Helicity and angular momentum: A symmetry-based framework for the study of light-matter interactions

Ivan Fernandez-Corbaton, Xavier Zambrana-Puyalto, and Gabriel Molina-Terriza

ARC Center of Excellence for Engineered Quantum Systems and Department of Physics & Astronomy,
Macquarie University, NSW 2109, Australia

We propose a theoretical and practical framework for the study of light-matter interactions and the angular momentum of light. Our proposal is based on helicity, total angular momentum, and the use of symmetries. We compare the framework to the current treatment, which is based on separately considering spin angular momentum and orbital angular momentum and using the transfer between the two in physical explanations. In our proposal, the fundamental problem of spin and orbital angular momentum separability is avoided, predictions are made based on the symmetries of the systems, and the practical application of the concepts is straightforward. Finally, the framework is used to show that the concept of spin to orbit transfer applied to focusing and scattering is masking two completely different physical phenomena related to the breaking of different fundamental symmetries: transverse translational symmetry in focusing and electromagnetic duality symmetry in scattering.

I. INTRODUCTION

In the last decade, the angular momentum of light has received much attention from very diverse areas of physics: From experimental astrophysics proposals for the detection of exotic cosmic objects [1] to the use of light beams to rotate atoms [2], through the exploitation of the infinite number of possible angular momentum values for increasing the capacity of optical communication networks [3], or the development of new concepts in quantum information [4]. The list of applications is long indeed [5].

The availability of appropriate theoretical tools for the study of light beams with angular momentum and their interactions with matter is crucial for the development of such a wide and promising range of applications. The current state-of-the-art theoretical framework is based on the separation of angular momentum $\mathbf{J}$ in its orbital ($\mathbf{L}$) and spin ($\mathbf{S}$) components: $\mathbf{J} = \mathbf{L} + \mathbf{S}$. In the paraxial approximation the value of the total angular momentum along the optical axis can be split into a term which depends on the azimuthal spatial phase of the field and a term which depends on the polarization [6]. Several efforts have been undertaken to rigorously extend this approach to the nonparaxial regime, but have encountered fundamental difficulties [7,8]. Strongly nonparaxial tightly focused light fields are the bread and butter of applications where light is made to interact with nanostructures, molecules, and atoms. The mechanism of spin to orbit angular momentum conversion (SAM to OAM), also referred to as spin-orbit interaction, is the explanation of choice for numerically obtained observations in focusing [9–11] and remarkable results in scattering experiments [12–14]. For example, the presence of optical vortices in tightly focused fields and in scattered fields is explained by conversions between the two types of angular momenta.

Rigorously speaking, though, the separation between SAM and OAM cannot be made on firm physical grounds. Consequently, a conversion between the two quantities is not a fully satisfactory explanation for physical phenomena. The separate consideration of $\mathbf{S}$ and $\mathbf{L}$ is known to pose fundamental problems for the electromagnetic field [7,8] and its quantum excitations [15,16]. From the point of view of quantum field theory only the total angular momentum operator is a valid observable property of the photon. Even more generally, the strict nonseparability is not restricted to photons. For example, it also applies to the electron since only in the nonrelativistic limit can the orbital and spin parts of its angular momentum be separately considered [15]. The fundamental reason for such nonseparability is the geometry of rotations of vectors and spinors. Such a restriction also applies to rotations of classical electromagnetic fields [17].

In this article we put forward an alternative theoretical framework for the general and rigorous treatment of the angular momentum of light and its role in light-matter interactions. Our approach solves the theoretical difficulties of the current framework, draws its predictive power from symmetry considerations, and can be simply applied in practice. By using it, we discover that the actual physical reason responsible for the presence of optical vortices in tightly focused beams is totally unrelated to the one responsible for the appearance of optical vortices in scattering experiments. In the current state-of-the-art framework, both cases are explained as SAM to OAM conversion.

Our proposal is based on total angular momentum and helicity. The role of helicity (Λ) in light matter interactions has recently been considered [18]. The macroscopic Maxwell equations have been shown to be invariant under generalized electromagnetic duality transformations, and helicity has been identified as the generator of those transformations. By exploiting this connection, helicity can be used within the powerful formalism of symmetries and conserved quantities for the study of light matter interactions when the approximations implicit in the macroscopic Maxwell equations hold [19]. The use of symmetries and conserved quantities for the study of electromagnetic problems is the paradigm used in this article for the development of its theoretical concepts and their application to practical situations.

In Sec. II we outline the paradigm, mathematical concepts, and notation used throughout the paper. In Sec. III we summarize the different aspects involved in the separation

of SAM and OAM. In Sec. IV we outline our proposal. First, we discuss the concept of helicity and its associated symmetry and comment on a result from [18], which shows that, upon scattering, helicity transforms independently of the geometry of the scatterer. Then we show that the combined use of angular momentum and helicity solves the problems associated with the separation of SAM and OAM in a way that is simpler than the existing theoretical solutions and comment on the practical applicability of our ideas. Finally, we establish a relationship between helicity eigenstates and the transverse electric (TE) and transverse magnetic (TM) components of the field. Using this relationship we express the conservation law for helicity as a function of the partial scattering matrices in the TE-TM basis. This expression becomes useful in the practical application of the framework. In Sec. V we use the developed ideas to revisit the concept of SAM to OAM conversion in focusing and scattering. We are able to clearly identify the underlying reasons for the presence of optical vortices in focused and scattered fields, which happen to be two totally different physical phenomena connected to the breaking of two independent fundamental symmetries: transverse translational symmetry in focusing and electromagnetic duality in scattering. Up to now, the two are explained by SAM to OAM conversion. During Sec. V, we provide the analytical tools necessary for the practical application of our framework. Section VI contains our conclusions and discussion.

II. PARADIGM, MATHEMATICAL SETTING, AND NOTATION

The paradigm that we follow in our work is the use of symmetries and conserved quantities for the study of electromagnetic problems. In this article we consider classical Maxwell fields. Symmetry operations like rotations and translations are linear transformations that apply to the fields. Similarly, we model the light-matter interactions as linear transformations of the free space fields. The fields themselves will hence always be transverse. These ideas are best formalized using the mathematical setting of Hilbert spaces.

Therefore, in this article we make extensive use of the concepts and tools associated with a vector space endowed with an inner product (i.e., a Hilbert space), and the linear operators acting within that vector space. The vector space we consider is the space of transverse solutions of Maxwell's equations, or transverse Maxwell fields, which we call $\mathbb{M}$. A linear operator within $\mathbb{M}$ takes one of its vectors, a transverse Maxwell field, and transforms it into another vector of the space, still a transverse solution of Maxwell's equations. We are interested in symmetry transformations of the fields: time translation, spatial translations, and rotations, etc. These transformations are operators acting within $\mathbb{M}$. Such continuous symmetries are generated by Hermitian operators, also acting within $\mathbb{M}$, which are associated with properties of the fields. For instance, energy generates time translations, the components of linear momentum generate spatial translations and the components of angular momentum generate spatial rotations. A Hermitian operator O generates a continuous

symmetry transformation $T(\theta)$ by means of

$$T(\theta) = \exp(-i\theta O) = \sum_n \frac{(-i\theta O)^n}{n!}. \qquad (1)$$

See [17] for the detailed study of classical Maxwell fields using angular momentum and its generated transformation, spatial rotations. The fact that $\mathbb{M}$ has an inner product allows us to speak of Hermitian operators. It also allows us to construct orthonormal basis into which any transverse Maxwell field can be expanded. The basis vectors can be chosen to be transverse fields which are simultaneous eigenvectors of four commuting Hermitian operators. The choice of the inner product is not necessarily unique, even though the one we have used, defined in Sec. IV C has a long tradition for vector fields [20].

The consideration of $\mathbb{M}$, the Hermitian operators associated with the fundamental quantities of the field and the transformations that these operators generate, allows one to study electromagnetic problems using Maxwell fields together with the powerful framework of symmetries and conserved quantities. When the electromagnetic equations of a given system are invariant under the transformations generated by a given Hermitian operator, the property of the field associated with that operator is a constant of the motion. Conversely, if the system does not possess that symmetry, we know that the associated property can, in general, change during evolution. The effect of the symmetry of the system is even stronger, as it must also preserve the eigenvectors and eigenvalues of the operator generating the symmetry, thus offering an adequate basis of vectors to solve the electromagnetic problem. As we will show, this is a simple yet insightful approach to electromagnetic problems. In this article, we want to exploit this approach for the study of the angular momentum of light. Consequently, for our purposes, we may only use properties of the electromagnetic field which are associated with a Hermitian operator in $\mathbb{M}$. Only then can we consider their associated symmetry. This rules out the separate use of the components of $\mathbf{L}$ and $\mathbf{S}$, since their associated operators transform a transverse Maxwell field into a nontransverse field [21] and [16]: They do not act within the required vector space. From this point of view, $\mathbf{L}$ and $\mathbf{S}$ are qualitatively different from $\mathbf{J}$, the linear momentum $\mathbf{P}$, the energy H or, as we will discuss, helicity Λ. In the most commonly used representation, the expressions corresponding to these operators are

$$H = i\frac{\partial}{\partial t}, \quad \mathbf{P} = -i\nabla, \quad \mathbf{J} = \mathbf{L} + \mathbf{S}, \quad \Lambda = \frac{\mathbf{J} \cdot \mathbf{P}}{|\mathbf{P}|}, \quad (2)$$

$$\mathbf{L} = -i\mathbf{r} \times \nabla, \quad S_{nm}^k = -i\epsilon_{knm}, \qquad (3)$$

where S^k, the kth component of $\mathbf{S}$, is a matrix of indexes nm defined using the totally antisymmetric tensor ϵ_{knm} with $\epsilon_{123} = 1$. As discussed above, the operators in Eq. (2) transform a transverse Maxwell field into a transverse Maxwell field, while those in Eq. (3) break the transversality of the fields.

In our notation we use capital letters like J_z and P_x to denote operators, and lower case letters like j_z (or n) and k_x to denote their eigenvalues. When we speak of a field having a "sharp" or "well-defined" value for an operator, we mean that the field is an eigenvector of that operator with eigenvalue equal to its "sharp" value. In the text, names like "helicity" or "third component of angular momentum" refer to the operators

unless it is clear from the context that this is not the case. In addition, all analytical calculations from Sec. IV C on, assume a time harmonic decomposition of the fields with an $\exp(-iwt)$ dependence.

Finally, we would like to mention that the main context of this work is that of classical Maxwell fields. Nevertheless, the approach here taken, which is based on the study of symmetries, is general and often used in quantum mechanics and other areas of modern physics.

III. THE SEPARATION OF L AND S

Serious theoretical difficulties are encountered when attempting to separately consider **L** and **S** for the electromagnetic fields ([7,8]) or its quanta ([15,16]). We also know [21] and [16] that the operators **L** and **S** break the transversality of the fields, taking a vector of $\mathbb{M}$ out of that space. As far as we are concerned, this prevents the separate use of **L** and **S** for studying electromagnetic problems using symmetries and conserved quantities: **L** and **S** do not generate symmetry transformations for vectors in $\mathbb{M}$.

