1, 1\}$, $\mathbf{F}_+(\theta_1) = (2^{-0.5}J_1(\beta\theta_1)(e^{i\Phi}\cos\alpha - ie^{-i\Phi}\sin\alpha), 0)$, and $\mathbf{F}_-(\theta_1) = (0, 2^{-0.5}J_1(\beta\theta_1)(e^{i\Phi}\cos\alpha + ie^{-i\Phi}\sin\alpha))$.

In Ref. [36], tight focusing of laser beams with hybrid circular-azimuthal polarization was theoretically studied. Mathematically, such a light field is expressed as $\mathbf{E}_0(\theta_1, \phi) = A(\theta_1)[-(i\sin m\phi)\mathbf{n}_x + (\cos m\phi)\mathbf{n}_y]$, where $A(\theta_1)$ is a scalar amplitude and m an integer constant. Exploiting Euler's formula to express the trigonometric functions in terms of complex exponents, we obtain $M = \{m, -m\}$, $\mathbf{F}_+(\theta_1) = -0.5iA(\theta_1)(e^{-i\pi/4}, e^{i\pi/4})$, and $\mathbf{F}_-(\theta_1) = 0.5iA(\theta_1)(e^{i\pi/4}, e^{-i\pi/4})$.

B. Logarithmlike series entrance field: Common investigation

Here we will consider an entrance fractional vortex beam. An extensive review of fractional vortex beams was performed, for example, in Ref. [37]. Such a beam can be described by any scalar amplitude but the simplest possible one is a fractional Bessel beam,

$$\mathbf{E}_0(\theta_1, \phi) = J_\alpha(2\sin\theta_1)e^{i\alpha\phi}\mathbf{n}_+, \tag{71}$$

with $J_\alpha(x)$ being a Bessel function of the first kind and α a noninteger positive number. Its scalar amplitude is decomposed as [38]

$$E_0^+(\theta_1, \phi) = \frac{\exp(i\pi\alpha)\sin(\pi\alpha)}{\pi}J_\alpha(2\sin\theta_1)\sum_{m=-\infty}^{\infty}\frac{e^{im\phi}}{\alpha - m} \tag{72}$$

and $M = \{-\infty, \ldots, -1, 0, 1, \ldots, \infty\}$, $\mathbf{F}_+(\theta_1) = [\exp(i\pi\alpha)\sin(\pi\alpha)/\pi]J_\alpha(2\sin\theta_1)(\ldots, \frac{1}{\alpha+1}, \frac{1}{\alpha}, \frac{1}{\alpha-1}, \ldots)$ with $F_m^+(\theta_1) = [\exp(i\pi\alpha)\sin(\pi\alpha)/\pi]J_\alpha(2\sin\theta_1)/(\alpha - m)$.

We will investigate the focusing of a beam with $\alpha = 0.4$ by an objective with an angular aperture $\theta_{\max} = 64°$ for the following two cases: (i) free space with the refractive index $n = 1.5$ and an excitation beam with the wavelength $\lambda = 0.5$ μm and (ii) a planar microcavity used in Refs. [39,40] with mirrors described by the matrices

$$M_1 = \begin{pmatrix} 3\text{ nm} & 30\text{ nm} & 5\text{ nm} & 20\text{ nm} \\ 1.47 & 1.5159 + 1.8844i & 0.10433 + 2.5279i & 2.2288 + 4.3757i \end{pmatrix}, \tag{73a}$$

$$M_2 = \begin{pmatrix} 3\text{ nm} & 60\text{ nm} & 5\text{ nm} & 20\text{ nm} \\ 1.47 & 1.5159 + 1.8844i & 0.10433 + 2.5279i & 2.2288 + 4.3757i \end{pmatrix}, \tag{73b}$$

the laser excitation wavelength $\lambda = 0.44$ μm, core layer refractive index $n = 1.34$ (water) and thickness $d = 1$ μm, surrounding media refractive indices $n_1 = 1.526$ and $n_2 = 1.526$ (glass), and core layer edge position $z_1 = 0.3$ μm. The mirrors described by Eqs. (73) are assumed to consist of the following four layers: silicon dioxide (SiO$_2$), gold, silver, and chromium. To calculate partial fields (28) normalized to the average entrance field amplitude (38), we developed a script in PYTHON 3.9. Another script was developed to compute the normalized focal field according to the double-integral procedure for comparison.

First, we analyzed the behavior of the residual field described by the left-hand side of Eq. (69) in free space. The respective results at the distance coordinate $\rho = 3\lambda = 1.5$ μm and longitudinal coordinate $z = 0$ are displayed in Table I for

the positive-valued branch of the index m. We need first to define the critical index m_0. As the entrance field is described by a monotonically decreasing series (72), the entrance-function decrease beginning index m_0^{en} does not play any role. It means that $m_0 = m_0^{bf}$ and the critical index should be calculated from Eq. (53) as

$$l(m_0) = \left[\frac{2\pi n_1}{\lambda}3\lambda\sin 64°\right] + 1 = 26. \tag{74}$$

The index $l(m_0)$ is expressed in terms of m_0 as $l = m_0$ for the A_+ component, $l = m_0 + 2$ for the A_- component, and $l = m_0 + 1$ for the A_z component. Hence, $m_0 = 26$ for the plus component, 24 for the minus one, and 25 for the longitudinal one. As $m = 26$ is the biggest number among these

TABLE I. The positive-branch residual field as a function of a series number in free space. The parameters are assumed to be as follows: $\lambda = 0.5$ μm, $n = 1.5$, $z = 0$, and $\rho = 3\lambda$.

m	$\max(|E_+|_{res})$	$\max(|E_-|_{res})$	$\max(|E_z|_{res})$
26	4.41×10^{-03}	4.57×10^{-04}	2.04×10^{-03}
27	1.29×10^{-03}	1.21×10^{-04}	5.69×10^{-04}
28	4.75×10^{-04}	4.08×10^{-05}	2.00×10^{-04}
29	1.87×10^{-04}	1.47×10^{-05}	7.52×10^{-05}
30	7.45×10^{-05}	5.40×10^{-06}	2.88×10^{-05}
31	2.96×10^{-05}	1.98×10^{-06}	1.10×10^{-05}
32	1.16×10^{-05}	7.18×10^{-07}	4.13×10^{-06}
33	4.42×10^{-06}	2.56×10^{-07}	1.52×10^{-06}
34	1.65×10^{-06}	8.94×10^{-08}	5.50×10^{-07}
35	6.02×10^{-07}	3.05×10^{-08}	1.94×10^{-07}
36	2.14×10^{-07}	1.02×10^{-08}	6.68×10^{-08}
37	7.39×10^{-08}	3.31×10^{-09}	2.24×10^{-08}
38	2.49×10^{-08}	1.05×10^{-09}	7.34×10^{-09}
39	8.19×10^{-09}	3.27×10^{-10}	2.34×10^{-09}
40	2.63×10^{-09}	9.93×10^{-11}	7.30×10^{-10}
41	8.20×10^{-10}	2.94×10^{-11}	2.22×10^{-10}
42	2.50×10^{-10}	8.53×10^{-12}	6.60×10^{-11}
43	7.45×10^{-11}	2.41×10^{-12}	1.92×10^{-11}
44	2.17×10^{-11}	6.69×10^{-13}	5.44×10^{-12}
45	6.15×10^{-12}	1.81×10^{-13}	1.51×10^{-12}
46	1.71×10^{-12}	4.81×10^{-14}	4.09×10^{-13}

indices, we start from it (one can start after but not before the critical index). Table II displays similar results for the microcavity described above. We can see that, according to the negligibility criterion (70), terms starting from $m = 37$ in free

