

Extinction, scattering and absorption of electromagnetic waves in the coupled-dipole approximation

Vadim A. Markel¹

Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

ARTICLE INFO

Keywords:

Coupled-dipole approximation
Coupled-dipole equations
Extinction
Absorption
Scattering
Extinction paradox
Purcell factor

ABSTRACT

In this article, we consider the theoretical underpinning of the coupled-dipole approximation as it is used in the multiple scattering theory. Specific topics include the definitions of the bare and renormalized polarizabilities, radiative and non-radiative corrections, coupled-dipole equations in the vicinity of a substrate, and rigorous derivation of the energy relations, particularly, in the case when the particle polarizabilities can be tensorial. It is shown how the Purcell factors can be related to the renormalized polarizabilities. As an application, several extinction-related paradoxes are considered by using the coupled-dipole approximation as the underlying physical model.

Published in *Journal of Quantitative Spectroscopy and Radiative Transfer* **236**, 106611 (2019)

doi: 10.1016/j.jqsrt.2019.106611

Data available from Mendeley Data, V1,

doi: 10.17632/277jsr32md.1

A Green's tensor in free space

23

Contents

1	Introduction	1
1.1	Motivation and review	1
1.2	Notations and units	3
2	Two versions of the coupled-dipole equations	3
2.1	First Approach	4
2.2	Second Approach	5
3	Non-radiative corrections	6
4	Polarizability in the presence of a substrate	8
5	Energy relations	9
5.1	General considerations	10
5.2	Application to a system of point particles	11
6	Radiative and non-radiative corrections in energy relations	13
6.1	Particles with Lorentz-type permittivity in the anomalous dispersion region	14
6.2	Particles with Drude permittivity	16
6.3	Beyond the dipole approximation	17
7	Paradoxes involving extinction	18
7.1	Classical extinction paradox	18
7.2	Non-additivity of extinction	19
7.3	Extinction for a collimated beam	20
8	Summary	21

 vmarkel@penmedicine.upenn.edu (V.A. Markel)

ORCID(s): 0000-0002-9748-6865 (V.A. Markel)

¹This work was performed while the author was a Visiting Scholar in the Department of Mathematics at the University of Michigan.

1. Introduction

The coupled-dipole approximation (CDA) is immensely popular in the optics literature due to its intuitive physical appeal, the wide variety of optical (more generally, electromagnetic) phenomena that this mathematical model encompasses and, not least, simplicity. There exist however two different, concurrently used versions of the coupled-dipole equations (CDEs). This can lead to confusion. It is useful therefore to state both versions of the CDEs in one place and explain how conversion from one formulation to another can be achieved. Doing so seems to be especially important for the case when the particle polarizabilities are tensorial since this possibility is rarely addressed in the literature (a few notable exceptions are mentioned below). Some closely-related subjects that can benefit from a self-contained exposition include the so-called radiative and non-radiative corrections (to the bare polarizabilities) and the definition of renormalized polarizabilities in the presence of large objects such as a substrate.

Another goal of this article is to provide a clean derivation of the energy relations such as the relation between the excited dipole moments and the extinguished, absorbed and scattered powers. In the case of tensorial polarizabilities, or particles of complicated shape and internal structure, this derivation is not straightforward and requires some care.

As an application, we use the CDA to illustrate some paradoxes related to extinction. The point of this exercise is to show that, for example, the classical extinction paradox (wherein the extinction cross section of an optically large sphere is roughly twice its geometrical cross section) is not just a peculiar mathematical feature of the Mie solution but, rather, a general phenomenon, which can be fully understood within the framework of CDA. We also note that only the absorbed power is a directly measurable physical quantity. Contrary to the popular belief, the extinguished and the scattered powers (or cross sections in the case of an incident plane wave) are mathematical constructs that cannot be

easily related to some measurable energy fluxes. This is the reason behind the various “extinction paradoxes” that are discussed below, some of which are well-known while others are not.

This Section contains introductory material. In the Section 2, we discuss two versions of the CDEs. Here the bare and the renormalized polarizabilities are introduced. Non-radiative corrections is an important topic that comes up frequently in the context of finding an accurate relation between the bare and the renormalized polarizabilities. These corrections are discussed in detail in Section 3. In Section 4, we illustrate the concepts of bare and renormalized polarizabilities by considering the CDEs near a planar substrate. It is shown how the renormalized polarizabilities can be used to compute the Purcell’s factors. In Section 5, we derive the energy relations while keeping the discussion as general as possible. In Section 6, we illustrate the effects of radiative and non-radiative corrections to the quasistatic polarizability of a particle (an elementary dipole in the CDA) with several numerical examples. The importance of accounting for the radiative corrections in systems with optical resonances is underscored. In Section 7, we discuss several extinction-related paradoxes using the CDA as the underlying physical model. Concluding remarks are given in Section 8.

1.1. Motivation and review

To start with, it is useful to distinguish between two general classes of applications in which the CDEs are used. The first class is the so-called *discrete-dipole approximation* (DDA), which is concerned with computing numerically the optical responses (typically, optical cross sections) of bulk non-spherical particles such as small ice crystals in the atmosphere. The particles are approximated by an array of point isotropic (in some rare cases, anisotropic) dipoles arranged, most typically, on a cubic (sometimes, rectangular) lattice and having approximately the same overall shape and size as the bulk particle. The basic assumption is that, with the proper choice of particle polarizabilities and with sufficiently dense sampling, the array of point dipoles would mimic the optical response of the bulk particle. The technique was introduced by Purcell and Pennypacker [1] and then refined by Draine [2] and Draine and Goodman [3]. Mathematically, the DDA can be derived by discretization of the macroscopic Maxwell’s equations written in the integral form [4, 5, 6, 7]. Further reviews of the technique were given by Draine and Flatau [8, 9] and, with an emphasis on a mathematically-rigorous derivation, by Yurkin and Hoekstra [7]. A recurring theme in the theory of DDA is the correct definition of polarizabilities of the fictitious dipoles that form the discrete backbone of a bulk particle [7, 10].

It should be kept in mind that, in the framework of DDA, the dipoles are purely mathematical constructs; they do not correspond to any physical particles. Perhaps, the dipoles can be said to represent the medium voxels. However, one can envisage a situation in which particles with the polarizabilities prescribed by the DDA actually exist and can even be arranged on the DDA lattice. In this case, the same

set of equations describes two different physical objects. This is not surprising if we recall that the electromagnetic properties of some composites can be well approximated by an *effective medium theory*. Then, assuming the wavelength is sufficiently large, the Clausius-Mossotti relation and its inverse – the Maxwell Garnett mixing formula – can be used to navigate between the bulk and the particulate descriptions of the medium.

Another class of applications, with which this article is primarily concerned, is multiple scattering and absorption of electromagnetic waves by an assembly of small particles. These particles are assumed to physically exist; they are not conjured to represent or approximate a bulk object. Consequently, the particles are not required to be arranged on any lattice or to form any geometrical structure or shape. The only requirements are that the particles are sufficiently small compared to the wavelength and do not approach each other too closely, lest the dipole approximation breaks down [11]. We refer to this approximation as to the *coupled-dipole approximation* (CDA), not to be confused with the DDA¹. Conceptually, the CDA is similar to what is known in the mathematical literature as the Foldy-Lax approximation [12, 13], although the latter was originally formulated for scalar waves.

While the goal of the CDA is to describe physical particles, at some level of consideration, the particle shape and size can disappear completely from the theory. All that can be said about a particle in this case is that it has a position (a point in space) and a dipole polarizability. Then one can consider the arising mathematical model in a purely formal manner. Even though the CDA theory can be constructed without any reference to the physical origin, shape and size of the particles, *application* of this theory requires that we take all those things into account in order to determine whether the CDA is applicable to a particular setting. Luckily, the CDA is not too restrictive. For example, if two spherical particles of radius a are small compared to the external wavelength (i.e., the optical size parameter satisfies $|m|ka \lesssim 0.2$, where $k = \omega/c$ and m is the complex refractive index of the particle [14]) and do not quite touch (i.e., if the surface-to-surface separation σ of two spherical particles of radius a is $\sigma \gtrsim 0.2a$ [15, 16]), the CDA is applicable. It was further argued by Khlebtsov that the above pairwise condition of CDA applicability remains valid for aggregates containing an arbitrary number of particles [17]. Of course, in each particular setting, the CDA applicability conditions can depend to some extent on the geometry and properties of the particles and the aggregate. A first principles check can be performed by comparing the CDA results to those of

¹The term discrete-dipole approximation (DDA) has been used in the literature quite broadly. In particular, it was used to refer to the multiple scattering theories in which several or many small particles interact via dipole electromagnetic fields. Here we wish to make a clear distinction and refer to the latter flavor of the theory as to the coupled-dipole approximation (CDA). DDA and CDA operate, formally, with the same set of equations, but the definitions of particle polarizabilities and the spatial arrangements of dipoles are substantially different in the two cases.

a more accurate approximation that involves several higher-order multipoles, i.e., see [18].

The motivation for writing this article is three-fold. First, most if not all available reviews of the subject, some of which have been cited above, are concerned specifically with the DDA. However, the DDA does not consider the most general problem of dipole interaction and rarely allows the particle polarizabilities to be tensorial. However, tensorial polarizabilities become important in some of the modern applications. The energy relations for the tensorial polarizabilities, or for the polarizabilities that correspond to complicated or internally inhomogeneous particles, are not obvious and the derivation in this case requires some care. The second motivation is the existence of two different versions of the CDEs, which differ from each other by the manner in which the diagonal element of the Green's tensor is used or denoted. The two versions are mathematically equivalent and related to each other by a simple algebraic transformation. However, this fact is not always obvious or transparent, especially in the case when more complicated Green's functions are used, i.e., for particles located near a substrate or another large object. This situation is discussed in Section 4 below. Third, we show how the CDA can be used to describe and understand various paradoxes related to extinction, including the classical extinction paradox.

We finally note that the CDA is an important research tool in optics and electromagnetic theory. To name just a few examples, the CDA has been applied to the description of superradiance [19], ferroelectricity [20], surface effects in organic molecular films [21], atmospheric optics [22, 23], and chemical sensing [24, 25]. Recently, applications related to optimizing optical transparency [26] or absorption [27, 28] have also emerged (the application of Ref. [27] explicitly requires the introduction of tensorial polarizabilities). There are therefore sufficient grounds to believe that the CDA is still a useful and viable physical theory.

1.2. Notations and units

Below we use two different notations for the polarizability of a particle. The upright symbol χ (tensor) or the slanted symbol χ (scalar) denote the bare polarizability, which does not account for the radiation reaction effects while the symbols α (tensor) or α (scalar) refer to the renormalized polarizability, which does.

Tensors are distinguished from scalars by the font type. The upright "typewriter" font such as G , I , etc., is used to represent tensors (3×3 matrices). Scalars are typeset using the usual mathematical italic font, such as k , Q , etc. Three-dimensional vectors are typeset using the bold-face straight font as in $\mathbf{d}$ or $\mathbf{E}$. No overhead decorations are used to distinguish tensors, vectors and scalars with one exception: unit vectors are decorated with an overhead hat as in $\hat{\mathbf{n}}$ or $\hat{\mathbf{r}}$. It is assumed throughout that all unit vectors have purely real Cartesian components so that, for example, $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1$ (no complex conjugation is involved). This assumption does not hold for other vectors; for example, the vectors of dipole moments $\mathbf{d}_n$ can have complex Cartesian components

of arbitrary relative phase; such vectors do not define a direction in space.

We will also define $3N$ -dimensional vectors. In this case, Dirac notations will be used as in $|d\rangle$, which is a $3N$ -dimensional vector of all Cartesian components of N dipole moments, or in $|E\rangle$, which is a similar $3N$ -dimensional vector of incident field components. Linear operators acting in the $3N$ -dimensional vector space ($3N \times 3N$ matrices) are denoted by capital letters using the mathematical italic font like G in the matrix element $\langle E|G|d\rangle$. Note that here the usual rules for Hermitian conjugation should be applied, i.e., $\langle E|$ is the Hermitian conjugate of $|E\rangle$.

We will use the same symbols for time-dependent quantities represented in time and frequency domains. For example, $\mathbf{E}(\mathbf{r}, t)$ is the electric field in time domain while $\mathbf{E}(\mathbf{r})$ is the electric field in frequency domain. The dependence of the frequency ω is omitted in the latter case. However, we should keep in mind that the frequency-domain solutions depend on the frequency. It should be clear from the context and the displayed lists of formal arguments whether the time-domain or the frequency-domain representation is used.

Gaussian system of units is used throughout this article.

2. Two versions of the coupled-dipole equations

In this Section, we describe two equivalent versions of the CDEs. The versions are related to each other by a trivial transformation but involve different definitions of the Green's function and the polarizabilities. We start however with describing the basic mathematical ingredients of the CDA that are common in both approaches.

In all cases, we consider N point particles located at the positions $\mathbf{r}_n$ in space and characterized by the (generally, tensorial) polarizabilities denoted below by α_n or χ_n (depending on the approach used) and by the induced dipole moments $\mathbf{d}_n$, $n = 1, 2, \dots, N$. We note that in the literature the polarizabilities are typically assumed to be scalar, although deviations from this assumption can be practically important. There are however some exceptions. For instance, tensor polarizabilities have been introduced by Smuneev, Chaumet and Yurkin [29] in the context of DDA for the so-called "rectangular voxels" (and for an anisotropic bulk material) and by Rasskazov, Karpov and Markel for the case of interacting differently oriented ellipsoids. [30].

The system is irradiated by a monochromatic external wave $\mathbf{E}_{\text{ext}}(\mathbf{r})$ whose electric field at the point $\mathbf{r}_n$ is denoted by $\mathbf{E}_n$, that is, $\mathbf{E}_n = \mathbf{E}_{\text{ext}}(\mathbf{r}_n)$. We assume that α_n or χ_n are symmetric and diagonalizable in some real and orthogonal principal axes, although we do not require that these axes are the same for all particles. We can write therefore (for economy of space, we adduce the formulas for α but exactly the same relations are valid for χ):

$$\alpha_n = \mathbf{R}_n^T \mathbf{D}_n \mathbf{R}_n, \quad (1)$$

where $\mathbf{R}_n$ are orthogonal matrices of rotation, the superscript T denotes transposition, and $\mathbf{D}_n$ are diagonal matrix of the

form

$$\mathbf{D}_n = \begin{bmatrix} \alpha_n^{(1)} & 0 & 0 \\ 0 & \alpha_n^{(2)} & 0 \\ 0 & 0 & \alpha_n^{(3)} \end{bmatrix}. \quad (2)$$

Here $\alpha_n^{(p)}$ ($p = 1, 2, 3$) are the principal values of α_n . An equivalent representation of the tensor α_n is

$$\alpha_n = \sum_{p=1}^3 \alpha_n^{(p)} \hat{\mathbf{u}}_n^{(p)} \otimes \hat{\mathbf{u}}_n^{(p)}, \quad (3)$$

where $\hat{\mathbf{u}}_n^{(1)}$, $\hat{\mathbf{u}}_n^{(2)}$ and $\hat{\mathbf{u}}_n^{(3)}$ are three mutually-orthogonal unit vectors with purely real Cartesian components and the symbol $\otimes$ denotes tensor product. It can be seen that

$$\mathbf{R}_n = \begin{bmatrix} u_{nx}^{(1)} & u_{ny}^{(1)} & u_{nz}^{(1)} \\ u_{nx}^{(2)} & u_{ny}^{(2)} & u_{nz}^{(2)} \\ u_{nx}^{(3)} & u_{ny}^{(3)} & u_{nz}^{(3)} \end{bmatrix}. \quad (4)$$

Equations (1) through (4) describe many particles that are of interest in applications. Note however that the above equations imply that the particles are reciprocal and non-chiral. In fact, the assumptions of chirality and zero size are at odds with each other and can not be easily reconciled. Here we choose the assumption of zero size as fundamental. On the other hand, it is possible to make a chiral object out of non-chiral point particles, i.e., by placing them on a helix. However, it is not possible to create a non-reciprocal object out of a collection of reciprocal particles. In other words, the T-matrix of any collection of interacting particles whose polarizability tensors are symmetric is also symmetric.

Further, we work in frequency domain and the time dependence of all quantities can be restored by writing, for example,

$$\mathbf{d}_n(t) = \text{Re}[\mathbf{d}_n e^{-i\omega t}], \quad (5)$$

and similarly for other time-dependent quantities.

The central piece of the CDA is the assumptions of linearity and locality of response, which is stated mathematically as

$$\mathbf{d}_n \propto \mathbf{e}_n, \quad (6)$$

where $\mathbf{e}_n$ is the electric field at the point $\mathbf{r}_n$, which is applied to the n -th dipole. Following the commonly-used terminology, we will refer to $\mathbf{e}_n$ as to the *local* field, but see the remark in the beginning of Section 5. The local field $\mathbf{e}_n$ is, obviously, different from the external field $\mathbf{E}_n$ since the former is a superposition of the latter and the scattered field. The proportionality coefficient between $\mathbf{e}_n$ and $\mathbf{d}_n$ is so far left out of consideration as it is different in the two different approaches considered below; this coefficient will be denoted by either α_n or χ_n depending on which approach is used.