On the other hand, the total integrated value of the angular momentum of the electromagnetic field, expressed here in convenient units as an integral over all space involving the electric and magnetic fields $\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$,

$$\langle \mathbf{J} \rangle = \int d\mathbf{r} \, \mathbf{r} \times [\mathbf{E}(\mathbf{r}) \times \mathbf{B}(\mathbf{r})], \tag{4}$$

can be separated into two gauge-invariant integrals. To achieve such separation one needs to consider only the transverse parts of the field, which are always gauge independent. This restriction is justifiable because the degrees of freedom associated with the longitudinal part of the electric field can always be combined with the degrees of freedom of the sources [16,22]. Then, in

$$\langle \mathbf{J_t} \rangle = \int d\mathbf{r} \, \mathbf{r} \times (\mathbf{E}_t(\mathbf{r}) \times \mathbf{B}(\mathbf{r})) \tag{5}$$

$$= \int d\mathbf{r} \, (\mathbf{E}_t(\mathbf{r}) \times \mathbf{A}_t(\mathbf{r})) + \int d\mathbf{r} \sum_{i=1}^{3} E_t^i(\mathbf{r})(\mathbf{r} \times \nabla) A_t^i(\mathbf{r})$$

$$= \langle \mathbf{S_t} \rangle + \langle \mathbf{L_t} \rangle, \tag{6}$$

where index i refers to the three spatial components, the two terms $\langle \mathbf{S_t} \rangle$ and $\langle \mathbf{L_t} \rangle$ are apparently gauge invariant. From now on, we will drop the $_t$ underscripts since in this article we only consider transverse Maxwell fields. The identification of the two parts of Eq. (6) with spin and orbital angular momenta is tempting due to the appearance of the operator $\mathbf{L} = -i\mathbf{r} \times \nabla$ and the relationship of the cross-product in $\langle \mathbf{S_t} \rangle$ with the spin-1 matrices representing **S**. But, since **L** and **S** are not operators in $\mathbb{M}$, the question arises of which are the operators corresponding to the two parts of Eq. (6). In his book on quantum mechanics, Messiah [22] offers an expression which corresponds to the second quantization of the first part of Eq. (6). In 1994, Van Enk and Nienhuis [21], in a more detailed study, arrived at the same result and also derived the expression for the second part of Eq. (6). In that work, they showed that the two operators are not angular momenta because they do not satisfy the commutation relations which define angular momentum operators. These

Fock space operators have their corresponding operators in $\mathbb{M}$ for classical fields (see Sec. IV B), and their third components commute [23]. It is hence possible to split the total angular momentum **J** into two operators $\widehat{\mathbf{S}} = \Lambda \frac{\mathbf{P}}{|\mathbf{P}|}$ and $\widehat{\mathbf{L}} = \mathbf{J} - \widehat{\mathbf{S}}$. Consequently $\langle \mathbf{L}_t \rangle$ and $\langle \mathbf{S}_t \rangle$ in (6) are in reality $\langle \widehat{\mathbf{L}} \rangle$ and $\langle \widehat{\mathbf{S}} \rangle$. Unfortunately, in terms of the separation of the angular momentum operator, this approach is not fully satisfactory because the resulting operators are not angular momentum operators. This can be easily proved by checking that their vectorial components do not fulfill the commutation relations required for angular momentum operators. The consequence of this is that they cannot separately generate meaningful rotations. In summary, **J** can always be decomposed into two meaningful operators, $\mathbf{J} = \widehat{\mathbf{L}} + \widehat{\mathbf{S}}$, independently of the paraxial approximation. These two operators are never angular momenta since they never fulfill the correct commutation relationships. This last statement is also independent of the paraxial approximation. See Sec. IV B for a more detailed discussion.

Following the paradigm discussed in Sec. II, we only consider properties of the field whose corresponding operators generate transformations in $\mathbb{M}$. As already mentioned several times, this disqualifies **L** and **S**. As substitutes, one may choose $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{S}}$. We prefer to disregard the question of the separation completely and use the total angular momentum **J** and the helicity instead. Our choice is based on the fact shown below (Sec. IV B) that **J** and helicity generate very simple symmetry transformations which lead to a simpler framework.

Finally, it is worth noting that expressions (4) and (5) can be interpreted and computed as a weighted average. When the electromagnetic field is decomposed in a basis of eigenvectors of J_z, the total integrated value $\langle J_z \rangle$ is equal to a weighted average [19]. The weights are the squares of the amplitudes in the linear decomposition of the field, and they multiply the different eigenvalues j_z of each mode in the basis. This connection is not restricted to angular momentum. All the integrals that are used to compute total integrated values of properties of the electromagnetic field can be interpreted and computed as weighted averages involving the squares of the expansion coefficients and the eigenvalues of the operator related to each particular property. This connection is the key step used for obtaining a Fock space operator from the classical spatial integral involving the fields. See, for instance, Ref. [24] for the cases of H and **P**, respectively. This connection relates the algebraic formalism introduced in Sec. II with the well-known spatial integrals involving the fields, like (4) for the case of angular momentum.

IV. A FRAMEWORK BASED ON HELICITY AND ANGULAR MOMENTUM

A. The helicity of light fields and its associated symmetry

Helicity is defined as the operator which results from projecting the total angular momentum onto the linear momentum [25], that is, $\Lambda = \mathbf{J} \cdot \mathbf{P}/|\mathbf{P}|$. Helicity commutes with all the generators of rotations **J** and translations **P** [25]. In the case of the photon, the helicity can only take the values ± 1 [26]. For the electromagnetic field, Λ has only two eigenvalues equal to ± 1. A useful interpretation of helicity is obtained by
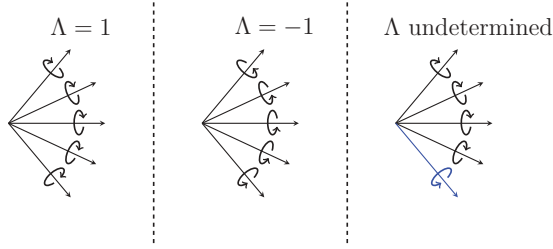
FIG. 1. (Color online) A field composed by the superposition of five plane waves has definite helicity equal to one if all the plane waves have left handedness (left part), equal to minus one if they all have right handedness (central part), and does not have a definite helicity if all the plane waves do not have the same handedness (right part).

considering the expansion of the field as a superposition of plane waves. In such a representation, helicity is associated with the handedness of each plane wave. For a particle to have a well-defined helicity, all the plane waves must be purely circularly polarized and have the same handedness with respect to its momentum vector. This is illustrated in Fig. 1.

Given that a general light beam can always be expanded as a sum of plane waves and that helicity is associated with the polarization handedness of those plane waves, helicity seems a very suitable candidate for representing the polarization degrees of freedom of a general light beam. Crucially, the action of helicity on a plane wave affects only its polarization vector: It leaves the plane-wave momentum vector invariant. This is also true for other types of light beams like multipolar fields or Bessel beams: The action of helicity does not change the quantities related to energy, linear or angular momenta which define those beams. Helicity acts on separated degrees of freedom, which, in this article, we refer to as the degrees of freedom of the polarization.

From its definition, it is readily checked that helicity is a Hermitian operator. With respect to related transformations, helicity is the generator of the generalized electromagnetic duality transformation of the fields:

$$
\begin{aligned}
\mathbf{E} &\to \cos(\theta)\mathbf{E} - \sin(\theta)\mathbf{H}, \\
\mathbf{H} &\to \sin(\theta)\mathbf{E} + \cos(\theta)\mathbf{H},
\end{aligned}
\tag{7}
$$

where θ is a real angle. That helicity generates duality is a remarkable fact which can hardly be deduced from the mathematical definition of helicity containing the angular and linear momenta.

Generalized duality (7) is well known to be a symmetry of the source free microscopic Maxwell equations. The results in [18] allow one to extend the relationship between helicity and duality to the macroscopic Maxwell equations in material systems. That work shows that the helicity of the light interacting with a piecewise homogeneous isotropic medium will transform independently of the shapes of the different material domains, and that helicity will be preserved by the interaction if and only if the ratio between the relative electric and magnetic constants of all the involved materials $i \in [1 \dots N]$ is constant, that is,

$$
\frac{\epsilon_i}{\mu_i} = \alpha \ \forall \ i.
\tag{8}
$$

When condition (8) is met, the macroscopic Maxwell's equations for the system are invariant under transformation (7).

The geometry independent character of the duality symmetry (helicity conservation) allows one to separately consider the transformations of the polarization degrees of freedom from the geometry of the scattering problem. Such notable simplification is very desirable in a framework for the study of light-matter interactions.

We propose the use of helicity, the generator of duality transformations, for treating the polarization degrees of freedom in electromagnetic problems. We will show that, when used in conjunction with the total angular momentum, the generator of rotations, we obtain a general framework for the study of problems involving the angular momentum of light and its interaction with matter using the language of symmetries and conserved quantities. It is important to recall that, in this article, we will remain within the approximations implicit in the macroscopic Maxwell's equations [19].

For completeness, an expression of the total integrated value of helicity in terms of electromagnetic fields can be found in [27]. In this article, though, we are not concerned with the total integrated value of helicity. Instead, we use modes with well-defined helicity and consider their transformations in particular situations, which will depend on whether the system is invariant under electromagnetic duality.

B. Theoretical and practical considerations when using Λ and J_z

The fundamental problem of the separability of S_z and L_z is avoided by using Λ and J_z instead. Both J_z and Λ are commuting Hermitian operators in $\mathbb{M}$, and generate two simple and independent symmetry transformations of the electromagnetic field: J_z generates rotations around the z axis and Λ generates the generalized duality transformation (7).

As previously mentioned, the two Fock space operators obtained from the second quantization of expression (6) studied in [21] (and also later in [28]) have their corresponding operators for classical fields: In [23] they are obtained in the momentum space. For instance, it can be seen that $\widehat{S}_z$, the operator substituting S_z, is $\Lambda P_z/|\mathbf{P}|$. This operator commutes with $\widehat{L}_z$ [23,28], whose expression is obviously $\widehat{L}_z = J_z - \Lambda P_z/|\mathbf{P}|$. Consequently, we could use $\widehat{S}_z$ and $\widehat{L}_z$ instead of J_z and Λ. Considering the symmetries generated by each pair of operators we prefer to choose J_z and Λ for reasons of simplicity. While J_z and Λ are related to the two simple symmetries indicated above, the symmetries related with $\widehat{S}_z$ and $\widehat{L}_z$ are more complicated. An explicit mathematical expression for the transformation generated by $\widehat{\mathbf{S}}$ can be found in [29]. For once, $\widehat{S}_z$ involves a combination of duality and translational symmetries, while the symmetry related to $\widehat{L}_z$ has not been properly studied as such, up to our knowledge. This suggests that using J_z and Λ should, in most situations, result in a simpler analysis. For instance, it is explicitly seen in Sec. V B that the conservation law associated with $\Lambda P_z/|\mathbf{P}|$ is not independent of the geometry of the problem.