TABLE II. The positive-branch residual field as a function of a series number in a metal-dielectric planar microcavity. The cavity parameters are described in the text; coordinates are $\rho = 3\lambda$, $z = 0$.

m	$\max(|E_+|_{res})$	$\max(|E_-|_{res})$	$\max(|E_z|_{res})$
26	3.84×10^{-04}	1.24×10^{-04}	4.24×10^{-04}
27	1.12×10^{-04}	3.28×10^{-05}	1.19×10^{-04}
28	4.11×10^{-05}	1.10×10^{-05}	4.24×10^{-05}
29	1.61×10^{-05}	3.97×10^{-06}	1.61×10^{-05}
30	6.42×10^{-06}	1.46×10^{-06}	6.20×10^{-06}
31	2.54×10^{-06}	5.38×10^{-07}	2.38×10^{-06}
32	9.90×10^{-07}	1.95×10^{-07}	9.04×10^{-07}
33	3.78×10^{-07}	6.98×10^{-08}	3.36×10^{-07}
34	1.41×10^{-07}	2.44×10^{-08}	1.22×10^{-07}
35	5.13×10^{-08}	8.33×10^{-09}	4.34×10^{-08}
36	1.82×10^{-08}	2.79×10^{-09}	1.50×10^{-08}
37	6.27×10^{-09}	7.58×10^{-10}	5.07×10^{-09}
38	2.10×10^{-09}	3.75×10^{-10}	1.67×10^{-09}
39	5.56×10^{-10}	1.15×10^{-10}	4.62×10^{-10}
40	2.85×10^{-10}	3.48×10^{-11}	1.44×10^{-10}
41	8.81×10^{-11}	1.02×10^{-11}	4.41×10^{-11}
42	2.66×10^{-11}	2.94×10^{-12}	1.32×10^{-11}
43	7.84×10^{-12}	8.27×10^{-13}	3.84×10^{-12}
44	2.25×10^{-12}	2.27×10^{-13}	1.09×10^{-12}
45	6.34×10^{-13}	6.12×10^{-14}	3.05×10^{-13}
46	1.74×10^{-13}	1.61×10^{-14}	8.34×10^{-14}

space and from $m = 35$ in the microcavity can be rejected. If one needs to calculate the field with a higher accuracy, more terms should be taken and we can see that the series converges rapidly: to increase the accuracy by one order, two additional terms are normally required. It has to be emphasized that the entrance logarithmiclike series in Eq. (72) is slowly and conditionally convergent: its convergence is the slowest possible. From this, we conclude that the focusing process transforms any entrance series into a rapidly convergent one and the focal field can be calculated using a pretty small number of terms: from several terms up to several tens of them. The only possible exception to this rule is an entrance series with rapidly increasing coefficients in a wide index range. But such a beam would be an exceptional—although theoretically possible—example and it is barely used in practice widely.

In the next step, we calculated the field inside the microcavity at several spatial points, using the single-integral series algorithm and the conventional double-integral algorithm to compare field values and computational times. The respective results are presented in Table III. The first column displays the distance coordinate measured in entrance beam wavelengths. The second one presents the set of terms M and calculated entrance-normalized field values as a vector $\mathbf{E} = (E_+, E_-, E_z)$. The longitudinal coordinate is $z = 0$ and the azimuthal angle $\varphi = 0$. The third and fourth columns display computational times for the double- and single-integral algorithms, respectively. The last column shows the difference vector $\delta\mathbf{E} = (\delta E_+, \delta E_-, \delta E_z)$ with

$$\delta E_j = \left| E_j^{(1)} - E_j^{(2)} \right|, \tag{75}$$

where $j = +, -, z$ and $E_j^{(1)}$ and $E_j^{(2)}$ are field values calculated using the single- or double-integral algorithm, respectively. The former was accomplished through the SciPy function *integrate.quad* with the absolute error (*epsabs*) set to be equal to 10^{-9} and the relative error (*epsrel*) 10^{-5}, the latter through *integrate.dblquad* with *epsabs* = 10^{-7} and *epsrel* = 10^{-4}; other parameters of both functions took on default values. Field values calculated by means of the double-integral procedure will be used as reference values to assess the error of the single-integral calculation. We can state that the error is less than both the absolute threshold 10^{-7} and the product of the relative error (*epsrel* = 10^{-4}) with a field absolute value ($|E_j| \geqslant 10^{-4}$) *epsrel*$|E_j| \geqslant 10^{-8}$. It demonstrates that the single-integral algorithm is able to reproduce field values obtained by the direct Richards-Wolf integration with unlimited completely controllable accuracy. The comparison of times t_1 and t_2 for each field component shows that the single-integral algorithm normally computes the field several times faster than the double-integral one. To compare these times more precisely, we collected them in Table IV. Table IV shows the times required to calculate the total three-component field. For example, Table III shows that, to compute the field through the double-integral procedure at $\rho = 0$, one needs 12 s for E_+, 11 s for E_-, and 5 s for E_z. Table IV collects it in the first row of the second column as $12 + 11 + 5 = 28$ s to calculate the total field. The fourth column shows that double-integral time is 17.5 times greater than single-integral one at $\rho = 0$, and about 5 times at ρ from 0.5λ to 3λ. Averaged over this

TABLE III. A normalized light field inside the metal-dielectric microcavity. The cavity parameters are described in the text, the distance coordinate is varied (column 1), $z = 0$, and times t_1 and t_2 correspond to the single- and double-integral algorithms, respectively; $\delta\mathbf{E}$ is the difference between field values calculated using the single- and double-integral algorithms.

