Dissertation

Design and Fundamentals of Optical Nanoantennas for High Intensity Enhancement or High Directivity

von

M.Sc. Andre Hildebrandt

Fakultät für Elektrotechnik, Informatik und Mathematik
Theoretische Elektrotechnik

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.).
Dem Department Physik der Fakultät für Naturwissenschaften der Universität Paderborn vorgelegt von Andre Hildebrandt.

Promotionskommission:

Vorsitzender:	Prof. Dr. Cedrik Meier
Erstgutachter:	Prof. Dr. Jens Förstner
Zweitgutachter:	Prof. Dr. Torsten Meier

Vorgelegt am 30. November 2016

Zusammenfassung

In dieser Arbeit geht es um das Design und die Eigenschaften verschiedener Antennen im optischen Frequenzbereich. Relevante Größen wie z.B. die Intensitätsverstärkung oder die Richtcharakteristik werden dabei durch verschiedene semi-analytische und numerische Methoden simuliert und mittels automatischer Optimieralgorithmen optimiert, was u.U. zu komplett neuartigen oder nicht intuitiven Strukturen führen kann. Da Ableitungen der benötigten Gleichungen nur für die einfachsten Systeme analytisch berechnet werden können, kommt es in dieser Arbeit hauptsächlich zur Anwendung von nicht-gradientenbasierten Algorithmen, was zu einer beträchtlichen Erhöhung der benötigten Iterationen und damit zu extremer Rechenleistung führt. Daher ist ein wesentlicher Bestandteil einer solchen Optimierung, die Rechenleistung jeder Iteration zu minimieren, z.B. durch geeignete Wahl von Randbedingungen oder Inkaufnahme systematischer Abweichungen, die nach der Optimierung bereinigt werden können.

Im wesentlichen gibt es zwei verschiedene Antennentypen: Empfangende Antennen, die die Energie von ebenen Wellen auf einen sehr kleinen Bereich fokussieren, und sendende Antennen, die aus einer Dipolquelle eine gewünschte Abstrahlung generieren, meistens möglichst in eine Richtung.

Zuerst werden bereits existierende, empfangende Antennen, die sich durch eine hohe Feld- oder Intensitätsverstärkung in einem sehr kleinen Bereich auszeichnen, als Start einer Optimierung betrachtet. Es ist bereits bekannt, dass solche Antennen besser funktionieren, wenn sie nicht isoliert, sondern in einem periodischen Gitter angeordnet sind. Hier wird jedoch gezeigt, dass manche Geometrien in solchen Gittern darüber hinaus Resonanzen besitzen, die auf kollektive Gittermoden zurückgeführt werden können. Dadurch wird nicht nur die maximale Intensitätsverstärkung drastisch erhöht, sondern diese Moden sind auch langlebiger, sodass mehr Energie aus der ebenen Welle verwendet werden kann. Danach wird eine freie Optimierung diskutiert, mit dem Ergebnis einer neuen Struktur, deren Leistung nochmals deutlich besser ist, als vergleichbare existierende Antennen.

Anschließend werden sendende Antennen behandelt, die sich durch eine hohe

Richtcharakteristik auszeichnen sollen. Diese basieren auf Interferenzeffekten von Dipolen (oder höherer Pole) und sollen optimiert werden. Die aus dem Rundfunk-Bereich bekannten Yagi-Uda Antennen besitzen vor allem zwei Probleme bei optischen Frequenzen: Erstens besitzen metallische Strukturen bei optischen Frequenzen hohe Verluste und zweitens strahlen Dipole, die auf einem Substrat sitzen, bevorzugt in ebendieses ein, wodurch die Anzahl an Dipolen (oder Antennenelementen), und damit die Leistung der Antenne, begrenzt ist. In dieser Arbeit wird aufgezeigt, dass beide Probleme durch geeignete, dielektrische Strukturen, die einem schwach führendem Wellenleiter entsprechen, gelöst werden können. Diese Antenne ist zwar größer, zeigt aber eine deutliche Verbesserung der Richtcharakteristik und ist robust bzgl. der Fabrikationstoleranzen, was experimentell nachgewiesen wurde. Daneben wird eine weitere dielektrische Antenne vorgestellt, die deutlich kleiner als die Wellenlänge des emittierten Lichtes ist, und mit der verschiedene gewünschte Abstrahlungsbilder erzielt werden können.

Abstract

This thesis focuses on the design and characteristics of different antennas in the optical frequency range. Relevant quantities such as intensity enhancement or directive gain are calculated by several semi-analytical and numerical methods and are optimized via automatic optimization algorithms, which might lead to new or unintuitive structures. Since the derivatives of the corresponding equations can be calculated analytically only for the most simple systems, in this work mainly non-gradient based algorithms are used, resulting in an increase of necessary iterations and thus also the computational effort. Therefore a main task is the minimization of the needed computation power for each iteration, e.g. by choice of boundary conditions or acceptance of systematic deviations, which can be resolved after the optimization.

In principle there are two different antenna types: Receiving antennas, which focus the energy of plane waves on a very small area, and sending antennas, which transform the nearly isotropic emission of a dipole source into a desired radiation pattern, mainly a beam into the desired direction.

First, already existing, receiving antennas, which are characterized by a high field or intensity enhancement in a very small area, are considered as the start of an optimization. It is already known that the performance of such antennas is increased, if they are not isolated, but are arranged in a periodic array. However, it is shown here that some geometries in such arrays also have resonances that can be traced back to collective array modes. As a result, not only the maximum intensity enhancement is drastically increased, but these modes also exhibit a higher lifetime, so that more energy from the plane wave can be used. After that a free optimization is discussed, resulting in a new structure, whose performance is significantly increased compared to already existing antennas.

Subsequently, transmitting antennas, which are characterized by a high directive gain, are discussed. They are based on interferences of dipoles (or higher poles) and are optimized in this work. The Yagi-Uda antennas, which are known from the radio

frequency regime, have two main problems in the optical frequency regime: Firstly, metallic structures exhibit high losses at optical frequencies, and secondly, dipoles sitting on a substrate radiate preferentially into the substrate, whereby the number of dipoles (or antenna elements), and thus the performance of the antenna, is limited. In this thesis it is shown that both problems can be solved by means of suitable, dielectric structures that correspond to a weakly guiding waveguide. Although this antenna is larger, it shows a significant improvement in the directive gain and is robust with respect to fabrication tolerances, which has been demonstrated experimentally. In addition, a dielectric antenna is proposed, which is significantly smaller than the wavelength of the emitted light, and with which various desired radiation patterns can be obtained.

Contents

1. Introduction

Antennas basically have two purposes: Either they receive electromagnetic waves from a remote source in order to extract contained information or to exploit the transmitted power, or they create a well-defined radiation pattern from a point source, enabling the electromagnetic waves to travel over large distances in free space. Today they are mainly used as an interface between locally processed information, consisting of electrical signals, and the free-space, wireless transmission of information, encoded in several degrees of freedom of electromagnetic waves, such as frequency, phase, or amplitude. There are ample approaches to design antennas in the radio frequency (RF) regime, which are understood well and found their way in almost every part of the current technology (e.g. radios, mobile phones, WLAN, Headphones, wireless bus systems [1], or for power transfer in modern TVs [2]).

There are several (semi-) analytic concepts for their design, mostly using the well-reasoned assumption, that the currents in such antennas appear only on its surfaces. That this is the case not only for ideal metals, known as perfect electric conductor (PEC), but also for real metals like copper, can be understood by comparing the skin depth - the penetration depth of the electric field into the metal - of real metals to the dimensions of the structure. The skin depth of an electromagnetic wave with a frequency of $f = 2.4GHz$ (one of the three WLAN frequency bands) is $\delta_{CU} = 1.3\mu m$. The vacuum wavelength of such a wave is then $\lambda = 12.5cm$. For antennas to work properly, their dimensions have to be (roughly) in the same order of magnitude as the wavelength. With the skin depth being smaller by a factor of around 100,000, one can safely assume, that the currents are surface currents only, leading to much simpler calculations of the properties of antennas.

Over the past decade the progress in structuring techniques enabled the fabrication of high-quality metallic nano structures with nearly arbitrary geometrical shapes [3]. Since many of the (surface-plasmon based) properties can be controlled via the geometrical shape and the dimensions of the structure, this field has gained a tremendous interest over the last years. Particularly the transfer of the knowledge of antennas

in RF regime to optical frequencies have been of interest, known as optical or nano antennas, leading to plenty of new applications [4], e.g. enhancing [5] and directing [6] the emission of single molecules. It can also be used for small volume spectroscopy [7], nonlinear spectroscopy [8], for holographic multiplexing [9], coherent control of surface plasmon polaritons (SPPs) [10], or coupling of free-space waves to surface plasmon waveguides [11].

However, since the skin depth becomes comparable to the dimensions of optical antennas, one cannot assume to have only surface currents any more, making the simulations more complicated, but also offering new properties of such. Since the optical field usually penetrates through the whole structure, they affect the free electrons within the metal simultaneously, leading to so called plasmonic resonances, which are very sensitive to the geometrical shape and the dimensions of the antenna. The usage of such a surface plasmon polariton (SPP) further leads to smaller dimensions, since their wavelength is much smaller than the wavelength of light with the same frequency in vacuum.

Using appropriate geometries, those structures can be used to focus optical fields to very small areas, acting as a receiving antenna. Using conventional lenses one can theoretically focus a Gaussian beam of light to a minimum radius of half of its wavelength. For light of the wavelength of $\lambda = 800nm$ the beam waist would be then $d = 400nm$, where nanoantennas can focus the light to areas with radius of a few tens of nanometers. When dealing with nanoantennas one usually measures the electric field or intensity enhancement in a defined point or area instead of the area of focus. Particularly for nonlinear processes in small structures like quantum dots or molecules such antennas provide a huge advantage, because for such experiments extremely high field strengths are required, which could, if distributed uniformly over a large area, evaporate the nano structures or the substrate. Typical antennas enhance the intensity by a factor of several hundred, so that, if the antenna and the nano structure are positioned appropriately, the intensity of the exciting optical fields can been reduced correspondingly, saving energy, increasing the life time of the investigated structures, and decreasing the experimental complexity, though increasing the effort in fabricating those structures significantly.

Many different geometries have been investigated so far, e.g. bow-ties [12], a dimer consisting of spheres [13], nanorods [14], or dumbbell-shaped nanorods [15], just to name a few. All have in common, that the designer had a structure in mind, where only existing parameters have been roughly optimized, mainly by trial-and-error. Models

containing design rules are only known for the most simple antennas [16]. In this work an automatic optimization process is used firstly to optimize already existing structures like the bow-tie and secondly to perform a free-form optimization. By the usage of such automatic optimization algorithms one often ends up with unintuitive structures, which are not limited by the creativeness of a designer, but also might give more insight into the mechanisms in such antennas. Therefore such automatic optimization routines can also be used as an analysis tool. Unfortunately they usually need extremely high computation power to get results and a convergence to a reasonable structure requires many constrains and therefore the optimization process becomes complicated. Also with increasing degrees of freedom one gets theoretically an increasing number of concurrent solutions, but ends up with only one, which might not be the best.

So far only receiving antennas, which couple to the far field and change the near field patterns, have been discussed. With the technology available nowadays, semiconductor structures can be fabricated with dimensions much smaller than the wavelength of waves in the optical frequency regime. A prominent representative of such a semiconductor heterostructure is the quantum dot (QD). They are already widely used, e.g. in QD lasers, diodes, or single electron transistors. Additionally, they are promising candidates for quantum information processing, e.g. for manipulation and storage of quantum information [17]. Since the charge carriers are confined in all three dimensions, the energy levels are discrete in all dimensions like in atomic systems. Thus they are often called artificial atoms. Optically excited they become a perfect point (dipole) source, since they are stable and some parameters can be controlled well (e.g. dimensions, energy levels), though the positioning of such small emitters is still a challenge.

Such quantum dots can be used as stable and efficient single photon sources [18], but they radiate like a point dipole, i.e. in all directions, where many of the generated photons cannot be used properly. However, this can be overcome by structuring the dielectric environment, e.g. in Ref. [19] little trumpets have been used to couple the single photons to a Gaussian beam. Also the usage of antennas may generate useful radiation patterns.

A QD (as well as nitrogen-vacancies in diamond) can also be used to store a quantum bit (qubit), the quantum mechanic analogue of a classical bit. To be able to optically address each single qubit, they have to be spatially separated, but entangling operations between qubits can be achieved by exploiting single-photon interference effects [20]. For that high performance nanoantennas would be extremely useful, since they are able to collimate the emission into a narrow beam. Such antennas can be divided into two

groups: plasmonic and dielectric antennas.

Plasmonic antennas usually consist of metals or other conducting materials (e.g. graphene) and exhibit large scattering cross sections, which means, that they have a strong effect on their environment. In principle one could transfer much of the knowledge of directive antennas from studies made in radio-frequency, e.g. classical Yagi-Uda antennas for TVs, but there are also many differences like plasmonic resonances or the parasitic impedance of the substrate, on which the antennas are positioned. In Ref. [21] a 5-Element Yagi-Uda like antenna has been investigated using cathodoluminescence spectroscopy. However, all plasmonic antennas show high intrinsic losses for optical frequencies, resulting in low efficiencies, rendering them almost useless for quantum computation, though they are still useful for spectroscopy.

Dielectric antennas consist of dielectric, i.e. non-conducting, materials, where either the dielectric environment is manipulated (e.g. using transformation optics [22]) or the interference of dipole (or higher order pole) resonances are utilized. Using appropriate materials one can design antennas almost without material losses, giving them a huge advantage over plasmonic based antennas, though often with the drawback, that such dielectric antennas become significantly larger than their plasmonic counterparts. In this work several improvements of directive nanoantennas are shown for the near-infrared (NIR) regime, starting with the optimization of the classical Yagi-Uda antenna and then turning to other materials and geometries.

1.1. Structure of this work

The structure of this thesis is divided into 5 chapters. The simulation of electromagnetic waves is shown in chapter 2, where the main properties of the finite integration technique (FIT) together with the dielectric function of gold are discussed. Although part of the work is obtained by the usage of the finite difference time domain (FDTD) method, it is not described here, because it is assumed, that this method is already known well or can be read upon, e.g. in the book in Ref. [23]. Additionally, the coupled dipole approximation (CDA) is shortly discussed, where only dipole fields are calculated, which is useful to identify the mechanisms of Yagi-Uda antennas.

In chapter 3 a very brief overview of the optimization theory in general and some algorithms is given. Though not developed in the scope of this work, it is important to understand how they work, because the results depend strongly on how an automatic

optimization algorithm is applied.

In chapter 4 optical antennas, which focus light on a small area, are discussed. As mentioned above the main quantity is the intensity enhancement in a defined point or area close to the antenna, which is dependent on many parameters like the material, geometrical shape, or the arrangement in an array of such antennas. A nearly free optimization is shown as well as fabrication related constraints.

In chapter 5 an overview of directive antennas is given, starting with the classical antenna theory, developed for the RF regime. Then the plasmonic based antennas are optimized for optical frequencies, switching to dielectric materials. At the end this chapter the design of an extremely small antenna, made of silicon, is discussed, allowing several useful radiation patterns by changing only the size of this antenna.

The last chapter 7 concludes this work and gives a short outlook on possible future works.

In the appendix obtained parameters are listed in section A. In section B the equations for the CDA, calculated for 5 particles, are given.

2. Simulation of electromagnetic waves

In this work nano structures, which modify light, are investigated. The frequencies lie in the optical and near-infrared regime and can be described via the macroscopic Maxwell's equations. These have to be solved numerically to calculate the desired properties of the nano structures. In this work the light and material are assumed to behave classically, which means neither electromagnetic fields nor the energy states in matter are quantized. This is a reasonable assumption, since the nano structures are large compared to e.g. a typical quantum dot and the density of photons is sufficiently high.

In the following chapter first the well known Maxwell's equations are summarized. Since analytical solutions are only available for simple structures, all results are achieved by numerical calculations. So one of the not so well known methods, the *finite integration technique (FIT)*, which has mainly been used in this work, is introduced in detail in this chapter as well. Apart from this method many results have been achieved by an in-house code of a standard finite difference time domain (FDTD) algorithm, which is mainly based on Ref. [23], the standard book for this method, so the formal description has been omitted.

Since the FIT and FDTD solve the full Maxwell's equations, it is useful to investigate some structures with a simplified model to get a simple picture of a given system. So a short introduction to the coupled dipole approximation (CDA) has been made as well, mostly to identify, if the investigated antenna can be described by dipoles only and therefore get a deeper insight into the mechanisms of the nanoantennas.

2.1. Maxwell's equations and propagation of electromagnetic waves

The propagation of electromagnetic waves can be described by the macroscopic Maxwell's equations, a system of four partial differential equations. The integral formulation is given in equations (2.1a) - (2.1d).

$$\oint_{\partial A} \boldsymbol{E}(\boldsymbol{r},t) \cdot d\boldsymbol{s} = -\int_A \frac{\partial}{\partial t}\boldsymbol{B}(\boldsymbol{r},t) \cdot d\boldsymbol{A} \qquad \text{(Faraday's law)} \qquad (2.1a)$$

$$\oint_{\partial A} \boldsymbol{H}(\boldsymbol{r},t) \cdot d\boldsymbol{s} = \int_A \left(\frac{\partial}{\partial t}\boldsymbol{D}(\boldsymbol{r},t) + \boldsymbol{J}(\boldsymbol{r},t)\right) \cdot d\boldsymbol{A} \qquad \text{(Ampère's law)} \qquad (2.1b)$$

$$\oint_A \boldsymbol{D}(\boldsymbol{r},t) \cdot d\boldsymbol{A} = \int_V \rho(\boldsymbol{r},t) dV \qquad \text{(Gauss's law)} \qquad (2.1c)$$

$$\oint_A \boldsymbol{B}(\boldsymbol{r},t) \cdot d\boldsymbol{A} = 0 \qquad (2.1d)$$

Using Gauss's and Stokes' law, one can derive the equivalent, differential formulation of Maxwell's equations (2.2).

$$\nabla \times \boldsymbol{E}(\boldsymbol{r},t) = -\frac{\partial}{\partial t}\boldsymbol{B}(\boldsymbol{r},t) \qquad (2.2a)$$

$$\nabla \times \boldsymbol{H}(\boldsymbol{r},t) = \frac{\partial}{\partial t}\boldsymbol{D}(\boldsymbol{r},t) + \boldsymbol{J}(\boldsymbol{r},t) \qquad (2.2b)$$

$$\nabla \cdot \boldsymbol{D}(\boldsymbol{r},t) = \rho(\boldsymbol{r},t) \qquad (2.2c)$$

$$\nabla \cdot \boldsymbol{B}(\boldsymbol{r},t) = 0 \qquad (2.2d)$$

In order to solve them, one needs the constitutive material relations in equations (2.4), which are given in a general form in time domain.

$$\boldsymbol{D}(\boldsymbol{r},t) = \varepsilon_0 \boldsymbol{E}(\boldsymbol{r},t) + \boldsymbol{P}(\boldsymbol{r},t) \qquad (2.3a)$$

$$\boldsymbol{B}(\boldsymbol{r},t) = \mu_0 \left(\boldsymbol{H}(\boldsymbol{r},t) + \boldsymbol{M}(\boldsymbol{r},t)\right) \qquad (2.3b)$$

$$\boldsymbol{J}(\boldsymbol{r},t) = \sigma(\boldsymbol{r},t)\boldsymbol{E}(\boldsymbol{r},t) + \boldsymbol{J}_{inc}(\boldsymbol{r},t) \qquad (2.3c)$$

Using the Fourier transform, or more accurately the differentiation theorem, one obtains Maxwell's equations in frequency domain. For homogeneous, linear and isotropic

media one can simplify the material relations in frequency domain to:

$$\boldsymbol{D}(\boldsymbol{r},\omega) = \varepsilon_0\varepsilon_r(\omega)\boldsymbol{E}(\boldsymbol{r},\omega) \tag{2.4a}$$
$$\boldsymbol{B}(\boldsymbol{r},\omega) = \mu_0\mu_r(\omega)\boldsymbol{H}(\boldsymbol{r},\omega) \tag{2.4b}$$
$$\boldsymbol{J}(\boldsymbol{r},\omega) = \sigma(\boldsymbol{r},\omega)\boldsymbol{E}(\boldsymbol{r},\omega) + \boldsymbol{J}_{inc}(\boldsymbol{r},\omega) \tag{2.4c}$$

If the relative permittivity ε_r and permeability μ_r are independent of the frequency ω, which can only be an approximation in a limited frequency range according to Kramers-Kronig relations, the first two material relations can be transformed back to time domain.

$$\boldsymbol{D}(\boldsymbol{r},t) = \varepsilon_0\varepsilon_r\boldsymbol{E}(\boldsymbol{r},t) \tag{2.5a}$$
$$\boldsymbol{B}(\boldsymbol{r},t) = \mu_0\mu_r\boldsymbol{H}(\boldsymbol{r},t) \tag{2.5b}$$
$$\tag{2.5c}$$

Taking the curl of the equations (2.2a) and (2.2b), leads to the so called wave equation for electromagnetic waves:

$$\nabla \times \nabla \times \boldsymbol{E}(\boldsymbol{r},t) = -\mu_0\frac{\partial^2}{\partial t^2}\boldsymbol{D}(\boldsymbol{r},t) \tag{2.6}$$

Transforming this equation into the Fourier domains, one obtains the generic dispersion relation for transversal waves ($\boldsymbol{k}\cdot\boldsymbol{E} = 0$)

$$k = \sqrt{\varepsilon_r(\omega,\boldsymbol{k})}\frac{\omega}{c_0}, \tag{2.7}$$

where $k = |\boldsymbol{k}|$ is the wave number and ω the angular frequency of the electromagnetic wave and $c_0^2 = \frac{1}{\varepsilon_0\mu_0}$ is the speed of light in free space.

One important property of Maxwell's equations is the implicitly stated *continuity equation*. It is derived by taking the divergence of eq. (2.2b) and using eq. (2.2c) to simplify the equation

$$\nabla\cdot\boldsymbol{J}(\boldsymbol{r},t) = -\frac{\partial}{\partial t}\rho(\boldsymbol{r},t), \tag{2.8}$$

which means, that a current changes the charge carrier density and vice versa.

2.2. The *Finite-Integration-Technique* (FIT)

The dynamics of electric and magnetic fields can be described via Maxwell's equations, together with the constitutive material equations and appropriate boundary conditions. They build a system of coupled partial differential equations, which can be solved analytically only for very simple problems. Since optical devices are usually of more complex structure, one has to solve these equations numerically. Within the last century several methods have been developed to do so, e.g. the finite element method (FEM), the FDTD or boundary element method (BEM).

During the 1970th the fundamentals of the FIT were proposed [24], but still much research is done to enhance accuracy and simulation time or to develop new features. As a result, the commercial software package CST Microwave Studio [25], which is based on the FIT, has been very successful over the last years, and is also used in this work. In the following chapter only the fundamental concept is summarized, with the notation mainly used from [26].

2.2.1. Discretization of the computational domain

As known from other numerical models, the computational domain $\Omega \in \mathbb{R}^3$ has to be decomposed into a finite number of simplicial cells V_i. Note, that the FIT is not restricted to Cartesian meshes, but allows unstructured meshes as well [27, 28]. In frequency domain an unstructured grid (consisting of e.g. tetrahedral cells) is often used, but in this work mainly a structured grid consisting of hexahedral cells has been used, since it is advantageous for time domain calculations. Similar to the Yee-Cube in the FDTD method [23] two orthogonal grids are used, the primary (here referred to as G) and the dual grid (here referred to as $\tilde{G}$), as sketched in Fig. 2.1.