In the paraxial limit, when $P_z \to P$, it can be shown that $\widehat{S}_z \to \Lambda$ and $\widehat{L}_z \to J_z - \Lambda$. Even in this limit, neither of these operators generates physical rotations for the full electromagnetic field. For a paraxial beam, the helicity may

be approximated by the real space circular polarization component perpendicular to the z axis and, the value of $J_z - \Lambda$ coincides [see Eqs. (17) and (18) below] with the azimuthal phase winding number of the dominant circular polarization. It is customary to use the paraxial correspondence between SAM and circular polarization and OAM and azimuthal phase. From the previous considerations, we think that it would be more insightful to use Λ and $J_z - \Lambda$, which have the advantage of retaining its meaning outside the paraxial approximation.

From the experimental point of view, Sec. V C shows that the preparation of general (nonparaxial) beams with well-defined helicity can be done using simple optics in a straightforward fashion. The measurement of the helicity state of a general beam is also shown to be easily achieved using simple optics.

In this article we focus in the combined use of Λ and J_z. However, it is worth mentioning that the fact that Λ commutes with all the generators of rotations $\mathbf{J}$ and translations $\mathbf{P}$, and that it generates a transformation which is independent of geometry, should allow its combined use with other degrees of freedom different from angular momentum when the particular problem requires it.

C. Helicity preservation conditions on the TE-TM scattering matrices

In this section we arrive at a relationship between helicity conservation and the scattering of the transverse electric and transverse magnetic components of the field. This relationship is useful in practical problems, as will be seen in Sec. V.

Let us consider the general scattering problem presented in Fig. 2. An incident electromagnetic field $\mathbf{E}_{in}$ impinges onto a scatterer of arbitrary shape S resulting in the scattered field $\mathbf{E}_{sc}$. Consider an orthonormal basis of transverse electromagnetic modes with well-defined helicity $\{\mathbf{A}_\nu^+, \mathbf{A}_\nu^-\}$ $\forall$ ν, where the superindex denotes the sign of Λ and the subindex ν is a composed index which contains three other commuting operators. For example, $\nu = [H, P_x, P_y]$ for plane waves, with H being the energy, and P_x, P_y the first two components of linear momentum.

Both incident and scattered fields can be expanded using the $\{\mathbf{A}_\nu^+, \mathbf{A}_\nu^-\}$ basis, and the scattering process is fully characterized
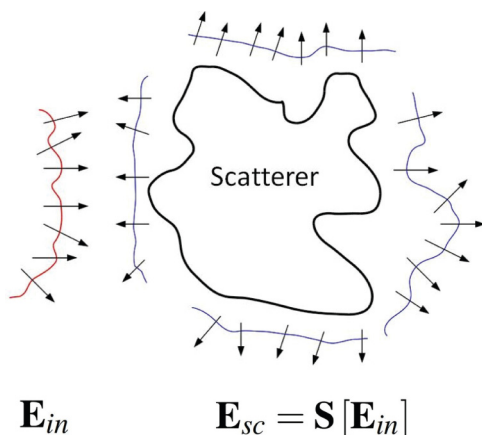

$\mathbf{E}_{in}$ $\quad$ $\mathbf{E}_{sc} = \mathbf{S}[\mathbf{E}_{in}]$

FIG. 2. (Color online) Scattering problem of arbitrary geometry.

by the following transformation of the incident field:

$$\mathbf{E}_{sc} = \mathbf{S}\mathbf{E}_{in},$$

$$\mathbf{E}_{sc} = \left[\int d\bar{\nu} \int d\nu \, S_{\nu,+}^{\bar{\nu},+} \mathbf{A}_{\bar{\nu}}^+ \mathbf{A}_\nu^{+\dagger} + S_{\nu,+}^{\bar{\nu},-} \mathbf{A}_{\bar{\nu}}^- \mathbf{A}_\nu^{+\dagger} + S_{\nu,-}^{\bar{\nu},+} \mathbf{A}_{\bar{\nu}}^+ \mathbf{A}_\nu^{-\dagger} \right.$$

$$\left. + S_{\nu,-}^{\bar{\nu},-} \mathbf{A}_{\bar{\nu}}^- \mathbf{A}_\nu^{-\dagger} \right] \mathbf{E}_{in}, \tag{9}$$

where $\mathbf{A}^\dagger$ is the Hermitian conjugate of $\mathbf{A}$. In (9), the transformation $\mathbf{S}$ is represented by a doubly infinite weighted sum of projection operators of the type $\mathbf{A}\mathbf{A}^\dagger$, whose action on the input field is $(\mathbf{A}\mathbf{A}^\dagger)\mathbf{E}_{sc} = \mathbf{A}(\mathbf{A}^\dagger\mathbf{E}_{in})$ and the inner product,

$$\mathbf{A}^\dagger \mathbf{E}_{in} = \int d\mathbf{r} \mathbf{A}(\mathbf{r})^\dagger \mathbf{E}_{in}(\mathbf{r}), \tag{10}$$

is assumed.

Transformation (9) is specified by the infinite set of 2×2 partial scattering matrices relating the coefficients of $\mathbf{E}_{in}$ in $\mathbf{A}_\nu^+, \mathbf{A}_\nu^-$ to the coefficients of $\mathbf{E}_{sc}$ in $\mathbf{A}_{\bar{\nu}}^+, \mathbf{A}_{\bar{\nu}}^-$ for all $\nu, \bar{\nu}$:

$$S_\nu^{\bar{\nu}} = \begin{pmatrix} S_{\nu,+}^{\bar{\nu},+} & S_{\nu,-}^{\bar{\nu},+} \\ S_{\nu,+}^{\bar{\nu},-} & S_{\nu,-}^{\bar{\nu},-} \end{pmatrix}. \tag{11}$$

Let us impose helicity conservation upon scattering, forcing all the partial scattering matrices to be diagonal: $S_{\nu,-}^{\bar{\nu},+} = S_{\nu,+}^{\bar{\nu},-} = 0$ for all $\nu, \bar{\nu}$.

Consider the following linear combinations of modes with well-defined helicity.

$$\mathbf{M}_\nu = \frac{1}{\sqrt{2}}(\mathbf{A}_\nu^+ + \mathbf{A}_\nu^-), \quad \mathbf{N}_\nu = \frac{1}{\sqrt{2}}(\mathbf{A}_\nu^+ - \mathbf{A}_\nu^-), \tag{12}$$

which transform into each other by the action of Λ,

$$\Lambda \mathbf{M}_\nu = \mathbf{N}_\nu, \quad \Lambda \mathbf{N}_\nu = \mathbf{M}_\nu. \tag{13}$$

In Appendix A we show that, for plane waves, Bessel beams, and multipoles, the sum and subtraction of modes with same ν and different helicity result in what are commonly known as TE and TM modes. We will adopt the TE-TM naming to denote general sum and subtractions of modes differing only by their sharp helicity eigenvalue.

Since $\{\mathbf{A}_\nu^+, \mathbf{A}_\nu^-\}$ is an orthonormal basis, so is $\{\mathbf{M}_\nu, \mathbf{N}_\nu\}$. After changing (9) to the $\{\mathbf{M}_\nu, \mathbf{N}_\nu\}$ basis, the condition for helicity preservation upon scattering, expressed in the TE-TM basis reads

$$\begin{pmatrix} S_{\nu,TE}^{\bar{\nu},TE} & S_{\nu,TE}^{\bar{\nu},TM} \\ S_{\nu,TM}^{\bar{\nu},TE} & S_{\nu,TM}^{\bar{\nu},TM} \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} S_{\nu,+}^{\bar{\nu},+} & 0 \\ 0 & S_{\nu,-}^{\bar{\nu},-} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1}$$

$$= \frac{1}{2} \begin{pmatrix} S_{\nu,+}^{\bar{\nu},+} + S_{\nu,-}^{\bar{\nu},-} & S_{\nu,+}^{\bar{\nu},+} - S_{\nu,-}^{\bar{\nu},-} \\ S_{\nu,+}^{\bar{\nu},+} - S_{\nu,-}^{\bar{\nu},-} & S_{\nu,+}^{\bar{\nu},+} + S_{\nu,-}^{\bar{\nu},-} \end{pmatrix} = \begin{pmatrix} a_\nu^{\bar{\nu}} & b_\nu^{\bar{\nu}} \\ b_\nu^{\bar{\nu}} & a_\nu^{\bar{\nu}} \end{pmatrix}, \tag{14}$$

for all $(\nu, \bar{\nu})$. Namely, that the scattering of TE and TM components is on an equal footing for all $(\nu, \bar{\nu})$, as can be seen from the scattering matrix having the same values $a_\nu^{\bar{\nu}}$ in the diagonal and $b_\nu^{\bar{\nu}}$ in the off-diagonal. Condition (14) is clearly a restriction which will not be met in general. We conclude that, in general, a scatterer will partially convert the helicity

of the incident field and that this partial helicity conversion is reflected in asymmetries of the scattering matrices with respect to the TE and TM modes.

Since helicity conservation is equivalent to invariance under generalized duality transformations, having information on the TE-TM scattering properties of a system can be used to assess its duality invariance and vice versa.

In systems with a high degree of symmetry, like a sphere or a planar multilayer system, a wise choice of basis simplifies condition (14). The symmetries of those two systems make the TE and TM multipoles [19] and the TE and TM plane waves (definition contained in Appendix B) the eigenmodes of the spherical and planar structures, respectively. This means that $S_\nu^{\bar{\nu}} = 0$ unless $\nu = \bar{\nu}$ and that $S_{\nu,\text{TE}}^{\nu,\text{TM}} = S_{\nu,\text{TM}}^{\nu,\text{TE}} = 0$ for all ν, $\bar{\nu}$. All the partial scattering matrices are diagonal. The preservation of the TE-TM components is related to the geometrical symmetries of the system. Using Mie scattering theory and Fresnel's formulas, it is an easy and interesting exercise to analytically verify that, in these two cases, when the materials meet (8), all the scattering matrices are proportional to the identity ($S_{\nu,\text{TE}}^{\nu,\text{TE}} = S_{\nu,\text{TM}}^{\nu,\text{TM}}$ for all ν), hence preserving helicity as well.

V. EXEMPLARY APPLICATION OF THE FRAMEWORK: REVISION OF SPIN TO ORBIT ANGULAR MOMENTUM CONVERSION IN FOCUSING AND SCATTERING

The conversion between spin and orbital angular momentum is widely used to explain phase singularities in numerical simulations of tightly focused fields [9–11], and in scattering experiments [12–14]. A detailed discussion of the SAM to OAM conversion can be found in [30].

We will now use symmetries and conserved quantities arguments, particularly those related to J_z and Λ, to identify

the actual physical reasons for those observations. We will demonstrate that the mechanism responsible for the presence of optical vortices in focused fields is totally different from the one responsible for the observation of optical vortices in scattering experiments. This, in our opinion, shows that the SAM-OAM formulation can be quite misleading: It assigns the same explanation to two distinct physical phenomena.

The analytical tools and methodology employed in this section allow a simple application of the ideas developed in the previous section to practical electromagnetic problems.