The above conventions are the same in both approaches. Now we will focus on the differences.

2.1. First Approach

In the First Approach, the local field at $\mathbf{r}_n$ is a superposition of the external field and the fields scattered by all dipoles except for the n -th. That is,

$$\mathbf{e}_n = \mathbf{E}_n + \sum_{m \neq n} \mathbf{G}_{nm} \mathbf{d}_m, \quad (7)$$

where $\mathbf{G}_{nm}$ is the frequency-domain Green's tensor for the electric field acting between the points $\mathbf{r}_m$ and $\mathbf{r}_n$. The expression for $\mathbf{G}_{nm}$ in free space is well known; for completeness, we adduce the relevant formulas and a brief derivation in Appendix A. However, we are by no means constrained to the consideration of free space; the Green's tensor in (7) can be applicable to more complicated geometries with boundaries or large objects present. A typical example of such a physical setting is the presence of a planar substrate, and this case is discussed in more detail in Section 4 below. What is important for us now is that the diagonal terms $\mathbf{G}_{nn}$ do not appear in (7). Still, as we will see below, it would be a mistake to assume that $\mathbf{G}_{nn} = 0$. This term is not zero even in free space and one can easily anticipate that it would not be zero in more complicated geometries, i.e., in a cavity or near a substrate.

The CDEs in the First Approach are obtained by combining (6) and (7) and are of the form

$$\mathbf{d}_n = \alpha_n \left[\mathbf{E}_n + \sum_{m \neq n} \mathbf{G}_{nm} \mathbf{d}_m \right]. \quad (8)$$

So, in the right-hand sides of (7) or (8), there is no self-action of the dipole. Instead, the self-action, which includes the radiative reaction, is accounted for in the expressions for the polarizabilities α_n . The polarizability in this case is defined as follows: it is the tensor coupling the dipole moment $\mathbf{d}_n$ of a small particle to the electric field at the particle location $\mathbf{r}_n$ that is produced by the external sources [the term $\mathbf{E}_n$] and by all other dipoles [the term $\sum_{m \neq n} \mathbf{G}_{nm} \mathbf{d}_m$]. A subtle point here is that the actual electric field at $\mathbf{r}_n$ can be different. This fact will become especially transparent when we consider a substrate in Section 4. However, it is true even in free space. We will refer to the linear coefficient determined precisely according to the above definition as to the “renormalized polarizability”. The justification behind this terminology will become apparent later.

We now state the following important fact. Energy conservation requires that all principal values of a renormalized polarizability tensor α_n satisfy some inequality. In free space, this inequality can be obtained by requiring that the absorption cross section of an isolated, optically-passive particle be non-negative. If one also introduces the extinction and scattering cross sections, the above condition becomes equivalent to the requirement that the extinction cross section be not smaller than the scattering cross section. The latter condition is simpler mathematically and has been used to find the constraint on α_n [2, 31]. Indeed, the extinction cross section of an isolated particle excited by a monochromatic plane wave with a complex amplitude $\mathbf{E}_0$ is

given by the well-known expression

$$\sigma_e = \frac{4\pi k}{\mathbf{E}_0^* \cdot \mathbf{E}_0} \text{Im}[\mathbf{E}_0^* \cdot \alpha \mathbf{E}_0], \quad (9)$$

where $k = \omega/c$. The scattering cross section is given by

$$\sigma_s = \frac{8\pi k}{3\mathbf{E}_0^* \cdot \mathbf{E}_0} (\alpha \mathbf{E}_0)^* \cdot (\alpha \mathbf{E}_0). \quad (10)$$

Note that, in these expressions, α is the renormalized polarizability according to the definition given above. By requiring that $\sigma_e \geq \sigma_s$ and allowing $\mathbf{E}_0$ to be arbitrary, we find that the following inequality must hold for all principal values of α :

$$\text{Im} \frac{1}{\alpha^{(p)}} \leq -\frac{2k^3}{3}, \quad k = \frac{\omega}{c} \quad (\text{free space}). \quad (11)$$

The term in the right-hand side of this inequality is known as the *radiative correction*. In free space, this term is quite simple. However, in geometries involving boundaries, interfaces or large objects, it can be of a more complicated form. In fact, the modification of this term by the interaction of a small particle (sometimes referred to as a point emitter or a quantum emitter) with a surface or a relatively large object can give rise to a modification of the characteristic time of radiative decay (the radiative lifetime); the ratio of this modified radiative lifetime to that in free space is known as the Purcell's factor [32]. In the literature, the Purcell's factor is more frequently associated with the imaginary part of the Green's tensor G_{nn} . However, it is equally valid to associate the Purcell's factor with the renormalized polarizability. An illustrative example of a calculation of the radiative lifetime is given, for example, in [33].

One reason why the use of renormalized polarizabilities can be viewed as natural is that (11) is consistent with the Mie solution for spheres in free space. Indeed, the total dipole moment of a dielectric sphere of permittivity $\epsilon = \epsilon' + i\epsilon''$ and radius a excited by a plane wave of the amplitude $\mathbf{E}_0$ is given by $\mathbf{d} = \alpha_{\text{Mie}} \mathbf{E}_0$, where

$$\alpha_{\text{Mie}} = \frac{3i}{2k^3} \frac{m\psi_1(mka)\psi_1'(ka) - \psi_1(ka)\psi_1'(mka)}{m\psi_1(mka)\xi_1'(ka) - \xi_1(ka)\psi_1'(mka)}. \quad (12)$$

Here $m = \sqrt{\epsilon}$ is the complex refractive index of the sphere, $\psi_1(x)$, $\xi_1(x)$ are the Riccati-Bessel functions and prime denoted derivative with respect to the argument in the parenthesis. The result (12) is exact for all values of the radius a . If we are interested in small particles, we can use the Laurent expansion of $\text{Im}(1/\alpha_{\text{Mie}})$ in powers of a , which reads

$$\begin{aligned} \text{Im} \frac{1}{\alpha_{\text{Mie}}} = & -\frac{3\epsilon''}{a^3|\epsilon-1|^2} - \frac{3k^2\epsilon''}{5a|\epsilon-1|^2} - \frac{2k^3}{3} \\ & - \frac{3k^4 a(8+|\epsilon|^2-2\epsilon')\epsilon''}{350|\epsilon-1|^2} + O(k^6 a^3). \end{aligned} \quad (13)$$

The expansion contains only odd powers of a (including zero) and it can be verified that each term in the expansion

is non-positive. Thus we see that (11) holds for α_{Mie} . The equality in this condition is achieved only for non-absorbing materials with $\epsilon'' = 0$.

Of course, (12) gives the polarizability of a finite-size sphere due to a very special form of excitation (a plane wave). In an aggregate, each particle is not necessarily a homogeneous sphere and it is not excited by a plane wave; as a result, the expression (12) is not generally applicable. The polarizability tensor α_n , as used in this article and in the vast literature utilizing the dipole approximation, is a more general and a more abstract mathematical object than α_{Mie} . However, the third-order term $-2k^3/3$ (the radiative correction) in the expansion (13) is universal; it does not depend on the type of excitation or the particle shape. This is why the condition (11) is also universal. On the other hand, the lower-order terms in the expansion are not universal; they can depend on the particle shape and form of excitation. These lower-order terms (so-called *non-radiative* corrections) are discussed in more detail in Section 3 below.

To summarize, the First Approach to the CDA makes use of the renormalized polarizabilities, which satisfy the inequality (11) in free space or its generalization in more complicated geometries, but does not use explicitly the diagonal elements G_{nn} of the Green's tensor. The First Approach might be viewed as natural because the property (11) is satisfied by exact solutions such as, for example, the Mie solution.

2.2. Second Approach

In Second Approach, the diagonal terms G_{nn} are used explicitly in the CDEs. In free space, the relevant terms are independent of n and given by

$$G_{nn} = i\frac{2k^3}{3} \mathbf{I} \quad (\text{free space}). \quad (14)$$

This expression is derived in a semi-qualitative manner in Appendix A; a more mathematically-rigorous derivation has been recently presented by Moskalensky and Yurkin [34] where it was pointed out that only the imaginary part of G_{nn} , which is finite and given by (14) in free space, is physically important. In more complicated geometries, G_{nn} can be tensorial and depend on n . What is important for us now is that the local field $\mathbf{e}_n$ in the Second Approach includes the self-action of a dipole and is given by

$$\mathbf{e}_n = \mathbf{E}_n + \sum_m G_{nm} \mathbf{d}_m \quad (15a)$$

$$= \mathbf{E}_n + \sum_{m \neq n} G_{nm} \mathbf{d}_m + G_{nn} \mathbf{d}_n \quad (15b)$$

$$= \mathbf{E}_n + \sum_{m \neq n} G_{nm} \mathbf{d}_m + i\frac{2k^3}{3} \mathbf{d}_n \quad (\text{free space}). \quad (15c)$$

The second difference is that the Second Approach utilizes the so-called “bare” polarizabilities, which we denote here by χ_n . Simply stated, the bare polarizabilities are the quasistatic polarizabilities; they can be obtained by computing the total dipole moment of a dielectric particle of known

shape and composition subjected to a spatially-uniform external electric field. Mathematically, this is achieved by solving the Laplace equation (rather than the full Maxwell's equations) with appropriate boundary conditions at the particle surface and at infinity. If the particle is a molecule or a quantum dot, one would solve the appropriate frequency-dependent density-matrix equations while assuming that the external (driving) field is spatially uniform over the extent of the particle. The bare polarizabilities therefore do not depend on any external boundaries or the environment in which the particle is placed. Consequently, bare polarizabilities convey no information about the Purcell's factor.

It is important to emphasize that the effects of frequency dispersion, i.e., on the material properties are retained in the quasistatic approximation, and the working frequency can still be quite large. In particular, the frequency can be in the visible spectral range. However, the characteristic size a of a particle under consideration should be vanishingly small compared to the free-space wavelength λ . More precisely, it is required that $ka = 2\pi a/\lambda \ll 1$. How strong this inequality should be for the quasistatic to be accurate depends on the material. If the permittivity is very large at the working frequency, as could be the case in metals, the inequality needs to be very strong. However, from a mathematical point of view, a sufficiently small size for which the quasistatic limit sets in always exists for any finite frequency, except, perhaps, in superconductors². When this limit is achieved, the particle polarizability does not depend on a except for the trivial overall factor a^3 . For example, the bare polarizability of a small dielectric sphere of radius a and permittivity ϵ is

$$\chi = a^3 \frac{\epsilon - 1}{\epsilon + 2}. \quad (16)$$

This expression can be obtained by taking the limit $a \rightarrow 0$ of (12). It is obvious that (16) does not generally satisfy the inequality (11). For example, if ϵ is purely real³, then χ given by (16) is also purely real and does not satisfy (11). One can come to the paradoxical conclusion that the scattering cross section for such a particle is larger than the extinction cross section and therefore absorption is negative. This paradox is resolved if we account for the self-field of the dipole – that is, if we either take the term G_{nn} into account or use the renormalized polarizability instead of the bare polarizability.

The CDEs in Second Approach are of the form

$$\mathbf{d}_n = \chi_n \left[\mathbf{E}_n + \sum_m G_{nm} \mathbf{d}_m \right]. \quad (17)$$

Here the summation is extended over all indexes m and χ_n are the bare polarizabilities. To see that (17) is equivalent to (8), we consider the terms with $m = n$ and $m \neq n$ in (17)

²It is a different matter whether this small size corresponds to a realistic physical particle.

³This can be the case with very high precision for some transparent materials such as water in the visible spectral range.

separately. That is, we write

$$\mathbf{d}_n = \chi_n \left[\mathbf{E}_n + \sum_{m \neq n} G_{nm} \mathbf{d}_m + G_{nn} \mathbf{d}_n \right]. \quad (18)$$

Moving the term $G_{nn} \mathbf{d}_n$ to the left-hand side of the equation, we obtain

$$[\mathbf{I} - \chi_n G_{nn}] \mathbf{d}_n = \chi_n \left[\mathbf{E}_n + \sum_{m \neq n} G_{nm} \mathbf{d}_m \right]. \quad (19)$$

We then multiply both sides of the equation by the tensor $[\mathbf{I} - G_{nn} \chi_n]^{-1}$, which is implicitly assumed to exist (and we would run into trouble if it does not), to obtain

$$\mathbf{d}_n = [\mathbf{I} - \chi_n G_{nn}]^{-1} \chi_n \left[\mathbf{E}_n + \sum_{m \neq n} G_{nm} \mathbf{d}_m \right]. \quad (20)$$

We then define the renormalized polarizability α_n according to

$$\alpha_n = [\mathbf{I} - \chi_n G_{nn}]^{-1} \chi_n. \quad (21)$$

The action of the operator $[\mathbf{I} - G_{nn} \chi_n]^{-1}$ can be seen as a renormalization operation applied to the bare polarizabilities. We can also write for the inverse polarizability tensors

$$\alpha_n^{-1} = \chi_n^{-1} - G_{nn} \quad (22a)$$

$$= \chi_n^{-1} - i \frac{2k^3}{3} \mathbf{I} \quad (\text{free space}). \quad (22b)$$

Upon adoption of the definition (21), the set of equations (20), which is equivalent to (17), becomes also equivalent to (8). We conclude that (8) can be derived from (17) and *vice versa* by using only invertible linear operations. Therefore, the two sets of CDEs are equivalent and the two approaches to the CDA are also equivalent. The relations (21) or (22) allow one to switch easily between First and Second Approaches and the corresponding sets of CDEs.

3. Non-radiative corrections

Unfortunately, the definitions of bare and renormalized polarizabilities have been badly muddled by the so-called *non-radiative* corrections. These corrections are of the orders $O(1/a^2)$ and $O(1/a)$ and primarily of interest in the framework of DDA. The corrections in question are applied to the bare polarizability; by using them, one goes beyond the quasistatic limit⁴ when computing χ . The main problem with these corrections is that they are not universal and, therefore, can not be introduced in a sufficiently general manner. The discussion of non-radiative corrections is one example wherein the physical setting of DDA (which is rather specific) should be distinguished from the physical

⁴In the context of DDA, it is equally correct to say that the non-radiative corrections are obtained by going beyond the Clausius-Mossotti approximation [3].

setting of the CDA and the multiple scattering theory (which is much more general). We also note that the non-radiative corrections discussed here should not be confused with the terms in G_{nn} that are proportional to k and k^2 but do not contain the particle size. Such terms can appear due to interaction with a substrate, as is discussed in Section 4 below.

It is convenient to introduce the non-radiative corrections by examining the Mie solution more closely. The expansion (13) has been written for the imaginary part of the inverse Mie polarizability, but we can write a similar expansion in powers of a for the complex quantity $1/\alpha_{\text{Mie}}$:

$$\frac{1}{\alpha_{\text{Mie}}} = \frac{1}{a^3} \frac{\epsilon + 2}{\epsilon - 1} - \frac{3k^2}{5a} \frac{\epsilon - 2}{\epsilon - 1} - i \frac{2k^3}{3} - \frac{3k^4 a}{350} \frac{\epsilon^2 - 24\epsilon + 16}{\epsilon - 1} + O(k^6 a^3). \quad (23)$$

The first term in this expansion is the inverse of the quasistatic polarizability of a sphere, which is given in (16). Clearly, this term is shape-dependent. A similar expansion, say, for a cube would have started with a different term. For more complicated shapes, this term can be a tensor with three different principal values. However, since this term does not depend explicitly on the wave number k , it is expected to be independent of the illuminating field. This point is important and worth re-iterating.

Indeed, the expansion (23) has been obtained under the assumption that the incident field is a plane wave with the wave number k or a superposition of such plane waves. However, due to the effects of multiple scattering, the field incident on any given particle in an aggregate can contain comparatively large near-field or intermediate-field contributions, unless the particles are very far apart. When expanded in a spatial Fourier integral, such fields have non-zero components with the wave numbers that are different from k . The polarizability defined for such an incident field would generally be different from α_{Mie} . However, the first term in expansion (23) would still be the same because it is independent of the wave number⁵.

Next consider the second term in the expansion. It depends explicitly on the wave number k , the size of the sphere a and its permittivity ϵ . This term would change not only if we change the size or shape of the particle but also if we vary the form of the incident field. We say therefore that this term is not universal; it depends on the minute details that are difficult to control in any specific application. We note in passing that the term proportional to k/a^2 is identically zero in (23), but it can be non-zero for other particle shapes.