The points $\boldsymbol{P}$ on the primary grid G can be addressed in an arbitrary coordinate system with the coordinates u, v and w. In eq. (2.9) the sampling of the grid is given by the maps of $u(i)$, $v(j)$, and $w(k)$.

$$\boldsymbol{P}(i,j,k) = u(i)\boldsymbol{e}_u + v(j)\boldsymbol{e}_v + w(k)\boldsymbol{e}_w \tag{2.9}$$

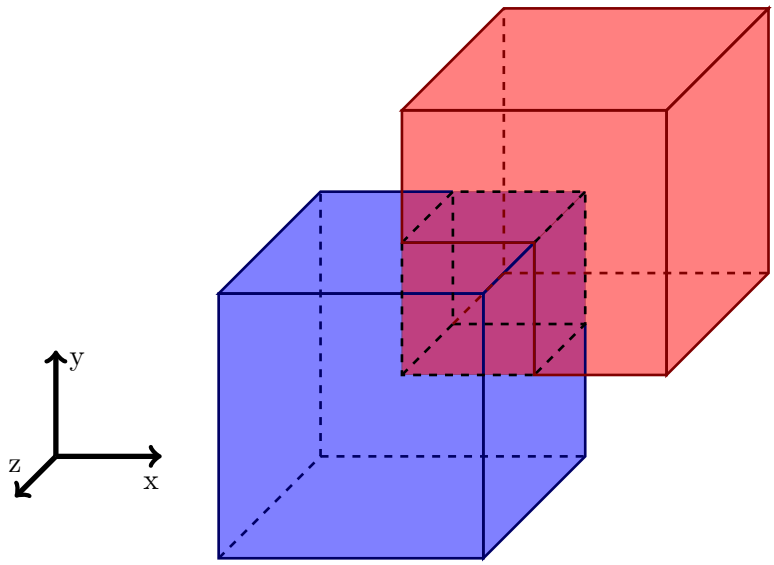


Fig. 2.1.: *The spatial allocation of a cell and a dual cell of the grid doublet* $\{G, \tilde{G}\}$ *is shown. The primary grid is colored red, the dual grid blue.*

For the implementation of the FIT it has been advantageous to replace the indexing (i,j,k) by one index n, defined in eq. (2.10).

$$n = i + (j-1)\underbrace{n_u}_{M_v} + (k-1)\underbrace{n_u n_v}_{M_w} \tag{2.10}$$

For a better readability an arbitrary coordinate is referred to as ν, which is ether u, v, or w. The corresponding two other coordinates are then ξ and ζ. With this definition the edges s_ν, facets A_ν, and volumes V within the computational domain can be expressed in a parametric representation depending on the index n, given in eqs. (2.11) to (2.13).

$$s_\nu(n) = \left\{ \boldsymbol{P}(n) + a_\nu h_\nu \Delta\nu_n \boldsymbol{e}_\nu \qquad |a_\nu \quad \in [0,1] \right\} \tag{2.11}$$

$$A_\nu(n) = \left\{ \boldsymbol{P}(n) + \sum_{\lambda=\xi,\zeta} a_\lambda h_\lambda \Delta\lambda_n \boldsymbol{e}_\lambda \qquad |a_{\xi,\zeta} \ \in [0,1] \right\} \tag{2.12}$$

$$V(n) = \left\{ \boldsymbol{P}(n) + \sum_{\lambda=\nu,\xi,\zeta} a_\lambda h_\lambda \Delta\lambda_n \boldsymbol{e}_\lambda \quad |a_{\nu,\xi,\zeta} \in [0,1] \right\} \tag{2.13}$$

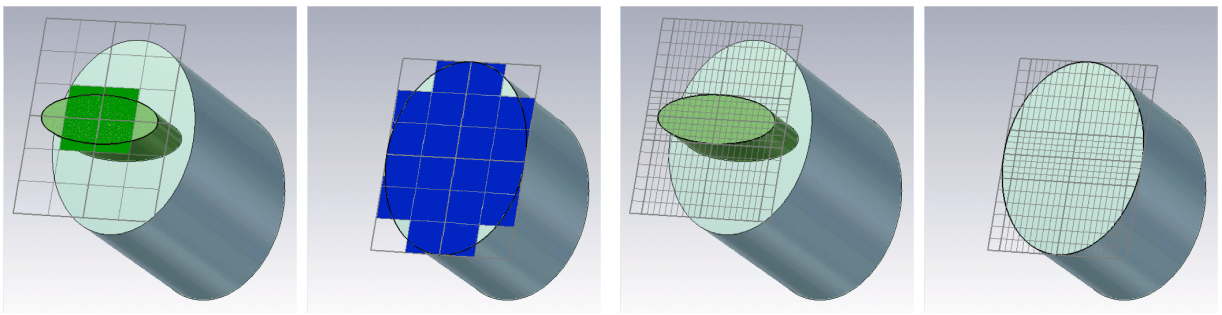
Please note, that the quantities s_v and A_v are often used for both, the grid edges (or facets) and the corresponding lengths (or areas), which usually is clear from the context. The variables h_i are the metrical factors of the corresponding coordinate system,

whereas $\Delta\lambda$ describes the step size of each coordinate axis. With this definition one obtains some redundant grid points, which lie outside of the computational domain. However, to maintain the symmetry of the resulting matrix these points are still allocated, but set to zero. Using this indexing one obtains $n_p = n_u n_v n_w$ grid points and volumes and $3n_p$ grid edges and facets.

In order to solve Maxwell's equations, each cell is filled homogeneously with one material, which means a spatially constant relative permittivity (ε_r) and permeability (μ_r), causing a geometrical discretization error. The mesh has to be generated not only to ensure a stable solution of Maxwell's equations, but also to ensure an accurate representation of the simulated structure. This is illustrated in Fig. 2.2 (a). This example consists of two elliptical cylinders of the same material, a large one (light blue) and a smaller one (green). Two cross sections are shown, one of each cylinder. The resolution of the mesh (gray) is a uniform mesh, which means that all cells have the same size. The resolution of this example is high enough to get a stable solution of Maxwell's equations, but to low to resolve the geometry. For the most simple case, that a cell is filled with the material, which covers more than half of the cell, one would get a square of four cells for the small cylinder.

In order to improve the resolution, one can either decrease the size of the cells uniformly or refine the mesh locally. The latter case is often preferred, when one has to deal with structures of different length scale such as a nano structure on substrate, surrounded by vacuum. Advantage compared to the first case is a reduction of memory and computation time. Disadvantage is an increased numerical dispersion, which might lead to some artifacts especially if the computational domain is large in comparison with the incident wavelength of the light. This is illustrated in Fig. 2.2 (b) for the two cylinders again. There one can see, that in a hexahedral grid the local mesh refinement extends in normal direction to the shown cut plane, which reduces the reduction of computation time.

For all four Maxwell's equations to be discretized over space an additional grid is required, the so called dual grid $\tilde{G}$. It is defined by taking the foci of the cells of the primary Grid G as its grid points (c.p. Fig. 2.1). For a uniform hexahedral grid, the points of $\tilde{\boldsymbol{P}}(n)$ of the dual grid $\tilde{G}$ are displaced by the half of the step width in direction of each coordinate. As a consequence facets $\tilde{A}_\nu(n)$ are each pierced through by the edges $s_\nu(n)$ of the primary grid, which are perpendicular to the facets. In contrast to the primary grid, all points $\tilde{\boldsymbol{P}}(n)$ of the dual grid lie in the computational domain.



(a) Uniform grid of two elliptical cylinders consiting of the same material. Two cross sections are shown, one in the small and one in the large cylinder.

(b) Non-uniform grid with local mesh refinement of two elliptical cylinders consiting of the same material. Two cross sections are shown, one in the small and one in the large cylinder.

Fig. 2.2.: *Illustration of a rectangular mesh for two cylinders of the same material. The gray lines show the primary grid only. In a) a cross section of each cylinder with a uniform grid is shown. For the case, that the cylinder covers more then half of a cell, the cell is filled with the material of the cylinder (indicated by dark green and blue). Though the size of the cells is small enough for a stable solution of Maxwell's equations, particularly the small cylinder is represented poorly, for it would look like a square. In b) the same geometry is shown, albeit with a non-uniform grid and a local mesh refinement. In order to better resolve the small cylinder, the cells inside of it are much smaller, than around.*

2.2.2. Discretization of Maxwell's equations

In contrast to the Finite Difference method, the integral Maxwell's equations are solved in the Finite Integration method. For that it is convenient not to evaluate the discretized electric and magnetic fields, but to use grid voltages and and fluxes instead, defined in equations (2.14).

$$\widehat{\mathrm{e}}_\nu(n) = \int_{s_\nu(n)} \boldsymbol{E}(\boldsymbol{r},t)\cdot d\boldsymbol{s} \qquad \text{(electric grid voltage)} \qquad (2.14a)$$

$$\widehat{\mathrm{h}}_\nu(n) = \int_{\tilde{s}_\nu(n)} \boldsymbol{H}(\boldsymbol{r},t)\cdot d\boldsymbol{s} \qquad \text{(magnetic grid voltage)} \qquad (2.14b)$$

$$\widehat{\widehat{\mathrm{d}}}_\nu(n) = \iint_{\tilde{A}_\nu(n)} \boldsymbol{D}(\boldsymbol{r},t)\cdot d\boldsymbol{A} \qquad \text{(electric surface flux)} \qquad (2.14c)$$

$$\widehat{\widehat{\mathrm{b}}}_\nu(n) = \iint_{A_\nu(n)} \boldsymbol{B}(\boldsymbol{r},t)\cdot d\boldsymbol{A} \qquad \text{(magnetic surface flux)} \qquad (2.14d)$$

$$\widehat{\widehat{\mathrm{j}}}_\nu(n) = \iint_{\tilde{A}_\nu(n)} \boldsymbol{J}(\boldsymbol{r},t)\cdot d\boldsymbol{A} \qquad \text{(grid current)} \qquad (2.14e)$$

$$q(n) = \iiint_{\tilde{V}(n)} \rho(\boldsymbol{r})dV \qquad \text{(grid charge)} \qquad (2.14f)$$

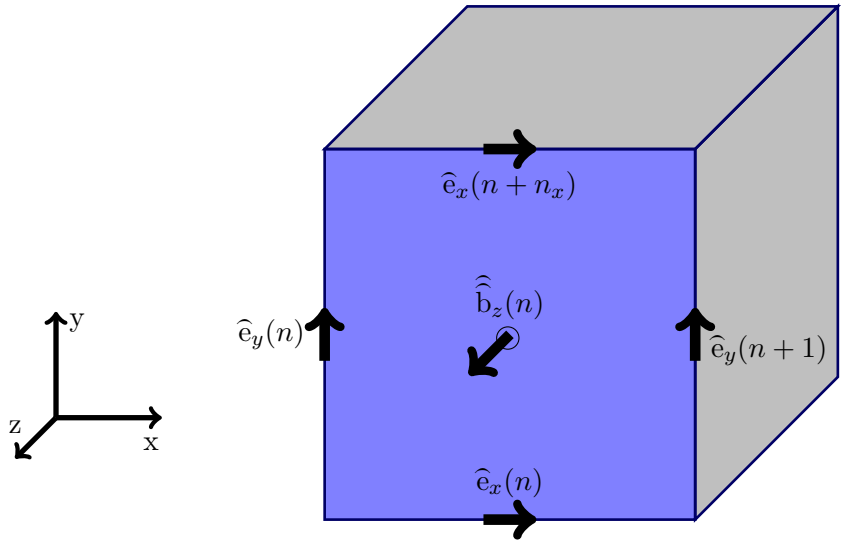


Fig. 2.3.: *A cell V_n of the primary grid G with the allocation of the grid voltages $\widehat{\mathrm{e}}$ on the edges of A_n and the magnetic facet flux $\widehat{\widehat{\mathrm{b}}}$ through the corresponding surface.*

Using these definitions and replacing the closed line integral by the sum of line integrals over the edges of the surface, one obtains the discretized representation of equations (2.1a) and (2.1b) for the volume V_n. Figure 2.3 illustrates the meaning of equation (2.15) using a standard Cartesian coordinate system. Please note, that the sign for the directions of the edges is in general arbitrary, but here only the common notation is shown.

$$\widehat{\mathrm{e}}_\xi(n) - \widehat{\mathrm{e}}_\xi(n+M_\zeta) + \widehat{\mathrm{e}}_\zeta(n+M_\xi) - \widehat{\mathrm{e}}_\zeta(n) = -\frac{d}{dt}\widehat{\widehat{\mathrm{b}}}_\nu(n) \tag{2.15}$$

$$\widehat{\mathrm{h}}_\xi(n) - \widehat{\mathrm{h}}_\xi(n-M_\zeta) + \widehat{\mathrm{h}}_\zeta(n-M_\xi) - \widehat{\mathrm{h}}_\zeta(n) = -\frac{d}{dt}\widehat{\widehat{\mathrm{d}}}_\nu(n) + \widehat{\widehat{\mathrm{j}}}_\nu(n) \tag{2.16}$$

In the same manner equations (2.1c) and (2.1d) can be discretized.

$$\widehat{\widehat{\mathrm{d}}}_u(n) - \widehat{\widehat{\mathrm{d}}}_u(n-1) + \widehat{\widehat{\mathrm{d}}}_v(n) - \widehat{\widehat{\mathrm{d}}}_v(n-M_v) + \widehat{\widehat{\mathrm{d}}}_w(n) - \widehat{\widehat{\mathrm{d}}}_w(n-M_w) = q(n) \tag{2.17}$$

$$\widehat{\widehat{\mathrm{b}}}_u(n) - \widehat{\widehat{\mathrm{b}}}_u(n+1) + \widehat{\widehat{\mathrm{b}}}_v(n) - \widehat{\widehat{\mathrm{b}}}_v(n+M_v) + \widehat{\widehat{\mathrm{b}}}_w(n) - \widehat{\widehat{\mathrm{b}}}_w(n+M_w) = 0 \tag{2.18}$$

For a compact notation the grid voltages ($\widehat{\mathrm{e}}(n)$, $\widehat{\mathrm{h}}(n)$) and fluxes ($\widehat{\widehat{\mathrm{d}}}(n)$, $\widehat{\widehat{\mathrm{b}}}(n)$, $\widehat{\widehat{\mathrm{j}}}(n)$) are condensed into vectors with $3n_p$ elements ($\widehat{\mathbf{e}}$, $\widehat{\mathbf{h}}$, $\widehat{\widehat{\mathbf{d}}}$, $\widehat{\widehat{\mathbf{b}}}$, $\widehat{\widehat{\mathbf{j}}}$). The corresponding grid charge $\boldsymbol{q}$ is then a vector of n_p elements. The vectors are exemplarily defined in eqs.

(2.19).

$$\widehat{\mathbf{e}}_\nu = (\widehat{\mathrm{e}}_\nu(1)\dots\widehat{\mathrm{e}}_\nu(n_p))^T \quad \widehat{\widehat{\mathbf{d}}}_\nu = \left(\widehat{\widehat{\mathrm{d}}}_\nu(1)\dots\widehat{\widehat{\mathrm{d}}}_\nu(n_p)\right)^T \quad \boldsymbol{q} = (q(1)\dots q(n_p))^T \tag{2.19a}$$

$$\widehat{\mathbf{e}} = \begin{pmatrix} \widehat{\mathrm{e}}_\nu \\ \widehat{\mathrm{e}}_\xi \\ \widehat{\mathrm{e}}_\zeta \end{pmatrix} \qquad \widehat{\widehat{\mathbf{d}}} = \begin{pmatrix} \widehat{\widehat{\mathrm{d}}}_\nu \\ \widehat{\widehat{\mathrm{d}}}_\xi \\ \widehat{\widehat{\mathrm{d}}}_\zeta \end{pmatrix} \tag{2.19b}$$

Additionally the matrices $\underline{\mathbf{P}}_\nu$ are defined as

$$P_\nu^{i,j} = -\delta_{i,j} + \delta_{i,j-M_\nu}, \tag{2.20}$$

where $M_v = 1$. The matrices $\underline{\mathbf{P}}_\nu$ can be interpreted as a discrete form of the partial derivative with respect to the coordinate ν. With this definition one can build the discrete curl operator $\underline{\mathbf{C}}$, as well as the divergence operator $\underline{\mathbf{S}}$, where the operators $\underline{\tilde{\mathbf{C}}}$ and $\underline{\tilde{\mathbf{S}}}$ are defined on the dual grid.

$$\underline{\mathbf{C}} = \begin{pmatrix} 0 & -\underline{\mathbf{P}}_w & \underline{\mathbf{P}}_v \\ \underline{\mathbf{P}}_w & 0 & -\underline{\mathbf{P}}_u \\ -\underline{\mathbf{P}}_v & \underline{\mathbf{P}}_u & 0 \end{pmatrix} \qquad \underline{\tilde{\mathbf{C}}} = \underline{\mathbf{C}}^T \tag{2.21}$$

$$\underline{\mathbf{S}} = \begin{pmatrix} \underline{\mathbf{P}}_u & \underline{\mathbf{P}}_v & \underline{\mathbf{P}}_w \end{pmatrix} \qquad \underline{\tilde{\mathbf{S}}} = \begin{pmatrix} -\underline{\mathbf{P}}_u & -\underline{\mathbf{P}}_v & -\underline{\mathbf{P}}_w \end{pmatrix} \tag{2.22}$$

In this compact representation the equations (2.15) to (2.18) are reduced to equations (2.23). Since this system of coupled equations is the discrete equivalent of Maxwell's equations, they often are called *Maxwell-Grid-Equations*.

$$\underline{\mathbf{C}}\widehat{\mathbf{e}} = -\frac{d}{dt}\widehat{\widehat{\mathbf{b}}} \tag{2.23a}$$

$$\underline{\tilde{\mathbf{C}}}\widehat{\mathbf{h}} = \frac{d}{dt}\widehat{\widehat{\mathbf{d}}} + \widehat{\widehat{\mathbf{j}}} \tag{2.23b}$$

$$\underline{\tilde{\mathbf{S}}}\widehat{\widehat{\mathbf{d}}} = \boldsymbol{q} \tag{2.23c}$$

$$\underline{\mathbf{S}}\widehat{\widehat{\mathbf{b}}} = \mathbf{0} \tag{2.23d}$$

An important feature of the Maxwell-Grid equations is the built-in continuity equation:

$$\tilde{\underline{\mathbf{S}}}(\tilde{\underline{\mathbf{C}}}\widehat{\mathbf{h}}) = 0 \tag{2.24}$$

$$= \tilde{\underline{\mathbf{S}}}\left(\frac{d}{dt}\widehat{\widehat{\mathbf{d}}} + \widehat{\widehat{\mathbf{j}}}\right) \tag{2.25}$$

$$= \frac{d}{dt}\boldsymbol{q} + \tilde{\underline{\mathbf{S}}}\,\widehat{\widehat{\mathbf{j}}} = 0 \tag{2.26}$$

The discrete continuity equation ensures that no spurious charges will occur. Such non-physical charges would result in additional static fields, contaminating discrete transient field solutions [29].

Please note, that so far no approximation has been made, i.e. the Maxwell-Grid-Equations are an exact representation of the integral Maxwell's equations on a grid doublet. However, when introducing the discretization of the constitutive material equations, approximations are inevitable.

2.2.3. Discretization of material relations

As mentioned the approximation of the Finite Integration method enters, when the integral state variables, which are allocated on the two different cell complexes G and $\tilde{G}$, are to be related to each other by the constitutive material equations. Since in this work only a Cartesian grid is considered, the cell complexes G and $\tilde{G}$ are dual orthogonal and represent a so called *Delaunay-Voronoi-Grid* doublet [29]. So the directions associated to the facet and to the dual edge penetrating this facet are identical.

For linear, isotropic, and lossless materials the local relation between $\widehat{\mathrm{e}}_\nu$ and $\widehat{\widehat{\mathrm{d}}}_\nu$ can be calculated by eq. (2.29).

$$\widehat{\widehat{\mathrm{d}}}_\nu(n) = \varepsilon_0 \sum_{i=1}^{i_\nu} \varepsilon_r \left(n - M_\nu^{(i)}\right) \int_{\tilde{A}_\nu^{(i)}(n)} \boldsymbol{E}(\boldsymbol{r}, t) \cdot d\boldsymbol{A} \tag{2.27}$$

$$\approx \varepsilon_0 \frac{\widehat{\mathrm{e}}_\nu(n)}{s_\nu(n)} \sum_{i=1}^{i_\nu} \tilde{A}_\nu^{(i)}(n) \varepsilon_r \left(n - M_\nu^{(i)}\right) \tag{2.28}$$

$$= \frac{\tilde{A}_\nu(n)}{s_\nu(n)} \overline{\varepsilon}_\nu(n) \widehat{\mathrm{e}}_\nu(n) \tag{2.29}$$

The identifier $M_\nu^{(i)}$ serves to index the neighboring primary volumes, whose intersections with the dual facets, which are directed in ν-direction, define the facets $\tilde{A}_\nu^{(i)}$:

$$\tilde{A}_\nu^{(i)}(n) = \tilde{A}_\nu(n) \cap V(n - M_\nu^{(i)}) \tag{2.30}$$

In eq. (2.29) the local, averaged permittivity is then given by

$$\overline{\varepsilon}_\nu(n) = \varepsilon_0 \sum_{i=1}^{i_\nu} \frac{\tilde{A}_\nu^{(i)}(n)}{\tilde{A}_\nu(n)} \varepsilon_r \left(n - M_\nu^{(i)}\right). \tag{2.31}$$

As for the state variables, the averaged permittivities of the cells can be written as vectors.

$$\boldsymbol{\varepsilon}_\nu = \begin{pmatrix} \overline{\varepsilon}_\nu(1) \\ \vdots \\ \overline{\varepsilon}_\nu(n_p) \end{pmatrix} \qquad \overline{\boldsymbol{\varepsilon}} = \begin{pmatrix} \boldsymbol{\varepsilon}_u \\ \boldsymbol{\varepsilon}_v \\ \boldsymbol{\varepsilon}_w \end{pmatrix} \tag{2.32}$$

Similarly the averaged permeability can be derived, though it is convenient to use the inverse of the permeability.

$$\widehat{\mathrm{h}}_\nu(n) \approx \frac{\tilde{s}_\nu(n)}{A_\nu(n)} \overline{\mu_\nu^{-1}}(n) \widehat{\widehat{\mathrm{b}}}_\nu(n) \tag{2.33}$$

$$\text{with } \overline{\mu_\nu^{-1}}(n) = \mu_0^{-1} \sum_{j=1}^{j_\nu} \frac{\tilde{s}_\nu^{(j)}(n)}{\tilde{s}_\nu(n)} \mu_r^{-1} \left(n + M_\nu^{(j)}\right) \tag{2.34}$$

Defining the diagonal permittivity and permeability matrices

$$\underline{\mathbf{D}}_\varepsilon = \mathrm{Diag}\left\{\overline{\boldsymbol{\varepsilon}}\right\}, \qquad \underline{\mathbf{D}}_{\mu^{-1}} = \mathrm{Diag}\left\{\overline{\boldsymbol{\mu^{-1}}}\right\} \tag{2.35}$$

and matrices containing the lengths and areas of the grid cells,

$$\underline{\mathbf{D}}_s = \mathrm{Diag}\left\{\begin{pmatrix} \boldsymbol{s}_u \\ \boldsymbol{s}_v \\ \boldsymbol{s}_w \end{pmatrix}\right\}, \qquad \boldsymbol{s}_\nu = \begin{pmatrix} s_\nu(1) \\ \vdots \\ s_\nu(n_p) \end{pmatrix} \tag{2.36}$$

$$\underline{\mathbf{D}}_A = \mathrm{Diag}\left\{\begin{pmatrix} \boldsymbol{A}_u \\ \boldsymbol{A}_v \\ \boldsymbol{A}_w \end{pmatrix}\right\}, \qquad \boldsymbol{A}_\nu = \begin{pmatrix} A_\nu(1) \\ \vdots \\ A_\nu(n_p) \end{pmatrix} \tag{2.37}$$

one obtains a compact form of the discretized material relations in eq. (2.38).

$$\widehat{\tilde{\mathbf{d}}} = \underline{\tilde{\mathbf{D}}}_A \underline{\mathbf{D}}_\varepsilon \underline{\mathbf{D}}_s^{-1} \widehat{\mathbf{e}} \qquad = \underline{\mathbf{M}}_\varepsilon \widehat{\mathbf{e}} \tag{2.38a}$$

$$\widehat{\mathbf{h}} = \underline{\tilde{\mathbf{D}}}_s \underline{\mathbf{D}}_{\mu^{-1}} \underline{\mathbf{D}}_A^{-1} \widehat{\tilde{\mathbf{b}}} \qquad = \underline{\mathbf{M}}_{\mu^{-1}} \widehat{\tilde{\mathbf{b}}} \tag{2.38b}$$

One consequence of using a hexahedral grid is, that the material matrices are in diagonal form, which significantly reduces the computational effort for solving the system of equations.