A. Bessel beams with well-defined angular momentum and helicity

As already mentioned, J_z and Λ commute. For our analysis we will need a basis of transverse electromagnetic modes which are simultaneous eigenvectors of J_z and Λ. One such set of modes is a particular type of Bessel beams. Bessel beams have been thoroughly studied. See, for instance, the series of papers [28,31,32].

In Appendix B, we constructively derive a complete orthonormal basis for transverse electromagnetic fields consisting of Bessel beams with a well-defined third component of angular momentum J_z and helicity Λ. These modes appear in [21], although their relationship with helicity is not considered in that paper. In [28], these type of electromagnetic modes are obtained in a different way as linear combinations of other types of more commonly used Bessel beams, the transverse electric (TE) and transverse magnetic (TM) modes. The constructive derivation in Appendix B shows clearly that the modes are eigenstates of Λ.

From Appendix B, Eq. (15) is the real space expressions in cylindrical coordinates $[\rho,\theta,z]$. An implicitly harmonic $\exp(-iwt)$ dependence has been assumed.

$$
\begin{aligned}
&\mathbf{C}_{nk_\rho}(\rho,\theta,z) \\
&= \sqrt{\frac{k_\rho}{2\pi}} i^n \exp\left(i(k_z z + n\theta)\right) \left(\frac{i}{\sqrt{2}}\left(\left(1 + \frac{k_z}{k}\right)\left(J_{n+1}(k_\rho\rho)\exp(i\theta)\hat{\mathbf{r}} + \left(1 - \frac{k_z}{k}\right)J_{n-1}(k_\rho\rho)\exp(-i\theta)\hat{\mathbf{l}}\right) - \frac{k_\rho}{k}J_n(k_\rho\rho)\hat{\mathbf{z}}\right), \\
&\mathbf{D}_{nk_\rho}(\rho,\theta,z) \\
&= \sqrt{\frac{k_\rho}{2\pi}} i^n \exp\left(i(k_z z + n\theta)\right) \left(\frac{i}{\sqrt{2}}\left(\left(1 - \frac{k_z}{k}\right)\left(J_{n+1}(k_\rho\rho)\exp(i\theta)\hat{\mathbf{r}} + \left(1 + \frac{k_z}{k}\right)J_{n-1}(k_\rho\rho)\exp(-i\theta)\hat{\mathbf{l}}\right) + \frac{k_\rho}{k}J_n(k_\rho\rho)\hat{\mathbf{z}}\right),
\end{aligned}
\tag{15}
$$

where

(i) $k_\rho = \sqrt{k_x^2 + k_y^2}$, $k^2 = k_\rho^2 + k_z^2$,

(ii) $\hat{\mathbf{l}} = \frac{\hat{\mathbf{x}}+i\hat{\mathbf{y}}}{\sqrt{2}}$, $\hat{\mathbf{r}} = \frac{\hat{\mathbf{x}}-i\hat{\mathbf{y}}}{\sqrt{2}}$.

By construction, the two types of vector wave functions $\mathbf{C}_{nk_\rho}$ and $\mathbf{D}_{nk_\rho}$ have a sharp value of the z component of angular momentum J_z equal to n and a sharp value of helicity Λ equal to -1 and $+1$, respectively. Additionally, they have well-defined values for the energy H and the z component of the linear momentum P_z proportional to k and k_z, respectively. For a given value of k, a well-defined value of P_z implies a well-defined value of the transverse momentum P_ρ proportional to k_ρ. Modes (15) form a complete orthonormal

basis of transverse electromagnetic fields when

$$
\begin{aligned}
&k \in [0, \infty), \quad n \in [0,\pm 1,\pm 2, \ldots], \quad k_\rho \in [0, \infty), \\
&\text{and } \Lambda = \pm 1,
\end{aligned}
\tag{16}
$$

and both signs of k_z in $k_z = \pm\sqrt{k^2 - k_\rho^2}$ are considered. In the following, the consideration of both signs of k_z is implicitly made.

B. Optical vortices in focusing

In order to study why optical vortices seem to appear in numerical studies of focalization of apparently vortex free

beams [9–11], we analyze the aplanatic lens model [33], which is routinely used to study the effects of high numerical aperture lenses.

As we show in Appendix C and has been discussed before [30], the action of an aplanatic lens preserves J_z and Λ. The cylindrical symmetry of the model can be reasonably expected, but its invariance under duality transformations is "hidden" in the assumption that the lens transmission coefficients for the two polarization components, TE and TM, are identical and that there is no crosstalk between input and output TE and TM components. That this assumption implies duality symmetry (helicity conservation) is obvious from condition (14) and the discussion at the end of Sec. IV C. The preservation of Λ by an aplanatic lens has been discussed in [30] without using its relationship to electromagnetic duality. In Appendix C we explicitly analyze the conservation of J_z, highlight the model's key assumption on TE and TM scattering, and show that the model conditions for the conservation of Λ and J_z are, as expected, independent of each other.

Even though both cylindrical and generalized duality symmetries are preserved by the model, the focalized beam is quite different from the input beam: Some other symmetry must be broken. The most obvious candidate is the lack of translational symmetry on the plane perpendicular to the optical axis of the lens. We know that the transverse momentum components P_x and P_y are the generators of that symmetry. Below we show that the differences between the input and focalized beams are due to changes in (k_x, k_y). We will prove this point using the basis introduced in (15).

Let us take a collimated right circularly (RC) polarized Gaussian beam going through an aplanatic lens of high numerical aperture. The linear momentum components of a collimated beam are all almost totally aligned with the propagation direction z: $k_z \approx k$. Consequently, a collimated beam only has components with small transverse momentum value k_ρ. With respect to (15), in the limit when $k_z \approx k$ ($\frac{k_\rho}{k} \to 0$), both $\mathbf{C}_{nk_\rho}$ and $\mathbf{D}_{nk_\rho}$ approach pure RC and LC polarized modes, respectively. This can be easily seen by setting a $\frac{k_\rho}{k} \to 0 \Rightarrow (k_z \approx k)$ in (15):

$$\mathbf{C}_{nk_\rho}(\rho,\theta,z) \approx \sqrt{\frac{k_\rho}{\pi}} i^{n+1} \exp[i(k_z z)] J_{n+1}(k_\rho \rho)$$
$$\times \exp[i\theta(n+1)]\hat{\mathbf{r}}, \qquad (17)$$

$$\mathbf{D}_{nk_\rho}(\rho,\theta,z) \approx \sqrt{\frac{k_\rho}{\pi}} i^{n+1}$$
$$\times \exp[i(k_z z)] J_{n-1}(k_\rho \rho) \exp[i\theta(n-1)]\hat{\mathbf{l}}. \qquad (18)$$

The other polarization components, the opposite circular and the longitudinal $\hat{\mathbf{z}}$ component, are strongly attenuated in this regime. Importantly though, they are actually present: without them, the modes are not solutions of Maxwell's equations, and its transformation properties cannot be consistently analyzed as general electromagnetic fields.

From (17), and since the collimated input RC Gaussian beam does not have a phase singularity in its $\hat{\mathbf{r}}$ dominant polarization, we can conclude that mostly C-type modes with

$n = -1$ will exist in its expansion in the (15) basis:

$$\mathbf{E}_{\text{input}} = \int_0^\infty dk_\rho c_{-1,k_\rho}^{\text{input}} \mathbf{C}_{nk_\rho}, \qquad (19)$$

where $c_{-1,k_\rho}^{\text{input}}$ is only significant when $\frac{k_\rho}{k} \to 0$. Equation (19) defines a beam with a sharp value of J_z, $n = -1$, and a sharp value of Λ, $\lambda = -1$. As per the above discussion regarding symmetries, the output beam must also have sharp J_z and Λ values of $n = -1$ and $\lambda = -1$. Focusing can hence only change the relative weight of k_ρ components. Intuitively, modes with higher transverse momentum are needed to expand the field after focusing.

$$\mathbf{E}_{\text{foc}} = \int_0^\infty dk_\rho c_{-1,k_\rho}^{\text{foc}} \mathbf{C}_{nk_\rho}, \qquad (20)$$

which is in line with the nonpreservation of (k_x, k_y) due to broken transverse translational symmetry. The fact that the change is limited to $k_\rho = \sqrt{k_x^2 + k_y^2}$ could have been foreseen: It stems from the cylindrical symmetry of the model.

Now, let's go back to Eq. (15) and check the spatial shape of $\mathbf{C}_{nk_\rho}$ modes when $n = -1$ and the condition $\frac{k_\rho}{k} \to 0$ is not met:

$$\mathbf{C}_{nk_\rho}(\rho,\theta,z) = -\sqrt{\frac{k_\rho}{2\pi}} i \exp(ik_z z) \left(\frac{i}{\sqrt{2}} \left(1 + \frac{k_z}{k} \right) J_0(k_\rho \rho)\hat{\mathbf{r}} \right.$$
$$+ \frac{i}{\sqrt{2}} \left(1 - \frac{k_z}{k} \right) J_{-2}(k_\rho \rho) \exp(-i2\theta)\hat{\mathbf{l}}$$
$$\left. - \frac{k_\rho}{k} J_{-1}(k_\rho \rho) \exp(-i\theta)\hat{\mathbf{z}} \right). \qquad (21)$$

The $\mathbf{C}_{nk_\rho}, \frac{k_\rho}{k} \to 0$ modes are almost purely right polarized, but when k_ρ increases, the terms multiplying $\hat{\mathbf{l}}$ and $\hat{\mathbf{z}}$ become significant. As it can be seen in (21), for $n = -1$ these newly enhanced terms have phase singularities of orders minus two and minus one, respectively. Figure 3 shows the radial spatial profiles of the three polarization components for two $\mathbf{C}_{-1,k_\rho}(\rho,\theta,z)$ modes, one with $\frac{k_\rho}{k} = 0.1$ and the other with $\frac{k_\rho}{k} = 0.9$. In the small k_ρ case [Fig. 3(a)], the dominant polarization component $\hat{\mathbf{r}}$ is much stronger than the $\hat{\mathbf{l}}$ and $\hat{\mathbf{z}}$ components (which are nonetheless present). In the large k_ρ case [Fig. 3(b)], the relative weight between the intensity of the different polarizations has shifted significantly. The vortices of charge -1 in $\hat{\mathbf{z}}$ and charge -2 in $\hat{\mathbf{l}}$ [see (21)] become relatively more important.