We can now define what we mean by the non-radiative corrections. These are the terms in the expansion of the exact inverse polarizability in powers of the size a (not necessarily

a radius) that are proportional to k/a^2 and k^2/a . In (23), only the $\propto k^2/a$ term is nonzero but, more generally, both terms can be present. The non-radiative corrections are called so because they can be introduced even in particles with zero radiative decay rate or in an approximation wherein this decay rate is zero. Thus, if we truncate expansion (23) at the order k^2/a , a particle with such a polarizability would have infinite radiative lifetime. Therefore, the non-radiative corrections are not caused by radiation. In the framework of CDA and the multiple-scattering theory (as opposed to the DDA), retaining the non-radiative corrections in the definition of the bare polarizability does not make much sense, as this is unlikely to improve the precision of the underlying approximation. The reason is that these corrections can be defined only by making some assumptions about the variation of the external (with respect to a given particle) electric field on the scale of a . This variation is usually not known in advance, i.e., before the CDEs are solved.

The third term in (23) was previously referred to as the radiative correction. Superficially, it appears to depend on the wave number k . However, this term contains no other physical scale to which k can be compared. In fact, it is more correct to say that this term depends on the frequency ω rather than on the wave number; the term is simply the inverse of the pre-factor in (12). Since it does not contain any physical scale related to the particle properties or size, it is also independent of the form of incident field. We do not give a rigorous proof of this statement but will see that this is the case in several examples. For this reason, we say that this term is universal.

We now briefly discuss the use of non-radiative corrections in the theoretical framework of DDA. Since the DDA is derived by a supposedly rigorous procedure of discretizing the integral Maxwell's equations, one can hope to derive the voxel polarizabilities rigorously. However, the discretization relies on some assumptions about the unknown solutions to Maxwell's equations. Different assumptions can be applicable under different conditions. Considering the theoretical uncertainty noted above, it is not surprising that the question of computing the non-radiative corrections "correctly" has a long history. A review of various approaches and further references have been given by Draine and Flatau [8] and Yurkin [10]. We mention here several frequently-used formulas.

Thus, Lakhtakia and Mulholland [35] proposed to define the renormalized polarizability (of an isotropic particle) according to the equation

$$\frac{1}{\alpha} = \frac{1}{\chi} + \frac{2}{a^3} [1 + (ika - 1)e^{ika}]. \quad (24a)$$

The expression (24a) can be obtained by replacing the regularized limit (67) in Appendix A with an integral over a finite sphere of radius a [36]. By expanding the right-hand side to third order in a , we obtain

$$\frac{1}{\alpha} = \frac{1}{\chi} - \frac{k^2}{a} - i \frac{2k^3}{3}. \quad (24b)$$

⁵One can still ask, what is the definition of the external field $\mathbf{E}_0$ if it is not related to the amplitude of a plane wave and, moreover, the external field can vary over the extent of the particle. The answer is that we can select any value of the external field inside the particle, or its volume average, and the differences will be small as at least ka [or $(ka)^2$ for particles with a center of symmetry]. So the first term in the expansion (23) is, in fact, independent of the form of the external field.

Therefore the right-hand side of (24a) contains both the non-radiative correction to the bare polarizability (the term k^2/a) and the radiative correction that converts the inverse bare polarizability into the inverse renormalized polarizability according to (22) [the term $-i(2k^3/3)$]. However, being non-universal, the non-radiative correction in (24b) is not the same as in (23). The reason for the discrepancy is the following: the derivation of (24a) assumes that the electric field within the sphere is spatially-uniform while the Mie solution does not rely on this assumption.

Draine and Goodman [8] proposed a different correction based on matching the dispersion relation in an infinite lattice of point dipoles and in the bulk material of the permittivity ϵ_{bulk} . The renormalized polarizability given by Draine and Goodman is of the form

$$\frac{1}{\alpha} = \frac{1}{\chi} - \frac{k^2}{a} [b_1 + \epsilon_{\text{bulk}}(b_2 + b_3 S)] - i \frac{2k^3}{3}. \quad (24c)$$

Here $b_1 \approx 1.73$, $b_2 \approx -0.102$ and $b_3 \approx 1.10$ are dimensionless *lattice sums*⁶ and S is related to the polarization state and propagation direction of the incident wave. The factor S changes from 0 for propagation along one of the crystallographic axes of the DDA lattice to 1/2 for in-plane propagation along one of the diagonals; the orientation average of S is 1/5. By construction, (24c) applies only to infinite lattices and therefore it can be expected to provide a useful approximation for objects that possess a lattice periodicity and then only locally (that is, for voxels not close to a boundary of the bulk object), as is the case for the digitized sphere in [8]. In the more general context of CDE, one is interested in objects of more general geometry such that the dipoles are not necessarily located in the nodes of a lattice, and even if they do, are not required to fill the lattice densely (an example of such an aggregate is considered in Section 6 below). Apart from that, the dispersion equation on an infinite lattice of point dipoles is not well-defined mathematically because the involved lattice sums are divergent. As is typically done in such cases, the divergences can be regularized. However, the regularization involves implicit assumptions about the particles shape and size or about the field variation near the nodes where the particles are located [37]. Finally, we note that ϵ_{bulk} in (24c) refers to the complex permittivity of the bulk object that is discretized by the DDA, not to the material of interacting small particles in the context of the multiple scattering theory. In the latter case, it is not clear how to determine ϵ_{bulk} in (24c). For these reasons, the result (24c) is also not universal and, in any event, it can not be applied to the multiple scattering problem involving arbitrarily located small particles.

Note that a generalization of (24c) to the case of a rectangular lattice and a minor correction that applies to the cubic lattice were given by Gutkowicz-Krusin and Draine

⁶Here we use a slightly different definition of b_k . In [8], these quantities were defined as the coefficients in front of k^2/h , where $h = (4\pi/3)^{1/3}a$ is the DDA lattice step. Here we express the non-radiative corrections as k^2/a times some dimensionless coefficients. For this reason the numerical values of b_k in (24c) are different from those in [8].

in [38]. The above comments fully apply to this corrected theory.

To sum up, accounting for the non-radiative corrections to the bare polarizability can provide a better approximation under the well-controlled conditions of the DDA [29]. However, in the context of multiple scattering by small particles, these corrections are not well-defined. On the other hand, accounting for the radiative correction is useful in all cases and can even be necessary, i.e., to enforce energy conservation and to guarantee that the CDEs have a physically-meaningful solution. It can be remarked that the radiative correction is the same in (23), (24b) and (24c) even though these equations have been derived from different principles and assumptions. This confirms the universality of the radiative correction.

The effects of radiative and non-radiative corrections on optical spectra of a collection of small particles is further illustrated in Section 6 below.

4. Polarizability in the presence of a substrate

The CDA in the presence of a substrate has been extensively studied in the literature [39, 40, 41]. A fundamental question one faces in this setting is how to compute G_{nm} . Indeed, Green's tensor is known analytically only in free space. In the half-space geometry, a formally exact result is given by the so-called Sommerfeld integral [42, 43]. However, the integral can be evaluated analytically only in some asymptotic regimes. Correspondingly, a combination of purely numerical and partially analytical approaches have been used. In this Section, we focus on the case when the dipoles are in the near-field zone of the substrate [33]. We will adduce below the results for the diagonal elements G_{nn} , which appear in the definition of the renormalized polarizability (21). Similar results for the off-diagonal terms G_{nm} are given in [33]. However, the asymptotic formulas of [33] should be used with caution in the framework of CDA since they can lose accuracy for large lateral separations of the dipoles.

Generally, in the presence of a substrate or of any other large object, we can write

$$G_{nm} = G_{nm}^F + G_{nm}^R, \quad (25a)$$

where G_{nm}^F is the free-space Green's tensor and G_{nm}^R is the reflected part originating due to the presence of the object. It is important to note that the reflected part of the Green's tensor has no singularity. That is, $G^R(\mathbf{r}, \mathbf{r}')$ is well defined when $\mathbf{r} \rightarrow \mathbf{r}'$. We can therefore define $G_{nn} = G(\mathbf{r}_n, \mathbf{r}_n)$ without resorting to any regularizing procedure of the sort that used in Appendix A for the free-space Green's tensor.

Let a substrate of some complex dielectric permittivity ϵ_s (at the working frequency) occupy the half-space $z < 0$. The particles are located in the upper half-space $z > 0$ but not too far from the interface $z = 0$. The directions of the X and Y axes are mutually orthogonal but otherwise arbitrary. Let z_n be the height of the n -th dipole above the interface.

Then we have the following expansion:

$$G_{nn}^R = \frac{1}{(2z_n)^3} \sum_{l=0}^{\infty} (2kz_n)^l K^{(l)}. \quad (25b)$$

where

$$K_{xx}^{(0)} = K_{yy}^{(0)} = \frac{\epsilon_s - 1}{\epsilon_s + 1}, \quad (25c)$$

$$K_{zz}^{(0)} = 2 \frac{\epsilon_s - 1}{\epsilon_s + 1}; \quad (25d)$$

$$K_{xx}^{(2)} = K_{yy}^{(2)} = \frac{1}{2} \frac{\epsilon_s(\epsilon_s - 1)}{(\epsilon_s + 1)^2}, \quad (25e)$$

$$K_{zz}^{(2)} = \frac{(\epsilon_s - 1)(2\epsilon_s + 1)}{(\epsilon_s + 1)^2}, \quad (25f)$$

$$K_{xx}^{(3)} = K_{yy}^{(3)} = \frac{i\epsilon_s}{3(\epsilon_s + 1)^2} [\Phi_{\parallel}(\epsilon_s) + 3\Lambda(\epsilon_s)], \quad (25g)$$

$$K_{zz}^{(3)} = -\frac{2i}{3(\epsilon_s + 1)^2} [\Phi_{\perp}(\epsilon_s) + 3\epsilon_s^2 \Lambda(\epsilon_s)], \quad (25h)$$

and

$$\Phi_{\parallel}(\epsilon_s) = \frac{1 - 3\epsilon_s^{1/2} + 3\epsilon_s + 2\epsilon_s^2}{\epsilon_s^{1/2} + 1}, \quad (25i)$$

$$\Phi_{\perp}(\epsilon_s) = \frac{1 + \epsilon_s^{1/2} + 2\epsilon_s + 2\epsilon_s^{3/2} - 2\epsilon_s^{5/2} - \epsilon_s^3}{\epsilon_s^{1/2} + 1}, \quad (25j)$$

$$\Lambda(\epsilon_s) = \frac{\epsilon_s}{(\epsilon_s - 1)(\epsilon_s + 1)^{1/2}} \ln \frac{1 + (\epsilon_s + 1)^{1/2}}{\epsilon_s + [\epsilon_s(\epsilon_s + 1)]^{1/2}}. \quad (25k)$$

Here we have listed only the non-zero matrix elements of $K^{(l)}$. The first-order tensor $K^{(1)}$ is identically zero and the off-diagonal tensor elements $K_{xy}^{(l)}$, $K_{xz}^{(l)}$ and $K_{yz}^{(l)}$ are zero to all orders due to symmetry. Also, the expansion coefficients were listed only up to third order; in Ref. [33], the expansion coefficients of $\text{Im}G_{nn}$ (for real-valued ϵ_s) are computed up to seventh order.

Combining all formulas in (25), we can write

$$(G_{nn})_{xx} = (G_{nn})_{yy} = \frac{1}{8z_n^3} \frac{\epsilon_s - 1}{\epsilon_s + 1} + \frac{k^2}{4z_n} \frac{\epsilon_s(\epsilon_s - 1)}{(\epsilon_s + 1)^2} + i \frac{2k^3}{3} \xi_{\parallel}(\epsilon_s). \quad (26a)$$

$$(G_{nn})_{zz} = \frac{1}{4z_n^3} \frac{\epsilon_s - 1}{\epsilon_s + 1} + \frac{k^2}{2z_n} \frac{(\epsilon_s - 1)(2\epsilon_s + 1)}{(\epsilon_s + 1)^2} + i \frac{2k^3}{3} \xi_{\perp}(\epsilon_s). \quad (26b)$$

Here

$$\xi_{\parallel}(\epsilon_s) = 1 + \frac{\epsilon_s}{2(\epsilon_s + 1)^2} [\Phi_{\parallel}(\epsilon_s) + 3\Lambda(\epsilon_s)], \quad (27a)$$

$$\xi_{\perp}(\epsilon_s) = 1 - \frac{1}{(\epsilon_s + 1)^2} [\Phi_{\perp}(\epsilon_s) + 3\epsilon_s^2 \Lambda(\epsilon_s)] \quad (27b)$$

are the Purcell's factors for parallel and orthogonal oscillations of an isolated point dipole in the vicinity of an interface.

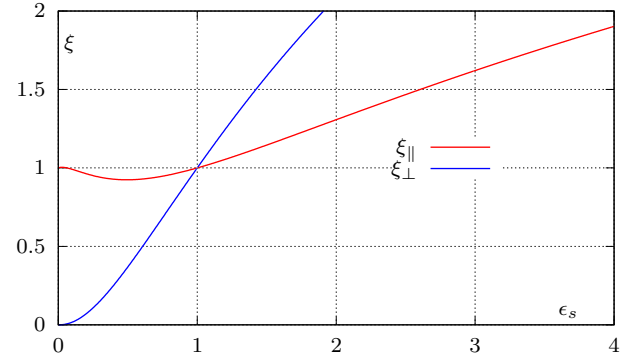


Figure 1: Purcell factors ξ for a single dipole near a dielectric interface with a purely-real permittivity ϵ_s and for the polarization of oscillations parallel to the interface ($\xi_{\parallel}$) and orthogonal to the interface ($\xi_{\perp}$).

The functions $\xi_{\parallel}(\epsilon_s)$, $\xi_{\perp}(\epsilon_s)$ are shown in Fig. 1 for real ϵ_s . It can be seen that, for realistic materials ($\epsilon_s > 1$), both Purcell's factors are greater than unity. This means that the presence of the substrate makes the radiative lifetime shorter. For $\epsilon_s < 1$, both factors are less than unity, and $\xi_{\perp}$ even reaches zero at $\epsilon_s = 0$. A hypothetical $\epsilon_s = 0$ substrate would make the radiative lifetime of the dipole oscillations polarized in the XY -plane infinite since $\xi_{\perp}(\epsilon_s)$ turns to zero at $\epsilon_s = 0$. This, however, is not an exact result. Rather, the above expressions are valid asymptotically in the limit $kz_n \rightarrow 0$. One should not expect truly infinite radiation lifetimes in any realistic physical setting.

We now make several comments.

First, the zero-order terms $K^{(0)}$ in (25c) describe the well-known contributions of the fictitious reflected (image) dipole, which is located below the interface at $z = -z_n$. The term $K^{(2)}$ is similar to the non-radiative corrections that were discussed in Section 3 above. Indeed, the corresponding contribution to G_{nn} in (26) is proportional to k^2/z_n . However, these terms do not depend on the particle size and shape nor on the exact form of the local field in the vicinity of $\mathbf{r}_n$. Therefore, we do not classify such terms as non-radiative corrections to the bare polarizability; rather, they are non-radiative contributions to G_{nn}^R . These terms are universal and should generally be taken into account. The third-order terms $K^{(3)}$ describe the first non-vanishing radiative contribution to G_{nn}^R ; together with the similar term in G_{nn}^F , they define the Purcell's factors for a dipole in the vicinity of a substrate according to (27a).

Second, even if the bare polarizabilities are isotropic, interaction with the substrate makes the renormalized polarizabilities tensorial.

Third, just as in the case of free space, we can use either the First Approach to the CDA in which the renormalized polarizabilities are used but the diagonal elements G_{nn} are left out of consideration or the Second Approach, in which only the bare polarizabilities are used but G_{nn} appear explicitly in the equations. The two approaches remain mathematically equivalent in the presence of a substrate.

Fourth, we should be careful to account for the presence of the substrate in computing the off-diagonal elements G_{nm} . In a general setting, when none of the previously explored asymptotic regimes are valid, this can be done only numerically.

Finally, the external fields $\mathbf{E}_n$ in either (8) or (17) should also account for the presence of a substrate. By definition, these are the fields that would have existed in the points $\mathbf{r}_n$ if the dipoles were removed to infinity. In particular, if we consider a plane wave incident from above, the fields $\mathbf{E}_n$ are superpositions of this plane wave and the reflected plane wave; the latter can be computed according to the usual Fresnel equations. If the excitation is performed in the total internal reflection geometry, then $\mathbf{E}_n$ are obtained by sampling the evanescent wave that exists above the interface.

5. Energy relations

Energy relations for the CDA are well known, although they are usually stated or derived in the context of DDA [2]. In particular, it is widely known that extinction is related to the work of the external field (denoted by $\mathbf{E}_n$ in this article) on the oscillating dipoles while absorption is related to the work of the local field (denoted by $\mathbf{e}_n$). While the above statements are generally correct, derivation of the result for absorption is not straightforward. Indeed, one might ask, what exactly is meant by the local field? From Joule's heat law, it should be the electric field acting on the moving charges. But the latter field is generally different from $\mathbf{e}_n$. To compute the "true" local field one needs a specific model for the particle.