2.2.4. Boundary conditions

Solving Maxwell's equations in finite domain requires some conditions for the boundaries, in order to simulate realistic systems accurately. In this chapter the most prominent conditions are briefly discussed, for they are well known in literature and their principles work for most numerical methods in a similar way.

Electric Boundary Conditions (PEC-BC)

The most simple boundary condition is the electric one, where the tangential components of the electric grid voltage at the boundaries are set to zero, i.e. $\boldsymbol{n} \times \widehat{\mathrm{e}} = 0$, with the normal vector $\boldsymbol{n}$ of the corresponding boundary. Consequence of this is, that the normal component of the magnetic grid flux has to be zero as well, i.e. $\boldsymbol{n} \cdot \widehat{\tilde{\mathrm{b}}} = 0$. This can be easily implemented in the FIT by setting the corresponding elements of the operator $\underline{\mathbf{C}}$ to zero. The electric boundary conditions are equivalent with an enclosure of the computational domain with a perfect electric conductor (PEC), thus the often used name PEC Boundary Condition.

Periodic Boundary Conditions (PBC)

Periodic boundary conditions are used to model an infinite array of identical structures. In the general case, the tangential components of the grid voltages of two edges, which lie on the opposite sides of the computational domain, have a fixed phase relation, i.e. $\widehat{\mathrm{e}}_{\parallel,1} = \widehat{\mathrm{e}}_{\parallel,2} e^{i\phi}$, which has already been implemented in 1986 [30]. Therefor the number of degrees of freedom is reduced, since the state variables at the boundary of one side

depends on the state variables on the other side. Thus such boundary conditions are called *Master-Slave-Boundary-Conditions* for the finite element method. It is clear then, that the mesh at opposite sides of the computational domain has to be identical.

Open Boundary Conditions

Open boundary conditions are used to simulate an isolated structure. In order to do that, they have to fulfill many requirements. Propagating and evanescent electromagnetic waves have to be able to penetrate through these boundaries with almost no reflection. This has to be the case for almost any angle of incidence and for wide range of frequencies simultaneously. Since the proposal in [31] the *perfectly matched layer* has been the most often used boundary condition. With that the computational domain is extended by a few layers of an impedance matched conducting material, which has no physical relevance, but absorbs an electromagnetic wave almost without reflection. A more detailed description of the PML in FIT can be found in [32, 33, 34].

2.2.5. Time integration of Maxwell's equations

While several time integration algorithms were developed, the *Leapfrog* algorithm has been the method of choice for solving Maxwell's equations since the publication of Yee in 1966 [35]. This has been developed for the Finite Difference method (FDTD), but is easily implemented within the FIT.

Main advantage of the Leapfrog algorithm is, that it is time invariant and, as consequence, the law of energy conservation is fulfilled. Also it is an explicit algorithm, i.e. for calculating the state variables of the next time step mainly matrix products are necessary, which saves much computational effort compared to implicit methods. Due to the sparsely populated material and operator matrices, the computational effort scales linearly with the degrees of freedom.

The two equations, which have to be integrated over time, can be written as in equations (2.39) and (2.40).

$$\frac{d}{dt}\widehat{\mathbf{h}}(t) = -\underline{\mathbf{M}}_{\mu^{-1}}\underline{\mathbf{C}}\widehat{\mathbf{e}}(t) \tag{2.39}$$

$$\frac{d}{dt}\widehat{\mathbf{e}}(t) = \underline{\mathbf{M}}_{\varepsilon}^{-1}\left(\underline{\tilde{\mathbf{C}}}\widehat{\mathbf{h}}(t) - \widehat{\tilde{\mathbf{j}}}(t)\right) \tag{2.40}$$

Since the surfaces and volumes of the grid cells remain constant over the simulation, one can take the time derivative out of the integrals, with which one obtains an ordinary derivative instead of a partial one with respect to time. This makes the integration more stable, since one has to care less about the characteristics, which play an important role for integrating a system of partial differential equations[1]. As a consequence one can use a constant step size for all time steps.

The Leapfrog algorithm employs a central difference scheme, i.e.

$$\frac{d}{dt} f^{(m+1/2)} = \frac{f^{(m+1)} - f^{(m)}}{\Delta t} + \mathcal{O}(\Delta t^2), \tag{2.41}$$

with the time step size Δt. Therefor it is convenient to describe the state variables $\widehat{\mathbf{h}}$, $\widehat{\widehat{\mathbf{b}}}$ and $\widehat{\widehat{\mathbf{j}}}$ on a time grid with the points $t_i = i\Delta t$. The other variables $\widehat{\mathbf{e}}$ and $\widehat{\widehat{\mathbf{d}}}$ are defined on the points in between, i.e. $t_i = (i+1/2)\Delta t$. With that, the time derivatives can be replaced by this central difference scheme:

$$\frac{d}{dt}\widehat{\mathbf{h}}^{(i+1/2)} \approx \frac{\widehat{\mathbf{h}}^{(i+1)} - \widehat{\mathbf{h}}^{(i)}}{\Delta t} \tag{2.42}$$

$$\frac{d}{dt}\widehat{\mathbf{e}}^{(i+1)} \approx \frac{\widehat{\mathbf{e}}^{(i+3/2)} - \widehat{\mathbf{e}}^{(i+1/2)}}{\Delta t} \tag{2.43}$$

Inserting the central differences into equations (2.39) and (2.40), one obtains the update scheme for the finite integration technique in time domain:

$$\widehat{\mathbf{e}}^{(i+1/2)} = \widehat{\mathbf{e}}^{(i-1/2)} + \Delta t \underline{\mathbf{M}}_{\varepsilon}^{-1} \left(\underline{\tilde{\mathbf{C}}} \widehat{\mathbf{h}}^{(i)} - \widehat{\widehat{\mathbf{j}}}^{(i)} \right) \tag{2.44}$$

$$\widehat{\mathbf{h}}^{(i+1)} = \widehat{\mathbf{h}}^{(i)} - \Delta t \underline{\mathbf{M}}_{\mu^{-1}} \underline{\mathbf{C}} \widehat{\mathbf{e}}^{(i+1/2)} \tag{2.45}$$

Please note, that the update equations for a hexahedral grid are identical with the ones known from the FDTD method, but the treatment of the boundaries is different. Also these state variables are different from FDTD, where the electric and magnetic fields are directly calculated.

[1] In this context the difference between integrating an ordinary and a partial derivative with respect to time is not significant, since the speed of light is constant in each cell. However, if electro-optical effects are included, i.e. the refractive index becomes dependent on the electric field strength, the speed of light in each cell is also dependent on the electric field. For such simulations the difference is not negligible any more. This is not a pathological example, since this effect is used e.g. in integrated-optical Mach-Zehnder interferometers, which are commonly used in telecommunications.

So far only the grid quantities have been calculated. To get the (physical) electric and magnetic fields, one has to divide the state variables by the edge length or face area, respectively.

$$e_\nu(n) = \frac{\widehat{\mathrm{e}}_\nu(n)}{s_\nu(n)} \qquad h_\nu(n) = \frac{\widehat{\mathrm{h}}_\nu(n)}{\tilde{s}_\nu(n)} \tag{2.46}$$

$$b_\nu(n) = \frac{\widehat{\widehat{\mathrm{b}}}_\nu(n)}{A_\nu(n)} \qquad d_\nu(n) = \frac{\widehat{\widehat{\mathrm{d}}}_\nu(n)}{\tilde{A}_\nu(n)} \tag{2.47}$$

Formulating these quantities in the condensed form, where all state variables were put in one vector (see chapter 2.2.2), and using the diagonal matrices containing the lengths and areas of the edges and faces of the cells, one obtains:

$$\mathbf{e} = \underline{\mathbf{D}}_s^{-1}\widehat{\mathbf{e}} \tag{2.48a}$$

$$\mathbf{h} = \underline{\tilde{\mathbf{D}}}_s^{-1}\widehat{\mathbf{h}} \tag{2.48b}$$

$$\mathbf{b} = \underline{\mathbf{D}}_A^{-1}\widehat{\widehat{\mathbf{b}}} \tag{2.48c}$$

$$\mathbf{d} = \underline{\tilde{\mathbf{D}}}_A^{-1}\widehat{\widehat{\mathbf{d}}} \tag{2.48d}$$

2.2.6. Numerical stability

In order to derive a stability criterion it is necessary to have a closer look at the dispersion relation of a plane, electromagnetic wave in vacuum. In the case of continuous Maxwell's equations (in vacuum) it is well known:

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2}, \tag{2.49}$$

with the angular frequency ω and the wave vector $\boldsymbol{k} = (k_x, k_y, k_z)^T$. After discretization of space and time, one can derive the dispersion relation for a Cartesian grid of space and time [26]:

$$\left(\frac{\sin(\frac{1}{2}k_x s_x)}{\frac{s_x}{2}}\right)^2 + \left(\frac{\sin(\frac{1}{2}k_y s_y)}{\frac{s_y}{2}}\right)^2 + \left(\frac{\sin(\frac{1}{2}k_z s_z)}{\frac{s_z}{2}}\right)^2 = \frac{1}{c^2}\left(\frac{\sin(\frac{1}{2}\omega\Delta t)}{\frac{\Delta t}{2}}\right)^2, \tag{2.50}$$

with the speed of light in vacuum c and the corresponding lengths of the cell edges s_i. It is required for all natural oscillations and their superposition that the field

energy remains bounded, which means that in case of harmonic oscillations the angular frequencies ω have to be real-valued. This can be achieved by the condition

$$\left(\sin(\frac{1}{2}\omega\Delta t)\right)^2 \leq 1. \tag{2.51}$$

Equivalent requirements for the three terms on the left side lead to a condition for the size of the time steps, which is called *Courant-Friedrichs-Lewy (CFL)*[2] condition:

$$\Delta t = a\frac{1}{c}\frac{1}{\sqrt{\frac{1}{s_x^2}+\frac{1}{s_y^2}+\frac{1}{s_z^2}}}, \tag{2.52}$$

with a constant $a \in [0, 1[$. This condition is only valid for homogeneous, equidistant grids. In practice one usually deals with several materials and non-uniform grids (often with a local mesh refinement), so that the condition above has to be modified:

$$\Delta t = a\frac{\sqrt{\varepsilon_i\mu_i}}{\sqrt{\frac{1}{s_{x,i}^2}+\frac{1}{s_{y,i}^2}+\frac{1}{s_{z,i}^2}}} \tag{2.53}$$

This expression is an empirical one, where the time step size depends on the size of the smallest cell within the computational domain.

This expression is only a compromise between to different errors made by the integration with respect to time. Consider one cell of an equidistant grid and a time step size higher than calculated by the expression above, then one would neglect influences by the fields in the cells next to the one considered, i.e. one would get a non-physical behavior of the fields, because one would underestimate the implicated speed of light. On the other hand, if the time step size is chosen way to small, one would consider influences of neighboring cells too much, i.e. one would overestimate the implicated speed of light. This results in an unwanted dispersion (the so called *numerical dispersion*), but mostly the simulation is stable. To systematically treat this problem, one would have to integrate each cell with a different time step size, which is unpractical. This is why one has to choose the maximum time step, where the simulation still runs stable. For most applications typical values for the parameter a lie in between $a = 0.9$ and

[2] Particularly in the community of the finite difference time domain method, this condition is often only called the *Courant* condition. Reducing this method to 1D and using an equidistant grid, one obtains the so called *magic time step*, where the relation between time step size and cell lengths is exact and no numerical dispersion occurs, because the time integration is done exactly on one of the so called characteristics. Further information can be found in [23].

$a = 0.99$. In this range the numerical dispersion can be neglected for most applications and small simulation spaces, but in some applications like resonators the numerical dispersion can be significant, which results in e.g. shifts in resonance frequencies.

2.3. Near-Field to Far-Field transformation

In this work directional antennas are optimized. Characteristic of such antennas are quantities such as directivity or gain, which are both determined in the far field. However, using a FDTD or FIT algorithm usually gives you information about the near field only, so a near to far field (NTFF) transformation has to be applied. Using the near-field data obtained in a single FDTD modeling run, this transformation efficiently and accurately calculates the complete far-field radiation pattern of such an antenna. In this section only a short summary of this transformation will be described. A more detailed derivation can be found in [23].

The standard NTFF-transformation formulation is based on the surface equivalence theorem. The idea is, that knowledge of the equivalent electric and magnetic currents tangential to any closed path, surrounding a two dimensional structure is sufficient to calculate the far field by integration of those currents [36].

Consider an arbitrary structure surrounded by vacuum, such as in Fig. 2.4 (a). It can be assumed, that all of the space is filled with the electric field $\boldsymbol{E}_1$ and magnetic field $\boldsymbol{H}_1$, generated by the electric and magnetic current sources $\boldsymbol{J}_1$ and $\boldsymbol{M}_1$, which are flowing on the surface of the structure. By removing the structure the fields inside of the space S change to $(\boldsymbol{E}, \boldsymbol{H})$.

Since the fields within S can be anything, it is useful to assume them to be identically zero. So the original structure can be replaced by an empty space with equivalent currents on its surface (Fig. 2.4 (b)), where the equivalent current densities can be expressed as

$$\boldsymbol{J}_S = \hat{\boldsymbol{n}} \times \boldsymbol{H}_1 \tag{2.54}$$

$$\boldsymbol{M}_S = -\hat{\boldsymbol{n}} \times \boldsymbol{E}_1 \tag{2.55}$$

with the unit outward normal vector $\hat{\boldsymbol{n}}$ to S. Following this concept and using inverse Fourier transform leads to the expression for the near-to-far-field in time domain:

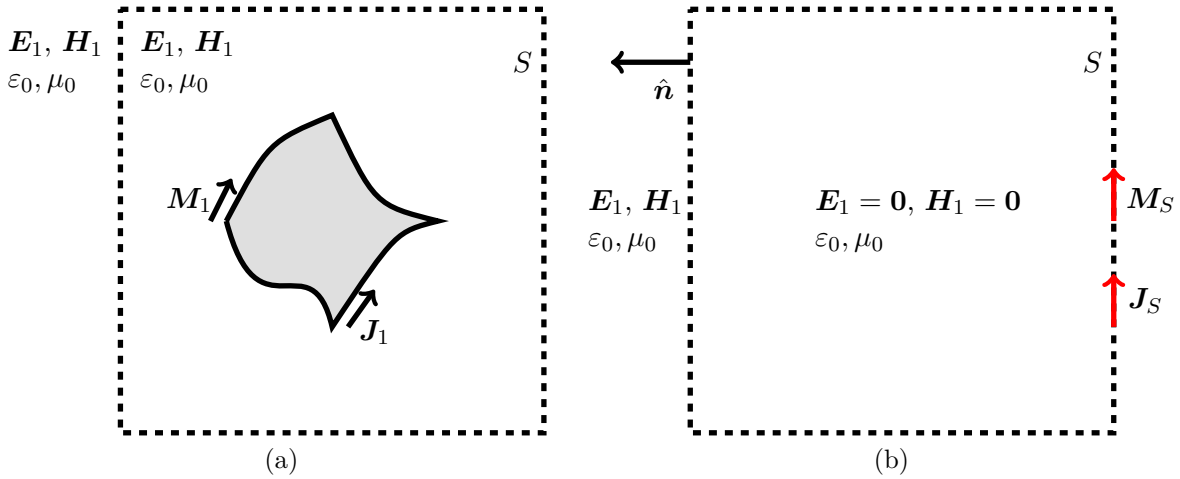


Fig. 2.4.: *Sketch of the electric and magnetic fields and the equivalent electric and magnetic virtual currents, used in the surface equivalence theorem. (a) The gray shape indicates an arbitrary scattering structure, with the corresponding surface currents $\boldsymbol{M}_1$ and $\boldsymbol{J}_1$. (b) The space now is empty and without fields, but on the surface are the equivalent surface currents.*

$$E_\theta(\boldsymbol{r}, t) = -Z_0 W_\theta(\boldsymbol{r}, t) - U_\phi(\boldsymbol{r}, t) \tag{2.56}$$

$$E_\phi(\boldsymbol{r}, t) = -Z_0 W_\phi(\boldsymbol{r}, t) + U_\theta(\boldsymbol{r}, t) \tag{2.57}$$

where $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ is the impedance of free space and the components of the electric field vector $E_i, i = r, \theta, \phi$ in spherical coordinates. Since in electromagnetic waves in free space are of almost pure transversal nature, the radial component of the electric field is $E_r(\boldsymbol{r}, t) \cong 0$. W_i and U_i are the angular components of the vector potentials, which can be determined by the following integrals of the equivalent current densities.

$$\boldsymbol{W}(\boldsymbol{r}, t) = \frac{1}{4\pi|\boldsymbol{r}|c_0} \frac{\partial}{\partial t} \left[\int_S \boldsymbol{J}_S \left(t - \frac{|\boldsymbol{r}| - \boldsymbol{r'} \cdot \hat{\boldsymbol{r}}}{c_0} \right) dS' \right] \tag{2.58}$$

$$\boldsymbol{U}(\boldsymbol{r}, t) = \frac{1}{4\pi|\boldsymbol{r}|c_0} \frac{\partial}{\partial t} \left[\int_S \boldsymbol{M}_S \left(t - \frac{|\boldsymbol{r}| - \boldsymbol{r'} \cdot \hat{\boldsymbol{r}}}{c_0} \right) dS' \right] \tag{2.59}$$

where $\boldsymbol{r}$ denotes the position of the observation point and $\boldsymbol{r'}$ the position of the source point in S. The unit vector $\hat{\boldsymbol{r}}$ then points into the direction of $\boldsymbol{r}$. Note, that in the argument of $\boldsymbol{J}_S$ and $\boldsymbol{M}_S$ a time delay can be seen. This is the propagation delay of an electromagnetic wave over the distance from the source $\boldsymbol{r'}$ to the observation point $\boldsymbol{r}$

in free space and is given by

$$\tau = \frac{|\boldsymbol{r}| - \boldsymbol{r'} \cdot \hat{\boldsymbol{r}}}{c_0} = \frac{|\boldsymbol{r}| - |\boldsymbol{r'}| \cos \Psi}{c_0}, \tag{2.60}$$

with the angle Ψ between the two vectors $\boldsymbol{r}$ and $\boldsymbol{r'}$.

This formalism is developed for the propagation of electromagnetic waves in free space. Simply by substituting the speed of light in the corresponding medium ($c = c_0/n$), which homogeneously surrounds the structure, one can also calculate the far fields in a dielectric medium. However, often a near-to-far-field transformation has to be done for a half space filled with two different media, which makes this approach much more complicated.

A formal correct way to obtain a NTFF for a half space is described in detail in [37], where a spectral dyadic Green's function is used, determined by several integrals and Fourier transforms. Since in this work automatic optimization plays an important role, where the calculation time is a critical point, a less costly approach is used. The near-to-far-field transformation is applied for both media and the results are simply combined. There might appear minor errors directly at the interface between both media, but since the interesting values are far away located from this interface, the errors can be neglected.

2.4. The dielectric function of metals (Drude model)

In order to describe the behavior of metals in optical fields, an appropriate model has to be introduced. In this work a simple model, known as Drude model, is used.

The main assumption within this model is, that many electrons are not bound by the ions of the metal, i.e. they can be treated as a free electron gas, but the electrons cannot propagate over the boundary of the metal either. They oscillate in response to the applied electromagnetic field, their motion is damped via collisions occurring with a characteristic collision frequency $\gamma = 1/\tau$, whereas τ is known as the relaxation time of the free electron gas. The resulting equation of motion for an electron with the charge of $e = -q$ in the metal is given in eq. (2.61).

$$\frac{d^2}{dt^2}\boldsymbol{r}(t) + \gamma \frac{d}{dt}\boldsymbol{r}(t) = -\frac{e}{m_e}\boldsymbol{E}(t) \tag{2.61}$$

with the electron mass m_e. By using the velocity $\boldsymbol{v} = \frac{d}{dt}\boldsymbol{r}$ of the electron the differential equation simplifies to

$$\frac{d}{dt}\boldsymbol{v}(t) + \gamma\boldsymbol{v}(t) = -\frac{e}{m_e}\boldsymbol{E}(t) \tag{2.62}$$

Fourier transform via differentiation theorem yields:

$$\boldsymbol{v}(\omega) = -\frac{e}{m_e\left(-i\omega + \gamma\right)}\boldsymbol{E}(\omega) \tag{2.63}$$

Using the relation of the current density $\boldsymbol{j}(\omega) = -en\boldsymbol{v} = \sigma(\omega)\boldsymbol{E}(\omega)$ for linear and isotropic media under the assumption, that the electrons don't interact with each other, one obtains the conductivity $\sigma(\omega)$.

$$\sigma(\omega) = \frac{ne^2}{m_e}\frac{1}{-i\omega + \gamma} \tag{2.64}$$

With the conductivity the generalized permittivity can then be calculated by

$$\tilde{\varepsilon}(\omega) = \varepsilon_0\varepsilon_\infty + \frac{i\sigma}{\omega} = \varepsilon_0\left(\varepsilon_\infty - \frac{\omega_p^2}{\omega^2 + i\omega\gamma}\right) \tag{2.65}$$

The parameter $\omega_p = \frac{ne^2}{\varepsilon_0 m_e}$ is the *plasma frequency* and γ the *collision frequency* of the electrons in the free electron gas. They are phenomenological parameters and need to be fitted to experimental measurements. The real-valued permittivity ε_∞ is called the background permittivity, since it is caused by the background ions, which are not considered in this free electron gas model. It is also often marked with an ∞, because it becomes the dominant term, when the frequency is very large, i.e. $\omega >> \omega_p$. The commonly used dielectric function consists of the generalized permittivity divided by ε_0:

$$\varepsilon_r(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{(\omega^2 + i\omega\gamma)} \tag{2.66}$$

It is insightful to have a closer look at the dielectric function for different frequency regimes. In Fig. 2.5 a fit of this Drude model to experimental data [38] of gold for frequencies between $250THz$ and $400THz$ is shown (for parameters see table A.1). In order to obtain the fit the deviation of the model to the experimental data in the

specified frequency regime has been calculated and then minimized. For the minimization the CRS2LM algorithm has been used, which is described in detail in chapter 3.3. This model is in good agreement for frequencies below $500THz$. One can see, that with increasing frequency the imaginary part becomes negligible, which doesn't correspond to measurements and cannot be modeled by a single Drude pole. In this region interband transitions within the d-shell of the molecules are possible, so the losses increase significantly, which is not included in the Drude model.

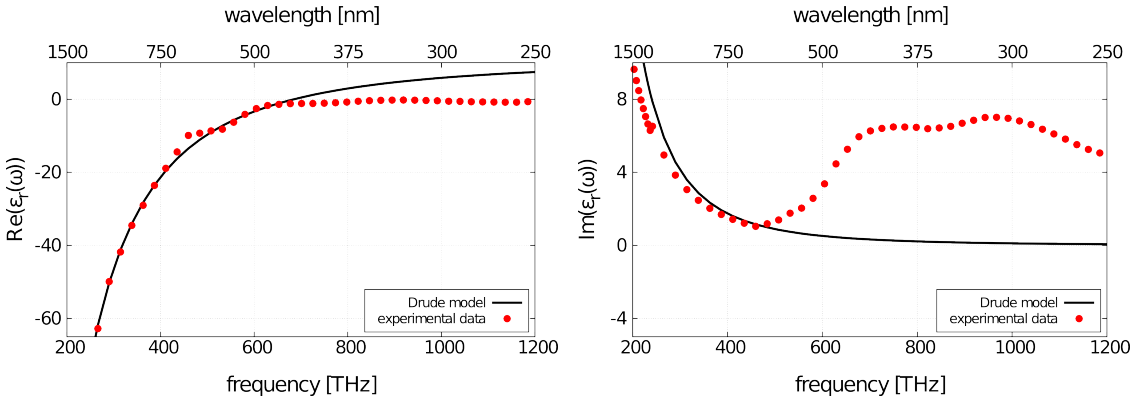


Fig. 2.5.: *Real part and imaginary part of the dielectric function of gold. The black line is given by the Drude model, with the parameters fitted to the experimental data (red dots).*

This region can be modeled by considering, that not all electrons have necessarily the same binding frequency, i.e. that one needs to take several of such poles into account, which are just added up [39]. In eq. (2.67) one can see an extension of this model, with the index p for each pole.

$$\varepsilon_r(\omega) = \varepsilon_\infty - \sum_{p=1}^{P} \frac{\omega_p^2}{(\omega^2 + i\omega\gamma_p)} \tag{2.67}$$

In Fig. 2.6 the Drude model with $P = 4$ poles is fitted to the experimental data. One can see now a good agreement even for higher frequencies. However, this model is more costly than the standard Drude model, since several poles have to be calculated all the time.