We argue that these are the optical vortices appearing in numerical simulations of focused beams, and that the correct explanation is not SAM to OAM conversion but just the inherent spatial properties of light modes with definite energy, P_z, J_z, and Λ propagating through a system that conserves energy, Λ and J_z while breaking transverse translational invariance. The lens shifts the weight distribution towards modes with larger k_ρ values and optical vortices already existing in the initially strongly attenuated polarization components of the input beam gain relative importance in the focalized beam. For the theoretical study of optical vortices in focused beams, [9] and [30] use solutions of the paraxial equation as the input modes, instead of using solutions of the full Maxwell equations
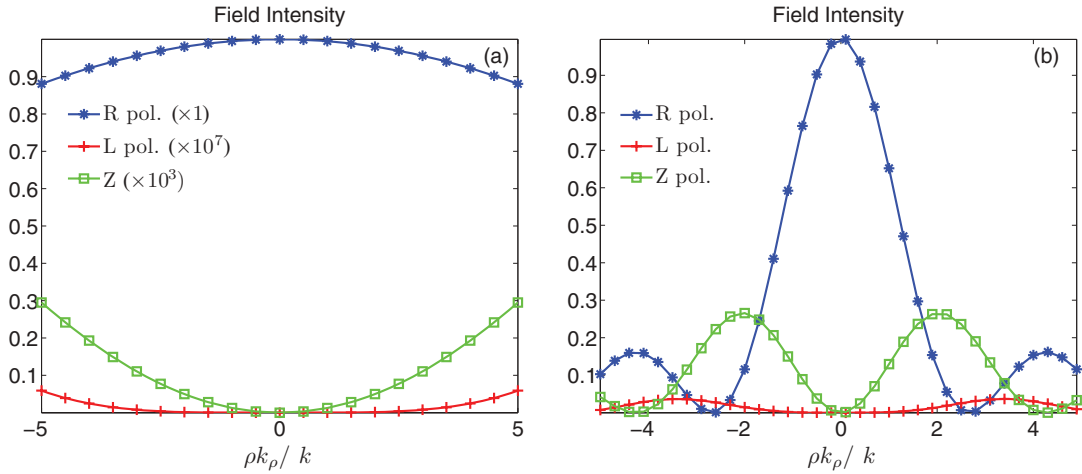
FIG. 3. (Color online) Normalized field intensity for the right, left, and longitudinal polarization components for two $\mathbf{C}_{nk_\rho}$, $n = -1$ modes, one with $\frac{k_\rho}{k} = 0.1$ (a) and the other with $\frac{k_\rho}{k} = 0.9$ (b). Note the scaling of the nondominant polarization components on the $\frac{k_\rho}{k} = 0.1$ case.

as we have done. Since paraxial input modes do not have the attenuated phase singularities in the other polarization components because only a single polarization component is nonzero, the appearance of optical vortices upon focusing was, contrary to this paper's explanation, attributed to SAM to OAM conversion.

It is interesting to note that the property associated with $\Lambda P_z/|\mathbf{P}|$ is not preserved upon focusing. Helicity is conserved but, due to the lack of translational symmetry in the transverse plane, P_ρ changes, which implies that P_z changes. The fact that the lack of a "geometrical" symmetry breaks the conservation law corresponding to $\Lambda P_z/|\mathbf{P}|$ indicates that the symmetry transformation generated by such "spin operator" is, in general, not independent of the geometry of the problem.

C. Optical vortices in scattering

The experimental observation of optical vortices in scattered fields has been reported in the literature [12–14]. In these papers, the observations are explained by means of spin to orbit angular momentum conversion during the interaction with the scatterer. Recently, similar observations have been analyzed in [18] using symmetries and conserved quantities, and reaching a very different conclusion. In line with the discussion of the last paper, we will show in this section that the reason for all these experimental observations is not SAM to OAM transfer, but a partial helicity change during the light-matter interaction due to the breaking of electromagnetic duality symmetry in the system. We will also argue that the helicity change is enhanced by physical effects which strongly break duality, beyond the inherent duality asymmetry of general planar multilayer structures.

We have already mentioned several times that helicity transforms independently of the geometry of the scatterer. In particular, it transforms independently of whether the considered system has cylindrical symmetry or not. Nevertheless, as seen below, a change in helicity is very clearly identifiable in the spatial patterns of the scattered fields when the system has cylindrical symmetry. Several of the experimental setups and input beams in the articles cited in this section have cylindrical

symmetry and are similar to the system in Fig. 4, which we will now analyze.

In Fig. 4, we distinguish the preparation and measurement apparatus from the focusing and scattering part of the setup. In the preparation stage, a collimated Gaussian beam goes through a linear polarizer (LP) and quarter wave plate (QWP), which results in a beam with well-defined values of J_z and Λ (see more details below). In the central part, the beam is focused onto a cylindrically symmetric target and the resulting scattered light is collected by another lens. The measurement part of the setup uses another QWP and LP to project light with different helicity depending on the setting of the LP (see more details below).

At first sight, the central part is cylindrically symmetric but lacks translational symmetry in the transverse plane. We should hence expect conservation of J_z and, as in Sec. (V B),
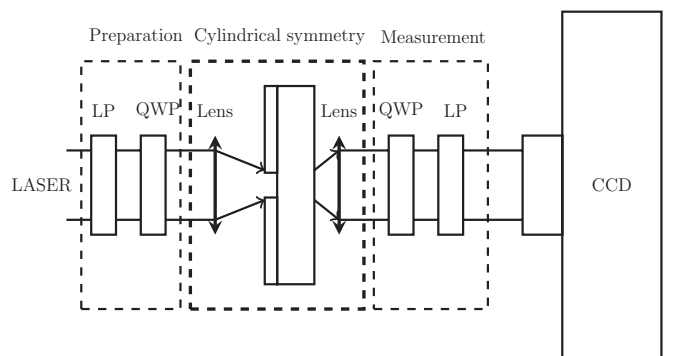

FIG. 4. Archetypal experimental setup. A collimated beam is circularly polarized by means of a linear polarizer (LP) and a quarter wave plate (QWP). After focusing, the beam interacts with a cylindrically symmetric target, in this example, a circular nanoaperture in a metallic film on a substrate. A portion of the scattered field is collected and collimated by a second lens, analyzed by a second set of QWP an LP, and its spatial profile is imaged into a charged couple Ddevice (CCD) camera. The two orthogonal settings of the last LP allow the projection of the two collimated helicity modes (see the text for details).

nonconservation of P_ρ. Since the lenses preserve helicity, the behavior of the system with respect to Λ depends on whether the target sample meets condition (8). Let us assume that it does not meet such condition.

Let us imagine that, after the first LP and QWP, the input to the first lens is a predominantly RC polarized Gaussian beam. Assuming perfect optical elements, and using the same arguments as in Sec. V B, we can see that the input beam can be expanded using basis (15) into modes with $\frac{k_\rho}{k} \to 0$ and sharp values of J_z, $n = -1$, and of Λ, $\lambda = -1$. The first lens will focus the beam into the sample without altering either J_z or Λ, then, since it is assumed that the sample does not meet (8), part of the light will change helicity upon interacting with the cylindrically symmetric target and, after collimation with the second lens, there will be two types of collimated modes, both with $n = -1$, but differing in the value of helicity. Schematically representing the actions of the lenses and the cylindrically symmetric scatterer as transformations of fields expanded in the basis (15) by coordinates c_{nk_ρ} and d_{nk_ρ}, we may summarize the whole sequence as

$$c_{-1,\frac{k_\rho}{k}\to0} \xrightarrow{\text{Lens}} c_{-1,k_\rho} \xrightarrow{\text{Scattering}} \left(c_{-1,k_\rho}, d_{-1,k_\rho}\right)$$
$$\xrightarrow{\text{Lens}} \left(c_{-1,\frac{k_\rho}{k}\to0}, d_{-1,\frac{k_\rho}{k}\to0}\right). \tag{22}$$

The corresponding modal expressions for the components of the two types of output modes are

$$\mathbf{C}_{-1\frac{k_\rho}{k}\to0}(\rho,\theta,z) \approx \sqrt{\frac{k_\rho}{\pi}} i^2 \exp(ik_z z) J_0(k_\rho\rho)\hat{\mathbf{r}}.$$
$$\mathbf{D}_{-1\frac{k_\rho}{k}\to0}(\rho,\theta,z) \approx \sqrt{\frac{k_\rho}{\pi}} i^2 \exp(ik_z z) J_{-2}(k_\rho\rho) \exp(-i2\theta)\hat{\mathbf{l}},$$
$$\tag{23}$$

that is, a predominantly RC beam similar to the input, without any phase singularity in $\hat{\mathbf{r}}$ and a predominantly LC vortex beam with a singularity of charge two in $\hat{\mathbf{l}}$. If we set the last LP to project the LC component of the output collimated beam, this vortex will appear in the CCD camera.

We argue that these are the vortices observed in the cylindrically symmetric scattering experiments of [[12], Fig. 4], [[13], Figs. 2(c) and 2(d)], and [14], and that the underlying reason for their appearance is not SAM to OAM transfer, but that the electromagnetic duality symmetry is broken in those systems.

The samples used in [12] and [13] consist of nanoapertures on metallic thin films. The breaking of duality in a general planar multilayer structure, and in particular when it includes a metal, is clear since the relative electric and magnetic constants of the system will not meet condition (8). Nevertheless, as shown in [18], the effect due to the multilayer alone is typically small in terms of helicity conversion. The reason why the helicity conversion is enhanced making it easily detectable, is that the nanoapertures allow for light to couple to the natural modes of the multilayer system, where duality is strongly broken. In a multilayer system, the natural modes are either TE or TM resonances. Since a mode with well-defined helicity has equal contributions from TE and TM components (12), a TE-only or TM-only resonance implies a strong breaking of electromagnetic duality. The interfaces between a metal and a

dielectric allow the existence of surface plasmon polaritons of TM-only character. As shown in [18], the influence of these modes in the transmitted light is responsible for the experimentally detectable helicity change.

In [14], optical vortices are observed upon propagation of light through a planar semiconductor microcavity, still a cylindrically symmetric system as noted in that work. In this case, duality is strongly broken in the multilayer itself by the energy splitting between TE and TM modes in the structure. This splitting is identified in that work as the enabler for SAM to OAM conversion.

References [12] and [13] contain also results for noncylindrically symmetric setups. Even though their study using J_z is not as simple, we are confident that the generality of the methodology that we propose can rid their analysis from the artificial concept of SAM to OAM conversion, possibly leading to further physical insights. For instance, for the squared nanoapertures studied in [13], the same argumentation about the nanoaperture mediated coupling onto strong duality breaking multilayer natural modes applies. We postulate that any instance where SAM to OAM conversion is invoked can be properly explained using our framework.

VI. CONCLUSIONS AND DISCUSSION

Using the helicity of light fields to represent the polarization degrees of freedom, we have introduced a general framework for the study of light beams with angular momentum and their interactions with matter. In particular, the framework does not depend on the applicability of the paraxial approximation. Our proposal is based on two Hermitian operators in the Hilbert space of transverse electromagnetic fields: the helicity Λ and a component of angular momentum J_z. These operators are the generators of simple transformations of the fields: generalized electromagnetic duality and rotations along the z axis, respectively. Since the generalized duality transformation is independent from rotations or translations [18], the use of Λ allows one to consider the changes in polarization independently from other degrees of freedom like angular momenta ($\mathbf{J}$) and linear momenta ($\mathbf{P}$). This simplification is crucial for the practical applicability of the framework. We recall that the approximations implicit in the macroscopic Maxwell's equations are assumed in this paper.