This can be illustrated with an example. Let the particle be a small dielectric sphere of the permittivity ϵ placed in the local field $\mathbf{e}$. Let us use the conventions of the Second Approach to the CDA so that $\mathbf{e}$ is determined with the account of dipole self-action. Then the dipole moment of the sphere is given by $\mathbf{d} = a^3[(\epsilon-1)/(\epsilon+2)]\mathbf{e}$. Here we have used the bare polarizability of the sphere, as is required by Second Approach. Now the field *inside* the sphere is $\mathbf{e}' = \mathbf{e} - \mathbf{e}_{\text{dep}}$, where $\mathbf{e}_{\text{dep}} = -[(\epsilon-1)/(\epsilon+2)]\mathbf{e}$ is the depolarizing field. It can be seen now that the induced current inside the sphere, $\mathbf{j} = -i\omega\mathbf{d}$, and the depolarizing field, $\mathbf{e}_{\text{dep}}$, are always out of phase (that is, the phase shift between the two oscillating quantities is strictly $\pi/2$). For this reason, the depolarizing field does no work on the induced current; consequently, the work exerted by $\mathbf{e}$ and $\mathbf{e}'$ on $\mathbf{j}$ is the same. We can compute the absorbed power as the work of the field $\mathbf{e}$, just as it was assumed above. The same can be demonstrated for ellipsoids.

So the conventional formula for absorbed energy is correct for ellipsoids. But is it correct generally? If the particle is not of ellipsoidal shape or not internally homogeneous or not even a dielectric particle, the depolarizing field is not known and not easily computable. The general result we wish to prove is that, within the quasistatics, the depolarizing field never does any work on the induced current. We will indeed prove this but by considering the energy fluxes at infinity. A

more direct proof based on the analysis of the quasistatic electric field inside a general particle can also be given, but the far-field argument appears to be more general. We will therefore present this argument below. We will follow a calculation given in [44] and generalize it to the case of tensorial polarizabilities.

We will perform all calculations in free space. However, as soon as it will be shown that both extinction and absorption are, essentially, local quantities, one can easily generalize the results to more complicated geometries including that of a half-space.

In this Section, we start by working in time domain. So, at least initially, all fields are assumed to depend on time and position in a rather general way. Later on, we will convert the fields to frequency domain and take the appropriate time averages of all quantities that are quadratic in the fields.

5.1. General considerations

In a typical formulation of a scattering problem, the electromagnetic fields everywhere in space are decomposed into the incident (labeled by the subscript i) and scattered (labeled by s) contributions, i.e.,

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_i(\mathbf{r}, t) + \mathbf{E}_s(\mathbf{r}, t), \quad (28a)$$

$$\mathbf{B}(\mathbf{r}, t) = \mathbf{B}_i(\mathbf{r}, t) + \mathbf{B}_s(\mathbf{r}, t), \quad (28b)$$

The extinguished, absorbed and scattered powers are defined most fundamentally as certain energy fluxes. Let the scattering medium be supported in some region of space Ω with the external boundary $\partial\Omega$. Although this is not essential, we can assume that $\partial\Omega$ is a sphere of sufficiently large radius. Then the absorbed (Q_a) and the scattered (Q_s) powers are given by the following surface integrals [14]:

$$Q_a(t) = - \oint_{\partial\Omega} [\mathbf{S}(\mathbf{r}, t) \cdot \hat{\mathbf{n}}] d^2r, \quad (29a)$$

$$Q_s(t) = \oint_{\partial\Omega} [\mathbf{S}_s(\mathbf{r}, t) \cdot \hat{\mathbf{n}}] d^2r. \quad (29b)$$

Here $\hat{\mathbf{n}}$ is the outward unit normal to $\partial\Omega$ and $\mathbf{S}$ and $\mathbf{S}_s$ are the Poynting vectors associated with the total field and the scattered field (taken alone, as if it could exist in the absence of the incident field), that is,

$$\mathbf{S}(\mathbf{r}, t) = \frac{c}{4\pi} \mathbf{E}(\mathbf{r}, t) \times \mathbf{B}(\mathbf{r}, t), \quad (30a)$$

$$\mathbf{S}_s(\mathbf{r}, t) = \frac{c}{4\pi} \mathbf{E}_s(\mathbf{r}, t) \times \mathbf{B}_s(\mathbf{r}, t). \quad (30b)$$

It can be seen that only the absorbed power has a well-defined physical meaning in the sense that it is measurable. The scattered power is not measurable directly because, in any conceivable experiment, the incident and the scattered fields are present simultaneously. The vector $\mathbf{S}_s$ is therefore not observable, at least not *everywhere* on $\partial\Omega$. We can view the scattered and extinguished powers, as well as the scattering and extinction cross sections ⁷ as auxiliary

⁷Cross sections can be introduced if the incident field is a plane wave as the ratio of the time-averaged extinguished, scattered or absorbed power fluxes to the incident power flux per unit surface. For more general illuminations, optical cross sections can not be defined.

quantities, which, under some approximations (but, perhaps, never exactly), can be related to measurable quantities.

The extinguished power Q_e is simply the sum of Q_a and Q_s . Adding the two equations in (29) together and accounting for the expressions in (30), we obtain

$$Q_e(t) = -\frac{c}{4\pi} \oint_{\partial\Omega} [\mathbf{E}_i(\mathbf{r}, t) \times \mathbf{B}(\mathbf{r}, t) + \mathbf{E}(\mathbf{r}, t) \times \mathbf{B}_i(\mathbf{r}, t)] \cdot \hat{\mathbf{n}} d^2r. \quad (31)$$

We thus see that extinction is related to a rather peculiar (and also unobservable) energy flux, which is created by the “interference” between the incident and the total fields. It should be noted that, in any stationary process, the total fields in (31) can be replaced with the scattered fields since the time-averaged energy flux created by the incident field through any closed surface is in this case zero. Therefore, it is often stated that extinction is given by the energy flux created by the interference of the incident and the scattered fields. This statement is valid only for stationary processes or for total time-integrated quantities in transient processes.

We can use Maxwell’s equations to express Q_a and Q_e as volume integrals. The equations obeyed by the total and incident fields are almost the same but differ by the presence of the induced current in the former case:

$$\nabla \times \mathbf{B} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \mathbf{J}, \quad \nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (32a)$$

$$\nabla \times \mathbf{B}_i = \frac{1}{c} \frac{\partial \mathbf{E}_i}{\partial t}, \quad \nabla \times \mathbf{E}_i = -\frac{1}{c} \frac{\partial \mathbf{B}_i}{\partial t}. \quad (32b)$$

Here $\mathbf{J}(\mathbf{r}, t)$ is the induced electric current density. In a medium, which can support electric polarization $\mathbf{P}$ and magnetization $\mathbf{M}$, the induced current is given by $\mathbf{J} = \partial \mathbf{P} / \partial t + c \nabla \times \mathbf{M}$. However, the exact form of the induced current or the constitutive relations that define the latter in terms of the fields are not important for us now. Without specifying the induced current in any detail, we can use the divergence theorem and the equations in (32a) to convert (29a) and (31) to volume integrals:

$$Q_a(t) = \int_{\Omega} \left[\frac{\partial U(\mathbf{r}, t)}{\partial t} + \mathbf{J}(\mathbf{r}, t) \cdot \mathbf{E}(\mathbf{r}, t) \right] d^3r, \quad (33a)$$

$$Q_e(t) = \int_{\Omega} \left[\frac{\partial W(\mathbf{r}, t)}{\partial t} + \mathbf{J}(\mathbf{r}, t) \cdot \mathbf{E}_i(\mathbf{r}, t) \right] d^3r. \quad (33b)$$

where

$$U(\mathbf{r}, t) = \frac{\mathbf{E}(\mathbf{r}, t) \cdot \mathbf{E}(\mathbf{r}, t) + \mathbf{B}(\mathbf{r}, t) \cdot \mathbf{B}(\mathbf{r}, t)}{8\pi}, \quad (34a)$$

$$W(\mathbf{r}, t) = \frac{\mathbf{E}(\mathbf{r}, t) \cdot \mathbf{E}_i(\mathbf{r}, t) + \mathbf{B}(\mathbf{r}, t) \cdot \mathbf{B}_i(\mathbf{r}, t)}{4\pi}. \quad (34b)$$

Equation (33a) is just the Poynting theorem; the field U in the right-hand side is commonly associated with the density of electromagnetic energy. The relation (33b) is somewhat less conventional. In particular, the field W can not be interpreted as energy density. However, an important observation is that, in any stationary process,

$$\langle \partial U / \partial t \rangle_t = \langle \partial W / \partial t \rangle_t = 0, \quad (35)$$

where $\langle \dots \rangle_t$ denotes time average. Conversely, in any transient process, we have

$$\int_{-\infty}^{\infty} [\partial U / \partial t] dt = \int_{-\infty}^{\infty} [\partial W / \partial t] dt = 0. \quad (36)$$

It is the availability of the simple properties (35) and (36) that makes the integral relations (33) useful.

We now assume that all fields are monochromatic and introduce frequency-domain notations similarly to (5). Then it is straightforward to derive for the time-averaged powers

$$\langle Q_a \rangle_t = \frac{1}{2} \text{Re} \int_{\Omega} \mathbf{J}(\mathbf{r}) \cdot \mathbf{E}^*(\mathbf{r}) d^3r, \quad (37a)$$

$$\langle Q_e \rangle_t = \frac{1}{2} \text{Re} \int_{\Omega} \mathbf{J}(\mathbf{r}) \cdot \mathbf{E}_i^*(\mathbf{r}) d^3r. \quad (37b)$$

In addition, assuming that the radius of the sphere $\partial\Omega$ is sufficiently large, we can write the scattered field on $\partial\Omega$ as

$$\mathbf{E}_s(\mathbf{r}) = \mathbf{f}(\hat{\mathbf{r}}) \frac{e^{ikr}}{r} \quad \text{for } \mathbf{r} \in \partial\Omega; \quad k = \frac{\omega}{c}. \quad (38)$$

Here $\mathbf{f}(\hat{\mathbf{r}})$ is the vector scattering amplitude. Since we have not made any choices regarding the form of the incident field and, in particular, we do not assume that it is a plane wave, the scattering amplitude depends only on one unit vector $\hat{\mathbf{r}}$, which points from the center of the sphere Ω to the point of observation on its surface $\partial\Omega$. Of course, we keep in mind that the scattering amplitude also depends on the form of the incident field, but this dependence is implicit in the notations. With this definition of the scattering amplitude, we can write

$$\langle Q_s \rangle_t = \frac{c}{8\pi} \int [\mathbf{f}(\hat{\mathbf{r}}) \cdot \mathbf{f}^*(\hat{\mathbf{r}})] d^2\hat{\mathbf{r}}. \quad (39)$$

In summary, for a monochromatic field, the absorbed power is given by the work of the *total* (local) electric field exerted on the induced current; extinguished power is the work exerted on the same current by the *incident* field and the scattered power can be obtained by integrating the squared scattering amplitude over all directions. We now proceed to applying these general results to a collection of point particles.

5.2. Application to a system of point particles

In the CDA, the shapes, internal structure and other small-scale properties of the particles do not enter the equations; only the positions and the polarizabilities do. But the knowledge of the polarizabilities or of the induced dipole moments seems to be not enough to apply (37a). This raises the question whether the CDA provides a complete description of the important physical phenomena such as absorption of energy. The answer to this question is positive but, to see this, it is more convenient to start with computing the extinguished and the scattered powers, and then define the absorbed power as the difference between the former two quantities. As we will see, the extinguished and the scattered powers can be expressed in terms of the dipole moments

$\mathbf{d}_n$ and do not require the knowledge of fields and current distributions inside the particles.

We start with extinction. The integral in the right-hand side can be decomposed as a sum of integral over each particle, viz,

$$\langle Q_e \rangle_t = \frac{1}{2} \text{Re} \sum_n \int_{V_n} \mathbf{J}(\mathbf{r}) \cdot \mathbf{E}_i^*(\mathbf{r}) d^3 r, \quad (40)$$

where V_n is the spatial region occupied by the n -th particle. Since the size of all particles is assumed to be vanishingly small compared to the characteristic scale of the field variation in the vicinity of $\mathbf{r}_n$ ⁸, we can replace the function $\mathbf{E}_i^*(\mathbf{r})$ with $\mathbf{E}_n^*$, where $\mathbf{E}_n$, as defined above in Section 2, is the incident field at the position of n -th particle. Therefore,

$$\langle Q_e \rangle_t = \frac{1}{2} \text{Re} \sum_n \int_{V_n} \mathbf{J}(\mathbf{r}) \cdot \mathbf{E}_n^* d^3 r. \quad (41)$$

Further, it is easy to see that, in the monochromatic case considered here,

$$\int_{V_n} \mathbf{J}(\mathbf{r}) d^3 r = -i\omega \mathbf{d}_n.$$

Therefore,

$$\langle Q_e \rangle_t = \frac{\omega}{2} \text{Im} \sum_n \mathbf{d}_n \cdot \mathbf{E}_n^*. \quad (42)$$

Introducing $3N$ -dimensional vectors $|d\rangle$ and $|E\rangle$, where

$$|d\rangle = (d_{1x}, d_{1y}, d_{1z}, \dots, d_{Nx}, d_{Ny}, d_{Nz})^T,$$

etc., we can re-write (42) compactly as

$$\langle Q_e \rangle_t = \frac{\omega}{2} \text{Im} \langle E | d \rangle. \quad (43)$$

This expression is independent of which approach to the CDA is used since the vectors $|d\rangle$ and $|E\rangle$ are the same in both approaches.

Now consider scattering. To proceed, we utilize the expression for the scattering amplitude of a collection of point dipoles, i.e., given in [44]:

$$\mathbf{f}(\hat{\mathbf{r}}) = k^2 \sum_n [\mathbf{d}_n - (\mathbf{d}_n \cdot \hat{\mathbf{r}}) \hat{\mathbf{r}}] e^{-ik\hat{\mathbf{r}} \cdot \mathbf{r}_n}. \quad (44)$$

Upon inserting this expression into (39), we obtain for the scattered power:

$$\begin{aligned} \langle Q_s \rangle_t &= \frac{ck^4}{8\pi} \sum_{n,m} \int [\mathbf{d}_n \cdot \mathbf{d}_m^* - (\mathbf{d}_n \cdot \hat{\mathbf{r}})(\mathbf{d}_m^* \cdot \hat{\mathbf{r}})] \\ &\quad \times e^{ik(\mathbf{r}_m - \mathbf{r}_n) \cdot \hat{\mathbf{r}}} d^2 \hat{\mathbf{r}}. \end{aligned} \quad (45)$$

The integrals can be computed directly and it turns out that they are expressed in terms of the Green's tensors G_{nm} .

⁸Note that the condition $a/\lambda \ll 1$ is insufficient; the field can contain near-field terms due to scattering from the neighboring particles. See the discussion on p. 2 and Refs. [15, 16].

Omitting tedious but straightforward integration⁹, we obtain the following result:

$$\langle Q_s \rangle_t = \frac{\omega}{2} \text{Im} \sum_{n,m} \mathbf{d}_n^* \cdot G_{nm} \mathbf{d}_m. \quad (46)$$

We emphasize that the summation here is performed over all values of n and m and the expression (46) contains the terms G_{nn} . We can therefore use the CDEs as formulated in Second Approach, (17), to express the sum over m in the right-hand side of (46) as

$$\sum_m G_{nm} \mathbf{d}_m = \chi_n^{-1} \mathbf{d}_n - \mathbf{E}_n.$$

Here χ_n are the bare polarizabilities. Substituting this result into (46), and upon some rearrangements, we arrive at the result [34, 44]

$$\langle Q_s \rangle_t = \frac{\omega}{2} \text{Im} \sum_n [\mathbf{d}_n^* \cdot \chi_n^{-1} \mathbf{d}_n + \mathbf{d}_n \cdot \mathbf{E}_n^*]. \quad (47)$$

We now identify the last term in the right-hand side of (47) as the extinguished power [compare to (42)]. From this it immediately follows that the absorbed power, determined as $\langle Q_a \rangle_t = \langle Q_e \rangle_t - \langle Q_s \rangle_t$, is given in terms of the dipole moments by the expression

$$\langle Q_a \rangle_t = -\frac{\omega}{2} \text{Im} \sum_n \mathbf{d}_n^* \cdot \chi_n^{-1} \mathbf{d}_n, \quad (48)$$

or, utilizing the $3N$ -dimensional notations and defining the $3N \times 3N$ matrix X , which contains the 3×3 blocks χ_n on the diagonal,

$$\langle Q_a \rangle_t = -\frac{\omega}{2} \text{Im} \langle d | X^{-1} | d \rangle, \quad (49)$$

Expressions (48),(49) are well-known for the case of scalar bare polarizabilities in the DDA literature, i.e., see [7]. In the special case when all polarizabilities are the same and scalar, expression (49) simplifies to

$$\langle Q_a \rangle_t = \frac{q\omega}{2} \langle d | d \rangle, \quad (50)$$

where

$$q = -\text{Im}(1/\chi). \quad (51)$$

is the parameter that characterizes the strength of absorption by one particle. Another convenient expression for $\langle Q_a \rangle_t$ can be obtained by noting that $\chi_n^{-1} \mathbf{d}_n = \mathbf{e}_n$, where $\mathbf{e}_n$ is the local field as defined in the Second Approach to the CDA, i.e., in (15a). Thus,

$$\langle Q_a \rangle_t = \frac{\omega}{2} \text{Im} \sum_n \mathbf{e}_n^* \cdot \mathbf{d}_n = \frac{\omega}{2} \text{Im} \langle e | d \rangle. \quad (52)$$

⁹The diagonal and off-diagonal terms in the sum should be integrated separately. For off-diagonal terms, it is convenient to choose a reference frame in which $\mathbf{r}_m - \mathbf{r}_n$ coincides with the direction of the Z -axis, and one should keep in mind that $\mathbf{d}_n$ are complex and do not necessarily possess a direction.