For lower frequencies, the model seems to be accurate. We consider now the regime of very low frequencies, i. e. $\omega \ll \gamma$. Dividing the complex permittivity into real and

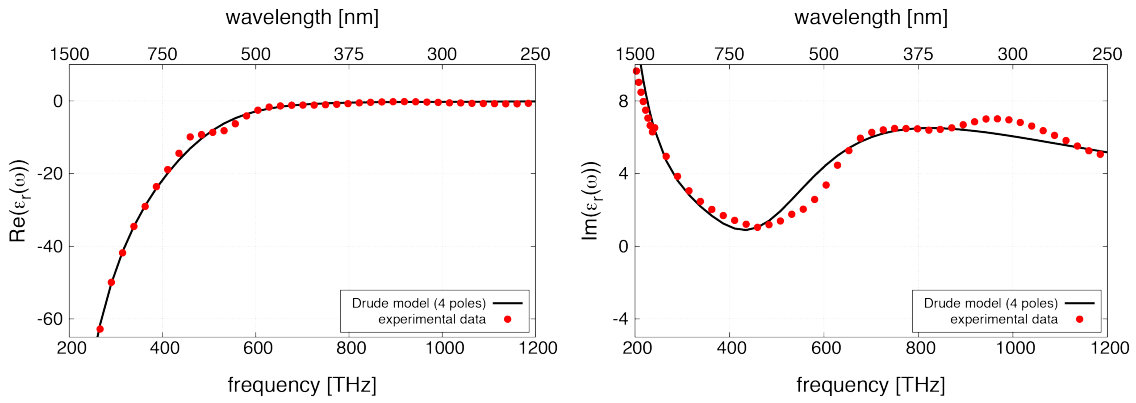


Fig. 2.6.: *Real part and imaginary part of the dielectric function of gold. The black line is given by the Drude model using $P = 4$ poles, with the parameters fitted to the experimental data (red dots).*

imaginary parts $\epsilon_r(\omega) = \varepsilon'(\omega) + i\varepsilon''(\omega)$, the real and the imaginary part of the complex index of refraction are of comparable magnitude:

$$n \approx \kappa = \sqrt{\frac{\varepsilon''}{2}} = \sqrt{\frac{\omega_p^2}{2\gamma\omega}} \tag{2.68}$$

In this region metals are mainly absorbing, with an absorption coefficient of

$$\alpha = \frac{\Im(n(\omega))\omega}{c_0} = \left(\frac{\omega_p^2\omega}{2\gamma c_0^2}\right)^{1/2}, \tag{2.69}$$

where c_0 is the speed of light in vacuum. Beer's law of absorption states, that for low frequencies the field inside the metal decreases exponentially with $\sim \exp(-x/\delta)$. δ then defines the skin depth and can be calculated by the inverse of the absorption coefficient, i.e.

$$\delta = \frac{1}{\alpha} = \left(\frac{2\gamma c_0^2}{\omega_p^2\omega}\right)^{1/2}. \tag{2.70}$$

For room temperature, the skin depth for typical metals is around $\delta \approx 100nm$. Since most devices for low frequencies are much larger than $100nm$, metals can be treated as PEC in e.g. the RF regime. A more detailed study of the low-frequency behavior shows, that this description is indeed valid, as long as the mean free path of the electrons is much smaller than the skin depth [40].

Integration of this dielectric function for metals in a time dependent solver for Maxwell's equations is straight forward[3]. In eq. (2.71) Ampere's law in the time domain can be expressed for this medium as

$$\nabla \times \boldsymbol{H} = \varepsilon_0 \varepsilon_\infty \frac{\partial \boldsymbol{E}}{\partial t} + \sigma \boldsymbol{E} + \sum_{p=1}^{P} \boldsymbol{J}_p \tag{2.71}$$

where $\boldsymbol{J}_p$ is the polarization current associated with the p'th Drude pole [23]. Since the dielectric function is obtained in the frequency domain, the corresponding current density, given in eq. (2.72), where $\tilde{}$ indicates the frequency dependence of the vectors, needs to be transformed.

$$\tilde{\boldsymbol{J}}_p = -i\omega\varepsilon_0 \left(\frac{\omega_p^2}{\omega^2 - i\omega\gamma_p} \right) \tilde{\boldsymbol{E}} \tag{2.72}$$

Applying the differentiation theorem and integrating once with respect to time, one obtains the equation of motion for the current density in time domain (eq. (2.73)), which can be used in eq. (2.71).

$$\frac{\partial \boldsymbol{J}_p}{\partial t} + \gamma_p \boldsymbol{J}_p = \varepsilon_0 \omega_p^2 \boldsymbol{E} \tag{2.73}$$

Using the semi-implicit scheme, this can be easily and accurately implemented in an FDTD code [23]. A simple finite-difference expression centered at time-step $n + 1/2$ for the current density yields

$$\left(\frac{\boldsymbol{J}_p^{n+1} - \boldsymbol{J}_p^n}{\Delta t} \right) + \left(\frac{\boldsymbol{J}_p^{n+1} + \boldsymbol{J}_p^n}{2} \right) = \varepsilon_0 \omega_p^2 \left(\frac{\boldsymbol{E}^{n+1} + \boldsymbol{E}^n}{2} \right) \tag{2.74}$$

Solving eq. (2.74) for $\boldsymbol{J}_p^{n+1}$, one obtains

$$\boldsymbol{J}_p^{n+1} = k_p \boldsymbol{J}_p^n + \beta_p \left(\boldsymbol{E}^{n+1} + \boldsymbol{E}^n \right) \tag{2.75}$$

with

[3] A detailed description can be found in [23], from which also the notation is taken for a better readability. For the case, that one uses a hexahedral grid in FIT, its update equations are exactly the same than in a standard FDTD, which is also described in [23].

$$k_p = \frac{1 - \frac{1}{2}\gamma_p \Delta t}{1 + \frac{1}{2}\gamma_p \Delta t} \qquad \beta_p = \frac{\frac{1}{2}\omega_p^2 \varepsilon_0 \Delta t}{1 + \frac{1}{2}\gamma_p \Delta t} \tag{2.76}$$

In order to obtain the $\boldsymbol{E}^{n+1}$ component, one uses again the semi-implicit scheme, which leads to the expression:

$$\begin{aligned} \boldsymbol{E}^{n+1} = & \left(\frac{2\varepsilon_0\varepsilon_\infty - \Delta t \sum_{p=1}^{P} \beta_p - \sigma \Delta t}{2\varepsilon_0\varepsilon_\infty + \Delta t \sum_{p=1}^{P} \beta_p + \sigma \Delta t} \right) \boldsymbol{E}^n \\ & + \left(\frac{2\Delta t}{2\varepsilon_0\varepsilon_\infty + \Delta t \sum_{p=1}^{P} \beta_p + \sigma \Delta t} \right) \cdot \left[\nabla \times \boldsymbol{H}^{n+1/2} - \frac{1}{2} \sum_{p=1}^{P} (1 + k_p) \boldsymbol{J}_p^n \right] \end{aligned} \tag{2.77}$$

2.5. Coupled Dipole Approximation (CDA)

For some applications it is insightful not to solve the full Maxwell's equations, but instead use a simplified model, for it is simpler then to distinguish between different effects, which lead to the performance of the investigated device. This is particularly useful when using automatic optimization routines in order to design the device, since one obtains only well performing solution of the objective function. However, in practice it is often much harder to understand the optimized result than to obtain it. In this section the model of the CDA is discussed.

When nano particles, which are surrounded by homogeneous, isotropic and linear material, are excited by e.g. a plane wave, the structure gets polarized. If the polarizability is known, one can derive an analytic expression for a dipole moment of the investigated particle. With the dipole moment one can calculate the electric field everywhere in space analytically, as well as other properties like extinction cross section, radar cross section, directivity, etc. As is clear from the name, it is assumed, that only dipoles are excited, i.e. all higher order poles are neglected, which can lead to significantly different results compared to full Maxwell simulations. A drawback of this model is, that there are few particle geometries, for which analytical expressions for the polarizability can be found. The concept of the CDA is taken from [41, 42], where it has been discussed in more detail.

2.5.1. System of equations for CDA

Consider a system of N nano particles, which are assumed to be dipoles. Each dipole has a polarization of $\boldsymbol{P}_i = \underline{\alpha}_i \boldsymbol{E}_i$. The electric field $\boldsymbol{E}_i$ at the position of each dipole is then the incident field plus the contribution of each of the other $N-1$ dipoles, with which one obtains the dipole moments:

$$\boldsymbol{p}_i = \underline{\alpha}_i \boldsymbol{E}_i = \underline{\alpha}_i \left(\boldsymbol{E}_{incident,i} - \sum_{j \neq i} \underline{\mathbf{A}}_{ij} \boldsymbol{p}_j \right) \tag{2.78}$$

The incident field can be chosen arbitrarily, as long as the field is homogeneous over the whole particle size. In this case a plane wave satisfies all requirements:

$$\boldsymbol{E}_{incident,i} = \boldsymbol{E}_0 \exp(i\boldsymbol{k} \cdot \boldsymbol{r}_i - i\omega t) \tag{2.79}$$

The coupling matrices $\underline{\mathbf{A}}_{ij}$ consist basically of dyadic Green's functions and can be analytically derived:

$$\underline{\mathbf{A}}_{ij} \overset{i \neq j}{=} \frac{\exp(ikr_{ij})}{r_{ij}} \left[k^2 (\hat{\boldsymbol{n}}_{ij} \otimes \hat{\boldsymbol{n}}_{ij} - \underline{\mathbf{I}}_3) + \frac{ikr_{ij} - 1}{r_{ij}^2} (3\hat{\boldsymbol{n}}_{ij} \otimes \hat{\boldsymbol{n}}_{ij} - \underline{\mathbf{I}}_3) \right] \tag{2.80}$$

with $r_{ij} = |\boldsymbol{r}_{ij}|$, $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, and $\hat{\boldsymbol{n}}_{ij} = \frac{\boldsymbol{r}_{ij}}{r_{ij}}$. The sign $\otimes$ denotes the dyadic product of two vectors, yielding a 3×3 matrix in this case. By defining the diagonal elements $\underline{\mathbf{A}}_{ii} = \underline{\alpha}_i^{-1}$, the scattering problem reduces to $3N$ complex linear equations:

$$\sum_{j=1}^{N_p} \underline{\mathbf{A}}_{ij} \boldsymbol{p}_j = \boldsymbol{E}_{incident,i} \tag{2.81}$$

That solved for two dipoles $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ yields:

$$\underbrace{(\underline{\mathbf{I}}_3 - \underline{\alpha}_1 \underline{\mathbf{A}}_{12} \underline{\alpha}_2 \underline{\mathbf{A}}_{21})}_{=\underline{\mathbf{M}}_1} \boldsymbol{p}_1 = \underline{\alpha}_1 \left(\boldsymbol{E}_{incident,1} - \underline{\mathbf{A}}_{12} \underline{\alpha}_2 \boldsymbol{E}_{incident,2} \right) \tag{2.82}$$

$$\boldsymbol{p}_1 = \underline{\mathbf{M}}_1^{-1} \underline{\alpha}_1 \left(\boldsymbol{E}_{incident,1} - \underline{\mathbf{A}}_{12} \underline{\alpha}_2 \boldsymbol{E}_{incident,2} \right) \tag{2.83}$$

$$\underbrace{(\underline{\mathbf{I}}_3 - \underline{\alpha}_2 \underline{\mathbf{A}}_{21} \underline{\alpha}_1 \underline{\mathbf{A}}_{12})}_{=\underline{\mathbf{M}}_2} \boldsymbol{p}_2 = \underline{\alpha}_2 \left(\boldsymbol{E}_{incident,2} - \underline{\mathbf{A}}_{21} \underline{\alpha}_1 \boldsymbol{E}_{incident,1} \right) \tag{2.84}$$

$$\boldsymbol{p}_2 = \underline{\mathbf{M}}_2^{-1} \underline{\alpha}_2 \left(\boldsymbol{E}_{incident,2} - \underline{\mathbf{A}}_{21} \underline{\alpha}_1 \boldsymbol{E}_{incident,1} \right) \tag{2.85}$$

With the dipole moments of the two interacting particles one can calculate the total electric field anywhere in space according to equation (2.86).

$$\boldsymbol{E}(\boldsymbol{r}) = \boldsymbol{E}_{incident}(\boldsymbol{r}) + \underline{\mathbf{A}}(\boldsymbol{r} - \boldsymbol{r}_1)\boldsymbol{p}_1 + \underline{\mathbf{A}}(\boldsymbol{r} - \boldsymbol{r}_2)\boldsymbol{p}_2 \tag{2.86}$$

2.5.2. Polarizability of spheroid particles

Assuming, the geometrical shapes of the metal particles are ellipsoids ($a \geq b \geq c$), whose surface is specified by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{2.87}$$

or - more specific - prolate spheroids ($b = c$) with the major axis along the x-axis, the polarizabilities can be analytically derived [43]. In equation (2.88) one gets the polarizability component for the direction of each of the three axis of the spheroid.

$$\alpha^j = 4\pi abc \frac{\varepsilon_m(\omega) - \varepsilon_d}{3\varepsilon_d + 3f_j(\varepsilon_m(\omega) - \varepsilon_d)}, \tag{2.88}$$

with $j = x, y, z$, the permittivity of the particle $\varepsilon_m(\omega)$ and the permittivity of the surrounding dielectric material ε_d. Unfortunately the surrounding material has to be homogeneous, since for a particle, which is placed on a substrate, a closed analytical expression is very hard (and for most cases not possible) to obtain. The parameters f_j are geometry dependent factors and are usually hard to determine, though for the special case of spheroids analytic expressions have been derived within a quasi-static limit. With a given eccentricity e

$$e^2 = 1 - \frac{b^2}{a^2}, \tag{2.89}$$

the x-component of the geometry factors is then

$$f_x = \frac{1 - e^2}{e^2}\left(-1 + \frac{1}{2e}\ln\left(\frac{1+e}{1-e}\right)\right). \tag{2.90}$$

Once one component of these factors are known, the others can be easily calculated via the relation

$$f_x + f_y + f_z = 1 \tag{2.91}$$

$$\Rightarrow f_x + 2f_y = 1. \tag{2.92}$$

In Fig. 2.7 one can see the geometry factor f_x (when x is the direction of the major axis of the ellipsoid) over the eccentricity. The insets show exemplary shapes for the eccentricities $e = 0$, $e = 0.5$, and $e = 0.9$. The eccentricity of $e = 1$ is not defined, but would correspond to an infinitesimal thin line, pointing in x-direction.

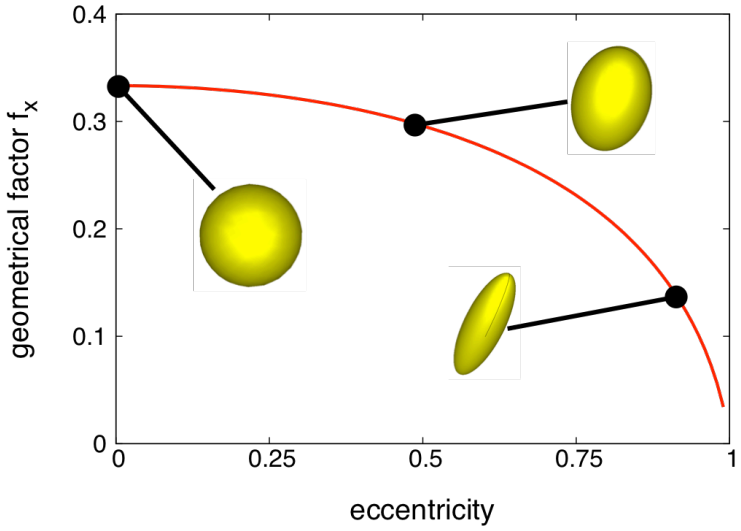


Fig. 2.7.: *The geometry factor f_x in dependency of the eccentricity e for prolate spheroids. The higher the eccentricity, i.e. the smaller the minor axis of the ellipsoid becomes compared to the major axis, the lower the value of f_x and therefore the lower the polarizability. The insets show examples of prolate spheroids with eccentricities of $e = 0.9$ (cigar shaped), $e = 0.5$, and $e = 0$ (sphere).*

This results then in a polarizability tensor for the prolate ellipsoid

$$\underline{\alpha}_i = \begin{pmatrix} \alpha_i^x & 0 & 0 \\ 0 & \alpha_i^y & 0 \\ 0 & 0 & \alpha_i^z \end{pmatrix} \tag{2.93}$$

If the axes of the prolate ellipsoid are not pointing into the directions of the Cartesian coordinate system, one can multiply the polarizability tensor with a rotational matrix for a proper alignment. For the special case, that the eccentricity is $e = 0$, one obtains

a spherical particle. The polarizability tensor then consist only of identical diagonal elements, thus it is often reduced to a scalar. When using the refractive index of the surrounding material instead of its permittivity, one obtains the so called *Clausius-Mosotti* polarizability:

$$\alpha_0 = 4\pi R^3 \frac{\varepsilon_m(\omega) - n^2}{\varepsilon_m(\omega) + 2n^2} \tag{2.94}$$

Often the particles are so small, that some finite size effects, which are included in the full dipolar Mie expansion coefficients, play a significant role, thus the polarizability has to be corrected. In [41] and [44] several of such corrections have been investigated, but for spherical particles only. For the case, that $kd \to 0$, with the wavenumber k and the largest dimension of the particle $d = a$, the polarizability is exact. But it has been shown, that for all other cases a radiative reaction correction of the order $\mathcal{O}[(kd)^3]$ should be included. In eq. (2.95) a correction is shown, which has been derived for the long-wavelength limit ($kd \ll 1$), but also works as long as $|n|kd < 1$, which has been shown empirically:

$$\tilde{\alpha} = \frac{\alpha}{1 + \frac{\alpha}{d^3}\left[(b_1 + n^2 b_2 + n^2 b_3 S)(kd)^2 - \frac{2}{3}i(kd)^3\right]}, \tag{2.95}$$

$$b_1 = -1.891531, \qquad b_2 = 0.1648469, \tag{2.96}$$

$$b_3 = -1.7700004, \qquad S = \sum_{j=1}^{3} (\hat{a}_j \hat{e}_j)^2, \tag{2.97}$$

where $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{e}}$ are unit vectors defining the incident direction and the polarization state. Comparison with FIT calculations for one particle have shown, that this correction also works for a low number of particles instead of an infinite array and also for prolate spheroids instead of spheres. A simpler version of this correction is used in [42] in which the constants b_2 and b_3 are zero, which works well for $kd < 1$, but causes some resonance shifts, if $kd \ll 1$.

3. Optimization theory

Focus of this chapter is the mathematical description of automatic optimization and some algorithms, which are often used in economy, natural science and technical applications. In general there are two types of algorithms, the local one, where only the nearest minimum is investigated, and the global one, where one is more concerned with the global minimum. All algorithms may also be divided in a gradient based and a gradient free algorithm, which describes how an algorithm is choosing it's function evaluations.

This chapter gives a short overview of the optimization theory and a summary of several algorithms, which have been used in this work and is based on [45, 46].

3.1. General description of optimization problems

A mathematical optimization problem (P) is the task to find the minimum (or maximum) of a continuous objective function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. The search region Ω (or parameter space) is usually a box, defined by bound constraints. The parameters (often called degrees of freedom (DOF)) then build a n-dimensional vector $\boldsymbol{x}$, so that this problem can be written as:

$$\min_{\text{s.t. } \boldsymbol{x}\in\Omega} f(\boldsymbol{x}) \tag{3.1}$$

A point $\boldsymbol{x}_{opt}$ is called a global minimizer, if

$$f(\boldsymbol{x}_{opt}) \leq f(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \Omega. \tag{3.2}$$

Maximizing the objective function is usually reduced to the minimizing problem. Consider an objective function $\tilde{f}(\boldsymbol{x})$, which should be maximized. This problem

$$\max_{\text{s.t. } \boldsymbol{x}\in\Omega} \tilde{f}(\boldsymbol{x}) \tag{3.3}$$

is equivalent to

$$\min_{\text{s.t. } \boldsymbol{x}\in\Omega} f(\boldsymbol{x}), \tag{3.4}$$

if the objective function f is defined as $f(\boldsymbol{x}) = -\tilde{f}(\boldsymbol{x})$. In this chapter only minimizing problems are considered, though in later chapters the objective function is often maximized.

Generally, unless both the objective function $f(\boldsymbol{x})$ and the parameter space Ω are convex, there may be several local minima. A local minimum then is defined as a point $\boldsymbol{x}_0$ for which exists some $\delta > 0$, so that for all $\boldsymbol{x}$ in a sub-domain of Ω

$$\|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \delta \tag{3.5}$$

the expression

$$f(\boldsymbol{x}_0) \leq f(\boldsymbol{x}) \tag{3.6}$$

holds. In other words, on some region around $\boldsymbol{x}_0$ all other function values are greater or equal to the found minimum. A large number of algorithms proposed for solving non-convex problems are not capable of making a distinction between local and global optimal solutions. If the objective function is not known, which is most often the case, if no analytic expression for it is known, one cannot distinguish between local and global optimum, i.e. whenever an optimum is found, there still might be a better solution. However, many of the developed algorithms can be classified into a local and a global optimization algorithm, which basically means that they have different chances to find the global optimum.

One also can have several objective functions, which adds to the problem's complexity, particularly if they are competing with each other. Then one has to weight these objective functions. However, it is hard to define an optimum then, because the favorite solution would be a compromise of both functions and depends on the problem at hand.

One of Fermat's theorems states, that optima of unconstrained problems are found at stationary points, where the gradient of the objective function is zero or undefined

(e.g. at the boundary of the parameter space). This is a necessary, but not a satisfying condition, and is therefore called a first order condition. If this is fulfilled, one may check the Hessian matrix, which is then called the second order condition. Are both conditions fulfilled, one has reached at least a local optimum. There are plenty of algorithms, which are based on evaluating the gradient of the objective function, thus they are called *gradient based* algorithms. Example of such an algorithm would be the BFGS-algorithm, which is based on the Newton method [45]. Since in this work the objective function has to be evaluated numerically, the computational effort in order to determine the gradients would be much too costly, so this chapter is restricted to gradient free algorithms only.

Often additional constraints for the parameters (apart from bound constraints) are required for a proper optimization. Constraints can be either *hard constraints*, which set conditions for the variables that are required to be satisfied, or *soft constraints*, which have some variable values that are penalized in the objective function, if the conditions on the variables are not satisfied. In general one distinguishes inequality and equality constraints, where the latter are harder to solve. Often the equality constraints can be converted into inequality constraints by the Lagrange multipliers, but still should be avoided, if possible.

Since the constraints effectively constrict the parameters, the search region Ω often becomes not-convex, which significantly influences the performance of many algorithms like simplex methods, where a simplex is contracting towards the (global) minimum of the objective function. Thus many algorithms do not provide the possibility to use constraints.

In many applications, e.g. in engineering, the objective function may be nonlinear, non-smooth, and / or simulation based. Therefor several methods were developed, which do not require much information about the function itself. Some of them are the DIRECT algorithm [47], simulated annealing [48], differential evolution [49], genetic algorithms [50] and the controlled random search [51]. In the following sections three algorithms, which have mainly been used in this work, are shortly summarized.

3.2. COBYLA

The *constrained optimization by linear approximation (COBYLA)*-algorithm has been developed for gradient-free optimization with nonlinear inequality and equality con-

straints, by M. J. D. Powell [52]. It constructs successive linear approximations of the objective function and constraints via a simplex of $n+1$ points (with an n-dimensional parameter space) and optimizes these approximations in a trust region at each step.