ρ/λ	$M, \mathbf{E}$	t_2 (s)	t_1 (s)	$\delta\mathbf{E}$
0	$\{-2, -1, 0\}, \begin{pmatrix} 0.011 - 0.034i \\ (1.5 + 0.8i) \times 10^{-3} \\ (-4.9 + 6.2i) \times 10^{-3} \end{pmatrix}$	$\begin{pmatrix} 12 \\ 11 \\ 5 \end{pmatrix}$	$\begin{pmatrix} 0.7 \\ 0.6 \\ 0.3 \end{pmatrix}$	$\begin{pmatrix} 3 \times 10^{-9} \\ 5 \times 10^{-10} \\ 2 \times 10^{-13} \end{pmatrix}$
0.5	$\{-13, \ldots, 11\}, \begin{pmatrix} (-3.8 - 3.0i) \times 10^{-3} \\ (-6.1 - 0.5i) \times 10^{-3} \\ (-8.9 - 2.2i) \times 10^{-3} \end{pmatrix}$	$\begin{pmatrix} 83 \\ 58 \\ 27 \end{pmatrix}$	$\begin{pmatrix} 13 \\ 13 \\ 5 \end{pmatrix}$	$\begin{pmatrix} 1 \times 10^{-9} \\ 8 \times 10^{-10} \\ 8 \times 10^{-11} \end{pmatrix}$
1	$\{-18, \ldots, 17\}, \begin{pmatrix} (-2.8 - 7.6i) \times 10^{-4} \\ (1.7 + 0.4i) \times 10^{-3} \\ (7.8 - 7.5i) \times 10^{-4} \end{pmatrix}$	$\begin{pmatrix} 84 \\ 81 \\ 33 \end{pmatrix}$	$\begin{pmatrix} 19 \\ 19 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 4 \times 10^{-10} \\ 2 \times 10^{-9} \\ 1 \times 10^{-9} \end{pmatrix}$
2	$\{-27, \ldots, 27\}, \begin{pmatrix} (-5.5 - 0.1i) \times 10^{-4} \\ (-2.2 - 3.5i) \times 10^{-4} \\ (-1.4 + 0.3i) \times 10^{-3} \end{pmatrix}$	$\begin{pmatrix} 179 \\ 173 \\ 72 \end{pmatrix}$	$\begin{pmatrix} 29 \\ 29 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 9 \times 10^{-11} \\ 4 \times 10^{-9} \\ 4 \times 10^{-9} \end{pmatrix}$
3	$\{-36, \ldots, 37\}, \begin{pmatrix} (-1.0 + 8.8i) \times 10^{-4} \\ (-9.8 - 1.9i) \times 10^{-4} \\ (8.8 + 6.7i) \times 10^{-4} \end{pmatrix}$	$\begin{pmatrix} 202 \\ 197 \\ 90 \end{pmatrix}$	$\begin{pmatrix} 40 \\ 39 \\ 16 \end{pmatrix}$	$\begin{pmatrix} 5 \times 10^{-10} \\ 5 \times 10^{-9} \\ 5 \times 10^{-9} \end{pmatrix}$

coordinate range ($\rho = 0, \ldots, 3\lambda$) times t_2 and t_1 are $\overline{t_2} = 303$ s and $\overline{t_1} = 56$ s and their relation is $\overline{t_2}/\overline{t_1} = 5.4$.

V. COMPUTATION OF FOCAL FIELDS INSIDE FINITE SPATIAL REGIONS

The procedure of computing a focal field in a given spatial region (for example, in a focal plane) is routinely exploited today and can be considered one of the most valuable tasks, where focal field calculation is accomplished. In this section calculations of focal fields in finite spatial regions will be studied and the single-integral algorithm will be adapted for this task.

A. Scan of a focal-region field: General remarks

The majority of practical applications use the following two types of focal-region scan: (i) a three-dimensional scan of a cubic region around the geometrical focus and (ii) a two-dimensional scan of a square in a focal plane. In some cases (but not often), a square scan of planes containing the z axis is used. Hence, the most important theoretical tasks are calculation of a field (i) inside a cubic region and (ii) inside a square in a focal plane.

TABLE IV. Computational times required for a total field, extracted from Table III.

ρ/λ	t_2 (s)	t_1 (s)	t_2/t_1
0	28	1.6	17.5
0.5	168	31	5.4
1	198	45	4.4
2	424	70	6.1
3	489	95	5.1

As one can see from Eq. (36), focal field components $E_j(\rho, \varphi, z)$ in the single-integral algorithm are proportional to the dot products $\mathbf{A}_{j,\pm}(\rho, z) \cdot \mathbf{e}(\varphi + \pi/2)$. As the dependence on the distance coordinates (ρ, z) and the azimuth φ is factorized, we are able (potentially) to calculate the field in one semiplane of constant azimuth and then spread it over the three-dimensional cylinder obtained by rotation of this plane around the z axis. Mathematically, it corresponds to calculation of the vectors $\mathbf{A}_{j,\pm}(\rho, z)$ in a given coordinate range $[z = z_1, \ldots, z_2, \rho = 0, \ldots, \rho_0]$ and multiplying them by the exponential vector $\mathbf{e}(\varphi + \pi/2)$ calculated in the range $\varphi = 0, \ldots, 2\pi$. Let us assume the numbers of steps to be N_z, N_ρ, and N_φ for z, ρ, and φ coordinates, respectively. As the computational time required to calculate the exponential vector and dot products is negligible, the computational complexity of the single-integral algorithm will be proportional to $N_z N_\rho$. The double-integral algorithm does not possess the advantage of the factorized coordinate dependence, which makes its computational cost proportional to $N_z N_\rho N_\varphi$. It may allow us to reduce computational time by orders of magnitude for three-dimensional and focal-plane field calculations.

B. Cylindrical-coordinate scanning

In physical experiments involving laser-scanning microscopy techniques, a focal plane is normally scanned line by line along x and y directions (xy-scan approach). A region of interest (ROI) is a square and the scanning step is fixed and equal for x and y directions. Three-dimensional regions are scanned plane by plane along the z direction with the xy scan of each plane. Following this approach, the rectangular scan is normally used in theoretical calculations of focal fields. In some situations this type of scan can be the only acceptable one, for example, when we need to compare a calculated field with that experimentally recorded on a given

rectangular mesh. But normally one needs to visualize the behavior of a field inside a given spatial region, whereas mesh structure does not affect this behavior. Hence, the scan can be performed in any coordinates, in particular cylindrical.

Let us consider the two-dimensional focal-plane xy scan. We have a square with side length L. The coordinates x and y run over the range $[-0.5L, \ldots, 0.5L]$ with the scanning steps $\Delta x = \Delta y = L/N$, where N is the number of pixels in the x and y directions. Such a square contains $N \times N = N^2$ pixels and can be covered by a circle of radius $L/\sqrt{2}$. It can be scanned in cylindrical coordinates with an equivalent quality circle by circle with circle radii $\rho_l = l\Delta\rho$, where $\Delta\rho = \Delta x = \Delta y = L/N$ is the step between circles, and $6\rho_l/\Delta\rho = 6l$ points on a circle; $l = 0, \ldots, M$ with $M = (L/\sqrt{2})/\Delta\rho \rightarrow [N/\sqrt{2}]$ being the number of circles, where $[\cdots]$ denotes the integer part. Such a scan allows one to keep distances from a given point to each of its four closest neighbor points equal to $\Delta\rho = \Delta x = \Delta y$. The total number of points inside the circle amounts to $3M(M+1)$ and inside the target square it is assessed as $3M(M+1)L^2/[\pi(L/\sqrt{2})^2] \approx (3/\pi)N^2 \approx 0.95N^2$. It is only 5% less than that for the xy scan and can be considered acceptable. Total computational times will be $T_2 = (3/\pi)N^2\overline{t_2} \approx 0.95N^2\overline{t_2}$ for the double-integral algorithm and $T_1 = M\overline{t_1} \approx N\overline{t_1}/\sqrt{2}$ for the single-integral one, where $\overline{t_2}$ and $\overline{t_1}$ are the respective average single-value calculation times. The computational speed enhancement will be

$$\frac{T_2}{T_1} \approx 1.35N\frac{\overline{t_2}}{\overline{t_1}}. \quad (76)$$

For the three-dimensional scan we have the same speed enhancement as three-dimensional regions are scanned plane by plane along the z direction. For the two-dimensional scan of a plane containing the z axis the enhancement is equal to $2\overline{t_2}/\overline{t_1}$.