This expression shows that the absorbed power is the work exerted by the local fields $\mathbf{e}_n$ on the oscillating dipole moments $\mathbf{d}_n$. However, it is important to keep in mind that we must use the definition of the Second Approach for the local fields.

We thus come to the following conclusion: absorbed power is the work of the local fields $\mathbf{e}_n$ on the oscillating dipoles, but the local fields must be defined according to the Second Approach, that is, with the account of dipole self-action. For this reason, absorbed power is more conveniently expressed in the Second Approach. We can, of course, convert all relevant formulas to First Approach. For example, we can compute all $\mathbf{d}_n$'s using the First Approach and renormalized polarizabilities, etc., and then apply (50) to compute absorbed power. However, (50) contains the absorption strength q , which is computed from the bare polarizabilities according to (51). In a sense, the bare polarizability and the absorption strength are the intrinsic properties of the particles while the renormalized polarizabilities take into account all kinds of extrinsic factors such as the presence of a substrate.

Thus, we have obtained the absorbed, extinguished and scattered powers in terms of the dipole moments $\mathbf{d}_n$ and the bare polarizabilities χ_n . It is significant that these results are quite general. In particular, they apply to tensorial polarizabilities of the rather general form as defined in Section 2. The particles can be non-ellipsoidal and intrinsically inhomogeneous, or not even describable as dielectric bodies. All that is required for the expressions derived in this section to hold are the underlying assumptions of the CDA itself. It is also significant that all expressions derived here are local and interpretable in terms of work exerted on the oscillating dipoles. We can therefore generalize these formulas to more general geometries, other than free space.

6. Radiative and non-radiative corrections in energy relations

Above, we have discussed at length the radiative and non-radiative corrections. However, one might think that this discussion is not truly relevant because the corrections in question are small under the typical experimental conditions. This is indeed so in many cases but not in the physically interesting case of strong multiple scattering and pronounced optical resonances. By optical resonances we mean the occurrences when α_n^{-1} in (8) or χ_n^{-1} in (17) are close to one of the generalized eigenvalues of the respective equation. We can define the generalized eigenvalues as the set of tensors λ_n for which the equation $\sum_m G_{nm} \mathbf{x}_m = \lambda_n \mathbf{x}_n$ [assuming (17) is used] has a non-trivial solution $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$. We emphasize that the equality $\chi_n^{-1} = \lambda_n$ is unphysical and never holds in a correctly constructed model. If such an equality could hold, the collection of dipoles would support infinite non-decaying oscillations in the absence of any external field.

Therefore, we expect on physical grounds that $\chi_n^{-1} \neq \lambda_n$. However, nothing prevents the above equality from holding

approximately. If this happens, the dipole fields experience strong multiple scattering and the optical spectra exhibit pronounced resonance peaks. Under the circumstances, accounting for the radiative correction to the inverse polarizability can become important. Essentially, this correction prevents the equality $\chi_n^{-1} = \lambda_n$ from holding exactly or, more precisely, it prevents the inverse polarizabilities from crossing into the region of unphysical parameters. On the other hand, the non-radiative corrections, being primarily the corrections to the real part of χ_n^{-1} , do not have the same dramatic effect. Rather, accounting (or not accounting) for the non-radiative corrections usually results in small spectral shifts; that is, the extinction, absorption and scattering spectra are slightly shifted along the frequency axis without changing their shape.

Of course, the significance of radiative corrections is best revealed when the particles are not too absorbing and not too small¹⁰. For dielectric or metallic macroscopic spherical particles, these conditions can be formulated in terms of the imaginary part of the complex permittivity ϵ and the particle radius a . The condition under which the radiative corrections become important is $3\text{Im}\epsilon/|\epsilon - 1|^2 \lesssim 2(ka)^3/3$. If the equality holds, not accounting for the radiative correction can result in singular matrices and diverging solutions. If a strict inequality holds and the radiative correction is not accounted for, a solution may exist but be unphysical: in particular, it can violate energy conservation. Steady state (i.e., monochromatic) solutions of this sort can mathematically satisfy the CDEs but be physically unreachable. That is, such solutions can not be attained by starting from any reasonable initial conditions and turning the external field slowly on.

In the forthcoming subsections, we illustrate the above qualitative arguments with a few numerical examples. We will compute the conventional optical spectra of a three-dimensional collection of dipoles whose positions are illustrated in Fig. 2. The spatial distribution of particles was obtained by the process of so-called diffusion-limited aggregation [45], but the physical nature of the aggregate is not important for us here. What is important is that the centers of all particles are located in the nodes of a simple cubic lattice of the pitch h and the particle polarizabilities are computed (with or without the corrections in question) for identical spheres with the radius $a = 0.25h$. Therefore, no two particles come closer (center-to-center) than two particle diameters. Equivalently, the minimum surface-to-surface separation is equal to one particle diameter ($2a$). Under these conditions, the dipole approximation is usually considered to be accurate. Although the ratio h/a was fixed in all simulations, the ratio of h to other physical length scales of the problem such as the resonance wavelength was varied.

¹⁰The two conditions are actually inter-related but it is not always clear how to extrapolate them beyond the applicability range of the classical macroscopic electrodynamics.

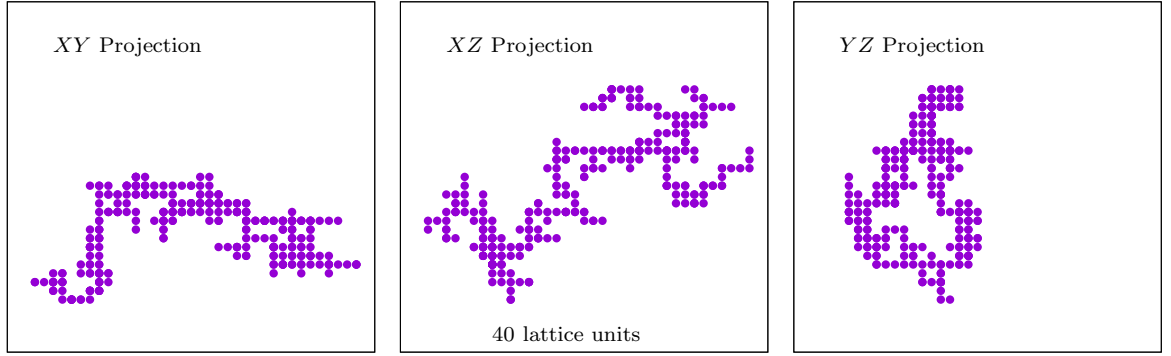


Figure 2: Geometry of the aggregate of small particles used in the simulation of Section 6. The aggregate consists of $N = 300$ particles whose centers are located in the nodes of a simple cubic lattice. The particles are positioned rather sparsely but connected into one aggregate so that one can travel between any two particles by making jumps that are no longer than one lattice step. In the simulations, it was assumed that the particles are spheres of the radius equal to $1/4$ of the lattice step and the surface-to-surface separation of any two neighboring particles is equal to the particle diameter. In the above drawing, the particle sizes are not shown to scale. Also, since the particles mask each other in each projection, the number of visible particles is smaller than 300.

In the plots below, we show the dimensionless efficiencies of extinction, scattering and absorption defined as

$$\eta_e = \frac{\langle Q_e \rangle_t}{\frac{c}{8\pi} |\mathbf{E}_0|^2 N \pi a^2} \quad (53)$$

for extinction, and similarly for scattering and absorption. Thus, the total extinguished (or scattered, or absorbed) power is divided by the incident energy flux of a plane wave of amplitude $\mathbf{E}_0$ and then by the total geometrical cross section of all particles. A linearly polarized plane wave was used for excitation, so that the incident energy flux was well defined. The plane wave was polarized along the X -axis (the horizontal direction in the left and center images in Fig. 2) and propagated along the Z -axis (the vertical direction in the center and right images in Fig. 2). The extinguished power was computed according to (42), which is a valid expression in both approaches to the CDA. The absorbed power was computed according to (50), which is also valid in both approaches, albeit we need to keep in mind that the coefficient q in (50) is defined through the bare polarizability according to (51). It is important for us that q is independent of any corrections to the bare polarizability, as long as the non-radiative correction is purely real. We have used (24b) to define both the radiative and the non-radiative corrections, so that the above condition is satisfied. The scattered power was computed as the difference between the extinguished and the absorbed powers. This can yield a negative result if the model is constructed incorrectly, i.e., if the radiative corrections are not accounted for.

In the simulations, we have used the First Approach to the CDA and, correspondingly, the set of equations (8). Accounting or not accounting for the corrections in question affected only the polarizabilities α_n that appear in (8) and, consequently, the computed dipole moments $\mathbf{d}_n$, but not the form of the equations (42) or (50). In all cases, we assumed that χ_n are given by the quasistatic expression (16) and independent of n . The renormalized polarizabilities

were determined according to (24b). If none of the terms $O(k^2/a)$ and $O(k^3)$ were retained in this expression (so that, trivially, $\alpha_n = \chi_n$), we say that no corrections were used. If only the $O(k^3)$ term was retained, we say that only the radiative correction was used. If only the $O(k^2/a)$ term was retained, we say that only the non-radiative correction was used. If both terms were retained, then both corrections were accounted for.

Note that we used (24b) rather than (23) to compute the non-radiative corrections. The difference is however rather minor.

6.1. Particles with Lorentz-type permittivity in the anomalous dispersion region

We start with the case when the particle material can be described by the Lorentz formula

$$\epsilon(\omega) = 1 + \frac{\omega_0^2(\epsilon_0 - 1)}{\omega_0^2 - \omega^2 - i\gamma\omega}. \quad (54)$$

Here $\epsilon_0 = \epsilon(0)$ is the static limit of the dielectric function, ω_0 is the resonance frequency and γ is the relaxation constant. Let $\lambda_0 = 2\pi c/\omega_0$ be the wavelength at the resonance frequency. Then, we can characterize the system completely by the five dimensionless parameters a/h , a/λ_0 , ω/ω_0 , γ/ω_0 and ϵ_0 .

For the simulations, we have fixed $a/h = 0.25$ and $\epsilon_0 = 4.0$. In this case, the optical resonances of the system occur when $1.32 \lesssim \omega/\omega_0 \lesssim 1.48$. As could be expected, this spectral window is located in the anomalous dispersion region where $\text{Re}[\epsilon(\omega)] < 0$. Further, we have considered two possible ratios $a/\lambda_0 = 0.01$ and $a/\lambda_0 = 0.02$ and two possible attenuation strengths, $\gamma/\omega_0 = 0.01$ and $\gamma/\omega_0 = 0.001$. The above parameters are typical for transparent dielectrics. Also, the selected values of a/h and a/λ_0 guarantee that the dipole approximation is reasonably accurate.

In Figs. 3,4 and 5, we plot the extinction, scattering, and absorption efficiencies for the parameter set described

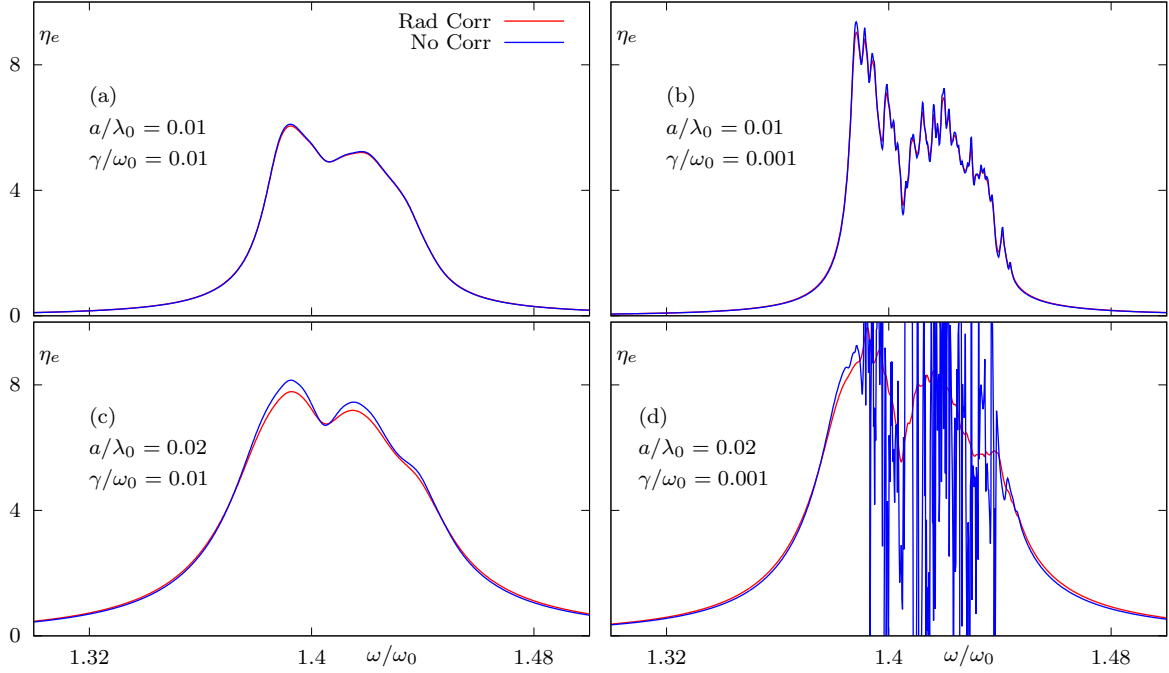


Figure 3: Extinction efficiency as a function of frequency for the particle geometry shown in Fig. 2. The particle material is described by the Lorentz formula (54). The ratio $a/h = 0.25$ is fixed while a/λ_0 and γ/ω_0 vary as labeled. Here $\lambda_0 = 2\pi c/\omega_0$.

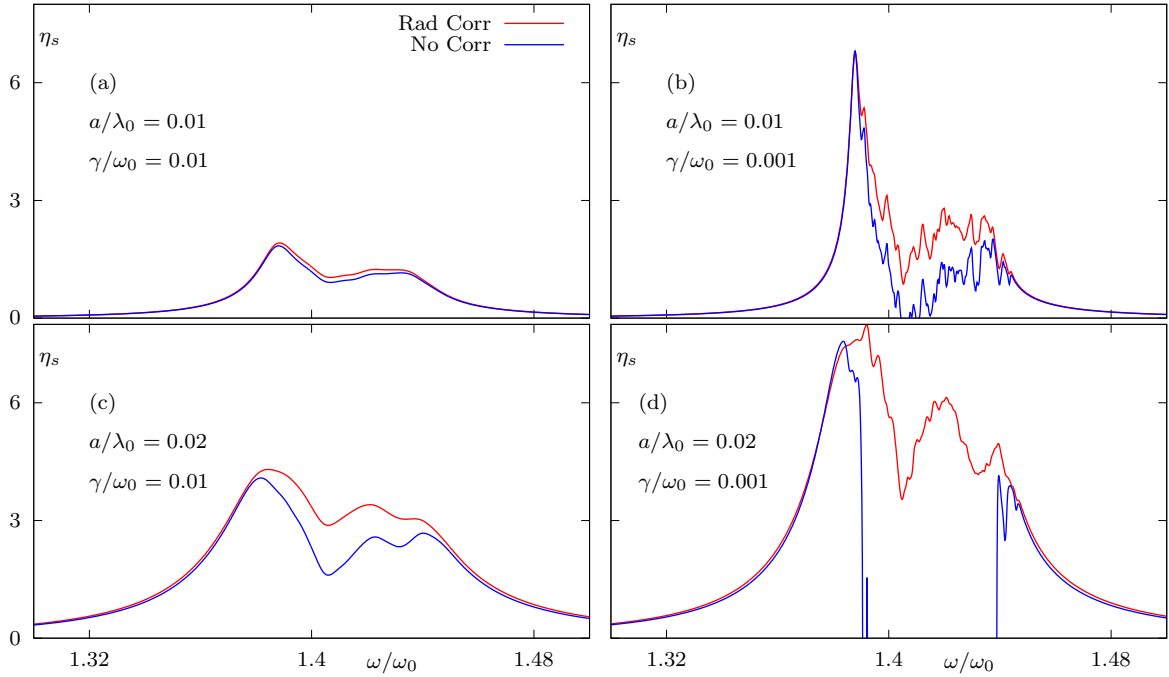


Figure 4: Same as in Fig. 3 but for the scattering efficiency.

above. In these three figures, we compare the case when the radiative correction was accounted for (the curves labeled as “Rad Corr”) to the case when no corrections were used at all so that $\chi_n = \alpha_n$ (the curves labeled “No Corr”). It can be seen that, for relatively small particles and relatively large losses