For that it employs the values of the objective function $f(\boldsymbol{x}_i)$, $i = 0, 1, ..., n$ at the vertices of a simplex, which is changed after each iteration. They are interpolated by a linear polynomial $L(\boldsymbol{x})$, $\boldsymbol{x} \in \Omega$, where each iteration requires a trust region radius $\Delta > 0$, which is adjusted automatically. The next vector of variables is then obtained by minimizing $L(\boldsymbol{x})$ with subject to $\|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \Delta$, where $f(\boldsymbol{x}_0)$ is the least value of the objective function that has been found so far. Thus, in the unconstrained case, $\boldsymbol{x}$ is the point

$$\boldsymbol{x} = \boldsymbol{x}_0 - \left(\frac{\Delta}{\|\nabla L\|}\right) \nabla L, \tag{3.7}$$

which is different from all the vertices of the current simplex, consisting of the vertices $\boldsymbol{x}_i$. The value $f(\boldsymbol{x})$ is calculated and then the simplex for the next iteration is formed by replacing just one value of the old simplex by $\boldsymbol{x}$. To avoid misleading gradients of the polynomial L, one usually defines a lower bound for Δ, depending on the requested accuracy of the algorithm.

The COBYLA algorithm is a so called local optimization algorithm, i.e. it only finds the minimum nearest to the initial parameter set. The version used in this work has some minor improvements, e.g. the simplex steps are pseudo-randomized, which improves the robustness by avoiding accidentally taking steps that don't improve the conditioning.

3.3. Controlled Random Search with Local Mutation

The controlled random search (CRS) algorithm is a direct search technique and is purely heuristic. It starts by initially filling a set S with a sample of N, $N \gg n$ points, which are uniformly distributed over Ω. The sample S is then gradually contracted by replacing the current worst point $\boldsymbol{x}_{high}$ with a better point $\boldsymbol{x}$, called the trial point.

In order to obtain the trial point, one has to form a simplex consisting of the points $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_{n+1}$. The point $\boldsymbol{x}_1$ is the current best point of S, whereas the other n points are randomly chosen from S. The trial point is now obtained by reflecting the point $\boldsymbol{x}_{n+1}$ in the centroid of the remaining n points of the simplex, such as given in eq.

(3.8).

$$\boldsymbol{x} = \frac{2}{n}\sum_{j=1}^{n}\boldsymbol{x}_j - \boldsymbol{x}_{n+1} \tag{3.8}$$

This process of finding a trial point $\boldsymbol{x}$ and replacing the current worst point $\boldsymbol{x}_{high}$ in S, if the $\boldsymbol{x}$ is better than $\boldsymbol{x}_{high}$, is repeated until a certain stopping condition is met.

The controlled random search with local mutation (CRSLM) modifies the point generation scheme by introducing a local mutation. In the standard CRS the trial point $\boldsymbol{x}$ is discarded, if the function value $f(\boldsymbol{x})$ is not better than the function value of the current worst point $f(\boldsymbol{x}_{high})$. This results in a waste of generating points. If a trial point fails to replace $\boldsymbol{x}_{high}$ in S, the local mutation scheme generates a second trial point $\tilde{\boldsymbol{x}}$, exploring the region around the current best point $\boldsymbol{x}_{low}$. In particular this is done by coordinate-wise reflection of $\boldsymbol{x}$ through $\boldsymbol{x}_{low}$:

$$\tilde{x}^i = (1+\omega_i)x^i_{low} - \omega_i x^i, \tag{3.9}$$

where the index i denotes the corresponding component of the vectors $\tilde{\boldsymbol{x}}$, $\boldsymbol{x}$, and $\boldsymbol{x}_{low}$ and ω_i is a random number in $[0,1]$ for each i. The new trial point then competes with the current worst point in S. The local mutation scheme results in more exploration in earlier stages of the algorithm, when the points in S are widely spread, i.e. it has a global effect.

Since the points in S are uniformly distributed over Ω, this algorithm explores not only the region around the initial parameter set, i.e. it is a global optimization algorithm. However, depending on the objective function, the CRSLM will probably find the best of some optima, but still not the globally best solution. Since it needs many function evaluations, it is often recommended to implement rough stopping criteria, and use a local optimization algorithm like COBYLA afterwards, to find the exact optimum. A more detailed description and some performance tests can be found in [53].

In this work the CRSLM has been slightly modified for a better processing of additional constraints. Using other inequality constraints as bound constraints often turns the parameter space into a non-convex space, which means, that the simplex cannot contract towards an optimum anymore. As a result such constraints are not fulfilled or the algorithm does not converge. To overcome this obstacle, such constraints have been implemented as soft constraints, which adds a penalty for the objective function, if a constraint is not fulfilled.

Another slight modification deals with the discretization of the simulation space within the FDTD, where the simulation space has been divided into e.g. 4nm × 4nm × 4nm cells. Since simplex algorithms cannot handle a discretized parameter space (or most of their strengths would then vanish), one gets often similar parameters after discretization (particularly if the algorithm is almost converged). In order to reduce redundant simulations, a lookup table has been created. Before each function evaluation it has been checked, if an identical simulation has been performed. If so, then instead of starting the simulation, the corresponding value of the objective function has been taken from the table. These evaluations have not been counted has iterations.

3.4. Particle Swarm Algorithm

The particle swarm (PSA) optimization is similar to a genetic algorithm [54] in the sense, that the system is initialized with a population of random solutions. Unlike a genetic algorithm, however, each potential solution is also assigned a randomized velocity and these potential solutions, called particles, are then flown through hyperspace. Each particle then keeps track of its coordinates, which are associated with the current best solution, and the corresponding values of the objective function. After each iteration the velocity of the particles is changed towards the best solution so far. The acceleration of each particle is weighted by a random term.

These particles behave like a swarm of, let's say, insects, roughly following their queen, but still showing a partial independent behavior. It is a so called global optimization algorithm, but has the advantage, that when all the particles move towards the best solution (which still moves), the immediate region around these optima is investigated as well, i.e. one can get some kind of sensitivity analysis during the optimization process. This is important, because often local minima are unstable with respect to slight changes of the parameters. A more detailed discussion and a variant of the PSA for local optimization can be found in [55].

3.5. Rough Benchmark studies for selected algorithms

In order to state some properties of different optimization algorithms, a simple performance test as been done. For that an objective function has been defined, of which

the global minimum can be derived analytically. The objective function consisted of N n-dimensional Gaussian minima with different depths Δ_i:

$$f(\boldsymbol{x}) = -\sum_{i=1}^{N} \Delta_i e^{-\left(\frac{(x_1-\mu_{i,1})^2}{\sigma_i^2}+\frac{(x_2-\mu_{i,2})^2}{\sigma_i^2}+\frac{(x_3)^2}{\sigma_i^2}+\cdots+\frac{(x_n)^2}{\sigma_i^2}\right)}, \tag{3.10}$$

with the corresponding component of the parameter vector in the subscript. With that this objective function has many local optima, but is still smooth, so all algorithms should be able to find at least a local minimum. In Fig. 3.1 the objective function for the case, that $n = 2$, can be seen.

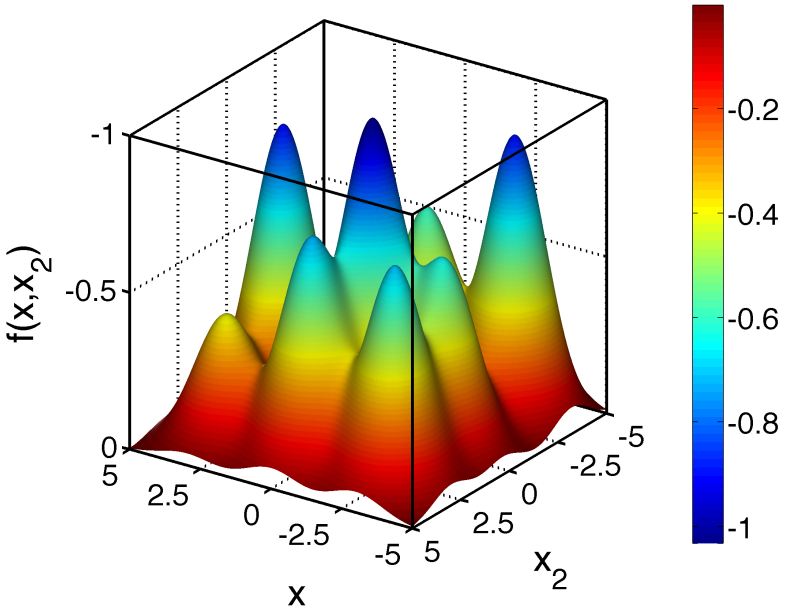


Fig. 3.1.: *The used objective function consists of several n-dimensional Gaussian functions and has therefore several local minima. This function is still smooth, so all algorithms should be able to find a (local) minimum.*

In Fig. 3.2 the convergence dependent on the degrees of freedom can be seen. The blue points show a complete random search, i.e. for a number of complete random points in Ω the value of the objective function has been calculated and the best value has been chosen. Unfortunately only for $n = 2$ and $n = 3$ the best value roughly hit the global minimum. In all other attempts the best value came not even close to any of the local optima. This is probably the most primitive optimization algorithm and gives a reference to the optimization problem at hand.

The green points show the results of a standard genetic algorithm (including

crossovers.) One can easily see the greatest disadvantage of this method, the extremely high number of function evaluation needed for convergence. An advantage of it is, that, with appropriate parameters, it most often finds the global minimum of the objective function. For $n \geq 25$ it didn't find the global optimum, because the maximum number of function evaluation was limited to 50,000. Another advantage is, though it is not shown in the graph, that this algorithm is very robust, i.e. it still finds at least local optima, if the objective function is non-smooth, has many sharp local minima and large areas, where a variation of parameters effects no change in its value. The particle swarm algorithm (PSA) is not included in this performance check, since it behaves much like the genetic scheme.

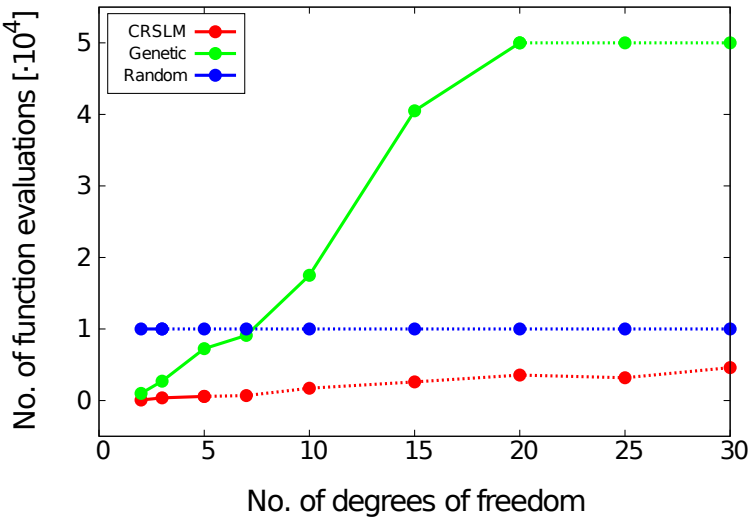


Fig. 3.2.: *Comparison of the CRSLM algorithm with a genetic one and a complete random search. All optimization processes have been converged, but many didn't find the global minimum (indicated by dashed lines), but often found a local minimum, in which the value of the objective function was close to the global optimum.*

The red points show the results of the CRSLM algorithm. Compared to the genetic scheme, it converges very fast[1], but finds the global optimum only for $n \leq 5$. In all other cases it still finds a local minimum, where the value of the objective function is still close to the global minimum. Like the genetic algorithm the CRSLM is very robust and yields good starting points for local optimizers, even if it is not converged. Additionally other inequality constraints than bound constraints can be easily implemented without

[1]Compared to gradient-based algorithms it still converges slower, which is not shown here.

slowing down the optimization process.

Note, that this little performance check gives only a hint to the properties of the algorithms, since the objective function is a very unproblematic one. In practice almost every optimization problem would be much more complex, which is the case when optimizing nano antennas. Additionally each function evaluation means a complete simulation of Maxwell's equations, i.e. depending on the computational domain and used materials, it takes between a few minutes and several hours of time. Since it is nearly impossible to calculate derivatives of the objective function with respect to the design variables, gradient-based algorithms cannot be used in this work. One option would have been an automatic differentiation algorithm, like ADOL-C [56], but the implementation of such a method would have required a completely new FDTD or FIT code and was therefore not practical in this work.

For optimizations with the FDTD method the library *nlopt* [77] has been used. This library contains a collection of the most used optimization algorithms. The aim of this section is to compare the CRSLM with the most commonly known (the genetic) algorithm and the worst one (pure random tries). We chose the CRSLM algorithm for optimization problems, where a nasty objective function is expected, i.e. an objective function with many local optima, steep gradients and large areas, where the gradient is (almost) equal to zero. Although this algorithm did not find the global maximum as often as the genetic one, it has always been close, and since the necessary number of iteration have been considerably smaller compared to the genetic algorithm, it suited the requirements. For simulations with CST Microwave Studio, where the parameter space Ω could be limited easily, the built-in particle swarm algorithm is used, which is not implemented in the *nlopt* library.

4. Optical and NIR antennas for light intensity enhancement

Nanoantennas are able to highly increase the electric field in small areas [57], which has a wide field of possible applications, e.g. solar energy conversion [58], spectroscopy [59, 8], sensing applications [60], or enhancement of excitation and emission of quantum emitters [61, 62, 63]. Due to the complex interplay of near-field interference, recursive coupling and plasmonic modes, the field enhancement is very sensitive to the geometrical shape [64]. The most commonly used geometries are bow-ties [65, 66], rods [67] and discs [8]. Therefore, it is essential to achieve a design, which is simple to fabricate and is capable of generating the highest intensity enhancement for the desired frequency. Optimization of these structures is often done by trial-and-error, which might not result in an optimal enhancement. Models, which contain design rules, are only known for the simplest antennas [16].

It is well known, that with sharp metal tips, the electric field near to those tips can be highly enhanced, so most investigated structures have such sharp tips. But the geometrical shapes are limited to the creativity of the designer, i.e. other surprising shapes might exist, which are not that intuitive in their functionality. So recently, automatic optimization of plasmonic antennas became more popular, for instance where the distances between several nano discs were varied [68] or an optimized structure was used as an inspiration for a new antenna shape [69]. Particularly in the latter paper, an optimization with many degrees of freedom has been performed, allowing an almost arbitrary shape of the resulting antenna. But, as is almost the case, the definition of the objective function and the constraints determines the outcome of the optimization.

In this section first some fundamentals and principles of nanoantennas are discussed. Then the popular bow-tie is optimized to be used as a reference, with which the following optimization of an (almost) arbitrary geometrical shape can be compared. The sections 4.4 and 4.5 are based on Refs. [63, 70].

4.1. Localized surface plasmon polaritons

In this section some fundamentals of surface plasmon polaritons are briefly discussed, taken from [71], where a much more detailed study can be found. Though the Drude model, which has been discussed in chapter 2.4, is not exact, some important features of metals in optical or NIR light fields can be derived. For large frequencies, that means $\gamma \ll \omega < \omega_p$, the metals still retain their metallic character, but the damping becomes negligible. In this region the permittivity $\varepsilon(\omega)$ is predominantly real-valued and becomes

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}, \tag{4.1}$$

with the plasma frequency ω_p. Inserting equation (4.1) into (2.7), one obtains

$$\omega = \sqrt{\omega_p^2 + |\boldsymbol{k}|^2 c_0^2} \tag{4.2}$$

the dispersion relation for traveling waves in metal. This relation is plotted in Fig. 4.1 for an arbitrary metal, which can be modeled via the free electron gas model. One can see, that for frequencies below the plasma frequency ω_p no propagation of electromagnetic waves is allowed. However, for frequencies above the plasma frequency a transverse wave is able to propagate through the metal, with the group velocity

$$v_g = \frac{d\omega}{dk}, \tag{4.3}$$

where $k = |\boldsymbol{k}|$ is the corresponding wave number. For the frequency $\omega = \omega_p$ the group velocity of the wave would be zero, i.e. this wave is not traveling. This can be understood as a collective longitudinal oscillation of the conduction electrons (which have been treated as free electrons) versus the background consisting of the ion cores and the bound electrons. This leads to the so called bulk plasmons.

In order to describe SPP, one usually starts with the wave equation in (2.6), for the case, that all external stimuli are absent, i.e. $\nabla \cdot \boldsymbol{D}(\boldsymbol{r}, t) = 0$. Using the identities

$$\nabla \times \nabla \times \boldsymbol{E}(\boldsymbol{r}, t) \equiv \nabla(\nabla \cdot \boldsymbol{E}(\boldsymbol{r}, t)) - \nabla^2 \boldsymbol{E}(\boldsymbol{r}, t), \tag{4.4}$$

$$\nabla \cdot (\varepsilon_r(\boldsymbol{r}) \boldsymbol{E}(\boldsymbol{r}, t)) \equiv \boldsymbol{E}(\boldsymbol{r}, t) \nabla \varepsilon_r(\boldsymbol{r}) + \varepsilon_r(\boldsymbol{r}) \nabla \cdot \boldsymbol{E}(\boldsymbol{r}, t), \tag{4.5}$$

the wave equation can be rewritten as

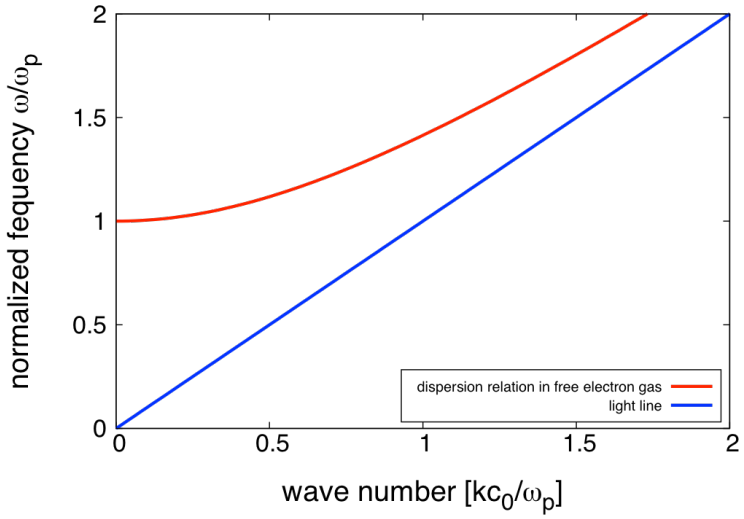


Fig. 4.1.: *Dispersion relation for a free electron gas (red) as given in eq.* (4.2) *compared to the dispersion relation in free space (blue). One can see, that an electromagnetic wave can only propagate, if* $\omega \geq \omega_p$*, though for the frequency* $\omega = \omega_p$ *the propagation constant of the wave would be zero.*

$$\nabla\left(-\frac{1}{\varepsilon_r(\boldsymbol{r})}\boldsymbol{E}_r(\boldsymbol{r},t)\cdot\nabla\varepsilon_r(\boldsymbol{r})\right)-\nabla^2\boldsymbol{E}(\boldsymbol{r},t)=-\mu_0\varepsilon_0\varepsilon_r(\boldsymbol{r})\frac{\partial^2}{\partial t^2}\boldsymbol{E}(\boldsymbol{r},\boldsymbol{t}). \tag{4.6}$$

If the material distribution is homogeneous, i.e. $\nabla\varepsilon_r(\boldsymbol{r}) = \boldsymbol{0}$, it simplifies to equation (4.7).

$$\nabla^2\boldsymbol{E}(\boldsymbol{r},t)=\mu_0\varepsilon_0\varepsilon_r(\boldsymbol{r})\frac{\partial^2}{\partial t^2}\boldsymbol{E}(\boldsymbol{r},\boldsymbol{t}) \tag{4.7}$$

This equation then has to be solved in each region with constant ε_r separately, and the solutions have to be matched via appropriate boundary conditions. Assuming now an electric field with harmonic time dependence, one obtains the so called Helmholtz equation. Analogous to the wave equation for the electric field, another one for the magnetic field can be derived.

For the geometry of an infinite metal plate, covered by an arbitrary insulator (e.g. air), the Helmholtz equation will result in a system of two sets of self-consistent solutions with different polarization properties, the so-called transversal magnetic (TM) and transversal electric (TE) waves, which are independent of each other. With an appropriate ansatz, it can be shown, that indeed no TE waves exist for surface plasmons.

The wave equation for a TM wave propagating in x-direction is then:

$$\frac{\partial^2}{\partial z^2} H_y + \left(k_0^2 \varepsilon_r - \beta^2\right) H_y = 0, \tag{4.8}$$

with the propagation constant $\beta = k_x$. For a surface plasmon polariton propagating at a single interface between a metal and a dielectric material, the dispersion relation can directly be calculated as

$$\beta = k_0 \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}}, \tag{4.9}$$

with the relative permittivity of the metal ε_m and the dielectric ε_d. For such traveling SPPs one needs a metal-dielectric interface. They are damped, so a propagation length $L_x = (2\Im(\beta))^{-1}$ has been established, typically between 10 and 100 μm in the visible regime. Since the SPP fields perpendicular to the interface fall off in the dielectric exponentially, i.e. $\sim \exp(-|k_z||z|)$, one can calculate the penetration depth, which is in this case $L_z = 1/k_z$, with

$$k_z = \sqrt{\beta^2 - \varepsilon_d \left(\frac{\omega}{c_0}\right)^2}. \tag{4.10}$$

For a gold-air interface for $\lambda = 800nm$ the penetration depth reaches around $L_z \approx 570nm$, which is important for simulations regarding the computational domain, which has to be calculated. The wavelength of such a surface plasmon is then

$$\lambda_{SPP} = \frac{2\pi}{\Re(\beta)}, \tag{4.11}$$

which is usually much smaller than the wavelength in free space. All those characteristic lengths are depending on the used material system. SPPs propagating along a flat interface are basically two-dimensional waves. Since their dispersion relation differs to the light line, the excitation by three-dimensional light beams is not possible, unless special techniques for optical phase matching are used, such as prism or grating coupling. More relevant in this work are the so called localized surface plasmon polaritons, which are non-propagating excitations of the conduction electrons of metallic particles, which are smaller than the wavelength.

In chapter 2.4 it has been shown, that for typical metals the skin depth in the optical regime can reach $\delta \approx 100nm$. If the particle has smaller dimensions d than the skin

depth in propagation direction of the incident field, one can excite an electric field inside of this sub-wavelength particle. If the wavelength is much larger than the particle size, the phase of the electric field within the particle can be assumed to be constant, i.e. all electrons within the particle are subject to the same external force given by the electric field. Thus the electrons collectively oscillate driven by the external field and a localized SPP is excited. The region with the relation of $d < \delta$ and $d \ll \lambda$, is usually called *electrostatic limit*.

Using the electrostatic potential ϕ and solving the Laplace equation, one can derive an expression for the scalar potential for a sphere, which is the most simple geometry [39]. The sphere is excited by a plane wave, and due to the symmetry the azimuthal angle can be neglected. The resulting scalar potential $\phi_{out},\ \forall r > a$ outside the sphere is then:

$$\phi_{out} = -E_0 r \cos\theta + \frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m + 2\varepsilon_d} E_0 a^3 \frac{\cos\theta}{r^2}, \tag{4.12}$$

with the radius of the sphere a and the polar angle θ. Introducing the dipole moment $\boldsymbol{p}$ and using the polarizability from equation (2.88), one obtains

$$\phi_{out} = -E_0 r \cos\theta + \frac{\boldsymbol{p} \cdot \boldsymbol{r}}{4\pi\varepsilon_0\varepsilon_d r^3}, \tag{4.13}$$

with the dipole moment and polarizability

$$\boldsymbol{p} = 4\pi\varepsilon_0\varepsilon_d a^3 \frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m + 2\varepsilon_d} \boldsymbol{E}_0 \tag{4.14}$$

$$= \varepsilon_0\varepsilon_d \alpha \boldsymbol{E}_0 \tag{4.15}$$

$$\alpha = 4\pi a^3 \frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m + 2\varepsilon_d}. \tag{4.16}$$

With the scalar potential one can calculate the distribution of the electric field:

$$\boldsymbol{E}_{out} = \boldsymbol{E}_0 + \frac{3\boldsymbol{n}(\boldsymbol{n} \cdot \boldsymbol{p}) - \boldsymbol{p}}{4\pi\varepsilon_0\varepsilon_d} \frac{1}{r^3} \tag{4.17}$$

In addition to the excitation with the strength $\boldsymbol{E}_0$, one gets an (dipolar) enhancement of the electric field around the particle. The dipole moment is basically the polarizability with some prefactors, so all of its properties are projected into the field enhancement. Since the polarizability strongly depends on the geometry and volume

of the particle (see chapter 2.5.2), the geometrical shape plays an important role for the desired performance of such particles.