We propose this new framework as a substitute of the current state of the art treatment, which is based on the separation of spin and orbital angular momenta for the description of the angular momentum of light, and on the mechanism of SAM to OAM transfer in light-matter interactions and other situations. We have shown that it avoids the fundamental problems associated with the separation of SAM and OAM in a simpler fashion than the current theoretical solutions. We have also demonstrated its predictive power based on analyzing the broken and unbroken symmetries of the system to predict which properties of the light will change and which ones will be preserved. The current framework lacks predictive power and, at this point, can only be qualified as a descriptive theory. This is illustrated by the fact that, using our methodology, we have shown that phenomena commonly explained as SAM to OAM transfer in focusing and scattering are actually due to two completely independent physical reasons, showing the

inconsistency of the current framework. The results of this paper suggest that such inconsistency may be related with the use of quantities like SAM and OAM, which are not valid operators for transverse Maxwell fields and hence cannot be related to meaningful transformations of the fields.

We are confident that the use of helicity and its associated transformation will become a powerful theoretical and practical tool which will improve the understanding and control of light-matter interactions at the nanoscale. Applications of the ideas presented in this paper can be foreseen in nano-optics, for the control of the shape and polarization of electromagnetic fields, in metrology, for probing the equivalence of the electric and magnetic responses of a system, and in quantum science and technology, where the identification of two truly independently observable degrees of freedom of the field like Λ and J_z should allow one to better understand the potential benefits of the use of angular momentum in quantum applications.

The association of helicity and duality with other degrees of freedom and their corresponding transformations constitutes a general and robust methodology to study practical light-matter interaction problems by using fundamental concepts.

ACKNOWLEDGMENTS

APPENDIX A: HELICITY EIGENSTATES AND TRANSVERSE ELECTRIC AND TRANSVERSE MAGNETIC MODES

We set out to proof a general relationship between electromagnetic states of well-defined helicity and the transverse electric and transverse magnetic modes. We start by making use of an elegant method for finding solutions of the monochromatic Maxwell equations in a source-free, isotropic, and homogeneous medium [20,34]. Under some suitable conditions of the coordinate system $\{u_1, u_2, u_3\}$, two transverse independent solutions of the vectorial Helmholtz equation of the medium can be obtained from a separable solution $\psi(u_1, u_2, u_3) = U_1(u_1)U_2(u_2)U_3(u_3)$ of the corresponding scalar Helmholtz equation. With $\hat{\mathbf{w}}$ a unit vector perpendicular to the surface of constant coordinate $u_1 = C$, the two vector solutions are obtained as

$$\mathbf{M}(\mathbf{r}) = \nabla \times (\hat{\mathbf{w}}\psi) \quad \text{and} \quad \mathbf{N}(\mathbf{r}) = \frac{\nabla \times \mathbf{M}(\mathbf{r})}{k}. \quad (A1)$$

There are six different coordinate systems for which a complete basis for transverse electromagnetic fields can be built in this way [20]. Plane waves, multipoles, and Bessel beams result from using Cartesian, spherical, and cylindrical coordinates, respectively [34]. Besides being eigenvectors of the Hamiltonian due to its monochromatic character, each coordinate system produces electromagnetic modes which are eigenvectors of a different set of operators:

(1) Plane waves with $\hat{\mathbf{w}} = \hat{\mathbf{z}}$: eigenvectors the of transverse momenta k_x and k_y.

(2) Multipoles with $\hat{\mathbf{w}} = \frac{\mathbf{r}}{|\mathbf{r}|}$: eigenvectors of the squared angular momentum norm J^2 and a component of angular momentum, for instance, J_z.

(3) Bessel beams with $\hat{\mathbf{w}} = \hat{\mathbf{z}}$: eigenvectors of the third components of the linear and angular momenta P_z and J_z.

From now on, we will lump the energy H, which is proportional to k, and these other degrees of freedom into a collective index ν and use $\mathbf{M}_\nu$ and $\mathbf{N}_\nu$. In all three reference systems, the $\mathbf{M}_\nu$ and $\mathbf{N}_\nu$ modes are commonly referred to as TE and TM modes, respectively.

Using $\nabla \times = \vec{\mathbf{S}} \cdot \vec{\mathbf{P}}$ [22], it can be proved that the helicity operator Λ can be written in real space as $\Lambda = \frac{\nabla \times}{k}$. In such formulation, it is also true that $\mathbf{M}_\nu = \Lambda \mathbf{N}_\nu$ [20]. Then, together with (A1), we see that

$$\mathbf{N}_\nu = \Lambda \mathbf{M}_\nu, \quad \mathbf{M}_\nu = \Lambda \mathbf{N}_\nu. \quad (A2)$$

Namely, the TE and TM modes are transformed into each other by the application of Λ. We can hence obtain electromagnetic fields with well-defined helicity as

$$\begin{aligned}
\mathbf{A}_\nu^+ &= \frac{1}{\sqrt{2}}(\mathbf{M}_\nu + \mathbf{N}_\nu), \quad \Lambda \mathbf{A}_\nu^+ = \mathbf{A}_\nu^+, \\
\mathbf{A}_\nu^- &= \frac{1}{\sqrt{2}}(\mathbf{M}_\nu - \mathbf{N}_\nu), \quad \Lambda \mathbf{A}_\nu^- = -\mathbf{A}_\nu^-.
\end{aligned} \quad (A3)$$

Equation (A3) already makes it obvious that helicity conservation will only happen for equivalent scattering of the TE and TM components of the field, which is needed in order to preserve their linear combinations.

It is clear that, since $\{\mathbf{M}_\nu, \mathbf{N}_\nu\}$ are an orthonormal basis for transverse fields when all values of ν are considered, so is $\{\mathbf{A}_\nu^+, \mathbf{A}_\nu^-\}$. This derivation of the TE, TM modes is valid for the systems of coordinates mentioned above. On the other hand, the helicity operator is well defined for any basis. Thus, we think that it is more natural to define the TE and TM modes as $\mathbf{M}_\nu = \frac{1}{\sqrt{2}}(\mathbf{A}_\nu^+ + \mathbf{A}_\nu^-)$, $\mathbf{N}_\nu = \frac{1}{\sqrt{2}}(\mathbf{A}_\nu^+ - \mathbf{A}_\nu^-)$, and as such we use them in the main text.

APPENDIX B: DERIVATION OF BESSEL BEAMS WITH WELL-DEFINED HELICITY

In this appendix we present a constructive derivation of electromagnetic modes with well-defined energy H, third component of angular and linear momenta J_z and P_z, and helicity Λ.

We start by considering TE and TM modes of well-defined energy and linear momentum. They are plane waves derived following the constructive procedure given in [20,34] which was already used in Sec. IV C. Their explicit expressions are, in the Cartesian $[\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}]$ basis,

$$\hat{\mathbf{s}} \cdot \exp(i\mathbf{k} \cdot \mathbf{r}) = \frac{i}{k_\rho}(k_y \hat{\mathbf{x}} - k_x \hat{\mathbf{y}}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (B1)$$

$$\hat{\mathbf{p}} \cdot \exp(i\mathbf{k} \cdot \mathbf{r}) = \left[\frac{-k_z(k_x \hat{\mathbf{x}} + k_y \hat{\mathbf{y}}) + k_\rho^2 \hat{\mathbf{z}}}{kk_\rho} \right] \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (B2)$$

where $k_\rho = \sqrt{k_x^2 + k_y^2}$, and $[k_x, k_y, k_z]$ and k are proportional to the linear momentum vector and energy of the plane waves. As per Eq. (A3), sum and subtraction of TE and TM modes

result in states of well-defined helicity:

$$
\widehat{\mathbf{e}}_+ \exp(i\mathbf{k}\cdot\mathbf{r}) = \frac{1}{\sqrt{2}}(\widehat{\mathbf{s}}+\widehat{\mathbf{p}})\exp(i\mathbf{k}\cdot\mathbf{r}),
$$
$$
\widehat{\mathbf{e}}_- \exp(i\mathbf{k}\cdot\mathbf{r}) = \frac{1}{\sqrt{2}}(\widehat{\mathbf{s}}-\widehat{\mathbf{p}})\exp(i\mathbf{k}\cdot\mathbf{r}). \tag{B3}
$$

Interestingly, any plane wave of well-defined helicity and momentum vector proportional to $[k_x,k_y,k_z]$ can also be obtained [35] by rotating a plane wave of the same helicity and momentum $[0,0,k]$:

$$
\widehat{\mathbf{e}}_+ \exp(i\mathbf{k}\cdot\mathbf{r}) = \mathbf{R}(\theta_k,\phi_k)(-\widehat{\mathbf{l}}\exp(ikz)),
$$
$$
\widehat{\mathbf{e}}_- \exp(i\mathbf{k}\cdot\mathbf{r}) = \mathbf{R}(\theta_k,\phi_k)(\widehat{\mathbf{r}}\exp(ikz)), \tag{B4}
$$

where $\widehat{\mathbf{l}}=\frac{\widehat{\mathbf{x}}+i\widehat{\mathbf{y}}}{\sqrt{2}}$, $\widehat{\mathbf{r}}=\frac{\widehat{\mathbf{x}}-i\widehat{\mathbf{y}}}{\sqrt{2}}$, $\theta_k=\arcsin\frac{k_\rho}{k}$, and $\phi_k=\arctan\frac{k_y}{k_x}$. Equation (B4) exploits the fact that helicity does not change under spatial rotations.

A rotation operation of a vectorial field $\mathbf{R}(\theta_k,\phi_k)(\mathbf{A}(\mathbf{r}))$ is, explicitly,

$$
\mathbf{A}(\mathbf{r}) \rightarrow R(\phi_k,\theta_k)\mathbf{A}(R^{-1}(\phi_k,\theta_k)\mathbf{r}), \tag{B5}
$$

where $R(\phi_k,\theta_k)$ is the rotation matrix,

$$
R(\phi_k,\theta_k) = R_z(\phi_k)R_y(\theta_k)
$$
$$
= \begin{pmatrix} \cos\theta_k\cos\phi_k & -\sin\phi_k & \sin\theta_k\cos\phi_k \\ \cos\theta_k\sin\phi_k & \cos\phi_k & \sin\theta_k\sin\phi_k \\ -\sin\theta_k & 0 & \cos\theta_k \end{pmatrix}, \tag{B6}
$$

where $R_z(\phi_k)$ and $R_y(\theta_k)$ are rotations around the $\widehat{\mathbf{z}}$ and $\widehat{\mathbf{y}}$ axis, respectively.

We now start our construction of modes with well-defined H, P_z, J_z, and Λ by expressing the most general monochromatic forward propagating transverse electromagnetic field as a combination of plane waves with well-defined momentum:

$$
\mathbf{A} = \int_0^{\frac{\pi}{2}}\sin\theta_k d\theta_k \int_0^{2\pi} d\phi_k \Big[\alpha(\theta_k,\phi_k)\mathbf{R}(\theta_k,\phi_k)(-\widehat{\mathbf{l}}\exp(ikz))
$$
$$
+ \beta(\theta_k,\phi_k)\mathbf{R}(\theta_k,\phi_k)(\widehat{\mathbf{r}}\exp(ikz))\Big]. \tag{B7}
$$

A backward propagating beam can be obtained by using $-\widehat{\mathbf{l}}\exp(-ikz)$ and $\widehat{\mathbf{r}}\exp(-ikz)$ as the initial plane waves instead.