[the case when $a/\lambda_0 = \gamma/\omega_0 = 0.01$ is shown in Panels (a) of all three Figures], the radiative correction is indeed not important. However, if we decrease losses by the factor of 10 [Panels (b)] or increase the particle size by the factor of 2 [Panels (c)], the effects of the radiative correction become

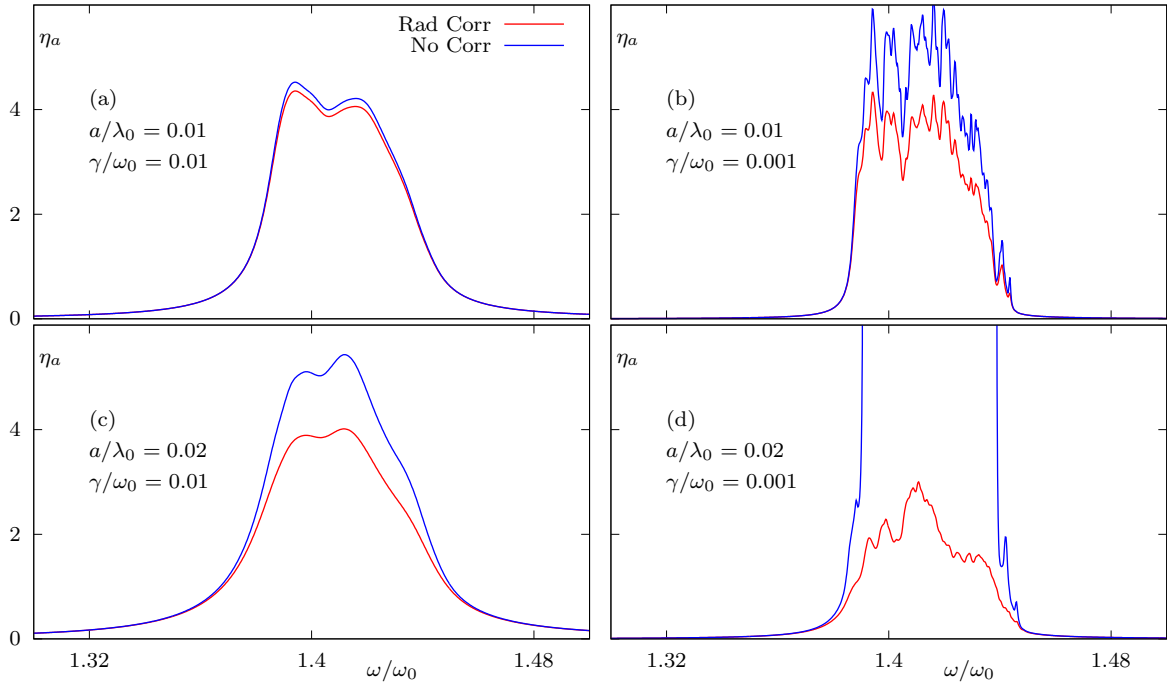


Figure 5: Same as in Fig. 3 but for the absorption efficiency.

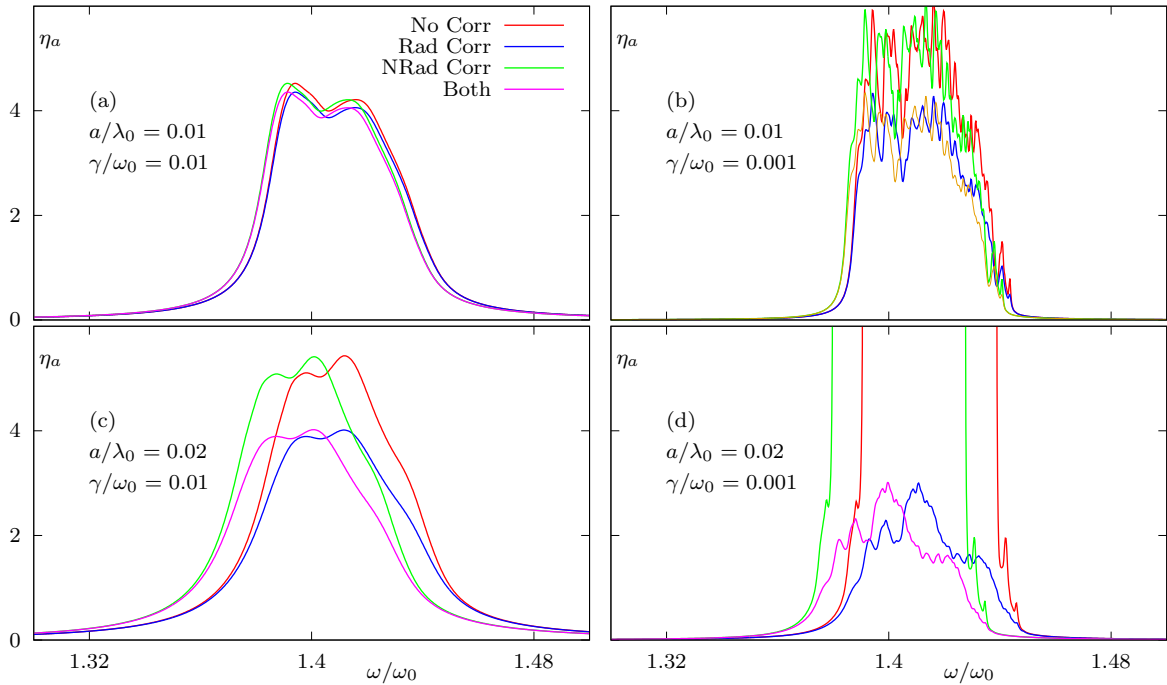


Figure 6: Same as in Fig. 5 but four different cases are compared: no corrections are included ("No Corr"), only radiative corrections are included ("Rad Corr"), only non-radiative corrections are included ("NRad Corr"), and both corrections are included ("Both").

quite noticeable, especially on absorption and scattering (extinction is not as affected). If we do both, that is, reduce losses and increase the particle size, as shown in Panels (d), the effect of the radiative correction becomes dramatic.

Without the correction, the absorption efficiency is grossly overestimated and the scattering efficiency is negative.

We now turn to the role of the non-radiative corrections. In Fig. 6, we compare the absorption efficiency for four different cases. First, we show the case where no corrections

are taken into account. Second, we show the case where only the radiative correction is accounted for. Third, we show the case where only the non-radiative correction is accounted for. Finally, we show the case when both the radiative and the non-radiative corrections are included according to (24b). It can be seen that, in the case $a/\lambda_0 = \gamma/\omega_0 = 0.01$ [Panel (a)], all corrections are rather insignificant. One can account or not account for any of these corrections, and the variation of the results is likely to be within the precision of any experimental measurement. However, the cases shown in Panels (b),(c) and (d) are affected by the corrections. It can be seen that accounting for the non-radiative correction results only in spectral shifts but does not fix the anomalies that can appear in the spectra if the radiative correction is ignored. It is difficult to tell how significant the above-mentioned spectral shifts are. This depends on the overall precision of the model, the degree to which the geometry of the aggregate is known, etc. We can only mention here that there exist other mechanisms that can result in spectral shifts such as the higher-order multipole interactions, and it may be impossible to disentangle all these contributions.

6.2. Particles with Drude permittivity

The importance of radiative corrections is most easily illustrated for metallic particles. In this Subsection, we consider exactly the same geometry of the aggregate as above, including the fixed ratio $a/h = 0.25$, but assume that the particles permittivity is given by the Drude formula

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i\gamma)}, \quad (55)$$

where ω_p is the plasma frequency and γ is the relaxation constant. The optical resonances in this case are located in the frequency range $0.50 \lesssim \omega/\omega_p \lesssim 0.65$. In the simulations, we have assumed that $\gamma/\omega_p = 0.002$. Further, we have considered particles of two different physical radii: $a = 5\text{nm}$ and $a = 10\text{nm}$ and defined ω_p so that the wavelength at the plasma frequency is $\lambda_p = 2\pi c/\omega_p \approx 136\text{nm}$. These parameters are characteristic of silver. We therefore have in the resonance spectral region $a/\lambda \approx 0.02$ for $a = 5\text{nm}$ and $a/\lambda \approx 0.04$ for $a = 10\text{nm}$.

In Fig. 7, we plot the absorption efficiency as a function of ω/ω_p . All four cases are displayed (no corrections used, only radiative corrections used, only non-radiative corrections used, both corrections used). It can be seen that, for $a = 5\text{nm}$, not accounting for the radiative corrections overestimates the absorption efficiency by approximately 50%. This can result in negative values for the scattering efficiency at some frequencies (data not shown). At the same time, not accounting for the non-radiative correction results in a small spectral shift. In the case $a = 10\text{nm}$, which is typical in plasmonics, the effect of not accounting for the radiative corrections is quite dramatic. The absorption efficiency becomes grossly over-estimated and unrealistically large; the scattering efficiency is negative in the whole resonance spectral range. However, the non-radiative corrections still

result only in a spectral shift, although, in this case, the shift is more pronounced.

In the above simulations, we assumed that the ratio γ/ω_p is fixed and did not account for the so-called finite-size effects, which are believed to increase γ when $a \rightarrow 0$. Although introduction of such size dependence of γ has a theoretical justification [46, 47, 48], the relevant theory has several *ad hoc* elements and becomes rather complicated for particles smaller than $\sim 5\text{nm}$. In particular, the relaxation constant is not reduced to the conventional expression $\gamma(a) = \gamma_\infty + Av_F/a$ for such very small particles [48]. Here γ_∞ is the bulk value of the relaxation constant, A is a numerical parameter, and v_F is the Fermi velocity for the degenerate electron gas. The homogeneous line-width of gold nanocylinders of the length $\sim 50\text{nm}$ and radius $\sim 8\text{nm}$, and the corresponding values of the parameter A , were investigated experimentally in [49]. It was found that the homogeneous spectral lines of isolated nanoparticles do not differ significantly from the theoretical predictions in which the bulk value γ_∞ was used. We have mentioned all this in order to argue that accounting for the finite-size effects is unlikely to change the conclusion that the radiative corrections are important in plasmonics and should be included in the model for most experimentally-relevant parameter sets.

6.3. Beyond the dipole approximation

Above, it was shown that not accounting for the radiative corrections can result in anomalous results in the spectral regions where the optical resonances occur. If the corrections are taken into account, the results appear to be “normal” but still there is a question whether these results are correct. The question boils down to determining the limits of applicability of the CDA and also to determining which changes one can expect when going beyond the CDA. The only way to answer these questions is to compare the CDA to exact solutions.

For aggregated spheres, exact solutions to the electromagnetic boundary-value problem can be obtained by considering higher-order multipoles of each sphere and solving the corresponding *coupled-multipole equations* (CMEs) [50, 51]. Obviously, the CDEs is a truncation of the CMEs in which all higher multipole moments of the particles are assumed to be zero. However, there exists a more fundamental difference. Namely, the various corrections to the dipole polarizability do not play a significant role in the more precise setting of the CMEs. Indeed, in the CMEs, the particles are no longer considered to be points, and one does not need to be concerned with the physical problems that the above assumption can cause.

In the simplified quasistatic version of the CMEs, when one assumes that the aggregate is small considered to the wavelength and (for the sake of simplicity) that all particles are made of the same material [52, 53, 54], the solutions

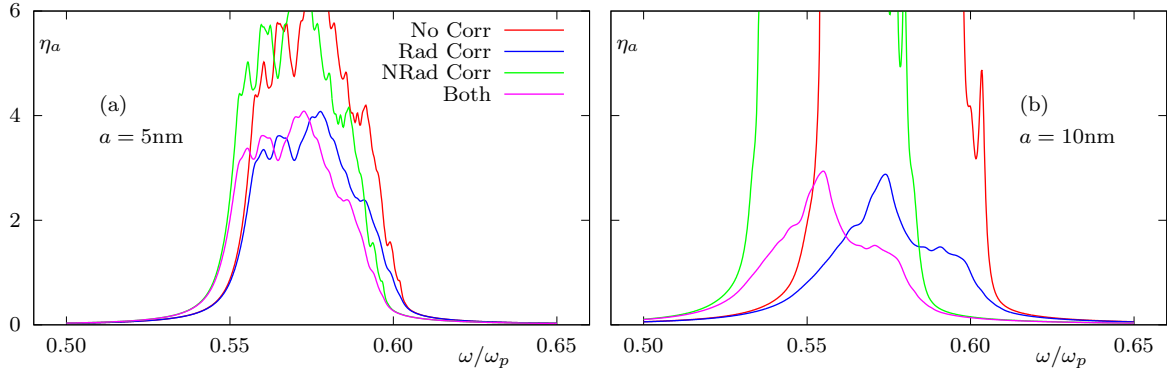


Figure 7: Absorption efficiency for the case when the particle permittivity is described by the Drude formula (55).

depend on the *spectral parameter*¹¹

$$z(\omega) = \frac{4\pi}{3} \frac{\epsilon(\omega) + 2}{\epsilon(\omega) - 1}, \quad (56)$$

and the optical resonances occur when $z(\omega)$ is close to one of the (purely real) eigenvalues w_n of the coupled-multipole equations. If, in addition to the quasistatic approximation, the dipole approximation is used, there are exactly $3N$ eigenvalues w_n , some of which can be degenerate or have very small or zero oscillator strengths (squared projections of the corresponding eigenvector onto the incident field). If we now remove the dipole approximation but still stay within the quasistatics, the number of linear equations in the CMEs would become formally infinite, and the same is true for the number of eigenvalues. However, the sum of all the oscillator strengths remains constant and finite (this is known as the sum rule). Nevertheless, it should be clear that account of higher-order multipoles can shift the optical resonances or give rise to new resonances that are not present in the dipole approximation. If the particles are allowed to touch, the surfaces of discontinuity in the boundary-value problem are no differentiable and the spectrum of w_n becomes a continuous distribution.

A convenient theoretical framework for investigating these effects (within the quasistatics) is based on the density of states formalism [57]. The density of states $\Gamma(w)$ defines a generalized spectrum of the aggregate. If the particles in the aggregate are only weakly absorbing and one can write $z(w) \approx x(w) - i\delta(w)$ and $\delta(w)$ is in some sense small, we have, for example, $\sigma_e \propto \Gamma[x(w)]$. More generally, the extinction cross section is given by a Hilbert transform of $\Gamma(w)$. In [57], examples are given of the changes in $\Gamma(w)$ for aggregates of touching spheres due to accounting for the higher-order multipole moments. The changes can be dramatic, i.e., two narrow spectral lines can be broadened into a relatively wide continuous spectrum (for the external field parallel to the axis of two touching spheres), or modest, when a resonance is spectrally shifted but remains of approximately the same shape (when the external polarization

is orthogonal to the above axis). Of course, these examples were given for the case of touching spheres when the CDA is not applicable. As the spheres move apart, the exact spectra approach those of the CDA, as expected.

More recently, a number of investigations of the CMEs for periodic two-dimensional arrangements of high-conductivity metal spheres have been carried out without the use of the quasistatic approximation [18, 58] (in these references, the higher-order multipoles are restricted to the magnetic dipole, electric and magnetic quadrupoles). The geometrical parameters in [18, 58] were somewhat similar to those used in the previous subsections: the particles were not touching and separated by approximately one particle diameter. It was found that the coupling of higher-order multipoles can give rise in such systems to additional sharp resonances. The resonances occur due to diffractive coupling in infinite (or just very large) periodic arrays. Spectral shifts of the dipole spectral lines were also observed. Note however that the particle size in [58, 18] was rather large compared to the wavelength (near the spectral features of interest); the typical ratio ranged from $\sim 1/4$ to $\sim 1/7$, which is significantly larger than in the simulations reported in this section.

In summary, we can hope that the red curves in Figs. 3-7 have been computed with a reasonable accuracy. Account of higher-order multipoles will of course change these curves but not in a significant way. The changes will be limited to emergence of new comparatively weak spectral lines and to spectral shifts of the lines that are already present in the dipole approximation. The spectral region where optical resonances are present can be moderately broadened. However, if we consider large periodic arrays, new interesting spectral features resulting from the diffractive coupling of the higher-order multipoles can emerge.

7. Paradoxes involving extinction

7.1. Classical extinction paradox

The classical extinction paradox is based on the observation that the extinction cross section of a large sphere of radius a is approximately twice its geometrical cross section, πa^2 . The paradox was stated for a sphere mainly because Maxwell's equations are exactly solvable in this geometry,

¹¹Or just $1/[\epsilon(\omega) - 1]$, which differs from (56) only by a real shift and scaling and is known as the Bergman-Milton spectral parameter and originally introduced in the theory of composites [55, 56].

so that the paradox can be demonstrated in computations. However, a similar mismatch can be obtained for a cube or any other shape. The paradox occurs for absorbing or non-absorbing objects and can be stated for the scattering or the extinction cross sections.