So far only the quasi-static behavior of such particles has been discussed. If the particle gets excited by an oscillating electric field, i.e. $\boldsymbol{E}(\boldsymbol{r},t) = \boldsymbol{E}_0(\boldsymbol{r})e^{-i\omega t}$, the dipole moment oscillates in phase with the exciting field. Thus the second term in equation (4.17) leads to a dipole-like radiation. Since the focus of this chapter lies on the field enhancement of metal nano particles close to it's surface, it is sufficient to discuss the quasi-static regime. Further details about the radiation of such induced dipoles can be found in [39].

Since fabricated structures usually show non-perfect geometries, with deformed shapes and a surface roughness, the electrostatic potential can most often not be described analytically. Apart from these difficulties, the derivation based on the quasi-static limit is only valid, if the particle is significantly smaller than the wavelength. Numerical simulations show, that for other particle geometries one might obtain not only a dipole moment, but also higher order poles like a quadrupole moment.

4.2. Intensity enhancement of a one-particle nanoantenna

An important statement of the previous section was, that one can achieve an enhancement of the electric field close to the surface of metal nano particles, if they are excited resonantly. In this section this is investigated in more detail and some properties of such a particle are discussed. Starting point is a single rod antenna consisting of gold, which is probably one of the most often fabricated nano structures, shown in Figure 4.2.

The nano rod consisting of gold (for parameters see table A.1) has a length of $L = 160nm$, a width of $w = 40nm$, and a height of $h = 40nm$. Since sharp edges are hard to fabricate and cause some numerical difficulties, the endings of the brick have been blended with a radius of $r_{vert} = 20nm$. Additionally the top edges have been blended with a radius of $r_{top} = 10nm$. This structure has been excited by a plane wave propagating in z-direction, with a polarization vector parallel to the y-axis. A field probe has been placed close to the surface of the rod, with a distance of only $5nm$. The resulting field enhancement is shown in Fig. 4.3.

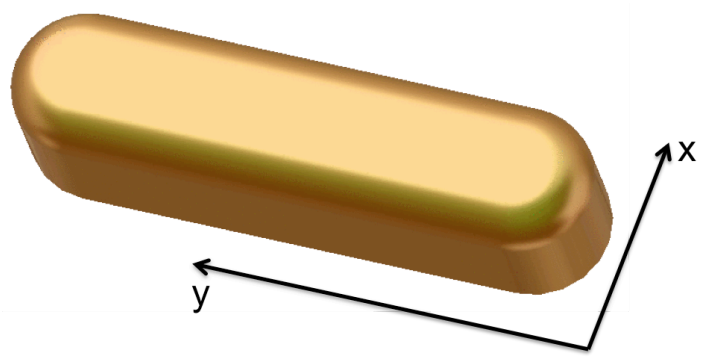


Fig. 4.2.: *Investigated geometrical shape of the rod with the definition of the corresponding coordinate system.*

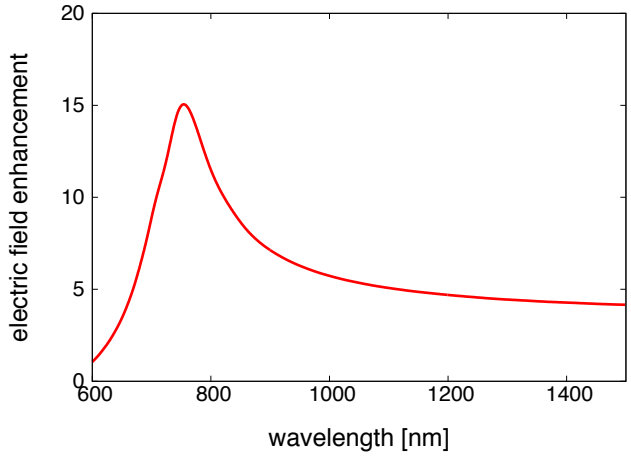
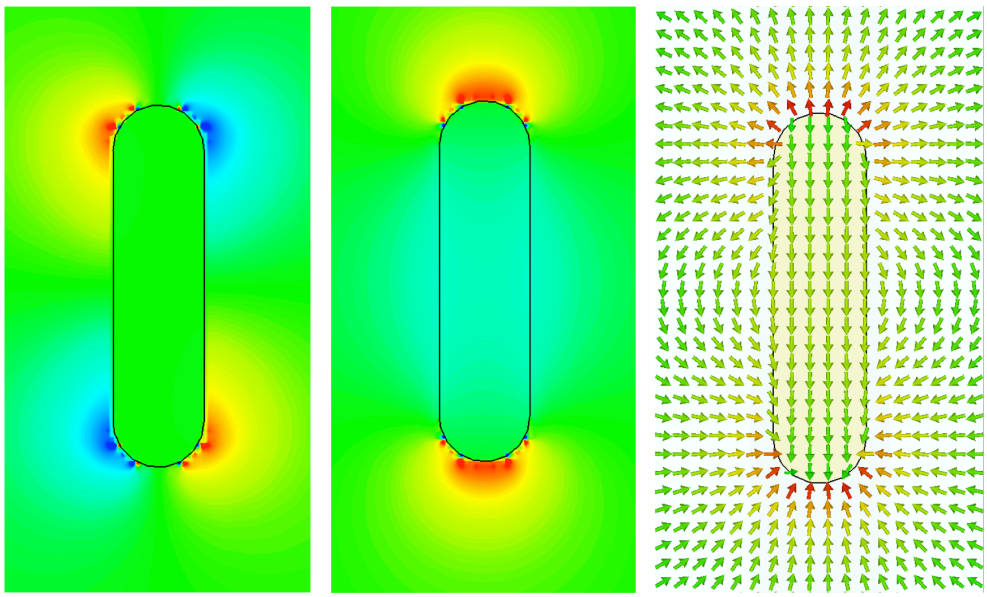


Fig. 4.3.: *Electric field enhancement close to the surface of the rod in dependence of the incident wavelength.*

The maximum field enhancement in this spot is $E_{enh} = 15$ with an excitation wavelength of $\lambda_{res} = 755nm$. As is usual for plasmonic antennas, the peak is spectrally broad, caused mainly by the dipolar emission into the far field.

In order to get a closer look at the plasmonic mode within this metallic rod, the electric field enhancement has been plotted for a cross section in the $z = 20nm$ plane (Fig. 4.4). For the case of an infinite, constant excitation with the wavelength of $\lambda = 755nm$, both the components perpendicular to the propagation direction of the incident wave and the vectorial depiction of the electric field enhancement are shown.

With a time domain solver such electric field plots are obtained by taking the Fourier transform of the time signal in each cell and picking the Fourier component of the



(a) x-component of the electric field enhancement.

(b) y-component of the electric field enhancement.

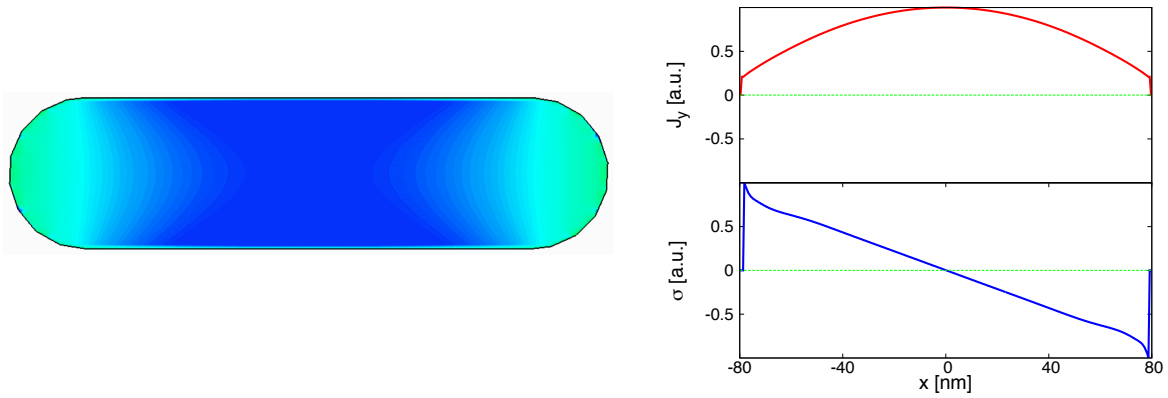
(c) electric field enhancement indicated by arrows.

Fig. 4.4.: *Electric field enhancement of a single rod antenna, excited by a plane wave polarized along the antenna axis. Within the cross section at the $z = 0$ plane, in a) the x-component, b) the y-component and in c) 2D vectors of the electric field enhancement is shown. The color scale for all three figures is the same.*

desired frequency. Thus one obtains a complex field value for each component of the electric field. It is convenient to use the polar form of the complex values, where one has the amplitude and the phase for each component of the field.

One can clearly see the dipolar characteristic of the electric field outside of the nano rod in Fig. 4.4 c), with the main direction of the electric field parallel to the y-axis. However, inside of the particle the electric field points into the opposite direction, parallel to the $-y$-axis.

As has been shown in chapter 2.4, one can also have a look to the current density. Using the continuity equation eq. (2.8) one can also calculate the charge carrier density. In Figure 4.5 the current density within the nano rod is shown. It is not surprising, that for a phase ϕ_{max}, where the electric field reaches its maximum, the absolute value of the current density is also maximal, but with opposite sign. The maximum is located exactly in the center of the nano rod. In contrast to the electric field, the x-component of the current density is significantly smaller than the y-component, but always points inwards.



(a) (b)

Fig. 4.5.: *a) Cross section of the current density J_y within the rod antenna, excited by a plane wave with polarization along the antenna axis in the $z = 20nm$ plane. Since the current density along the other axes are one order of magnitude smaller, their influence can be neglected. b) The current density along a line parallel to the antenna axis in the middle of the particle. Below the calculated charge carrier density is shown along the same line.*

Another important class of geometries is the so called ellipsoid, or more specifically the spheroid. In order to compare another geometry to the rod, in the following a prolate spheroid with similar length is discussed. The spheroid consists of gold, and the radii are $r_x = 20nm$, $r_y = 80nm$, and $r_z = 20nm$ (Fig. 4.6). The spheroid is more difficult to fabricate, but the dipole moment can be calculated analytically. Since the endings appear to be sharper compared to the rod, a higher field enhancement is expected.

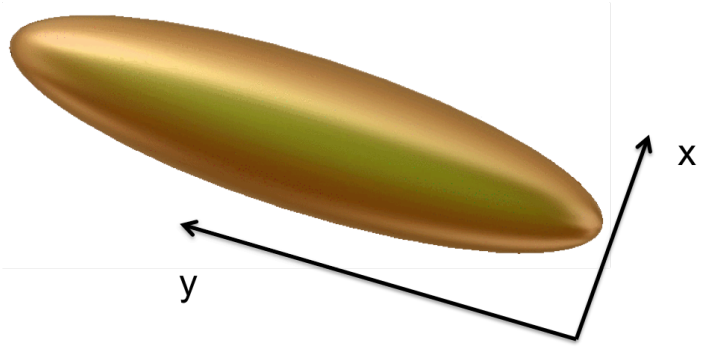


Fig. 4.6.: *Investigated geometrical shape of the rod with the definition of the corresponding coordinate system.*

In order to compare both structures, the same simulation (and also the distance between field probe and surface of the particle) has been performed. In general the

electric field distribution looks similar, apart from minor changes due to the different geometry. The spectra of both antennas are shown in Fig. 4.7. First of all, the electric field enhancement is almost twice as high than for a rod, which of course can partially attributed to the sharper ending.

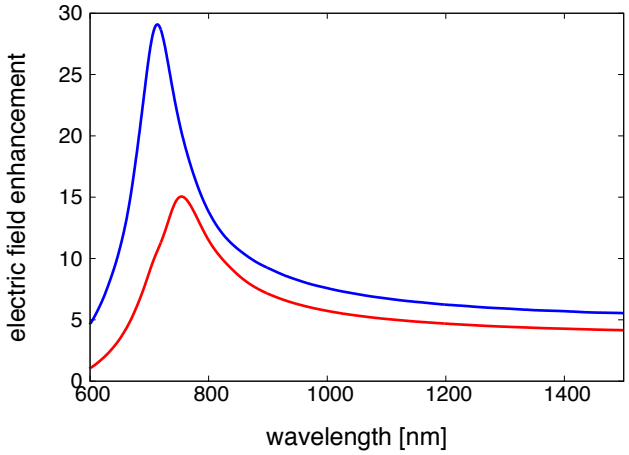


Fig. 4.7.: *Electric field enhancement close to the surface of the spheroid (blue) and rod (red) in dependence of the incident wavelength.*

Interestingly, the resonance of the spheroid is blue shifted compared to the rod, though both particles have the same length. With increasing volume of a particle with fixed length, e.g. if it becomes broader, the main plasmon resonance experiences a slight blue shift. Since the rod has a larger volume than the spheroid, one would have expected a red shift, which is not the case here. This can be attributed to the different geometry, for the resonance of the polarizability strongly depends on both the geometrical shape of the particle and it's volume.

An important feature of such antennas is the spatial distribution of the field enhancement, because for nonlinear applications high fields have to be generated. On the one hand it is advantageous, to have as small a field enhanced spot as possible, on the other hand that would make the positioning of the nano emitter more difficult. In Fig. 4.8 the field enhancement is shown in dependence of the distance to the surface of the particle.

The red curve shows the field enhancement calculated with the coupled dipole approximation. In this model it is assumed, the particle behaves like a dipole, i.e. in the near field the electric field is proportional to $1/r^3$. The FIT simulations otherwise show,

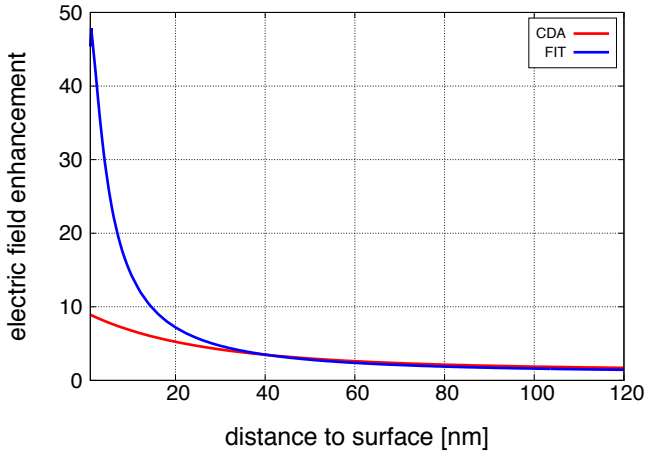


Fig. 4.8.: *Electric field enhancement in dependence of the distance to the surface of the particle. The blue curve shows the simulation of the full Maxwell equations, the red one the simplified coupled dipole model.*

that additionally to the dipole radiation the evanescent field of the plasmonic mode contribute to the field enhancement. As discussed in chapter 4.1, such evanescent fields can penetrate far into the surrounding space, but are still bound to the particle. The radiated field is a perfectly dipolar, as can be seen in Fig. 4.9, thus the name for such antennas.

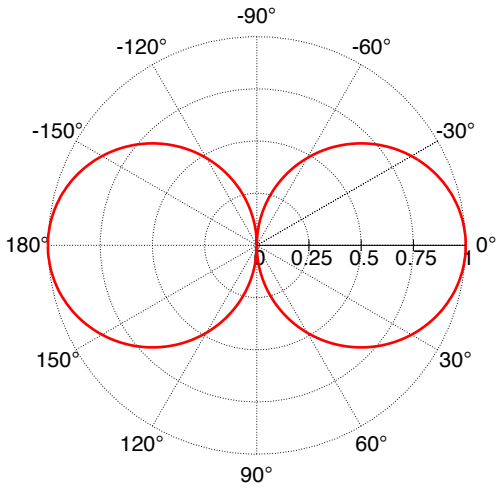


Fig. 4.9.: *Radiated far field pattern for the spheroid. Since the radiation is constant for all polar angles θ, only the azimuthal angle ϕ is shown here. The magnitude of the radiated field is normalized to 1.*

Since the magnetic field in the near field region is significantly smaller than the electric field, it is sufficient to characterize the electric field only. Many applications, in which such nano antennas can be used, depend on the intensity of a light field [71], which is in the near field region[1] for high frequencies (such as NIR or optical)

$$I = \frac{c_0 n \varepsilon_0}{2} |\boldsymbol{E}|^2. \tag{4.18}$$

The intensity enhancement is then, with the electric field in the region of interest $\boldsymbol{E}_{ROI}$,

$$I_{enh} = \left| \frac{\boldsymbol{E}_{ROI}}{\boldsymbol{E}_0} \right|^2. \tag{4.19}$$

Thus in the following sections only the intensity enhancement is investigated.

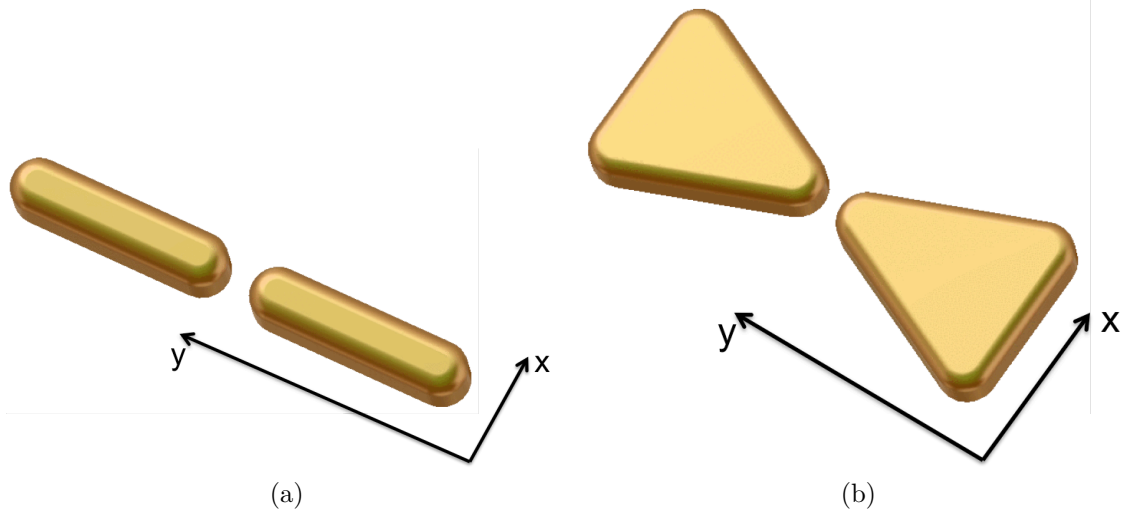
4.3. Intensity enhancement of a two-particle nanoantenna

Placing two or more particles close to each other leads to an increased intensity enhancement within the gap of such an antenna, if they are properly aligned. In this section two of the most prominent geometries, the rod and bow-tie antenna, and some of their properties will be discussed.

4.3.1. Single nano rod and Bow-Tie antennas

In this section the main properties of dipole antennas consisting of two rods and two triangles are discussed. In Fig. 4.10 a) a standard rod antenna with the corresponding coordinates is shown. It consists of two rods like in the previous section, though with a length of each rod of $L_y = 130nm$. The gap between those rods has been chosen to $L_{gap} = 16nm$, because it is too large for quantum mechanical effects like the tunnel-effect to occur and still yields a high intensity enhancement. In Fig. 4.10 b) a so called

[1] The space around an antenna is usually divided into different regions: the near field region, which is very close to the antenna, the far field region, and an intermediate region, which has both characteristics. The bounds of such regions can be defined in very different ways. Most often one uses the emission pattern of a dipole. In the near field the term proportional to $1/r^3$ dominates, which is definitely the case for distances smaller than the half of the wavelength. In the far field region the term proportional to $1/r$ dominates, for distances greater than 2 wavelengths. In the intermediate region both terms yield a significant contribution to the electric field.



(a) (b)

Fig. 4.10.: *Sketch of the geometrical shape of a) a rod and b) a bow-tie antenna, with a gap of* $16nm$.

bow-tie antenna is shown, with the tips pointing to each other. Each arm consists of an isosceles triangle with a tip angle of $\phi = 45°$ and a length of $L_y = 145nm$, with a gap of the same size as for the rod antenna in between. Since fabrication of such structures never leads to sharp edges, all vertical edges have been blended with a radius of $r_{vert} = 20nm$, and all horizontal edges on the top of the structure with a radius of $r_{top} = 10nm$. The height of both antennas has been $L_z = 28nm$. The surrounding material has been vacuum. All parameters are summarized in A.3.

These antennas have been excited via a plane wave propagating along the z-direction, with a polarization parallel to the y-axis. Again a transient solver has been used, for to obtain the complete spectrum only one simulation is needed. For that the incident plane wave has been chosen to be of the form of a Gaussian pulse in time domain. The spectrum of the point in the middle of the gap has then been divided by the spectrum of the incident field. In Fig. 4.11 the spectrally resolved intensity enhancement of both antennas can bee seen. They both perform similarly, with the bow-tie yielding a slightly higher intensity enhancement. The parameters have been chosen such that the main resonance peak appears at a vacuum wavelength of $\lambda_{res} = 800nm$. The resonance of such antennas usually exhibits spectrally broad peaks and therefore low Q-factors, which is not due to ohmic losses, caused by collisions of the electrons within the material.

Since each particle can be considered a resonator, one can qualitatively describe such an antenna by a coupled resonator model (often referred to as mass and spring model

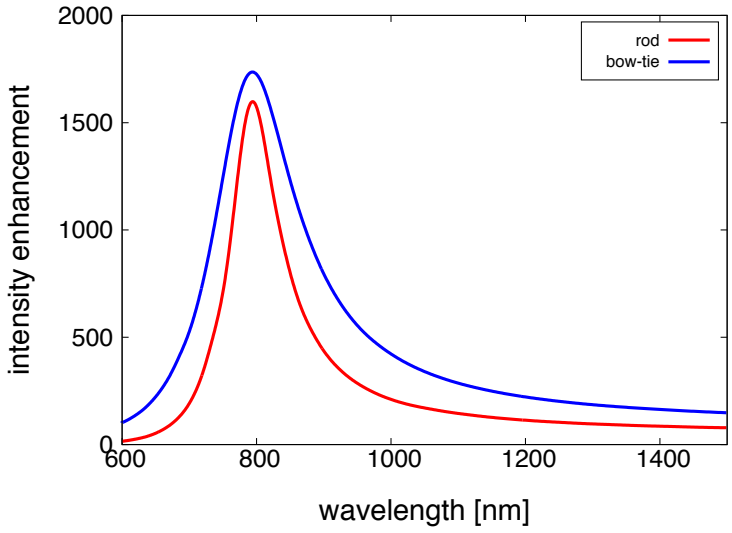
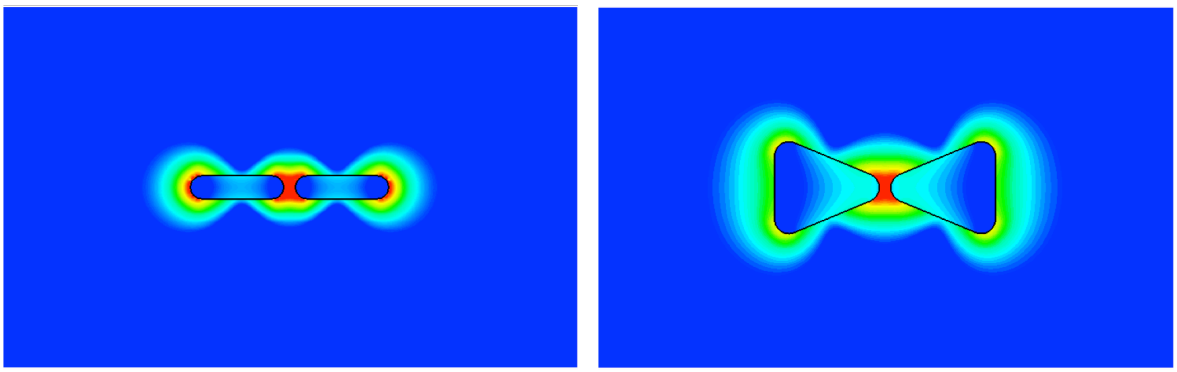


Fig. 4.11.: *Spectrally resolved intensity enhancement for a rod and a bow-tie antenna.*

[57]). When arguing with this simple model it becomes apparent, that two different modes appear due to the mode splitting - the so called bonding mode at lower and the anti-bonding mode at higher frequencies. The resonance depicted in Fig. 4.11 of both antennas correspond to the bonding mode. Since it consists of dipole-like charge oscillations, it can easily be excited by a plane wave illumination polarized along the antenna axis. If the impedance of such antennas is around the vacuum impedance, the excited plasmon leads to an effective emission of dipolar radiation, reducing the energy within this antenna to zero within femto seconds. As a result the resonance of this mode becomes broad and very bright in spectroscopic measurements. In Fig. 4.12 the normalized electric field enhancement for the main resonance has been plotted. The color scale is in dB and reaches from -40 (blue) to 0 (red).