For the fractional vortex entrance beam (71) we obtained previously the times $\overline{t_2} = 303$ s and $\overline{t_1} = 56$ s with $\overline{t_2}/\overline{t_1} = 5.4$. Substituting them into Eq. (76), one obtains

$$T_2^{\log}/T_1^{\log} \approx 7N. \quad (77)$$

We denoted the total times as $T_2^{\log}$ and $T_1^{\log}$ to emphasize their relationship to the logarithmiclike series. Typical dimensions of a focal area vary from hundreds of nanometers to several micrometers. Scanning steps vary from several nanometers to about 30 nm. So, the number of pixels, N, takes on values from about 50 to 1000. From this and Eq. (77), the speed enhancement will be $T_2^{\log}/T_1^{\log} = 350$–7000: from two to about four orders. As an example, at $N = 100$ the times are $T_2 \approx 800$ h and $T_1 \approx 1.1$ h.

In the next step, we will perform the same calculational time analysis for some pure entrance CPVVBs with Bessel beam profiles

$$\mathbf{E}_0(\theta, \phi) = J_m(2\sin\theta)e^{im\phi}\mathbf{n}_+ \quad (78)$$

in free space as described in Sec. IV B. These beams are basic elements of any entrance beam Fourier series representation and their computational times are the least possible. The results for computational times are collected in Table V, where m is the topological charge, t_1 and t_2 are average computational times over the region $\rho = [0, \ldots, 3\lambda]$ with the step of 0.6λ. From the table, the relation t_2/t_1 is greater than 100 for

TABLE V. Average computation times for circularly polarized vector vortex Bessel beams.

m	t_2 (s)	t_1 (s)	t_2/t_1
0	4.2	0.040	105
1	4.0	0.038	105
2	4.0	0.037	108
3	4.0	0.037	108
4	4.1	0.035	117
5	4.1	0.035	117
10	4.2	0.033	127

all the considered beams. The equation for speed enhancement takes the form

$$T_2^0/T_1^0 \approx 150N, \quad (79)$$

and substituting the number of steps, N, derived above, we obtain $T_2^0/T_1^0 = 7.5 \times 10^3$ to 1.5×10^5: the single-integral algorithm is from about four up to five orders faster than double-integral one.

C. Spatial-point expansion algorithm

In situations when the cylindrical-coordinate scan is inapplicable, the advantage of azimuth-distance factorization cannot be effectively used. However, as lateral coordinate scanning steps are finite (and normally small) and focal light field components are continuous differentiable functions of distance coordinates, we can exploit a power-series expansion of light fields.

The problem is as follows. We scan a square spatial ROI described in the previous section: $x_i = -L/2 + i\delta$, $y_j = -L/2 + j\delta$, $\delta = L/N$ is the scanning step with L being the square side length and N the number of steps, and the indices i and j run over the range $[0, \ldots, N]$. A focal light field has to be calculated at each point of the ROI (x_i, y_j). In this case, the distance coordinate ρ will run over the range $[0, \ldots, L/\sqrt{2}]$ and take on a set of values $\rho_{ij} = \sqrt{x_i^2 + y_j^2}$. Such a set will contain a number of elements of the order of N^2, which prevents time saving based on distance-azimuth factorization in Eqs. (36). To solve this problem, one should reformulate the task as follows. We have a distance coordinate range $\rho = [0, \ldots, L/\sqrt{2}]$ and need to calculate the amplitude-expansion vectors $\mathbf{A}_{j,\pm}(\rho, z)$ given by Eq. (37) at *any* point of this range. Thus, the task has to be solved first for a vector component $A_m(\rho, z)$ given by Eq. (46) (the indices j and $\pm$ are omitted for simplicity).

So, we have a set of the distance coordinate values $S_\rho = [\rho_0, \rho_1, \rho_2, \ldots, \rho_M]$ containing M elements, where $\rho_0 = 0$, $\rho_M = L/\sqrt{2}$, $\Delta\rho = L/(\sqrt{2}M)$, $\rho_j = j\Delta\rho$, and $j = [0, \ldots, M]$, and need to calculate $A_m(\rho, z = 0)$ at an arbitrary $\rho = \rho_\alpha + \delta\rho$ from this range. Here, ρ_α is an element of the set S_ρ, whose numerical value is the closest to the target ρ, and the difference $\delta\rho$ can take on any value from the range $[-\Delta\rho/2, \ldots, \Delta\rho/2]$. The target function is calculated as

$$A_m(\rho, z) = A_m(\rho_\alpha, z) + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k A_m(\rho, z)}{\partial\rho^k}\bigg|_{\rho=\rho_\alpha} \delta\rho^k. \quad (80)$$

As ρ dependence of this function is defined only by the Bessel function $J_m(k_1\rho\sin\theta_1)$ [see Eq. (46)], we have to explore the behavior of the Bessel function's derivative of an arbitrary order. It may be considered to depend only on ρ with constant θ_1 and k_1 to replace the partial derivative by a total one. Exploiting, for example, Eq. (26) as a Bessel function representation, one can readily express the Bessel function kth-order derivative $J_l^{(k)}(x)$ as

$$J_l^{(k)}(x) = 2^{-k}\sum_{i=0}^{k}\frac{(-1)^i k!}{i!(k-i)!}J_{l+2i-k}(x).\qquad(81)$$

Finally, Eq. (80) can be transformed into

$$A_m(\rho,z) = A_m(\rho_\alpha,z) + \sum_{p=1}^{\infty}C_m^p(\rho_\alpha,z)\frac{(k_1\delta\rho)^p}{p!}\qquad(82)$$

with the coefficient

$$C_m^p(\rho_\alpha,z) = \int_0^{\theta_{\max}}F_m(\theta_1)\cos\theta_1(\sin\theta_1)^{p+1}$$
$$\times K(\theta_1,z)J_{l(m)}^{(p)}(k_1\rho_\alpha\sin\theta_1)d\theta_1,\qquad(83)$$

derived from Eq. (46). If $k_1\delta\rho \leqslant 1$, the multiplier $(k_1\delta\rho)^p/p!$ decreases monotonically with increasing p and the coefficient $C_m^p(\rho_\alpha,z)$ satisfies the inequality

$$|C_m^p(\rho_\alpha,z)| \leqslant \int_0^{\theta_{\max}}U(\theta_1)|K(\theta_1,z)|(\sin\theta_1)^p d\theta_1\qquad(84)$$

as $|J_{l(m)}^{(p)}(k_1\rho_\alpha\sin\theta_1)| \leqslant 1$, which can be obtained from Eq. (26); the function $U(\theta_1)$ is explained after Eq. (66). Introducing a p-independent quantity

$$\beta = \int_0^{\theta_{\max}}U(\theta_1)|K(\theta_1,z)|d\theta_1,\qquad(85)$$

we can write

$$\left|C_m^p(\rho_\alpha,z)\right| \leqslant \beta.\qquad(86)$$

Now, it is possible to leave terms from $p=1$ to $p=K$ in Eq. (82) and assess the rejected part as

$$\frac{|A_m(\rho,z)|}{A_{en}} = \left|\sum_{p=K+1}^{\infty}\frac{C_m^p(\rho_\alpha,z)}{A_{en}}\frac{(k_1\delta\rho)^p}{p!}\right|$$
$$\leqslant \frac{\beta}{A_{en}}\frac{(k_1\delta\rho)^{K+1}}{(K+1)!}\frac{1}{1-(k_1\delta\rho)/(K+2)}.\qquad(87)$$

Finally, Eq. (82) can be rewritten as

$$A_m(\rho,z) = \sum_{p=0}^{L}C_m^p(\rho_\alpha,z)\frac{(k_1\delta\rho)^p}{p!}\qquad(88)$$

with the rejected part, the absolute value of which is assessed as

$$\frac{\left|A_m^{\rm rest}(\rho,z)\right|}{A_{en}} \leqslant \frac{2\beta}{A_{en}}\frac{(k_1\delta\rho)^{K+1}}{(K+1)!}.\qquad(89)$$

The multiplier 2 appeared because $K \geqslant 0$ and $|k_1\delta\rho| \leqslant 1$.