The extinction paradox was explained by Brillouin in 1949 [59]. Modern expositions including extensive numerical demonstrations and detailed theoretical reasoning can be found in [60] (in 2D geometry) and in [61] (for the full 3D problem). However, the very fact that the mismatch between the optical and the geometrical cross sections was viewed as a paradox, and that numerous attempts to explain it by edge diffraction effects have been made, is very indicative. All this activity is based on the erroneous assumption that the extinguished energy is somehow “removed” from the incident beam. In fact, as was discussed in the previous Section, neither the extinguished nor the scattered powers can be associated with a measurable power flux.

In fact, the extinction paradox is easily explained within elementary geometrical optics. This is illustrated in Fig. 8. Here a wide beam of rays is incident on a mirror of some area S , which is slightly tilted for convenience of measurements. The total power incident on the mirror is W and all of it is reflected back at a small angle to the original propagation direction. Due to this non-zero angle, the incident and the reflected beams can be spatially separated and the back-reflected power can be measured. But this is not all of the scattered field. Since there is a geometrical shadow behind the mirror, and accounting for the fundamental superposition principle, we must conclude that the mirror also produces a forward-scattered radiation, which cancels the incident radiation to produce the shadow. This field was referred to as the Ewald-Oseen field in [61]; it was also remarked in that reference that the Ewald-Oseen field (and the theorem that explains the formation of the shadow) is seemingly unrelated to the scattering experiments that are considered in conjunction with the classical extinction paradox. However, the Ewald-Oseen field is just some part of the scattered field that can not be separated spatially from the incident field. Therefore the Ewald-Oseen field can not be measured. The counter-intuitive fact here is that a large fraction of the scattered power is not measurable; correspondingly, the total scattered power is also not measurable.

The example of Fig. 8 illustrates the basic explanation of the extinction paradox: the scattered and the incident field can not be spatially separated. The scattered field for this reason is not physically related to any measurable power flux and, in particular, it is not required to be equal to the power flux incident on the mirror. Of course, the actual wave phenomena involved in the scattering process are much more complicated. The extinction paradox can be observed even in the case of large transparent objects when no well-defined geometrical shadow is formed, or in small objects with a large permittivity. This is discussed in more detail in Ref. [61]. Here we emphasize the basic principle of the extinction paradox can be easily demonstrated with the use of the CDA.

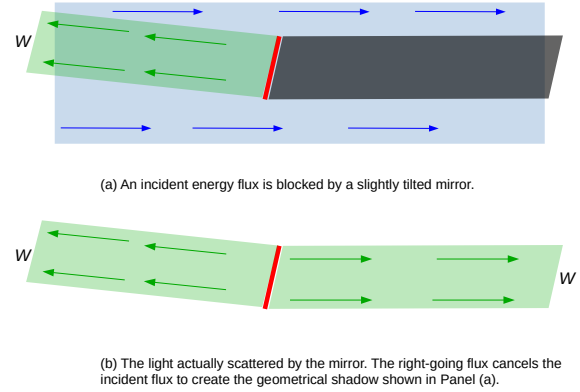


Figure 8: Illustration of the classical extinction paradox. Power flux W is incident on a slightly tilted mirror, which reflects the flux back at some small angle where the reflected beam can be detected and measured. The reflected power is W for an ideal mirror. The mirror also scatters the energy flux W in the forward direction to cancel the incident light and create the shadow area. The total power scattered by the mirror is $2W$ and the total extinction cross section is twice its geometrical cross section.

In Fig. 9, we show two examples of computing the extinction cross section for a collection of polarizable dipoles. In all cases, the incident field was assumed to be a plane wave propagating in the Z -direction. The dipoles formed four equidistant layers in the planes $z = (n - 1)h$, where $n = 1, 2, 3, 4$ and h is the separation between the layers. Each layer consisted of a square $L \times L$ lattice of step h , so that the frontal geometrical cross section of the structure is a square of the area $\sigma_g = (L - 1)^2 h^2$. This expression is approximate because the edge effects are difficult to account for. In one modification, the four layers formed a simple cubic lattice of period h . In another modification, the $n = 2$ and $n = 4$ lattices were shifted by $h/2$ in both X and Y directions relative to the planes $n = 1$ and $n = 3$. In this way, a distorted (extended in the Z -direction by the factor of 2) BCC lattice was formed. In all cases, the wavelength of the incident radiation was taken to be $\lambda = 10h$. The bare polarizabilities of all particles were computed according to (16) where $\epsilon = -2 + 0.002i$ and $a = h/4$. This corresponds to some high-conductivity metal particles.

It can be seen that, indeed, the ratio of the geometric and the extinction cross sections approaches the value of 2 as the size of the square is increased. The result is not much affected by the choice of the simple cubic lattice or the distorted BCC lattice, which can be naively expected to better block the incident rays. This confirms that the extinction paradox is a wave phenomenon. Moreover, we see that the extinction cross section can be more than twice larger than the geometrical cross section. This occurs when the size of the square is relatively small and the edge diffraction effects still play a significant role. However, as the size of the square is increased, the edge diffraction effects become

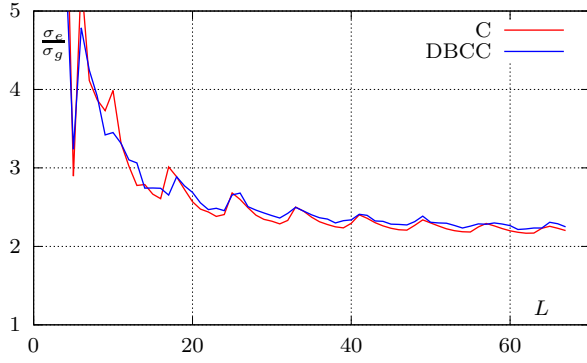


Figure 9: The ratio of the extinction and geometrical cross sections of the arrays of dipoles described in the text. L is the linear dimension of a square that is oriented perpendicularly to the direction of propagation. C denotes simple cubic lattice and DBCC denotes distorted body-centered cubic lattice.

progressively less important and the simple geometrical optics picture emerges, similar to the one illustrated in Fig. 8. The collection of metal particles act effectively as perfect mirror.

7.2. Non-additivity of extinction

Consider a scattering experiment in which two orthogonally-propagating beams are scattered by a system of dipoles or by just one single dipole. Let $\langle W_e^{(1)} \rangle_t$ be the extinguished power for the case when only one of the two beams is on and similarly for $\langle W_e^{(2)} \rangle_t$. However, the total extinguished power when both beams are on simultaneously is not a direct sum of the two quantities. We have for the total extinguished power (for just one dipole):

$$\begin{aligned} \langle W_e^{(\text{tot})} \rangle_t &= \frac{\omega}{2} \text{Im} [(\mathbf{E}_1^* + \mathbf{E}_2^*) \cdot \alpha(\mathbf{E}_1 + \mathbf{E}_2)] \\ &= \langle W_e^{(1)} \rangle_t + \langle W_e^{(2)} \rangle_t \\ &\quad + \frac{\omega}{2} \text{Im} (\mathbf{E}_1^* \cdot \alpha \mathbf{E}_2 + \mathbf{E}_2^* \cdot \alpha \mathbf{E}_1). \end{aligned} \quad (57)$$

Here $\mathbf{E}_1$ and $\mathbf{E}_2$ are the complex amplitudes of the electric field of beam 1 and beam 2, respectively, at the location of the dipole. Also recall that the renormalized polarizability must be used in (57) since $\mathbf{E}_n$ do not include the dipole self-action.

Thus, the total extinguished power is generally not a sum of the extinguished powers for each beam. Of course, this much could be easily anticipated, since extinction is a quadratic quantity (in the fields). However, what exactly does this mean for the physical interpretation of extinction? How the interference terms in (57) can be measured?

Let us assume that only one beam, say, the first, is initially on. Then, according to the common interpretation, the power $\langle W_e^{(1)} \rangle_t$ is removed from this beam. If we measure the total beam power before it encounters the scatterer and after it encounters the scatterer, the difference should be equal to $\langle W_e^{(1)} \rangle_t$. But this consideration is independent of the presence of the second beam. Indeed, even though the

second beam creates some additional scattered field, this scattered field can be arbitrarily small at the small apertures where the intensity of beam 1 is measured. Moreover, one can ask, if (57) gives the total extinguished power when both beams are on, which of the two beams is this power removed from?

The non-additivity of extinction suggests that characterizing particles by their extinction cross sections is not always useful for monochromatic fields. However, we said nothing so far about the more practical situation when the fields are partially coherent. One can hope, for example, that in the latter case the cross-terms in (57) will average out to zero.

Consideration of partially-coherent fields is beyond the scope of this article, but it is indeed reasonable to expect that extinguished powers become additive in some limit, just like the intensities of incoherent sources are additive. However, in the case of extinction, there is one additional complication. So far, we have left the spectral dependence of $\alpha(\omega)$ out of consideration. In many practical cases, i.e., for particles in the atmosphere, $\alpha(\omega)$ does not have very sharp spectral peaks. In this case incoherent addition of intensities can be expected. However, if sharp spectral peaks are present, the cross-terms in (57) will not average to zero even for completely incoherent incident beams. This is already clear for the case when $\alpha(\omega) \propto \delta(\omega - \omega_0)$. Indeed, in this case only monochromatic components of each beam (at the frequency ω_0) are important, and strictly monochromatic fields are perfectly coherent. Therefore, additivity of extinction in a practical setting depends on a complex interplay between the mutual coherence properties of the two beams and the spectral properties of $\alpha(\omega)$. In the case when many interacting dipoles are present, spectral properties of solutions to the CDEs come into play, which can also have sharp resonances.

7.3. Extinction for a collimated beam

This paradox is a variant of the paradox discussed in the previous subsection. Consider just one incident beam and one dipole. We assume however that the incident beam is very tightly collimated. The situation seems to be simple yet it involves an apparent paradox as is illustrated in Fig. 10.

Indeed, let us draw a spherical surface of radius R around the scatterer. This surface is denoted by $\partial\Omega$ in Fig. 10. We assume that R is large and at any rate $R \gg \lambda$. The two areas where the incident beam crosses $\partial\Omega$ are denoted by S_1 and S_2 . We first note that the incident and the scattered fields do not overlap on $\partial\Omega^{(-)} \equiv \partial\Omega \setminus (S_1 \cup S_2)$. The only parts of the big surface $\partial\Omega$ where the two fields overlap are S_1 and S_2 (according to the assumption of a non-divergent beam). Then, the outward power flux through $\partial\Omega^{(-)}$ is very close to $\langle W_s \rangle_t$, which is the power scattered by the small particle. Now let us estimate the power flux through the areas S_1 and S_2 . In these areas, the total field is the superposition of the incident and the scattered fields. The total flux of incident power through S_1 and S_2 is obviously zero (the power of the beam is preserved with propagation). The power flux of the scattered field through S_1 and S_2 approaches zero as $1/R^2$. Indeed, the areas of S_1 and S_2 are independent of R while

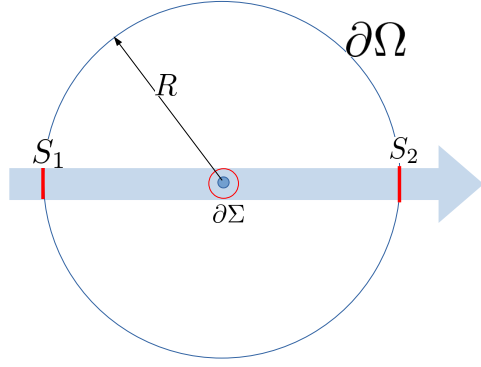


Figure 10: Illustration of the extinction paradox involving a perfectly collimated beam. The power flux through the large spherical surface minus two small areas S_1 and S_2 , $\partial\Omega^{(-)} \equiv \partial\Omega \setminus (S_1 \cup S_2)$, is, approximately, the scattered power $\langle W_s \rangle_t$. At the small areas S_1 and S_2 the incident field is dominating so that the total power flux through $S_1 \cup S_2$ is zero. The total outward flux through $\partial\Omega$ appears to be equal to $\langle W_s \rangle_t$ as if this energy is generated inside the sphere.

the scattered field decays as $1/R$. Finally, the cross terms involving the incident and the scattered field decay as $1/R$ for the same reason as above. So, in the limit $R \rightarrow \infty$, we obtain the following result: the outward power flux through $\partial\Omega$ is $\langle W_s \rangle_t > 0$. But this contradicts conservation of energy. The flux must be negative or zero (for a non-absorbing particle).

We have obtained the above paradox due to the unrealistic assumption about the incident beam, which is equivalent to the assumption that the total scattered energy can be measured precisely. The paradox shows that this is never truly the case. It is not possible to make a beam so tightly collimated that the scattered and the incident fields are entirely spatially separated. Such a possibility would entail a violation of energy conservation. Therefore, measurement of the integral scattering and extinction cross sections is a surprisingly non-trivial problem. Various approaches have been considered in the literature [62, 63] (these references consider a wide-front incident plane wave), but a general measurement scheme that does not rely on approximations or is applicable to all forms of the incident field is surprisingly difficult to find. We note that one possible explanation of the above paradox is that Gaussian beams are not as simple as commonly thought. The familiar expression that is super-exponentially localized near the optical axis is usually obtained in the scalar wave and paraxial approximations. Once these two approximations are relaxed, the form of a Gaussian beam is anything but simple [64]; in particular, energy considerations are affected by this complex mathematical structure.

To gain a further insight into the paradox, consider the small sphere denoted in Fig. 10 as $\partial\Sigma$. The sphere is drawn so that everywhere on $\partial\Sigma$ the scattered and the incident fields overlap. A sphere of this sort, or perhaps a more complicated

surface, can always be drawn. All the energy considerations developed by us previously in Section 5 apply to this surface. Now consider the space between $\partial\Sigma$ and $\partial\Omega$. Since this region is empty, the total flux of power that enters it in any stationary process is zero. So if a negative outward flux through $\partial\Sigma$ exists, exactly the same negative flux must exist through $\partial\Omega$, as long as Maxwell's equations in free space hold. The paradox of this subsection was obtained because we have made an assumption about the incident beam that is inconsistent with Maxwell's equations.

The resolution of this paradox has, in fact, far reaching consequences. What we have shown is that it is not possible to separate the scattered and the incident fields. All paradoxes of extinction are based on the incorrect implicit assumption that this is possible, as well as on the traditional interpretations of the scattered and extinguished powers. The fundamental impossibility to separate the incident and the scattered fields indicates that these interpretations should be used with caution.

8. Summary

The article was written with several goals in mind. First, we have clarified several confusing or contradictory points that related to the widely-used coupled-dipole approximation (CDA). An effort was made to keep the discussion as general and at the same time as simple as possible. Two different but equivalent variants of the coupled-dipole equations (CDEs) were described in detail. Fundamental to understanding these two variants of CDEs is the distinction between the bare and the renormalized polarizabilities. We have introduced this distinction for the most general case of tensorial polarizabilities, and illustrated it further with the example of CDEs in the vicinity of a substrate. Another important goal was to provide a general and rigorous derivation of the energy relations that are applicable within the framework of CDA. Here, however, one is encountered with some long-standing difficulties, which are not really specific to the CDA. The difficulties stem from the fundamental impossibility to separate spatially the scattered and the incident fields. Correspondingly, the scattered and the extinguished power fluxes are generally not measurable directly and assuming that they are can lead to various paradoxes, the most well-known of which is the classical extinction paradox. In the concluding parts of this article, we have discussed several such paradoxes within the theoretical framework of CDA and multiple scattering.

Data availability

All research data are for this paper are available from Mendeley Data, V1, doi: 10.17632/277jsr32md.1.

References

- [1] E. M. Purcell, C. R. Pennypacker, Scattering and absorption of light by nonspherical dielectric grains, *Astrophys. J.* 186 (1973) 705–714.
- [2] B. T. Draine, The discrete-dipole approximation and its application to interstellar graphite grains, *Astrophys. J.* 333 (1988) 848–872.