Slightly outside of the range of the spectrum in Fig. 4.11 at a wavelength of $\lambda \approx 550nm$ another peak arises, which correspond to the anti-bonding mode. It is characterized by a charge distribution which exhibits mirror symmetry, with the mirror plane in the middle of the gap. This mode doesn't emit into the far field as efficiently as the bonding mode, because one gets the radiation of two dipoles, oscillating out of phase and therefore interfering destructively. As a consequence it is much more difficult to excite such a mode, as long as the illumination path remains fully symmetric [57]. Therefor when illuminating with a plane wave, the field enhancement of the anti-bonding mode is much smaller than that of the bonding mode. The anti-bonding



(a) (b)

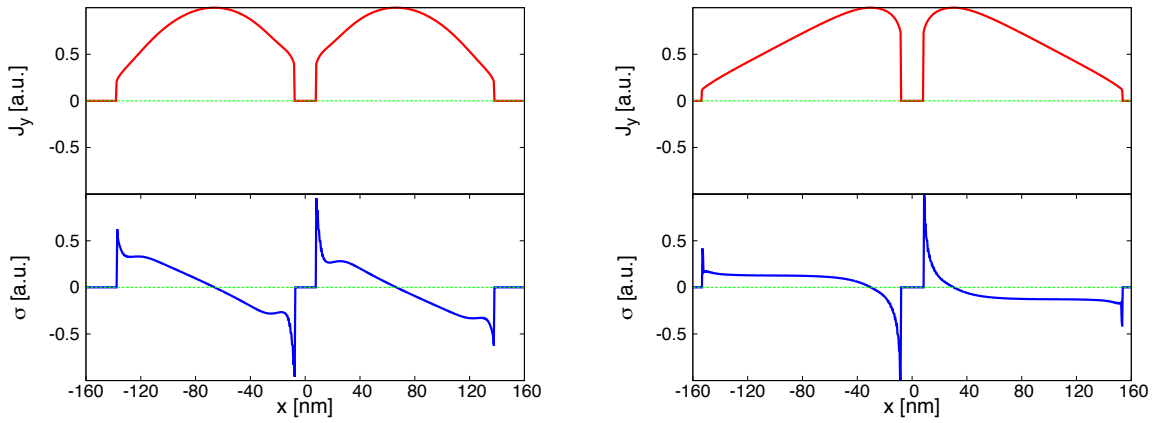
Fig. 4.12.: *Normalized electric field enhancement of a) a rod and b) a bow-tie antenna at the resonance frequency of the bright mode. The color scale is in dB and reaches from −40 dB (blue) to 0 dB (red).*

mode is designated as dark mode.

Compared to the resonance of a single particle, the bonding mode is red-shifted[2], which cannot be explained by the mass and spring model [72]. But the inter-particle coupling can also be described by a hybridization model, which relates to molecular orbital theory, where the overlap of wave functions is taken into account to derive the energy splitting [73]. In general one can state for ensembles of plasmonic particles, that whenever modes are spectrally and spatially overlapping, this coupling results in mode splitting including resonance shifts, analogous to atomic orbital hybridization [57].

For comparison in Fig. 4.13 a) the normalized current density of the bright mode has been plotted for the rod along a line parallel to the antenna axis. Below the corresponding charge carrier density has been calculated using the continuity equation in eq. (2.8). One can see, that both particles exhibit qualitatively the same behavior of the single particle rod in Fig. 4.5, though the maxima are shifted towards the gap. Therefor the difference of charge carriers at both sides of the gap is higher, which results in a stronger intensity enhancement. In Fig. 4.13 b) both, the current and charge carrier density have been plotted for the bow-tie. On the one hand, the maximum of the current density is much closer to the gap compared to the rod antenna, on the other hand the absolute value of the current density is smaller, so that the difference of the charge carrier density between both antennas is comparable. Thus the intensity

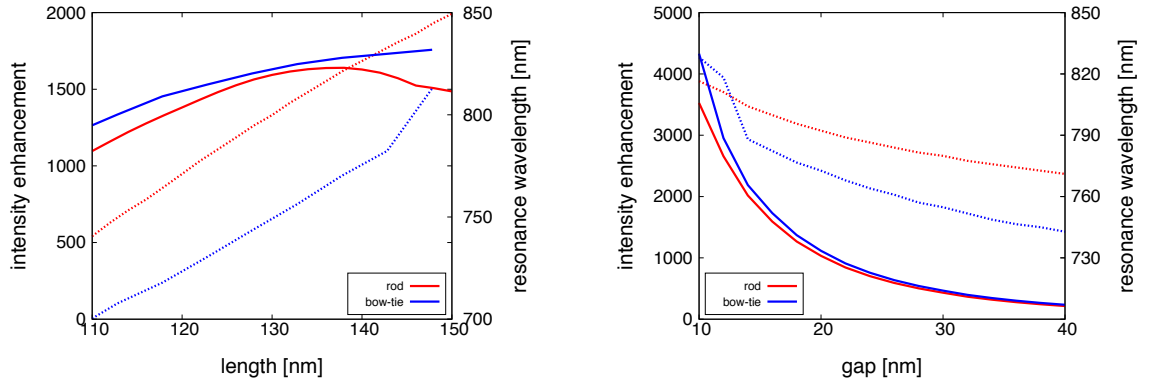
[2]For comparison the resonances of the two-particle antennas should be around the frequency of the resonance of the one-particle antenna. In order to compensate the red shift, the lengths of the antenna arms have been reduced.



(a) (b)

Fig. 4.13.: *Normalized current density and the resulting charge carrier density along a line through the center of a) the rod and b) the bow-tie antenna, parallel to the antenna axis.*

enhancement of both, bow-tie and the rod antenna, are close to each other.



(a) (b)

Fig. 4.14.: *Maximum intensity enhancement (solid line) and the corresponding resonance wavelength (dashed line) of a rod and a bow-tie antenna in dependence of a) the length of each antenna arm and b) the size of the gap in between the two antenna arms.*

The dependence of the maximum intensity enhancement and the corresponding resonance wavelength to the length of the antenna is depicted in Fig. 4.14 a). There all parameters have remained fixed except the length of each antenna arm. As predicted by Novotny (Ref. [16]) the resonance wavelength grows linearly for increasing length

of each antenna arm. At a length of around $L_y \approx 145nm$ the resonance of the bow-tie grows stronger, which is due to a deformed resonance peak. However, the maximum intensity enhancement is not increasing linearly as one would have expected. The reason is, that while the amount of material, which "collects" light is increasing linearly, the ohmic losses within the structure are increasing nonlinearly, since the permittivity is a nonlinear function with respect to the frequency.

In Fig. 4.14 b) one can see the maximum intensity enhancement and the corresponding wavelength in dependence of the gap between the two particles. While the resonance shifts only slightly compared to the spectral width of the peak, the intensity enhancement changes dramatically, since the middle of the gap is closer to the surface of both particles and thus more influenced by the evanescent field of the localized surface plasmon. If a high intensity enhancement is needed, one needs the smallest possible gap, in which another particle can be positioned. If the gap becomes smaller than $2nm$, the intensity enhancement cannot be calculated by classical methods alone, since at least the tunnel effect, where the electrons may tunnel from one particle to another, has to be considered. This might result in a dramatical decrease of the performance of such an antenna [74]. Apart from that, placing e.g. a quantum dot into the gap of an antenna will cause a shift of the resonance, which will be discussed later.

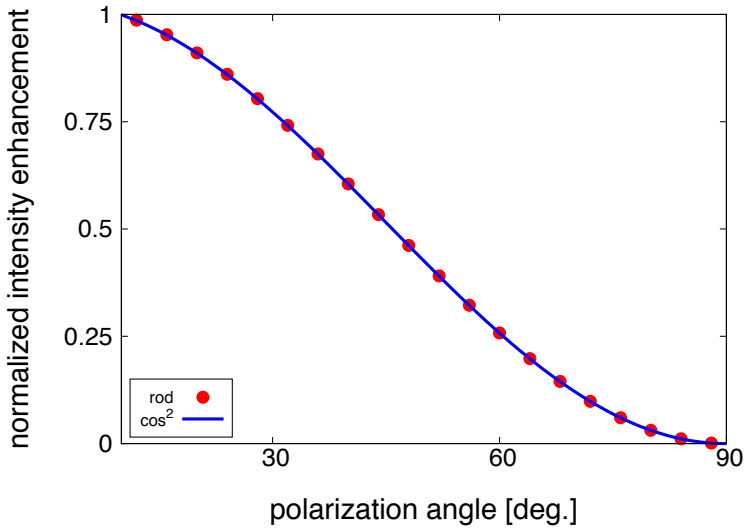


Fig. 4.15.: *Normalized intensity enhancement (red) in dependence of the polarization angle ϕ of the incident light field for a rod and bow-tie, which are identical. The blue line indicates a* $\cos(\phi)^2$ *fit.*

This section is closed with a short look at the polarization of the incident light field. For this the incident light field has the form of

$$\boldsymbol{E}(\boldsymbol{r},t) = \begin{pmatrix} \sin(\phi) \\ \cos(\phi) \\ 0 \end{pmatrix} E_0 e^{k_z z - \omega t}, \tag{4.20}$$

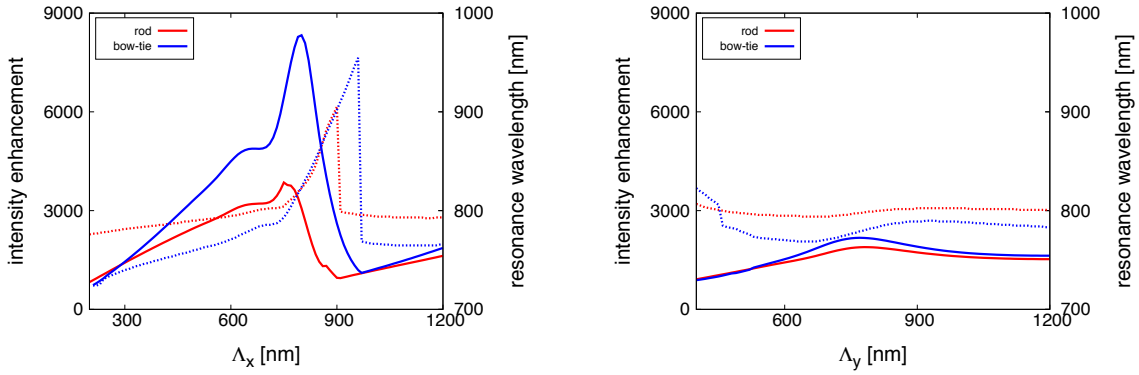
with the polarization angle ϕ. In Fig. 4.15 the normalized intensity enhancement in dependence of the polarization angle is shown (red dots), which is identical for both antennas. Additionally the function $f(\phi) = \cos^2(\phi)$ (blue) is plotted, which corresponds to the y-component of the electric field. It might be surprising that all three curves are identical, since the bow-tie consists of two triangles, so it should be more robust regarding the polarization of the incident light field. It is actually not, since both antennas exhibit a mirror symmetry regarding the $x = 0$ mirror plane. Thus for the excitation of the whole plasmonic mode only the y-component of the electric field is responsible. This can be further affirmed, by looking at the resulting mode in a field plot, in which only the strength, not the shape of the mode is changed with the polarization angle.

4.3.2. Nano antenna arrays

So far only an isolated dimer antenna has been discussed. Since these antennas emit dipolar radiation, it is expected, that an array of antennas will strongly couple, changing the intensity enhancement and shifting the resonance frequency. Some properties of antenna arrays will be discussed in this section.

Starting point is a single two-particle antenna with the same parameters as in the previous section. In order to simulate an infinite array of antennas, periodic boundary conditions (PBC) for the two opposite sides of the computational domain have been used. In Fig. 4.16 a) PBC have been used only for the x-direction, which is perpendicular to the antenna axis and the propagation direction of the incident light field and in b) only for the y-axis. This corresponds to a single line of antennas with infinite length.

An infinite line of antennas parallel to the antenna axis, and therefore parallel to the dipole moment of the antenna, causes a difference for the intensity enhancement only in the region, where the periodicity is smaller than $\Lambda_y < 1000nm$. While a small periodicity, compared to the wavelength of the incident light, leads to destructive interference,



(a) (b)

Fig. 4.16.: *Maximum intensity enhancement and the corresponding resonance wavelength of a rod and a bow-tie antenna in dependence of the periodicity in a) x-direction, which is perpendicular of the antenna axis and the propagation direction of the incident light field, and b) y-direction, which is parallel to the antenna axis.*

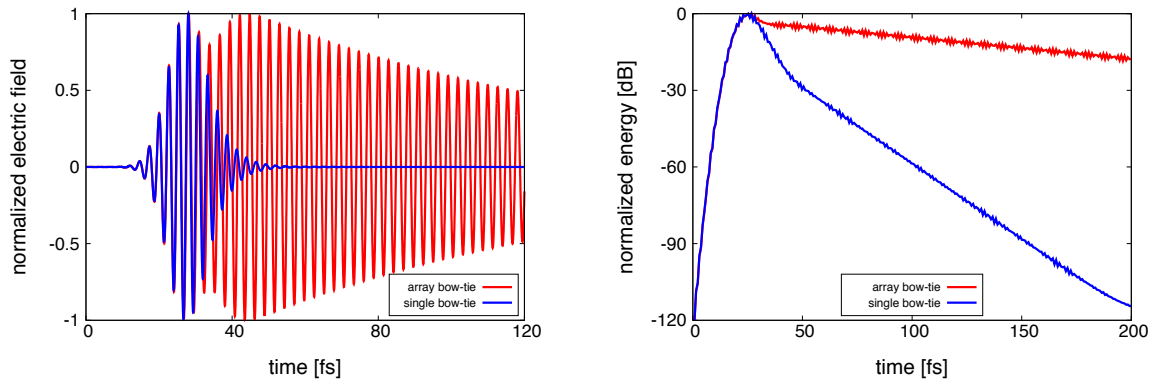
which results in a lower intensity enhancement compared to a single antenna, a periodicity around the wavelength results in an increased intensity enhancement. With increasing periodicity then the behavior matches that of a single antenna. Both the change of the intensity enhancement and the corresponding resonance wavelength are not that significant.

When varying the periodicity in x-direction, which is perpendicular to the antenna axis and the propagation direction of the incident light, both the maximum intensity enhancement and the corresponding resonance wavelength change dramatically. In Fig. 4.16 a) one can observe a strong increase of the intensity enhancement for a periodicity $\Lambda_x < 850nm$, compared to the single antenna. For higher period length the intensity enhancement first drops to a value of $I_{enh} \approx 1000$, an then asymptotically goes up to the value of a single antenna. One can also see, that the bow-tie performs much better than the rod, reaching an intensity enhancement of around $I_{enh} \approx 8300$ (compared to the rod, whose maximum is $I_{enh} \approx 3800$).

Responsible for this behavior are two effects. First, the antennas radiate like dipoles, i.e. the small gaps of an array of such antennas can be seen as a grid of dipole sources. Since it takes the radiation only around $3fs$ to cover the distance to the next antenna, it can interfere with the incident pulse. If the distance of these gaps is roughly around the wavelength, the radiation of the next neighbors is in phase with the incident light

field (both are polarized parallel to the y-axis at these points), and therefore interfere constructively. So some of the radiated energy is reabsorbed by neighboring antennas, which adds to the intensity enhancement. Such a behavior is known for other resonant dipole emitters like quantum dots [75], but, to my knowledge, has not been reported for antennas.

The second effect is, that such an array of antennas can be seen as many coupled resonators, which will exhibit a collective mode. This usually leads to a mode splitting as discussed in section 4.3. Since many particles are involved in this mode, it is much more complicated to describe than the mode splitting of just two particles, but one would expect a redshift of the resonance, which can be clearly seen in Fig. 4.16. Additionally, much of the energy is then stored between the antennas, so the damping becomes small and the radiation insignificant. This can be further affirmed by looking more closely to the time dependence of such a mode. For this a short Gaussian pulse has been used for the incident wave.



(a) (b)

Fig. 4.17.: *a) Calculated normalized electric field in dependence of time, taken in the middle of the gap of an array of bow-tie antennas with lattice constants $\Lambda_x = 780nm$ and $\Lambda_y = 770nm$ (red) and a single bow-tie antenna (blue). b) The corresponding electromagnetic energy in the whole computational domain for both, an array (red) and a single bow-tie antenna (blue).*

In Fig. 4.17 a) one can see the normalized electric field, taken in the middle of the gap of the antenna, for both, a single (blue) and the array (red) of bow-ties. The field for the single antenna has almost the same form as the incident light field with only short oscillations afterwards. This antenna gets excited and almost immediately loses the energy, stored in the localized SPP, in terms of radiation. The damping of the material

is not that important here. In contrast to this behavior, though the electric field of the array of antennas starts similarly, one can see an additional rise of the electric field strength a few femto seconds after the first maximum. This corresponds to the time needed for light to travel between two antennas within this array. Afterwards the electric field strength is decreasing much slower, than is the case for a single antenna. This can be explained by a collective mode of the array of antennas, using the coupled resonator model. Much of the energy in this mode is stored between the antennas, which leads to decreased radiation losses.

That these calculations were indeed stable[3] can be seen in Fig. 4.17 b), where the total electromagnetic energy in the computational domain is plotted vs. time. One can see the increase of energy with the excitation pulse, followed by an exponential decrease afterwards, which is expected. One can also see that right after the energy maximum the energy of both, the single and the array of antennas decreases more rapidly, though it is barely visible for the array. The reason is, that in order to establish this mode some time is needed, in which each antenna radiates individually, until a collective oscillation is achieved. Thus the wavelength dependence of the antenna array cannot be calculated by dividing the obtained spectrum in the gap by the spectrum of the incident wave. For this a continuous wave, which increases linearly in strength at the beginning has been used. After the electric field within the gap has been settled, the amplitude divided by the amplitude of the incident wave has been picked for each wavelength.

The spectral performance of the antenna arrays are shown in Fig. 4.18, for comparison the single antennas are included. One can see that the bow-tie array achieves a much higher intensity enhancement than the array of rod antennas, for such the effect of a collective mode is much smaller compared to the bow-tie antenna array. Also the resonance is red-shifted, which again indicates a collective mode of coupled resonators (c.p. chapter 4.3).

One can also see, that due to the coupling of an array of antennas, the resonance peaks becomes much more narrow. This is can be affirmed by the determination of the quality factor (or short Q-factor), which can be estimated through

$$Q \approx \frac{\lambda_0}{\Delta\lambda_{FWHM}}, \tag{4.21}$$

[3] A convergence test has been made once in the beginning and then checked for every unexpected result. To increase the number of cells here had no effect on the performance of this antenna.

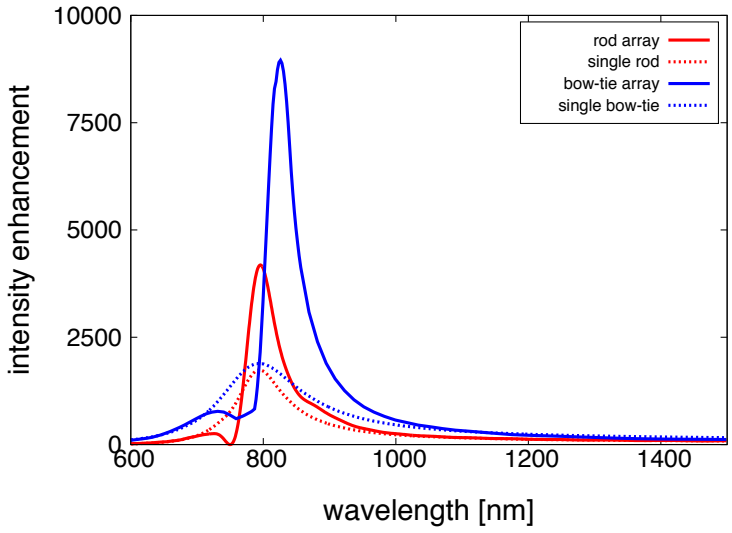


Fig. 4.18.: *Spectrally resolved intensity enhancement of a two-rod antenna (red) and a bow-tie (blue) isolated (dashed lines) compared to corresponding arrays (solid line).*

with the wavelength of the maximum of the peak λ_0 and the spectral width at half maximum $\Delta\lambda_{FWHM}$. In table 4.1 the estimated Q-factors for the rod antenna and bow-tie are shown. When comparing the Q-factors of the single rod antenna and the single bow-tie, one can deduce, that the bow-tie has a broader resonance peak. This is one of the said advantages of the bow-ties, because the two triangles represent the two-dimensional analog of a biconical antenna, which also leads to lower Q-factors. Another often said advantage of the bow-tie compared to the rod antenna is a higher intensity enhancement, because of a higher lightning-rod effect at the apex. However, while the resonance of the bow-tie is spectrally broader, compared to a rod antenna, a significant higher intensity enhancement for single antennas could not be observed as long as the sharp edges have been blended, which agrees with Ref. [57].

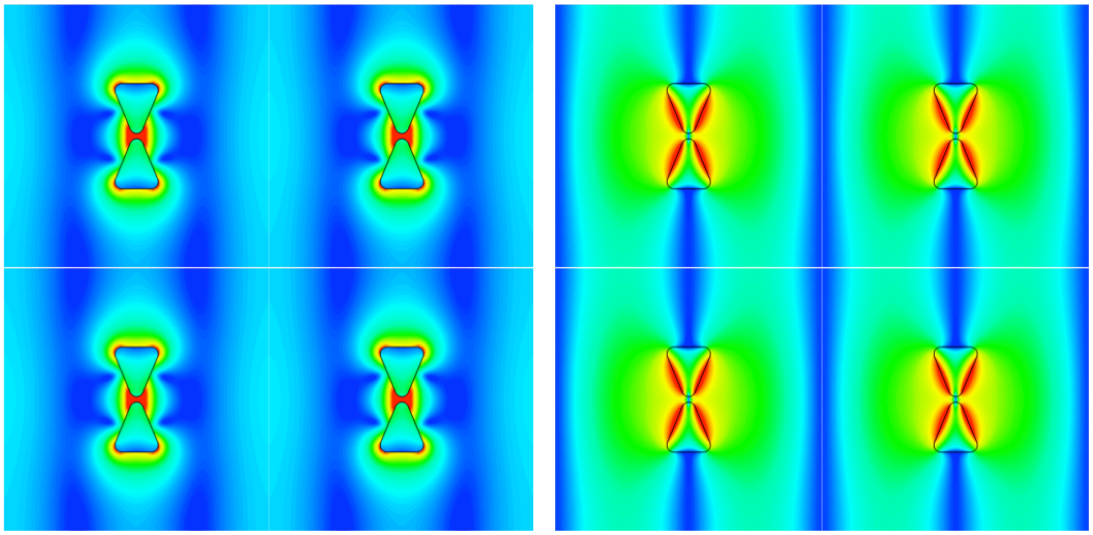
It turns out, that the Q-factors for the antenna arrays are significantly higher, i.e. narrower resonance peaks compared to the corresponding single antennas. And due to the coupled mode of the bow-tie array, its Q-factor is even higher, than the Q-factor of the rod antenna array.