Having restricted (B7) to a single wave number k assures that the resulting mode has a definite energy H proportional to k. With respect to Λ, it is clear from (B7) that, if we desire a field $\mathbf{A}$ with well-defined helicity, we need only to set either $\alpha(\theta_k,\phi_k)$ or $\beta(\theta_k,\phi_k)$ equal to zero for all (θ_k,ϕ_k), so that only plane waves of the same helicity type are present in (B7). With respect to having a well-defined P_z, and because, in a plane wave, its value is proportional to $k_z = k\cos\theta_k$, we must include only a single value of θ_k in (B7).

The conditions needed in (B7) to specify a beam with well-defined J_z are not as apparent as in the case of H, P_z, and Λ. The solution can be reached by applying the operator J_z to the general mode $\mathbf{A}$. In order to do that, we use the following definition of J_z as an infinitesimal rotation operation:

$$
J_z = \lim_{d\phi_k\to 0}\frac{I-\mathbf{R}(0,d\phi_k)}{id\phi_k}, \tag{B8}
$$

where I is the identity operator.

Applying (B8) to (B7) and making use of the properties of rotation operators we obtain

$$
J_z[\mathbf{A}] = \lim_{d\phi_k\to 0}\int_0^{\frac{\pi}{2}}\sin\theta_k d\theta_k \int_0^{2\pi} d\phi_k \alpha(\theta_k,\phi_k)
$$
$$
\times \frac{\mathbf{R}(\theta_k,\phi_k)-\mathbf{R}(\theta_k,\phi_k+d\phi_k)}{id\phi_k}(-\widehat{\mathbf{l}}\exp(ikz))
$$
$$
+ \beta(\theta_k,\phi_k)\frac{\mathbf{R}(\theta_k,\phi_k)-\mathbf{R}(\theta_k,\phi_k+d\phi_k)}{id\phi_k}(\widehat{\mathbf{r}}\exp(ikz)), \tag{B9}
$$

which, after integrating by parts can be reduced to

$$
J_z[\mathbf{A}] = \int_0^{\frac{\pi}{2}}\sin\theta_k d\theta_k \int_0^{2\pi} d\phi_k \frac{\partial\alpha(\theta_k,\phi_k)}{i\partial\phi_k}\mathbf{R}(\theta_k,\phi_k)
$$
$$
\times (-\widehat{\mathbf{l}}\exp(ikz)) + \frac{\partial\beta(\theta_k,\phi_k)}{i\partial\phi_k}\mathbf{R}(\theta_k,\phi_k)(\widehat{\mathbf{r}}\exp(ikz)). \tag{B10}
$$

So, in this representation, the operator J_z acts by taking the partial derivative of the coordinates $(\alpha(\theta_k,\phi_k),\beta(\theta_k,\phi_k))$ with respect to ϕ_k and dividing by i.

Gathering together all the above, we can assert that

$$
\alpha(\theta_k,\phi_k) = \frac{1}{2\pi}\delta(\theta_k-\theta_k')\exp(in\phi_k), \quad \beta(\theta_k,\phi_k) = 0, \tag{B11}
$$

specifies a mode with well-defined energy H proportional to k, well-defined Λ with value $\lambda = 1$, well-defined J_z with value n, and well-defined P_z proportional to $k_z = k\cos(\theta_k')$, and that

$$
\alpha(\theta_k,\phi_k) = 0, \quad \beta(\theta_k,\phi_k) = \frac{1}{2\pi}\delta(\theta_k-\theta_k')\exp(in\phi_k) \tag{B12}
$$

does the same for the opposite helicity.

After inserting the specified coordinates into Eq. (B7), substituting the rotated plane wave by their explicit expressions as linear combinations of (B1) and (B2), changing basis from $[\widehat{\mathbf{x}},\widehat{\mathbf{y}},\widehat{\mathbf{z}}]$ to $[\widehat{\mathbf{r}},\widehat{\mathbf{l}},\widehat{\mathbf{z}}]$ and using $\exp(i\mathbf{k}\cdot\mathbf{r}) = \sum_m i^m J_m(k_\rho\rho)\exp(im(\phi-\phi_k))\exp(ik_z z)$ before performing the integral in $d\phi_k$, we finally obtain the real space expressions of the modes in cylindrical coordinates that have been extensively used in the main text:

$$
\mathbf{C}_{nk_\rho}(\rho,\phi,z)
$$
$$
= \sqrt{\frac{k_\rho}{2\pi}}i^n \exp(i(k_z z+n\phi))\left(\frac{i}{\sqrt{2}}\left(\left(1+\frac{k_z}{k}\right)\left(J_{n+1}(k_\rho\rho)\exp(i\phi)\widehat{\mathbf{r}} + \left(1-\frac{k_z}{k}\right)J_{n-1}(k_\rho\rho)\exp(-i\phi)\widehat{\mathbf{l}}\right) - \frac{k_\rho}{k}J_n(k_\rho\rho)\widehat{\mathbf{z}}\right),
$$

$$\mathbf{D}_{nk_\rho}(\rho,\phi,z)$$
$$= \sqrt{\frac{k_\rho}{2\pi}} i^n \exp\left(i(k_z z + n\phi)\right) \left(\frac{i}{\sqrt{2}}\left(\left(1 - \frac{k_z}{k}\right)\left(J_{n+1}(k_\rho\rho)\exp(i\phi)\hat{\mathbf{r}} + \left(1 + \frac{k_z}{k}\right)J_{n-1}(k_\rho\rho)\exp(-i\phi)\hat{\mathbf{l}}\right) + \frac{k_\rho}{k}J_n(k_\rho\rho)\hat{\mathbf{z}}\right).$$

$$(B13)$$

Note that, due to the integration limits of θ_k in (B7), the derivation is restricted to propagating modes. In reality, values of $k_\rho > k$ in Eq. (B13) are possible, specifying nonpropagating modes with well-defined H, P_z, J_z, and Λ. Including nonpropagating modes from the start of the derivation can be done by using imaginary values of θ_k in order to obtain values of $\sin(\theta)$ bigger than one.

APPENDIX C: PRESERVATION OF J_z AND Λ BY THE ACTION OF AN APLANATIC LENS

The aplanatic or spherical lens model allows one to relate the real space field profile of the collimated input beam to the angular spectrum of the focused beam. Originally developed by Richards and Wolf in [33], we can find an explanation of the model and its formulas in [36], which we reproduce here using a slightly different notation using that $\exp(ikz\cos\theta_k)\exp(ik\rho\sin\theta_k\cos(\phi - \phi_k)) = \exp(i\mathbf{k}\cdot\mathbf{r})$:

$$\mathbf{E}_{\text{out}}(\rho,\phi,z) = \frac{ikf\exp(-ikf)}{2\pi}\int_0^{\theta_k^m}\int_0^{2\pi}\sin\theta_k d\theta_k d\phi_k$$
$$\times \mathbf{E}_\infty(\theta_k,\phi_k)\exp(i\mathbf{k}\cdot\mathbf{r}), \quad (C1)$$

$$\mathbf{E}_\infty(\theta_k,\phi_k) = \left[t^s(\theta_k)(\widehat{\mathbf{s}}_{0,\phi_k}\cdot\mathbf{E}_{inc}(f\sin(\theta_k),\phi_k))\widehat{\mathbf{s}}_{\theta_k,\phi_k}\right.$$
$$\left. + t^p(\theta_k)(\widehat{\mathbf{p}}_{0,\phi_k}\cdot\mathbf{E}_{inc}(f\sin(\theta_k),\phi_k))\widehat{\mathbf{p}}_{\theta_k,\phi_k}\right]$$
$$\times \sqrt{\frac{n_1}{n_2}}(\cos\theta_k)^{1/2}, \quad (C2)$$

where

(1) $\mathbf{E}_{\text{out}}(\rho,\phi,z)$ is the focused field in real space expressed in cylindrical coordinates $[\rho,\phi,z]$.

(2) f is the focal distance of the lens, $\theta_k^m = \arcsin(\text{NA})$, where NA is the numerical aperture of the lens, and k the wave number of the field.

(3) $[\rho,\phi,z = z_0]$, with $\rho = f\sin(\theta_k)$ and $\phi = \phi_k$ are both the real space cylindrical coordinate system for the input beam $\mathbf{E}_{inc}(\rho,\phi,z = z_0)$ and the spherical coordinates in momentum space of $\mathbf{E}_\infty(\theta_k,\phi_k)$, the angular spectrum of the focalized output beam, with $\theta_k = \arcsin(\frac{k_\rho}{k}), \phi_k = \arctan(\frac{k_y}{k_x})$. This dual role of the coordinates is the essence of the model.

(4) $(t^s(\theta_k),t^p(\theta_k))$ are the lenses TE and TM transmission coefficients and (n_1,n_2) the indexes of refraction of the input and output media.

We emphasize that the model is valid for a collimated input only.

The definitions of the polarization vectors $\widehat{\mathbf{s}}_{\alpha,\psi}$ and $\widehat{\mathbf{p}}_{\alpha,\psi}$ are those of (B1) and (B2) with $\alpha = \arccos(\frac{k_z}{k})$ and $\psi = \arctan(\frac{k_y}{k_x})$.

We now analyze the properties of the model with respect to conservation of J_z and Λ. We start by using some of the ideas and techniques of Appendix B in order to calculate the angular momentum of the focused beam when the input beam has a definite value of J_z.

Using (B7), expression (C1) can also be written as

$$\mathbf{E}_{\text{out}}(\rho,\phi,z) = \int_0^\pi \sin(\theta_k)d\theta_k \int_0^{2\pi} d\phi[g_s(\theta_k,\phi_k)\mathbf{R}(\theta_k,\phi_k)$$
$$\times (-i\hat{\mathbf{y}}\exp(i\mathbf{k}\cdot\mathbf{r}))$$
$$+ g_p(\theta_k,\phi_k)\mathbf{R}(\theta_k,\phi_k)(-\hat{\mathbf{x}}\exp(i\mathbf{k}\cdot\mathbf{r}))], \quad (C3)$$

where

$$g_s(\theta_k,\phi_k) = t^s(\theta_k)(\widehat{\mathbf{s}}_{0,\phi_k}\cdot\mathbf{E}_{inc}(f\sin(\theta_k),\phi_k)),$$
$$g_p(\theta_k,\phi_k) = t^p(\theta_k)(\widehat{\mathbf{p}}_{0,\phi_k}\cdot\mathbf{E}_{inc}(f\sin(\theta_k),\phi_k)). \quad (C4)$$

As seen in Appendix B, in order to apply the operator J_z to the output focused beam in (C3), we need to take the partial derivative of its coordinates $(g_s(\theta_k,\phi_k),g_p(\theta_k,\phi_k))$ with respect to ϕ_k and divide by i. Equation (C4) relates $(g_s(\theta_k,\phi_k),g_p(\theta_k,\phi_k))$ to the input beam. Note that the effect of the lens, which is given by the transmittivities $t^s(\theta_k),t^p(\theta_k)$, does not add any azimuthal dependence to $g_s(\theta_k,\phi_k)$ and $g_p(\theta_k,\phi_k)$. Thus, with this model, the lens will not affect the azimuthal dependence of the input beam, hence keeping the angular momentum constant. More explicitly, since vectors $\widehat{\mathbf{s}}_{0,\phi_k},\widehat{\mathbf{p}}_{0,\phi_k}$ are proportional to the azimuthal and radial polarization vectors defined as

$$\hat{\phi} = \begin{pmatrix} -\sin\phi_k \\ \cos\phi_k \\ 0 \end{pmatrix}, \quad \hat{\rho} = \begin{pmatrix} \cos\phi_k \\ \sin\phi_k \\ 0 \end{pmatrix}, \quad (C5)$$

the focalized field angular spectrum coordinates at point (θ_k,ϕ_k), $(g_s(\theta_k,\phi_k),g_p(\theta_k,\phi_k))$, are the real space azimuthal $E_{inc}^{\hat{\phi}}$ and radial $E_{inc}^{\hat{\rho}}$ components of the input field at point $\rho = f\sin(\theta_k),\phi = \phi_k$. In order to compute the angular momentum of the focused field we need to take the partial derivatives of the radial and azimuthal components of the input field with respect to $\phi_k = \phi$ and divide by i. Let us now find explicit expressions for $E_{inc}^{\hat{\phi}}$ and $E_{inc}^{\hat{\rho}}$.