Now we need to (i) divide the range of the distance coordinate $[0,\ldots,L/\sqrt{2}]$ into M segments (which corresponds to $M+1$ points) and (ii) calculate vectors $A_{j,\pm}(\rho_\alpha,z)$ and K

of their derivatives at each point. So, we have $M+1$ points and $K+1$ vectors to calculate at each point. According to Eq. (89), the number of derivatives at a given point $\rho = \rho_\alpha$ does not depend on ρ_α. Although the series (82) converges at any value of $k_1\delta\rho$, we consider $|k_1\delta\rho| \leqslant 1$ to be preferable. The parameter $\delta\rho$ is expressed in terms of the number of points, M, as

$$|\delta\rho| \leqslant \frac{L}{2^{1.5}M}.\qquad(90)$$

To minimize the number of vectors to be calculated, we need to solve the system of equations

$$\frac{2\beta}{A_{en}}\left(\frac{k_1 L}{2^{1.5}}\right)^{K+1}\frac{M^{-(K+1)}}{(K+1)!} \leqslant 10^{-8},\qquad(91a)$$

$$(M+1)(K+1) = \min,\qquad(91b)$$

$$M \geqslant [2^{-1.5}k_1 L] + 1.\qquad(91c)$$

The first equation describes the negligibility criterion fulfillment [see Eq. (44)] and the threshold is set one order greater than that in Eq. (44) to prevent a possible accumulative error effect; this threshold can be corrected if necessary. The second equation sets the minimization of the number of vectors that we need to compute and the third one guarantees that the expansion parameter $k_1\delta\rho$ does not exceed unity in absolute magnitude. In the case of equal computational times for all vectors $\mathbf{C}^p(\rho,z)$ given by Eq. (83) regardless of their order p, the minimization of the number of vectors would lead to the computational time minimization. But in general, these times can increase because the number of Bessel functions to be calculated is proportional to p [see Eq. (81)] and this problem has to be studied.

We developed a PYTHON script to compute vectors $\mathbf{C}^p(\rho,z)$, which calculates Bessel function derivatives directly according to Eq. (81), and investigated computation speeds of $\mathbf{C}^p(\rho,z)$ for basic CPVVBs (78) in free space. Computational time was averaged over the range $\rho = [0,\ldots,3\lambda]$. For the vector $\mathbf{C}^0(\rho, z=0)$ the following behavior of computational times was observed: 0.043 s at $m=0$, 0.040 s at $m=3$, and 0.036 s at $m=10$. With increasing the vector order p the times increase continuously and for $\mathbf{C}^{20}(\rho, z=0)$ reach 0.062 s at $m=0$, 0.061 s at $m=3$, and 0.065 s at $m=10$. Such an increase is substantial but it can be readily overcome by exploiting the recursive relation

$$x^2(n-1)J_{n+2}(x) = -x^2(n+1)J_{n-2}(x)$$
$$+ 2n(2n^2 - x^2 - 2)J_n(x),\qquad(92)$$

which allows one to compute only two Bessel functions appearing in Eq. (81) and obtain others from them using only simple binary operations. As vectors $\mathbf{C}^1(\rho,z)$ and $\mathbf{C}^2(\rho,z)$ contain two and three Bessel function terms, respectively, their computational time is a good approximation of that of an arbitrary vector $\mathbf{C}^p(\rho,z)$ calculated with the use of Eq. (92). We investigated computation times of $\mathbf{C}^1(\rho,z)$ and $\mathbf{C}^2(\rho,z)$ for basic CPVVBs (78) at $m = [0,\ldots,10]$ in free space and compared them with those for the $\mathbf{A}$ vector, displayed in Table V. It was observed that the vectors $\mathbf{C}^1(\rho,z)$ and $\mathbf{C}^2(\rho,z)$ consume about 10–15 % more time than $\mathbf{A}$ vectors. Denoting as t_a and t_c computational times for an $\mathbf{A}$ vector and an

TABLE VI. The number of vectors to be calculated as a function of the threshold.

ε_{th}	M	K	$(K+1)(M+1)$
10^{-6}	14	8	135
10^{-7}	14	9	150
10^{-8}	14	10	165
10^{-9}	15	10	176
10^{-10}	15	11	192
10^{-11}	15	12	208

arbitrary-order **C** vector, respectively, we will assume for the practical use $t_c < 1.5t_a$, and set $t_c = 1.5t_a$ in Eqs. (91).

Now the time efficiency of the spatial-point expansion algorithm can be assessed by comparing it with that of the cylindrical scanning algorithm. Computational times of three-dimensional and focal-plane scans obey

$$T_{spe} \propto 1.5(K+1)(M+1) \qquad (93)$$

for the spatial-point expansion algorithm and

$$T_{cs} \propto N/\sqrt{2} \qquad (94)$$

for the cylindrical scanning algorithm. As T_{spe} depends on ROI dimensions [from Eqs. (91)], we have to define the area side length L. So, we will consider a square with x and y running over the range $[-2\lambda, \ldots, 2\lambda]$: $L = 4\lambda$. The product $k_1 L/2^{1.5}$ amounts to $2\sqrt{2}\pi n_1 \approx 13.6$ (at $n_1 = 1.526$). We calculated the factor $2\beta/A_{en}$ for the beams considered above and discovered that it amounts to about 10^{-2}–10^{-1}. In general it can be greater for other beams and we explored Eqs. (91) at different thresholds $\varepsilon_{th} = 0.5A_{en}\varepsilon/\beta$. The parameter ε was set to be equal to 10^{-8} in the present case, but it can be varied in general. The results are displayed in Table VI. One can see that the number of vectors to be calculated increases continuously from 135 at $\varepsilon_{th} = 10^{-6}$ to 208 at $\varepsilon_{th} = 10^{-11}$, which, first, demonstrates that it increases slowly and, second, at normal values of the parameter $2\beta_m^{max}/A_{en}$, it is completely enough to calculate up to 210 vectors to reach an acceptable precision. Taking the reasonable range of the number of steps, $N = [50, \ldots, 1000]$, for the cylindrical scanning algorithm and the number of vectors from Table VI $(K+1)(M+1) = $ 135–208 for the spatial-point expansion algorithm, we assess the times, using Eqs. (93) and (94), as $T_{cs} \propto [35, \ldots, 710]$ and $T_{spe} \propto [200, \ldots, 310]$. So, the spatial-point expansion algorithm may be either several times slower or faster than the cylindrical scanning algorithm, depending on the number of steps, N, and the factor β/A_{en}, and they are of the same order on average. In general, the former is more complicated than the latter and the cylindrical scanning algorithm is preferable if rectangular scanning is not strictly necessary.

VI. SUMMARY AND DISCUSSION

In the present section, we summarize the results, analyze possible problems and practical aspects of the suggested single-integral algorithm, and compare it with existing algorithms.