- [3] B. T. Draine, J. Goodman, Beyond Clausius-Mossotti: Wave propagation on a polarizable point lattice and the discrete dipole approximation, *Astrophys. J.* 405 (1993) 685–697.
- [4] G. H. Goedecke, S. G. O'Brien, Scattering by irregular inhomogeneous particles via the digitized green's function algorithm, *Appl. Opt.* 27 (12) (1988) 2431–2438.
- [5] A. Lakhtakia, Strong and weak forms of the method of moments and the coupled dipole method for scattering of time-harmonic electromagnetic fields, *Int. J. Mod. Phys. 3* (3) (1992) 583–603.
- [6] P. C. Chaumet, A. Sentenac, A. Rahmani, Coupled dipole method for scatterers with large permittivity, *Phys. Rev. E* 70 (2004) 036606.
- [7] M. A. Yurkin, A. G. Hoekstra, The discrete dipole approximation: An overview and recent developments, *J. Quant. Spectrosc. Radiat. Transfer* 106 (2007) 558–589.
- [8] B. Draine, P. Flatau, Discrete-dipole approximation for scattering calculations, *J. Opt. Soc. Am. A* 11 (1994) 1491–1499.
- [9] B. T. Draine, P. J. Flatau, Discrete-dipole approximation for periodic targets: theory and tests, *J. Opt. Soc. Am. A* 25 (2008) 2593–2703.
- [10] M. A. Yurkin, *Handbook of Molecular Plasmonics*, Pan Stanford Pub., Singapore, 2013, Ch. Computational approaches for plasmonics, pp. 83–135.
- [11] D. W. Mackowski, Calculation of total cross sections of multiple-sphere clusters, *J. Opt. Soc. Am. A* 11 (1994) 2851–2861.
- [12] L. L. Foldy, The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers, *Phys. Rev.* 67 (1945) 107.
- [13] M. Lax, Multiple scattering of waves, *Rev. Mod. Phys.* 23 (1951) 287.
- [14] C. F. Bohren, D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, John Wiley & Sons, New York, 1998.
- [15] R. Ruppin, Optical absorption of two spheres, *J. Phys. Soc. Japan* 58 (4) (1989) 1446–1451.
- [16] I. E. Mazets, Polarization of two close metal spheres in an external homogeneous electric field, *Technical Phys.* 45 (10) (2000) 1238–1240.
- [17] B. Khlebtsov, A. Melnikov, V. Zharov, N. Khlebtsov, Absorption and scattering of light by a dimer of metal nanospheres: comparison of dipole and multipole approaches, *Nanotechnology* 17 (2006) 1437–1445.
- [18] S. D. Swiecicki, J. E. Sipe, Surface-lattice resonances in two-dimensional arrays of spheres: Multipolar interactions and amode-analysis, *Phys. Rev. B* 95 (2017) 195406.
- [19] Y. A. Avetisyan, A. I. Zaitsev, V. A. Malyshev, Theory of superradiance of multiatomic systems with account of the resonance dipole-dipole interaction of the atoms, *Opt. Spectrosc.* 59 (5) (1986) 582–586.
- [20] P. B. Allen, Dipole interactions and electrical polarity in nanosystems: The Clausius-Mossotti and related models, *J. Chem. Phys.* 120 (2004) 2951–2962.
- [21] D. Vanzo, B. J. Topham, Z. G. Soos, Dipole-field sums, Lorentz factors and dielectric properties of organic molecular films modeled as crystalline arrays of polarizable points, *Adv. Funct. Mater.* 25 (2015) 2004–2012.
- [22] M. V. Berry, I. C. Percival, Optics of fractal clusters such as smoke, *Optica Acta* 33 (5) (1986) 577–591.
- [23] G. W. Mulholland, R. D. Mountain, Coupled dipole calculation of extinction coefficient and polarization ratio for smoke agglomerates, *Combustion and Flame* 119 (1-2) (1999) 56–68.
- [24] S. Zou, N. Janel, G. C. Schatz, Silver nanoparticle array structures that produce remarkably narrow plasmon lineshapes, *J. Chem. Phys.* 120 (23) (2004) 10871–10875.
- [25] V. G. Kravets, F. Schedin, R. Jalil, L. Britnell, R. V. Gorbachev, D. Ansell, B. Thackray, K. S. Novoselov, A. K. Geim, A. V. Kabashin, A. N. Grigorenko, Singular phase nano-optics in plasmonic metamaterials for label-free single-molecule detection, *Nature Materials* 12 (2013) 304–309.
- [26] O. Leseur, R. P. R. Pierrat, R. Carminati, High-density hyperuniform materials can be transparent, *Optica* 3 (2016) 763.
- [27] J.-P. Hugonin, M. Besbes, P. Ben-Abdallah, Fundamental limits for light absorption and scattering induced by cooperative electromagnetic interactions, *Phys. Rev. B* 91 (2015) 180202.
- [28] F. Bigourdan, R. Pierrat, R. Carminati, Enhanced absorption of waves in stealth hyperuniform disordered media, *Opt. Expr.* 19 (2019) 8666.
- [29] D. A. Smuneev, P. C. Chaumet, M. A. Yurkin, Rectangular dipoles in the discrete dipole approximation, *J. Quant. Spectrosc. Radiat. Transfer* 156 (2015) 67–79.
- [30] I. L. Rasskazov, S. V. Karpov, V. A. Markel, Surface plasmon polaritons in curved chains of metal nanoparticles, *Phys. Rev. B* 90 (2014) 075405.
- [31] V. A. Markel, Scattering of light from two interacting spherical particles, *J. Mod. Opt.* 39 (4) (1992) 853–861.
- [32] T. V. Shahbazyan, Spontaneous decay of a quantum emitter near a plasmonic nanostructure, *Phys. Rev. B* 98 (2018) 115401.
- [33] G. Y. Panasyuk, J. C. Schotland, V. A. Markel, Short-distance expansion for the electromagnetic half-space green's tensor: general results and an application to radiative lifetime computations, *J. Phys. A* 42 (27) (2009) 275203.
- [34] A. E. Moskalensky, M. A. Yurkin, Energy budget and optical theorem for scattering of source-induced fields, *Phys. Rev. A* 99 (2019) 053824.
- [35] A. Lakhtakia, G. W. Mulholland, On two numerical techniques for light scattering by dielectric agglomerated structures, *J. Res. Natl. Inst. Stand. Technol.* 98 (6) (1993) 699–716.
- [36] Y. E. Danilova, A. I. Plekhanov, V. P. Safonov, Experimental study of polarization-selective holes, burning in absorption spectra of metal fractal clusters, *Physica A* 185 (1992) 61–65.
- [37] V. A. Markel, Maxwell Garnett approximation (advanced topics): tutorial, *J. Opt. Soc. Am. A* 33 (2016) 2237–2255.
- [38] D. Gutkowicz-Krusin, B. T. Draine, Propagation of electromagnetic waves on a rectangular lattice of polarizable points, *arXiv:astro-ph/0403082* (2004).
- [39] M. A. Taubenblatt, T. K. Tran, Calculation of light scattering from particles and structures on a surface by the coupled-dipole method, *J. Opt. Soc. Am. A* 10 (5) (1993) 912–919.
- [40] L. Novotny, Allowed and forbidden light in near-field optics. ii. interacting dipolar particles, *J. Opt. Soc. Am. A* 14 (1997) 105–113.
- [41] B. J. Soller, D. G. Hall, Scattering enhancement from an array of interacting dipoles near a planar waveguide, *J. Opt. Soc. Am. B* 19 (2002) 2437–2448.
- [42] A. Sommerfeld, Über die ausbreitung der wellen in der drahtlosen telegraphie, *Annalen Der Physik* 28 (1909) 665–736.
- [43] A. A. Maradudin, D. L. Mills, Scattering and absorption of electromagnetic radiation by semi-infinite medium in the presence of surface roughness, *Phys. Rev. B* 11 (4) (1975) 1392–1415.
- [44] V. A. Markel, Antisymmetrical optical states, *J. Opt. Soc. Am. B* 12 (10) (1995) 1783–1791.
- [45] P. Meakin, Formation of fractal clusters and networks by irreversible diffusion-limited aggregation, *Phys. Rev. Lett.* 51 (13) (1983) 1119–1122.
- [46] F. Hache, D. Ricard, C. Flytzanis, Optical nonlinearities of small metal particles: surface-mediated resonance and quantum size effects, *J. Opt. Soc. Am. B* 3 (12) (1986) 1647–1655.
- [47] S. G. Rautian, Nonlinear saturation spectroscopy of the degenerate electron gas in spherical metallic particles, *J. Exp. Theor. Phys.* 85 (3) (1997) 451–461.
- [48] A. A. Gomyadinov, G. Y. Panasyuk, J. C. Schotland, V. A. Markel, Theoretical and numerical investigation of the size-dependent optical effects in metal nanoparticles, *Phys. Rev. B* 64 (2001) 155461.
- [49] L. Qiu, T. A. Larson, D. Smith, E. Vitkin, M. D. Modell, B. A. Korgel, K. V. Sokolov, E. B. Hanlon, I. Itzkan, L. T. Perelman, Observation of plasmon line broadening in single gold nanorods, *Appl. Phys. Lett.* 93 (2008) 153106.
- [50] D. W. Mackowski, M. Mishchenko, Calculation of the t matrix and the scattering matrix for ensembles of spheres, *J. Opt. Soc. Am. A* 13 (11) (1996) 2266–2278.

- [51] K. A. Fuller, D. W. Mackowski, Light Scattering by Nonspherical Particles, Academic Press, 2000, Ch. Electromagnetic scattering by compounded spherical particles, pp. 226–273.
- [52] J. M. Gerardy, M. Ausloos, Absorption spectrum of clusters of spheres from the general solution of Maxwell's equations. the long-wave limit, Phys. Rev. B 22 (10) (1980) 4950–4959.
- [53] F. Claro, Theory of resonant modes in particulate matter, Phys. Rev. B 30 (9) (1984) 4989–4999.
- [54] D. W. Mackowski, Electrostatics analysis of radiative absorption by sphere clusters in the rayleigh limit: Application to soot particles, Appl. Opt. 34 (18) (1995) 3535–3545.
- [55] D. J. Bergman, Dielectric constant of a two-component granular composite: A practical scheme for calculating the pole spectrum, Phys. Rev. B 19 (4) (1979) 2359–2368.
- [56] G. W. Milton, Bounds on the complex dielectric constant of a composite material, Appl. Phys. Lett. 37 (1980) 300.
- [57] V. A. Markel, V. N. Pustovit, S. V. Karpov, A. V. Obuschenko, V. S. Gerasimov, I. L. Isaev, Electromagnetic density of states and absorption of radiation by aggregates of nanospheres with multipole interactions, Phys. Rev. B 70 (5) (2004) 054202.
- [58] A. B. Evlyukhin, C. Reinhardt, U. Zywietz, B. N. Chichkov, Collective resonances in metal nanoparticle arrays with dipole-quadrupole interactions, Phys. Rev. B 85 (2012) 245411.
- [59] L. Brillouin, The scattering cross section of spheres for electromagnetic waves, J. Appl. Phys. 20 (1949) 1110–1125.
- [60] H. M. Lai, W. Y. Wong, W. H. Wong, Extinction paradox and actual power scattered in light beam scattering: a two-dimensional study, J. Opt. Soc. Am. A 21 (2004) 2324–2333.
- [61] M. J. Berg, C. M. Sorensen, A. Chakrabarti, A new explanation of the extinction paradox, J. Quant. Spectrosc. Radiat. Transfer 112 (2011) 1170–1181.
- [62] M. J. Berg, N. R. Subedi, P. A. Anderson, N. F. Fowler, Using holography to measure extinction, Opt. Lett. 39 (2014) 3993–3996.
- [63] M. J. Berg, N. R. Subedi, P. A. Anderson, Measuring extinction with digital holography: nonspherical particles and experimental validation, Opt. Lett. 42 (2017) 1011–1014.
- [64] Y. I. Salamin, Fields of a focused linearly polarized gaussian beam: truncated series versus the complex-source-point spherical-wave representation, Opt. Lett. 34 (2009) 684–685.

A. Green's tensor in free space

Here we follow the notations of [37], see Eqs. 35 through 42. Note that some of the symbols used below such as $\mathbf{K}$ should be interpreted only within this Appendix; in other sections, they have different meaning or use.

The free-space, frequency-domain Green's tensor is defined as the solution to the following equation

$$[(\nabla \times \nabla \times) - k^2] \mathbf{G}(\mathbf{r}, \mathbf{r}') = 4\pi k^2 \mathbf{I} \delta(\mathbf{r} - \mathbf{r}'), \quad (58)$$

where $k = \omega/c$ and $\mathbf{I}$ is the identity tensor. The Green's tensor thus defined describes the radiation produced by an oscillating electric polarization in a physically-small volume. More specifically, for a general spatial distribution of polarization $\mathbf{P}(\mathbf{r})$, the electric field $\mathbf{E}_s(\mathbf{r})$ that is radiated or produced by this polarization is given by

$$\mathbf{E}_s(\mathbf{r}) = \int \mathbf{G}(\mathbf{r}, \mathbf{r}') \mathbf{P}(\mathbf{r}') d^3 r'. \quad (59)$$

We can use the Fourier transform technique and the translational invariance of free space to compute $\mathbf{G}(\mathbf{r}, \mathbf{r}')$. To this end, we start with the Fourier expansion of the form

$$\mathbf{G}(\mathbf{r}, \mathbf{r}') = \int \mathbf{K}(\mathbf{p}) e^{i\mathbf{p} \cdot (\mathbf{r} - \mathbf{r}')} \frac{d^3 p}{(2\pi)^3} \quad (60)$$

and substitute this expression into (58). This results in the momentum-space equation

$$[(\mathbf{p} \times \mathbf{p} \times) + k^2 \mathbf{I}] \mathbf{K}(\mathbf{p}) = -4\pi k^2 \mathbf{I}, \quad (61)$$

where we have referred to the Fourier variable $\mathbf{p}$ as to the “momentum”, although the analogy here is loose. We can solve the algebraic equation (61) directly with the result

$$\mathbf{K}(\mathbf{p}) = 4\pi \frac{k^2 \mathbf{I} - \mathbf{p} \otimes \mathbf{p}}{p^2 - k^2}, \quad (62)$$

where $\otimes$ denotes tensor product. Note that $\text{Tr}[\mathbf{K}(\mathbf{p})]$ does not approach zero when $|\mathbf{p}| \rightarrow \infty$. This means that the integral (60) is singular. We can however extract the singular part analytically by re-writing (62) identically as

$$\mathbf{K}(\mathbf{p}) = -\frac{4\pi}{3} \mathbf{I} + \mathbf{K}_R(\mathbf{p}), \quad (63a)$$

where

$$\mathbf{K}_R(\mathbf{p}) = \frac{4\pi}{3} \frac{(2k^2 + p^2)\mathbf{I} - 3\mathbf{p} \otimes \mathbf{p}}{p^2 - k^2}. \quad (63b)$$

It can be seen that $\text{Tr}[\mathbf{K}_R(\mathbf{p})] \rightarrow 0$ when $|\mathbf{p}| \rightarrow \infty$; therefore, $\mathbf{K}_R(\mathbf{p})$ is the Fourier transform of the regular part of the Green's tensor. Substituting (63) into (60) and integrating, we find the real-space representation of the Green's tensor:

$$\mathbf{G}(\mathbf{r}, \mathbf{r}') = -\frac{4\pi}{3} \mathbf{I} \delta(\mathbf{r} - \mathbf{r}') + \mathbf{G}_R(\mathbf{r}, \mathbf{r}'), \quad (64a)$$

$$\begin{aligned} \mathbf{G}_R(\mathbf{r}, \mathbf{r}') = & \left[\left(\frac{k^2}{|\mathbf{r} - \mathbf{r}'|} + \frac{ik}{|\mathbf{r} - \mathbf{r}'|^2} - \frac{1}{|\mathbf{r} - \mathbf{r}'|^3} \right) \mathbf{I} \right. \\ & + \left(-\frac{k^2}{|\mathbf{r} - \mathbf{r}'|} - \frac{3ik}{|\mathbf{r} - \mathbf{r}'|^2} + \frac{3}{|\mathbf{r} - \mathbf{r}'|^3} \right) \\ & \times \left. \frac{(\mathbf{r} - \mathbf{r}') \otimes (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^2} \right] e^{ik|\mathbf{r} - \mathbf{r}'|}. \end{aligned} \quad (64b)$$

Here $\mathbf{G}_R(\mathbf{r}, \mathbf{r}')$ is the regular part of the Green's tensor, which is used in the CDEs. In particular, the quantities G_{nm} (for $n \neq m$) in (8) and (17) are defined as

$$G_{nm} = G_R(\mathbf{r}_n, \mathbf{r}_m) \text{ (for } n \neq m \text{)}. \quad (65)$$

To determine the diagonal terms G_{nn} , we still use the regular part of the Green's tensor defined in (64b) and consider the small-distance expansions of the form

$$\begin{aligned} \text{Re}[G_R(0, \mathbf{r})] = & \left(-\frac{1}{r^3} + \frac{k^2}{2r} \right) \mathbf{I} \\ & + \left(\frac{3}{r^3} + \frac{k^2}{2r} \right) \frac{\mathbf{r} \otimes \mathbf{r}}{r^2} + O(r), \end{aligned} \quad (66a)$$

$$\begin{aligned} \text{Im}[G_R(0, \mathbf{r})] = & \left(\frac{2k^3}{3} - \frac{2k^5 r^2}{15} \right) \mathbf{I} \\ & + \frac{k^5 r^2}{15} \frac{\mathbf{r} \otimes \mathbf{r}}{r^2} + O(r^4). \end{aligned} \quad (66b)$$

Using the expansion (66), we can define G_{nn} as the *regularized limit*

$$G_{nn} = \lim_{a \rightarrow 0} \frac{3}{4\pi a^3} \int_{|\mathbf{r}-\mathbf{r}_n| < a} G(\mathbf{r}, \mathbf{r}_n) d^3r = i \frac{2k^3}{3} \mathbf{I} . \quad (67)$$

This result is applicable to free space and is typically used in the Second Approach to CDA. Note that a more rigorous derivation of (67) has been given in [34].