The collimated input beam can be expanded in the plane $z = z_0$ into modes of the types (17) and (18) with $\frac{k_\rho}{k} \to 0$, which we rewrite here after the change of basis $\hat{\mathbf{r}} = \frac{1}{\sqrt{2}}(\hat{\rho} - i\hat{\phi})\exp(-i\phi), \hat{\mathbf{l}} = \frac{1}{\sqrt{2}}(\hat{\rho} + i\hat{\phi})\exp(i\phi)$:

$$\mathbf{C}_{nk_\rho}(\rho,\phi,z = z_0)$$
$$\approx \sqrt{\frac{k_\rho}{\pi}} i^{n+1} \exp(ik_z z_0) J_{n+1}(k_\rho\rho)\exp(i\phi n)\frac{1}{\sqrt{2}}(\hat{\rho} + i\hat{\phi}),$$

$$\mathbf{D}_{nk_\rho}(\rho, \phi, z = z_0)$$

$$\approx \sqrt{\frac{k_\rho}{\pi}} i^{n+1} \exp(ik_z z_0) J_{n-1}(k_\rho \rho) \exp(i\phi n) \frac{1}{\sqrt{2}} (\hat{\rho} - i\hat{\phi}).$$

$$(C6)$$

Note that in order to avoid confusion with θ_k, the letter for the input beam coordinate angle $\arctan(y, x)$ is now ϕ, instead of the original letter θ in expressions (17) and (18).

The sum and subtraction of the above modes result in approximately pure radially and azimuthally polarized modes, which separates the radial and azimuthal components of the field:

$$\mathbf{A}_{nk_\rho}(\rho, \phi, z = z_0)$$

$$\approx \sqrt{\frac{2k_\rho}{\pi}} i^{n+1} \exp(ik_z z_0)[J_{n+1}(k_\rho \rho) + J_{n+1}(k_\rho \rho)]$$

$$\times \exp(i\phi n) \frac{1}{\sqrt{2}} \hat{\rho} = A_{nk_\rho}^{\hat{\rho}}(\rho, \phi, z_0)\hat{\rho},$$

$$\mathbf{B}_{nk_\rho}(\rho, \phi, z = z_0)$$

$$\approx -\sqrt{\frac{2k_\rho}{\pi}} i^{n+2} \exp(ik_z z_0)[J_{n+1}(k_\rho \rho) - J_{n+1}(k_\rho \rho)]$$

$$\times \exp(i\phi n) \frac{1}{\sqrt{2}} \hat{\phi} = B_{nk_\rho}^{\hat{\phi}}(\rho, \phi, z_0)\hat{\phi}, \qquad (C7)$$

which means that (C4) is explicitly

$$g_s(\theta_k, \phi_k) = t^s(\theta_k) E_{inc}^{\hat{\phi}} = t^s(\theta_k) B_{nk_\rho}^{\hat{\phi}}(f \sin(\theta_k), \phi_k, z_0),$$

$$g_p(\theta_k, \phi_k) = t^p(\theta_k) E_{inc}^{\hat{\rho}} = t^p(\theta_k) A_{nk_\rho}^{\hat{\rho}}(f \sin(\theta_k), \phi_k, z_0). \qquad (C8)$$

Then, we need to take the partial derivative of the azimuthal and radial input field components $B_{nk_\rho}^{\hat{\phi}}(f \sin(\theta_k), \phi_k, z = z_0)$

and $A_{nk_\rho}^{\hat{\rho}}(f \sin(\theta_k), \phi_k, z = z_0)$ with respect to ϕ_k and divide by i. Let us now assume that the input is a field with definite angular momentum $J_z = n$. Inspection of (C7) reveals that such operation will leave those components invariant except for a multiplicative factor equal to n. So, the output field is also a field with definite angular momentum equal to n. We have just proved that the aplanatic lens model transfers the J_z value of the input beam to the output beam without changing it.

We now study the behavior of the model with respect to helicity. From (C4), it becomes clear that the TE and TM content of the output field are independently determined by the real azimuthal and radial components (C7) of the input field. Having taken linear combinations of modes with defined helicity, and using the results in Appendix A, it follows that the modes in (C7) are also pure TE and TM modes, respectively. Using the results from Sec. IV C about helicity preservation in the TE-TM basis, we now see that, since there is no crosstalk between the TE and TM components of the input and focalized fields, the key condition for helicity preservation in the aplanatic lens model is

$$t_s(\theta_k) = t_p(\theta_k) \, \forall \, \theta_k. \qquad (C9)$$

In real manufacturing of microscope objectives, this condition is related to the coating of the lens surfaces [36], that is, a property of the materials and not the geometry of the system.

Let us assume for a moment a more general dependence of the transmission coefficients $t_s(\theta_k, \phi_k), t_p(\theta_k, \phi_k)$. The new ϕ_k dependence could destroy the J_z preservation since now more terms will be involved in the partial derivatives of $g_s(\theta_k, \phi_k)$ and $g_p(\theta_k, \phi_k)$. On the other hand, as long as $g_s(\theta_k, \phi_k) = g_p(\theta_k, \phi_k)$ for all (θ_k, ϕ_k), helicity will be preserved. This is another example of the independence of the conservation laws of J_z and Λ.

[1] F. Tamburini, B. Thide, G. Molina-Terriza, and G. Anzolin, Nat. Phys. **7**, 195 (2011).

[2] M. F. Andersen, C. Ryu, P. Cladé, V. Natarajan, A. Vaziri, K. Helmerson, and W. D. Phillips, Phys. Rev. Lett. **97**, 170406 (2006).

[3] I. B. Djordjevic, M. Arabaci, L. Xu, and T. Wang, Opt. Express **19**, 6845 (2011).

[4] G. Molina-Terriza, A. Vaziri, J. Rehácek, Z. Hradil, and A. Zeilinger, Phys. Rev. Lett. **92**, 167903 (2004).

[5] J. P. Torres and L. Torner, *Twisted Photons. Applications of Light with Orbital Angular Momentum* (Wiley, Verlag, 2011).

[6] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, Phys. Rev. A **45**, 8185 (1992).

[7] S. M. Barnett and L. Allen, Opt. Commun. **110**, 670 (1994).

[8] L. Allen, M. Padgett, and M. Babiker, Prog. Opt. **39**, 291 (1999).

[9] Y. Zhao, J. S. Edgar, G. D. M. Jeffries, D. McGloin, and D. T. Chiu, Phys. Rev. Lett. **99**, 073901 (2007).

[10] T. A. Nieminen, A. B. Stilgoe, N. R. Heckenberg, and H. Rubinsztein-Dunlop, J. Opt A-Pure Appl. Op. **10**, 115005 (2008).

[11] A. M. Yao and M. J. Padgett, Adv. Opt. Photon. **3**, 161 (2011).

[12] Y. Gorodetski, N. Shitrit, I. Bretner, V. Kleiner, and E. Hasman, Nano Letters **9**, 3016 (2009).

[13] L. T. Vuong, A. J. L. Adam, J. M. Brok, P. C. M. Planken, and H. P. Urbach, Phys. Rev. Lett. **104**, 083903 (2010).

[14] F. Manni, K. G. Lagoudakis, T. K. Paraïso, R. Cerna, Y. Léger, T. C. H. Liew, I. A. Shelykh, A. V. Kavokin, F. Morier-Genoud, and B. Deveaud-Plédran, Phys. Rev. B **83**, 241307 (2011).

[15] V. B. Berestetskii, L. P. Pitaevskii, and E. M. Lifshitz, *Quantum Electrodynamics*, Vol. 4, 2nd ed. (Butterworth-Heinemann, Oxford, 1982).

[16] C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, *Photons and Atoms–Introduction to Quantum Electrodynamics (Wiley Professional)* (Wiley-Interscience, New York, 1997).

[17] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).

[18] I. Fernandez-Corbaton, X. Zambrana-Puyalto, N. Tischler, A. Minovich, X. Vidal, M. L. Juan, and G. Molina-Terriza, arXiv:1206.0868.

[19] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1998).

[20] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill and Kogakusha Book Companies, New York, 1953).

[21] S. J. Van Enk and G. Nienhuis, J. Mod. Opt. **41**, 963 (1994).

[22] A. Messiah, *Quantum Mechanics* (Dover, Mineola, 1999).

[23] K. Y. Bliokh, M. Alonso, E. Ostrovskaya, and A. Aiello, Phys. Rev. A **82**, 063825 (2010).

[24] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995).

[25] W.-K. Tung, *Group Theory in Physics* (World Scientific, Singapore, 1985).

[26] S. Weinberg, *The Quantum Theory of Fields*, Vol. 1, 1st ed. (Cambridge University Press, Cambridge, 1995).

[27] G. Afanasiev and Y. Stepanovsky, Il Nuovo Cimento A **109**, 271 (1996).

[28] R. Jáuregui and S. Hacyan, Phys. Rev. A **71**, 033411 (2005).

[29] R. P. Cameron, S. M. Barnett, and A. M. Yao, New J. Phys. **14**, 053050 (2012).

[30] K. Y. Bliokh, E. A. Ostrovskaya, M. A. Alonso, O. G. Rodríguez-Herrera, D. Lara, and C. Dainty, Opt. Express **19**, 26132 (2011).

[31] R. Jáuregui, Phys. Rev. A **70**, 033415 (2004).

[32] S. Hacyan and R. Jáuregui, J. Phys. B: At. Mol. Opt. Phys. **39**, 1669 (2006).

[33] B. Richards and E. Wolf, Proc. R. Soc. London **253**, 358 (1959).

[34] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).

[35] G. Molina-Terriza, Phys. Rev. A **78**, 053819 (2008).

[36] L. Novotny and B. Hecht, *Principles of Nano-optics* (Cambridge University Press, Cambridge, 2006).