A. Single-integral algorithm: Issues and limitations

For the single-integral algorithm to be effective, the following two conditions should be met: (i) a focused laser beam should be Fourier-series expandable and (ii) series representation should provide more effective calculations than the direct Richards-Wolf double-integral procedure. Simple widely used entrance beams—like Laguerre-Gaussian, Hermite–Gaussian, Bessel, radially and azimuthally polarized beams, and many others—are represented by finite superpositions of several CPVVBs and the algorithm efficiency is obvious for them. For complicated beams one needs to calculate series coefficients according to Eq. (34) and it may be an issue in general. Following Eq. (34), an mth Fourier coefficient of an α-polarized beam $E_0^\alpha(\theta_1, \phi)$ (the index α defines right- or left-hand polarization) takes the form

$$F_m^\alpha(\theta_1) = \frac{1}{2\pi} \int_0^{2\pi} E_0^\alpha(\theta_1, \phi) \exp(-im\phi)d\phi. \qquad (95)$$

The first problem is concerned with possible divergence of the integral, but as the amplitude $E_0^\alpha(\theta_1, \phi)$ must be square integrable, this problem is irrelevant. The problem of a divergent series may arise if the integral does not decrease with increasing number m. If $E_0^\alpha(\theta_1, \phi)$ is a slowly varying function of ϕ, obviously, the higher are oscillations of the complex exponent in Eq. (95), the less absolute value of the integral will be. The coefficient $F_m^\alpha(\theta_1)$ is a constant with respect to m only if $E_0^\alpha(\theta_1, \phi)$ is the Dirac delta function $\delta(\phi)$ and increases if $E_0^\alpha(\theta_1, \phi)$ is a derivative of the δ function [30]. Thus, if an entrance amplitude is close to the δ function or a similar singularity function, the problem of poorly convergent focal series can arise. A mathematical example of a problem function is

$$f(\theta_1) = \frac{1 - \alpha^N \exp(iN\phi)}{1 - \alpha \exp(i\phi)} \qquad (96)$$

with $\alpha \geqslant 1$ being a real number. Although it corresponds to a finite sum, its decomposition can contain an unlimited number of terms as we can set as big N as we want ($N = 10, 100, 1000$, and so on) and the number of terms is proportional to N. At $\alpha = 1$ we have a δ-function-like high-peaked function and at $\alpha > 1$ a product of a highly oscillatory function with a slowly varying one. So, the problem can take place if and only if an amplitude $E_0^\alpha(\theta_1, \phi)$ is either a highly oscillatory multifrequency or singularity function of ϕ. We will call this class of functions *nonvanishing series coefficient functions* and study the problem later on. Finally, a coefficient $F_m^\alpha(\theta_1)$ in Eq. (95) can be a complicated function but it can be calculated numerically under any conditions and, in principle, approximated by a simple function of θ_1.

In Sec. IV B we studied an entrance beam, the Fourier-series representation of which is a slowly and conditionally convergent logarithmlike series. An absolute value of the series coefficient can be assessed as $|F_m^\alpha(\theta_1)| \approx 1/|m|$: the convergence of the series is the slowest possible. So, calculation of tight focusing of such a series is nearly the most time consuming for the single-integral algorithm as a number of series terms is maximal. For any other entrance beam [except for anomalous ones like in Eq. (96)] the single-integral algorithm will compute field values faster, while the double-

integral one will compute approximately with the same speed as the function $\exp(i\phi/2)$ is not complicated. Tables I and II show that the focal field numerical series converges rapidly: the residual field amplitude normally decreases by an order of magnitude per two steps of the number m, while the entrance coefficient decreases from $\approx 1/26$ to $\approx 1/46$ by about the factor of $46/26 \approx 1.8$ in the whole range, which is significantly less even than one order of magnitude. From this, we conclude that the decreasing behavior is caused not by the entrance coefficient but the Bessel functions in integrals in Eqs. (28). Neither can the intracavity interference coefficients be responsible for this behavior as Table I displays results for a free space and Table II for a complicated planar microcavity, which are absolutely different conditions.

Another nuance is that coefficients of the considered logarithmlike series decrease monotonically, while they can be unstable. Such a situation would be a bit worse as we would have to exploit a function $U_+^c(\theta_1) = 0.5\sin\theta_{\max}[A_{en}^2 \sin^2\theta_{\max} - 2F_{en}^2]^{0.5}$ instead of $U_+^m(\theta_1) = \sin\theta_1\cos\theta_1 |F_{m_0+p}^+(\theta_1)|$ in Eq. (66), which would have led to a greater number of series terms and, consequently, greater computational times. To simulate such a situation, we can recalculate data in Tables I and II through division by the average value of the function $U_+^m(\theta_1)$ in the range $\theta_1 = 0, \ldots, \theta_{\max}$ and multiplication by $U_+^c(\theta_1)$. Calculation of the average value gives $U_{+,av}^m \approx 0.07/m$ and taking $F_{en}^2 = 0$ [as it corresponds to the maximum possible value of $U_+^c(\theta_1)$], we obtain $U_+^c = 0.11$ and the ratio $U_+^c/U_{+,av}^m = 1.57m$. At values of the index $m = 26, \ldots, 46$ presented in Table II this relation takes on the biggest value equal to 72 at $m = 46$ and such a multiplier will increase the residual field values in the table by two orders of magnitude. It means that now one needs to take four additional terms (in the positive direction) to guarantee the desirable precision. From Table III, 74 terms were used to calculate the values of the field at $\rho = 3\lambda$. Four additional terms in each (positive and negative) direction will give eight additional terms and the number of terms will increase by $8/74 = 11\%$. It will increase the computational time by about 10%, respectively. At $\rho = 0$ only the three terms possessing the zero-order Bessel function will be in use and computational time will not increase at all. Computational times at ρ values between zero and 3λ will increase from zero to 10% and the total computational time in the range $\rho = 0, \ldots, 3\lambda$ will increase by less than 10%, which is insignificant.

So, a tight focusing calculation of any beam described by a convergent Fourier series will be at least several times faster for the single-integral algorithm at a given spatial point. A problem can arise only for a beam whose amplitude is described by a nonvanishing coefficient series explained after Eq. (96). A Fourier series of such an amplitude corresponds to a vector

$$\mathbf{F}(\theta_1) = \left(F_{m_1}(\theta_1), F_{m_2}(\theta_1), \ldots, F_{m_p}(\theta_1)\right) \quad (97)$$

[see Eq. (33)], the average absolute values of components of which are of mutually comparable magnitude and a number of terms must be big enough. To assess this number, we should compare computational times for one single-integral term and double-integral calculation. From Table V, a single-integral time amounts to about 1% of a double-integral one. The

table presents data for free-space calculations, which correspond to the simplest conditions. As the difference between single-integral and double-integral times cannot be decreased by intracavity interference coefficients, the value of 1% is the maximal possible. So, the single-integral algorithm can become inefficient if and only if the focused beam is described by a series that contains more than 100 CPVVBs. But such numbers are big enough to make the respective scalar amplitudes either nearly singular high-peaked or highly oscillatory functions, the numerical double integration of which, in turn, is orders of magnitude more complicated and time consuming.

Thus, the single-integral algorithm calculates a focal field at a given spatial point from several times to two orders faster than the double-integral algorithm for *any* focused collimated beam.

B. Single-integral algorithm as an alternative to DFT-based algorithms

Finally, we will compare the single-integral algorithm with those based on the DFT method. The following characteristics of algorithms will be used: computational time, calculation accuracy, and error level controllability. The key point of DFT-based algorithms is to calculate the function $\mathbf{E}_{rs}(\theta_1, \phi)\sin\theta_1$ in Eq. (3) once inside a discrete range $[\theta_1 = \theta_{1,0}, \ldots, \theta_{1,N}, \phi = \phi_0, \ldots, \phi_M]$ and then use it at different focal coordinates, substituting an integral by the sum

$$\mathbf{E}_f(\mathbf{r}) \propto \sum_{n=0}^{N} \sum_{m=0}^{M} \mathbf{E}_{rs}(\theta_{1n}, \phi_m) e^{i\mathbf{k}(\theta_{1n}, \phi_m)\cdot\mathbf{r}} \sin\theta_{1n}. \quad (98)$$

It is very time effective for calculations in focal regions containing a big number of spatial points but meaningless for single-point calculations. In fact, DFT-based algorithms calculate integrals using the two-dimensional rectangle method and, in particular, for a single-point calculation Eq. (98) is just a numerical expression to compute the double integral. Although DFT-based algorithms are able, in principle, to provide good accuracy, the use of the only given (not the most precise) numerical integration method spoils the accuracy and makes the error level poorly controllable.

FFT-based algorithm computational speeds exceed those of the direct Richards-Wolf double-integral algorithm by about two orders of magnitude for typical ROIs and entrance beams [20,22]. DFT computation using the Bluestein method enhances calculation speeds up to five orders compared to the direct Richards-Wolf method [41]. In Ref. [41], in particular, the following situation was studied: a $2\lambda \times 2\lambda$ focal-plane square, the number of points in one direction $N \approx \sqrt{10^5} \approx 300$, and a CPVVB with the topological charge $m = 1$. The following computational times were obtained: the direct Richards-Wolf method, 57 min; FFT, 2.9 s; and Bluestein, 0.2 s. The comparison of FFT and Bluestein methods with the direct Richards-Wolf calculation gives $57\text{ min}/2.9\text{ s} \approx 1.2 \times 10^3$ and $57\text{ min}/0.2\text{ s} \approx 1.7 \times 10^4$, respectively. For the three-dimensional calculations with 100 z slices the times were 95 hours for Richards-Wolf, 280 s for FFT, and 12 s for Bluestein. The respective enhancements are $95\text{ h}/280\text{ s} \approx 1.2 \times 10^3$ and $95\text{ h}/12\text{ s} \approx 2.8 \times 10^4$. Similar enhancements for the single-integral algorithm will be $7 \times 300 = 2.1 \times 10^3$

for a beam represented by a logarithmlike series according to Eq. (77) and $150 \times 300 = 4.5 \times 10^4$ for a CPVVB according to Eq. (79). Thus, at $N = 300$ the single-integral algorithm calculation speed is either the same as that of the FFT-based algorithm or one order greater, depending on the focused beam, and, similarly, either one order less or the same as that of the Bluestein algorithm. Similar analyses at $N = 1000$ give the enhancements of 3.3×10^3 for FFT and 5×10^5 for Bluestein. For the single-integral algorithm they will be from 7×10^3 to 1.5×10^5, depending on the beam. So, the single-integral algorithm is up to two orders faster than direct FFT-based algorithms and up to two orders slower than the Bluestein algorithm.

The single-integral algorithm is able to reproduce field values obtained by the basic Richards-Wolf integration with any given accuracy and completely controllable error level according to the negligibility criterion developed in Sec. III D. The accuracy of DFT-based algorithms is normally good but significantly less than that of the single-integral algorithm with poorly controllable error level. It is very difficult (if possible) to assign an error level similar to Eq. (70) for a DFT-based algorithm and it is rarely assessed in practice. In Ref. [22] an average error level was calculated and amounted to about 4% of the mean-square value of the focal field. In Ref. [41] at $N \approx 300$ the error of the Bluestein-method algorithm amounts from zero up to 15% of the average focal field value for a radially polarized beam. So, we can state that the accuracy of DFT-based algorithms amounts normally to several percent with uncontrollable error level. Although the result can be noticeably distorted, such accuracy is acceptable for many practical applications. But the uncontrollable error level impairs the reliability.

To summarize, the Bluestein-method-based algorithm [41] is up to two orders faster than the single-integral algorithm at the cost of lower accuracy and uncontrollable error level. Additionally, the computational speed of the single-integral algorithm can be further enhanced in principle by substituting an integral by a sum similarly to Eq. (98) and calculating a discrete set of single integrals in a way similar to the DFT method. This approach will significantly reduce the accuracy and error level controllability but it is potentially able to enhance computational speed by orders of magnitude and definitely provides faster calculations than the Bluestein-method-based or any other two-dimensional DFT-based algorithm.

VII. CONCLUSION

We developed the mathematical basis for a generalized time-effective algorithm to calculate tight focusing of laser beams with arbitrary cross-section light vector distribution. Focusing into an arbitrary planar microcavity is the most generalized geometrical configuration which can be treated analytically and we investigated this case. It was disclosed that a circularly polarized vortex vector beam series is an optimal basis to decompose an entrance beam. Based on this, a generalized single-integral algorithm was suggested.

A fractional vortex Bessel beam was treated as an example to test the algorithm. This beam's circularly polarized vortex vector beam representation is a logarithmlike conditionally convergent infinite series. However, the focal field induced by this beam is presented by a rapidly convergent series, and to provide acceptable precision of the calculated focal field, one needs to take from several terms to several tens of terms, depending on the focal distance coordinate value. As the entrance beam is represented by a slowly convergent series with slowly decreasing coefficients, the test results can be readily generalized to any entrance beam. It was shown that the suggested algorithm is, on average, about five times faster than the double-integral one based on the direct use of the Richards-Wolf method for a single spatial point calculation. This number (the fivefold calculation speed enhancement) depends on focusing conditions and may be a bit smaller for simpler conditions, or larger for more complicated planar structures, but at least several-times speed enhancement can be guaranteed for any conditions. Similar investigation for the basic circularly polarized vortex entrance beams in free space shows that the single-integral algorithm is more than 100 times faster for them.

Finally, we tested the algorithm for calculations of focal fields in finite two- and three-dimensional spatial regions as this task is one of the most widely applicable. In this case, the single-integral algorithm is from two to five orders faster than the double-integral algorithm. Algorithms based on the discrete Fourier transform are an existing alternative to the single-integral algorithm to calculate focal fields in finite spatial regions. The Bluestein-method-based algorithm is the fastest among them and is able to provide computational speeds from the same up to two orders greater than the single-integral algorithm. But its accuracy is orders of magnitude less and the error level is poorly controllable, whereas the accuracy of the single-integral algorithm is unlimited and the error level completely controllable.

[1] B. Sick, B. Hecht, and L. Novotny, Orientational imaging of single molecules by annular illumination, Phys. Rev. Lett. **85**, 4482 (2000).

[2] A. M. Chizhik, R. Jäger, A. I. Chizhik, S. Bär, H.-G. Mack, M. Sackrow, C. Stanciu, A. Lyubimtsev, M. Hanack, and A. J. Meixner, Optical imaging of excited-state tautomerization in single molecules, Phys. Chem. Chem. Phys. **13**, 1722 (2011).

[3] Z. Földes-Papp, U. Demel, and G. P. Tilz, Laser scanning confocal fluorescence microscopy: An overview, Int. Immunopharmacol. **3**, 1715 (2003).

[4] J. Pawley, *Handbook of Biological Confocal Microscopy*, Vol. 236 (Springer Science & Business Media, New York, 2006).

[5] S. W. Hell and J. Wichmann, Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy, Opt. Lett. **19**, 780 (1994).

[6] G. Vicidomini, P. Bianchini, and A. Diaspro, STED super-resolved microscopy, Nat. Methods **15**, 173 (2018).

[7] H. Blom and J. Widengren, Stimulated emission depletion microscopy, Chem. Rev. **117**, 7377 (2017).

[8] A. M. Larson, Multiphoton microscopy, Nat. Photonics **5**, 1 (2011).

[9] K. König, Multiphoton microscopy in life sciences, J. Microsc. **200**, 83 (2000).

[10] L. Novotny and B. Hecht, *Principles of Nano-optics* (Cambridge University Press, Cambridge, UK, 2012).

[11] A. Döbarre, R. Jaffiol, C. Julien, D. Nutarelli, A. Richard, P. Tchénio, F. Chaput, and J.-P. Boilot, Quantitative determination of the 3D dipole orientation of single molecules, Eur. Phys. J. D **28**, 67 (2004).

[12] L. Novotny, M. R. Beversluis, K. S. Youngworth, and T. G. Brown, Longitudinal field modes probed by single molecules, Phys. Rev. Lett. **86**, 5251 (2001).

[13] P. R. Dolan, X. Li, J. Storteboom, and M. Gu, Complete determination of the orientation of NV centers with radially polarized beams, Opt. Express **22**, 4379 (2014).

[14] H. Rigneault, S. Monneret, and C. I. Westbrook, Resonant focusing in a planar microcavity, J. Opt. Soc. Am. B **15**, 2712 (1998).

[15] D. Khoptyar, R. Gutbrod, A. Chizhik, J. Enderlein, F. Schleifenbaum, M. Steiner, and A. J. Meixner, Tight focusing of laser beams in a $\lambda/2$-microcavity, Opt. Express **16**, 9907 (2008).

[16] L. Wang, Q. Liu, F. Wackenhut, M. Brecht, P.-M. Adam, J. Gierschner, and A. J. Meixner, Monitoring tautomerization of single hypericin molecules in a tunable optical $\lambda/2$ microcavity, J. Chem. Phys. **156**, 014203 (2022).

[17] N. Kovalets, E. Kozhina, I. Razumovskaya, S. Bedin, A. Piryazev, Y. V. Grigoriev, and A. Naumov, Toward single-molecule surface-enhanced raman scattering with novel type of metasurfaces synthesized by crack-stretching of metallized track-etched membranes, J. Chem. Phys. **156**, 034902 (2022).

[18] B. Richards and E. Wolf, Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system, Proc. R. Soc. London A **253**, 358 (1959).

[19] Q. Zhan, Properties of circularly polarized vortex beams, Opt. Lett. **31**, 867 (2006).

[20] M. Leutenegger, R. Rao, R. A. Leitgeb, and T. Lasser, Fast focus field calculations, Opt. Express **14**, 11277 (2006).

[21] B. Boruah and M. Neil, Focal field computation of an arbitrarily polarized beam using fast Fourier transforms, Opt. Commun. **282**, 4660 (2009).

[22] J. Lin, O. Rodríguez-Herrera, F. Kenny, D. Lara, and J. Dainty, Fast vectorial calculation of the volumetric focused field distribution by using a three-dimensional Fourier transform, Opt. Express **20**, 1060 (2012).

[23] R. Tong, Z. Dong, Y. Chen, F. Wang, Y. Cai, and T. Setälä, Fast calculation of tightly focused random electromagnetic beams: Controlling the focal field by spatial coherence, Opt. Express **28**, 9713 (2020).

[24] M. Aviñoá, R. Martínez-Herrero, and A. Carnicer, Efficient calculation of highly focused electromagnetic Schell-model beams, Opt. Express **29**, 26220 (2021).

[25] S. Boichenko, Theoretical investigation of confocal microscopy using an elliptically polarized cylindrical vector laser beam: Visualization of quantum emitters near interfaces, Phys. Rev. A **97**, 043825 (2018).

[26] I. Basistiy, V. Pas'Ko, V. Slyusar, M. Soskin, and M. Vasnetsov, Synthesis and analysis of optical vortices with fractional topological charges, J. Opt. A: Pure Appl. Opt. **6**, S166 (2004).

[27] J. Wen, B. Gao, G. Zhu, Y. Cheng, S.-Y. Zhu, and L.-G. Wang, Observation of multiramp fractional vortex beams and their total vortex strength in free space, Opt. Laser Technol. **131**, 106411 (2020).

[28] S. Boichenko, Orientational imaging of single fluorescent molecules: What laser beam can provide azimuth-rotation-invariant image shape? Opt. Commun. **498**, 127256 (2021).

[29] D. McGloin and K. Dholakia, Bessel beams: Diffraction in a new light, Contemp. Phys. **46**, 15 (2005).

[30] A. Zygmund, *Trigonometric Series*, Vol. 1 (Cambridge University Press, Cambridge, UK, 2002).

[31] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Elsevier, Amsterdam, 2013).

[32] Q. Zhan, Cylindrical vector beams: From mathematical concepts to applications, Adv. Opt. Photonics **1**, 1 (2009).

[33] Z. Xiaoqiang, C. Ruishan, and W. Anting, Focusing properties of cylindrical vector vortex beams, Opt. Commun. **414**, 10 (2018).

[34] S. Stafeev, A. Nalimov, V. Zaitsev, and V. Kotlyar, Tight focusing cylindrical vector beams with fractional order, J. Opt. Soc. Am. B **38**, 1090 (2021).

[35] S. Boichenko, Toward super-resolution fluorescent microscopy of arbitrarily oriented single molecules, Phys. Rev. A **101**, 043823 (2020).

[36] V. V. Kotlyar and S. S. Stafeev, A transverse energy flow at the tight focus of light with higher-order circular-azimuthal polarization, Comput. Optics **45**, 311 (2021).

[37] H. Zhang, J. Zeng, X. Lu, Z. Wang, C. Zhao, and Y. Cai, Review on fractional vortex beam, Nanophotonics **11**, 241 (2022).

[38] M. Berry, Optical vortices evolving from helicoidal integer and fractional phase steps, J. Opt. A: Pure Appl. Opt. **6**, 259 (2004).

[39] T. Rammler, F. Wackenhut, S. zur Oven-Krockhaus, J. Rapp, K. Forchhammer, K. Harter, and A. J. Meixner, Strong coupling between an optical microcavity and photosystems in single living cyanobacteria, J. Biophoton. **15**, e202100136 (2022).

[40] T. Rammler, F. Wackenhut, J. Rapp, S. zur Oven-Krockhaus, K. Forchhammer, A. J. Meixner, and K. Harter, Analysis of fast fluorescence kinetics of a single cyanobacterium trapped in an optical microcavity, Plants **12**, 607 (2023).

[41] Y. Hu, Z. Wang, X. Wang, S. Ji, C. Zhang, J. Li, W. Zhu, D. Wu, and J. Chu, Efficient full-path optical calculation of scalar and vector diffraction using the Bluestein method, Light: Sci. Appl. **9**, 119 (2020).