In this configuration the bow-tie array reaches an intensity enhancement in the middle of the gap of $I_{enh} \backsimeq 9000$, which is more than twice the intensity enhancement of the rod antenna array. In Fig. 4.19 the absolute value of the normalized a) electric and b) magnetic field strength of this mode is shown. In contrast to the field of a single antenna

antenna	Q-factor
single rod antenna	8.3
rod antenna array	14.9
single bow-tie	4.9
bow-tie array	16.7

Tab. 4.1.: *Estimated Q-factors for the calculated antennas.*

in Fig. 4.12 b) a high field strength can be seen between the antennas, particularly the magnetic field. Thus the impact of the damping of the antenna material becomes less important for the performance.



(a) normalized electric field (b) normalized magnetic field

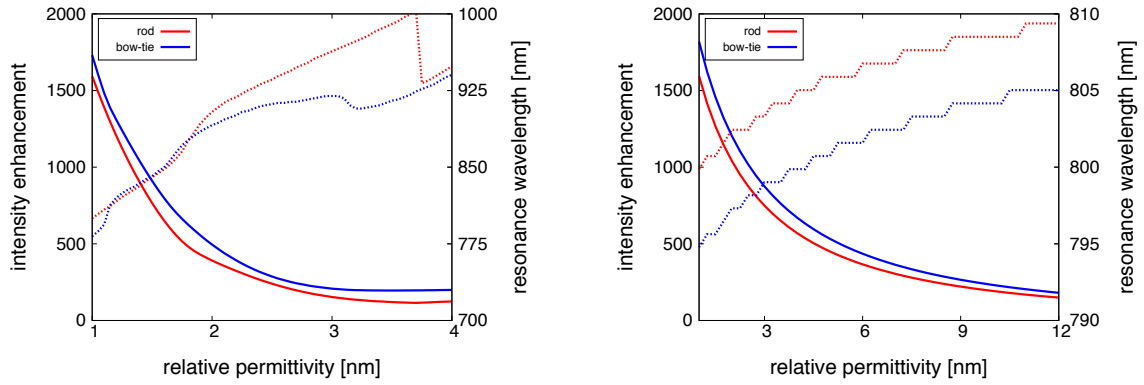
Fig. 4.19.: *a) Electric and b) magnetic field plot for the coupled mode of an array of nano antennas. The frequency of the incident wave is* $f = 363THz$*, which corresponds to the maximum in Fig. 4.16 a). The color scale is in dBi (so normalized) reaching from* -40 *dB (blue) to* 0 *dB (red).*

4.3.3. Influence of the substrate and surrounding particles

So far the antennas have been surrounded by vacuum. Since fabricated structures usually lie on a substrate, in this chapter a small overview of the effect of the substrate is given.

In equation (4.10), the penetration depth of the evanescent field of a surface plasmon polariton into the surrounding material is calculated. Thus it is apparent, that the permittivity of the environment has a strong effect on antennas, where localized SPPs are generated. Unfortunately the equation is only valid, if the particle is surrounded by only one material, which is not the case, if a nanoantenna is placed on substrate.

In Fig. 4.20 a) the maximum intensity enhancement and the corresponding resonance wavelength have been calculated for a varying permittivity of the substrate. The intensity has been taken again in the middle of the gap, $10nm$ above the substrate. One can see, that with increasing permittivity the maximum intensity enhancement decreases rapidly, while the corresponding resonance wavelength is increasing almost linearly. In practice often silicon dioxide (SiO_2) is used as substrate, where from the initial intensity enhancement of $I_{enh} = 1700$ only a value of $I_{enh} = 500$ remains for the bow-tie. The performance might be improved by optimization of the geometrical parameters though.



(a) (b)

Fig. 4.20.: *Maximum intensity enhancement and the corresponding resonance wavelength of a rod and a bow-tie antenna in dependence of a) the relative permittivity of the substrate, on which each antenna is grown and b) the relative permittivity of a sphere with a radius of $r = 6nm$ in between the two antenna arms, which is surrounded by vacuum.*

An often used application of these antennas is the usage of the intensity enhancement for nonlinear processes like second harmonic generation (SHG) or third harmonic generation (THG), where very high fields are used. For that one can place a nano particle with high nonlinear coefficient within the gap of the dipole antenna in order to increase the efficiency of these processes, as has been done in [76]. Such an incorporation of a

particle, however, changes the properties of the antenna itself.

In Fig. 4.20 b) the effect of a sphere with a radius of $r_sph = 6nm$ in dependence of the permittivity of the sphere is shown, while the antenna itself is surrounded by vacuum. Although the particle is relatively small compared to the antenna, the maximum intensity enhancement drops dramatically, the resonance wavelength though remains almost constant.

4.4. Optimization of the intensity enhancement of an arbitrary nanoantenna

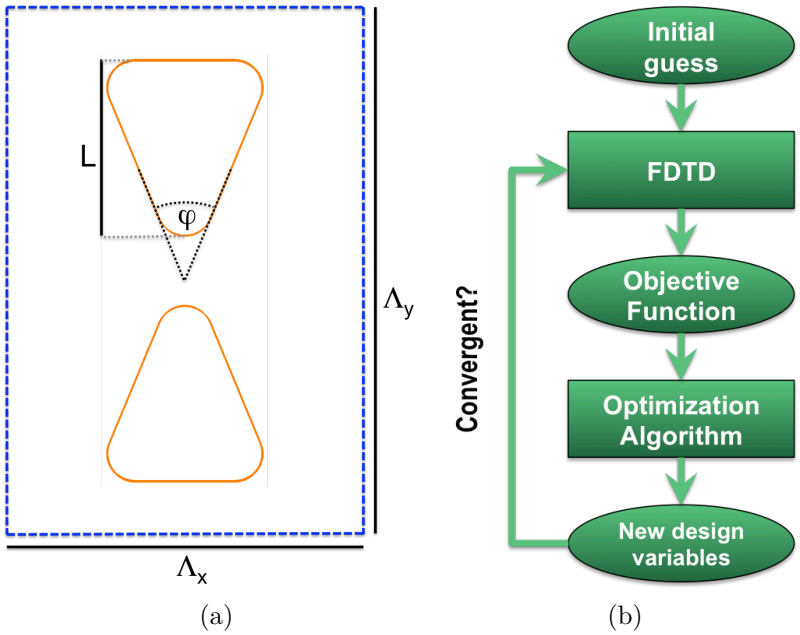
As stated before there has been done almost no optimization of these nanoantennas. In this section first the already known bow-tie antenna will be optimized and then afterwards an optimization of an arbitrary geometry will be performed.

4.4.1. Optimization of the intensity enhancement of a bow-tie

Starting point is the bow tie, which is again illuminated from top with a linearly polarized plane wave of wavelength $\lambda = 800nm$ and the electric field in the middle of the gap is monitored. The propagation of light for the optimization is numerically calculated with a standard FDTD [23] code and the metal is treated with the same Drude parameters as before (c.p. table A.1). In order to validate the optimization process the design variables of the bow-tie have been reduced as much as possible. Therefore, the distance between the two antennas has been fixed to $d_{gap} = 16nm$ and the thickness (in propagation direction) to $h = 32nm$. The only two degrees of freedom left are the opening angle φ of the underlying triangle and the length L. As before, all vertical edges have been blended with a radius of $r_{vert} = 20nm$ and the horizontal edges with a radius of $r_{top} = 10nm$. The periodicity parameters were set to $\Lambda_x = 500nm$ and $\Lambda_y = 700nm$, which is not close to the resonance discussed in section 4.3. A sketch of the design variables can be found in Fig. 4.21 a). Note, that in this approach only one arm is investigated, whereas the other arm is a mirrored version of the optimized one. The reason for that is, that the excitation of the antenna is symmetrical with respect to the y-axis.

The goal of the first optimization is to find the geometrical parameters for which the intensity enhancement in the gap is at maximum assuming an excitation wavelength of

$\lambda = 800nm$. For that the *nlopt*-library [77] has been included into the FDTD solver[4].



(a) (b)

Fig. 4.21.: *a) Sketch of a bow-tie with the definition of the design variables, which can be optimized. b) Sketch of the optimization process using the CRSLM algorithm included in a FDTD code.*

The optimization process is sketched in Fig. 4.21 b). The process is started by an initial guess of the design variables, with which the current value of the objective function is numerically calculated. This value is fed to the optimization algorithm, which first checks, if the convergence condition has been reached, and, if this is not the case, determines a new set of parameters. This routine is repeated until either the convergence has been reached or a predefined number of function evaluations has been performed.

The objective function in this case is simply the maximum intensity enhancement, so it can be formally expressed:

$$\min_{\text{s.t. } \boldsymbol{x}\in\Omega} -\left|\frac{\boldsymbol{E}_{ROI}}{\boldsymbol{E}_0}\right|^2(\boldsymbol{x}), \tag{4.22}$$

with the vector of design variables $\boldsymbol{x} = (L, \varphi)^T$ and the corresponding bound constraints

[4]The FDTD solver *maexle*, which has been used in this work, is written by Prof. Dr. Jens Förstner. It was parallelized via OpenMP and was usually running on 8 cores simultaneously.

$$x_1 \in [L_{min}, L_{max}], \tag{4.23}$$

$$x_2 \in [\varphi_{min}, \varphi_{max}]. \tag{4.24}$$

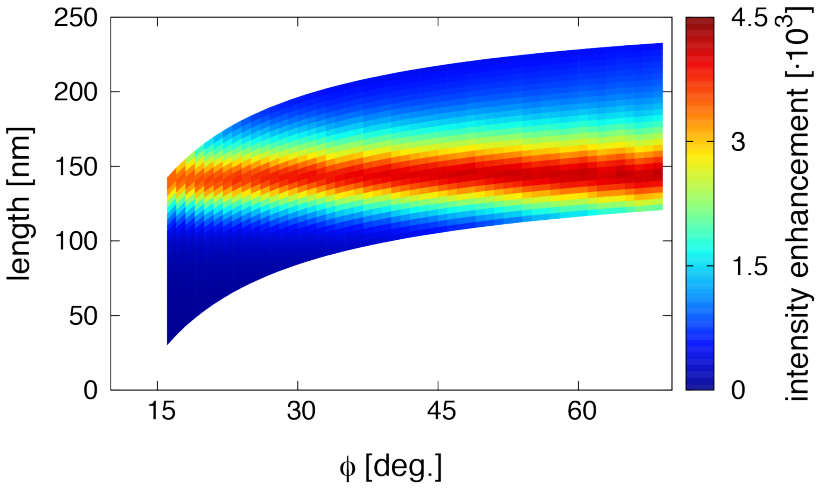


Fig. 4.22.: *Systematic parameter sweep for the case of constant periodicity in x- and y-direction. It is apparent, that the resonance length of the bow-tie does not depend on the opening angle of the underlying triangle.*

For the case of two design variables the objective function can be calculated systematically, as has been done for comparison in Fig. 4.22. For this plot over 1000 calculations have been done, each taking at least 5 minutes of computation time, which means a total time of 84 hours. One can see that the resonance length of the bow-tie does not depend on the opening angle of the underlying triangle, though magnitude of the intensity enhancement does. With increasing opening angle the resonance peak becomes broader, which leads to lower Q-factors, but makes the antenna more robust against a deviation of the length. The optimal opening angle of the bow-tie in this configuration (including the periodicity) is then $\varphi_{2DV} = 70deg.$ and the optimal length $L_{2DV} = 145nm$.

The optimization itself converged after around 140 iterations, leading to an intensity enhancement a bit over $I_{2DV} \approx 4000$, which is actually the global maximum, with an angle of $\varphi = 70deg.$ and a length of $L = 144nm$. For two parameters this seems a great number of iterations, but compared to the number of calculations done in order to obtain the data for the systematic parameter sweep, this number seems to be trivial. In Fig. 4.23 a) the optimization history is plotted. There one can see the random behavior of the CRSLM algorithm. Though a parameter set at least close to the global maximum has been found already in the first stage of the optimization process, still

many iterations are needed for convergence. The rough stopping condition used in this process is apparent, since the objective function still visibly varies for each run shortly before the optimization was finished.

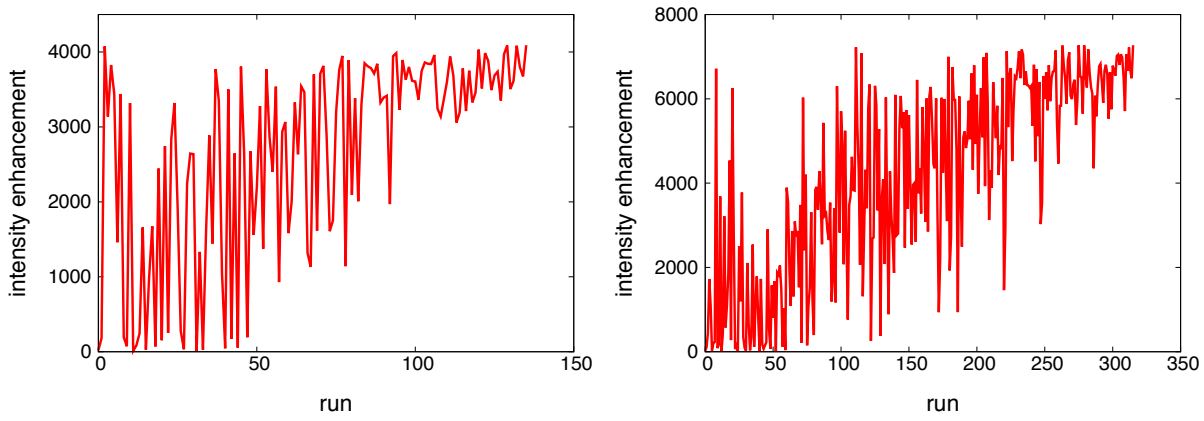


Fig. 4.23.: *Optimization history for the optimization of a bow-tie for a) 2 and b) 4 design variables using the CRSLM algorithm with a rough stopping condition. In both graphs the random behavior of the algorithm is apparent, where the global maximum is found in the early stage of the process, but still needs many function evaluations for convergence due to the global search within the parameter space.*

For the case that the periodicity parameters Λ_x and Λ_y are turned to design variables another optimization has been applied, for which in Fig. 4.23 b) the optimization history is shown. For 4 design variables the optimization took around 340 function evaluations for convergence, still with a relatively rough stopping condition. In the ideal case one would take the result and apply a local optimization afterwards in order to find the exact maximum. In contrast to the case with 2 design variables, for this optimization it is likely that several local maxima exist, since the periodicity may depend on the opening angle of the triangles. Still, with this rough optimization a maximum intensity enhancement of $I_{4DV} = 7500$ has been found, which is close to the predicted maximum in the previous chapter, with almost the same parameters.

The optimization of a bow-tie is a relatively simple task, since a good parameter set was already known, so the bound constraints could be set close to the expected values of the design variables. However, it is apparent at least for the case of 4 design variables, that in order to get a sufficient parameter set, the optimization routine is very efficient. A small estimation helps to emphasize this point. If for every design variable 10 values are simulated, one ends, if only 4 design variables are investigated, with 10,000 simulations, for all permutations have to be considered. However, when doing such optimization,

it is mandatory to reduce both, the number of design variables and the computation time as much as possible.

4.4.2. Optimization of the intensity enhancement of a nanoantenna with arbitrary geometrical shape

As stated before, many different geometrical shapes have been investigated in the past, e.g. bow-ties, ellipsoids, bi-pyramidal structures, discs and rods. All structures have in common, that they depend on the creativity of the developer, who comes up with a (new) idea. In this section the optimization of an almost free geometrical shape is performed. Though the results did not lead to a satisfying geometry, it is a convenient example to address some other issues, which are important for all optimizations.

Starting point for this approach is a polygon with n points. Since both, the periodicity and the height of the structure are kept fixed, one obtains $N = 2n$ design variables. Again only one antenna arm is optimized, while the other one is mirrored. In this approach, a polygon with 10 edges has been optimized, leading to 20 design parameters. In Fig. 4.24 a) a sketch of the optimization approach is shown. The gray region indicates the bound constraints, in which all edges of the polygon have been placed. In the middle of the gap again the intensity enhancement has been used for the objective function.

Note, that a slight variation of the structure near the gap results in a significant change of the objective function, whereas a variation far away of the gap influences the objective function only, if the resulting intensity enhancement of this structure is near to a (local) maximum. This means, that in a large part of the parameter space Ω a variation of the parameters lead to no change in the objective function itself. Since most gradient based algorithms would stop the process then, which is one reason why a non-gradient based algorithm, or more specific the CRSLM, has been chosen, which is known to be robust.

The CRSLM converged after around 12,000 iterations and results in a complex structure, consisting of several particles, caused by intersections of the polygon, and a very sharp tip. In Fig. 4.24 b) the resulting structure is pictured together with the intensity enhancement, when excited resonantly. This structure reached an intensity enhancement of around $I_{poly} = 10,000$, but is highly unrealistic. Firstly it is very challenging to fabricate small structures, and the smallest particles had a size of around $10nm$,

which is at the current stage of technology impossible to fabricate in a controlled manner. Secondly this structure is very sensitive to slight changes, e.g. if the smallest particle has been moved $10nm$ upwards or its size changed by $5nm$, the intensity enhancement would be reduced to around 2,000. Even the increase of the grid resolution, which causes a slight change in the geometrical representation of the structure, led to a significant change of the objective function.

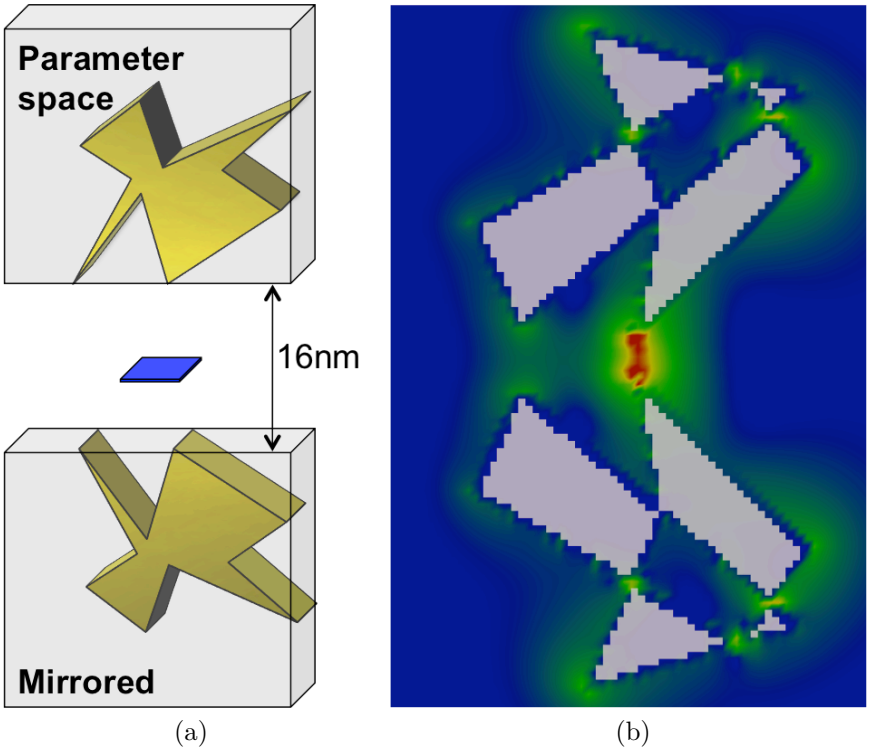


Fig. 4.24.: *a) Sketch of the optimization approach with the parameter space Ω indicated by the gray shaded box. b) Resulting structure of the optimization and the intensity enhancement for a resonant excitation in logarithmic scale, with the lowest value 1 (blue) and the highest value 10,000 (red).*

Unfortunately most maxima in this approach were that unstable, so in order to follow this approach, which has some advantages, one would have to implement several restrictions and / or constraints. One has to ensure, that no "small" particles are used, with a proper definition of "small". One way to do this would be to allow only convex polygons, but this restricts the parameter space drastically and the simplex might not contract around the global maximum, which means most algorithms would not be able to perform the optimization successfully. Another way could be the usage of a density filter, rounding all sharp edges and ensuring that all particles have a minimum size, but the structure would then be modified to beyond recognition and the computational

effort would be very high.

The sharp tip close to the gap, however, is the result of the definition of the objective function, since the intensity enhancement has been determined only in one point. It can be partially avoided by defining a face with a constant width and height, on which the intensity enhancement is spatially averaged. Then the high values caused by a sharp tip become less significant and more flattened tips might lead to better results.

For this work the most efficient solution was an approach using simple geometries, which have been combined to more complex structures.

4.4.3. Optimization of the intensity enhancement of a nanoantenna consisting of bricks

A more promising alternative to using the positioning of polygon points, is the usage of several simple geometries, which can be combined to more complex structures. Here a region, which can contain up to ten blocks of gold material of arbitrary size and position, is introduced, which results in 40 design variables. A sketch of this approach can be seen in Fig. 4.25 a). Additionally, constraints were added, which ensured, that too small structures, i.e., side lengths smaller than $16nm$, were not considered. There are implicit constraints by using boxes, e.g. all angles in the structure are fixed to $90°$, which perfectly matched the FDTD grid. Unfortunately it might lead to weird geometries and is computationally demanding to blend the edges of all structures during the optimization process, so they have been left sharp.

Since even more needed function evaluations were expected than for the polygon approach, the computational domain in propagation direction of the incident light field has been decreased to $\Lambda_z = 400nm$, i.e. the distance from the surface of the antenna to the open boundary (perfectly matched layers (PML)) has been roughly $200nm$. The penetration depth of the electric field of a surface plasmon at a gold-air interface can be up to $570nm$. Though the penetration depth of localized surface plasmons can be much smaller, one still should expect part of the field penetrating into the boundary, i.e. that part of the evanescent fields, which are not radiated, get damped by the boundaries. Calculations have shown, that the intensity enhancement has been reduced roughly by less than 10%, which is a systematical error, where all structures have been affected similarly.

Again PBC have been used in lateral directions in order to reduce the computation

time. The periodicity has been fixed to $\Lambda_x = 500nm$ and $\Lambda_y = 700nm$, which is still far away from the collective mode discussed in chapter 4.3.

As mentioned in the previous section, the objective function now was modified in the sense, that not only one point was taken into account, but a face Ω_f with the side lengths $\Delta w = 16nm$ and $\Delta h = 28nm$, where the intensity enhancement was spatially averaged. It was expected that the hot spots, caused by sharp edges of the bricks, would become less significant then. The optimization problem can then be formulated by (4.25).

$$\min_{\text{s.t. } \boldsymbol{x}\in\Omega} -\left|\frac{\frac{1}{N}\sum_i^N \boldsymbol{E}_i}{\boldsymbol{E}_0}\right|^2 (\boldsymbol{x}), \tag{4.25}$$

with the electric fields $\boldsymbol{E}_i$ monitored at the positions $\boldsymbol{p}_i \in \Omega_f$.

Additionally, a simplicity constraint has been implemented in order to avoid overly complex structures. For that first the minimum number of bricks have been calculated, with which the structure could be replaced, without changing it. It was possible for the bricks to overlap and therefore some bricks could be fully enclosed by others. For every brick used more than 3 a penalty of p has been attributed to the objective function, so it was actually a so called *soft constraint*. If all 10 bricks had been needed to obtain the current structure and the penalty had been $p = 5\%$ for each brick, the intensity enhancement has been reduced by 35%. Several optimization processed have been started with values for p between $0\% \leq p \leq 5\%$, which needed between 6500 and 8000 iterations to converge.

Interestingly, the code always converged towards a geometrical setup built of only two blocks. Each antenna arm consisted of a sharp tip plus an additional bigger block. It is not surprising that adjacent to the gap region a very thin metal tip has been found as sharp edges are known to give rise to large field gradients. It is interesting though that the intensity enhancement can be further increased by supporting the "small" antennas with a bigger block. The parameter sets from each optimization differed, but yielded almost identical performances, so it is likely, that only a local maximum has been found and even better solutions might exist. One parameter set can be found in table 4.2 a).

In Fig. 4.25 b) one example of the optimized antenna can be seen. The intensity enhancement with respect to a plane wave traveling in free space is logarithmically scaled and covers four orders of magnitude. The intensity enhancement in the middle

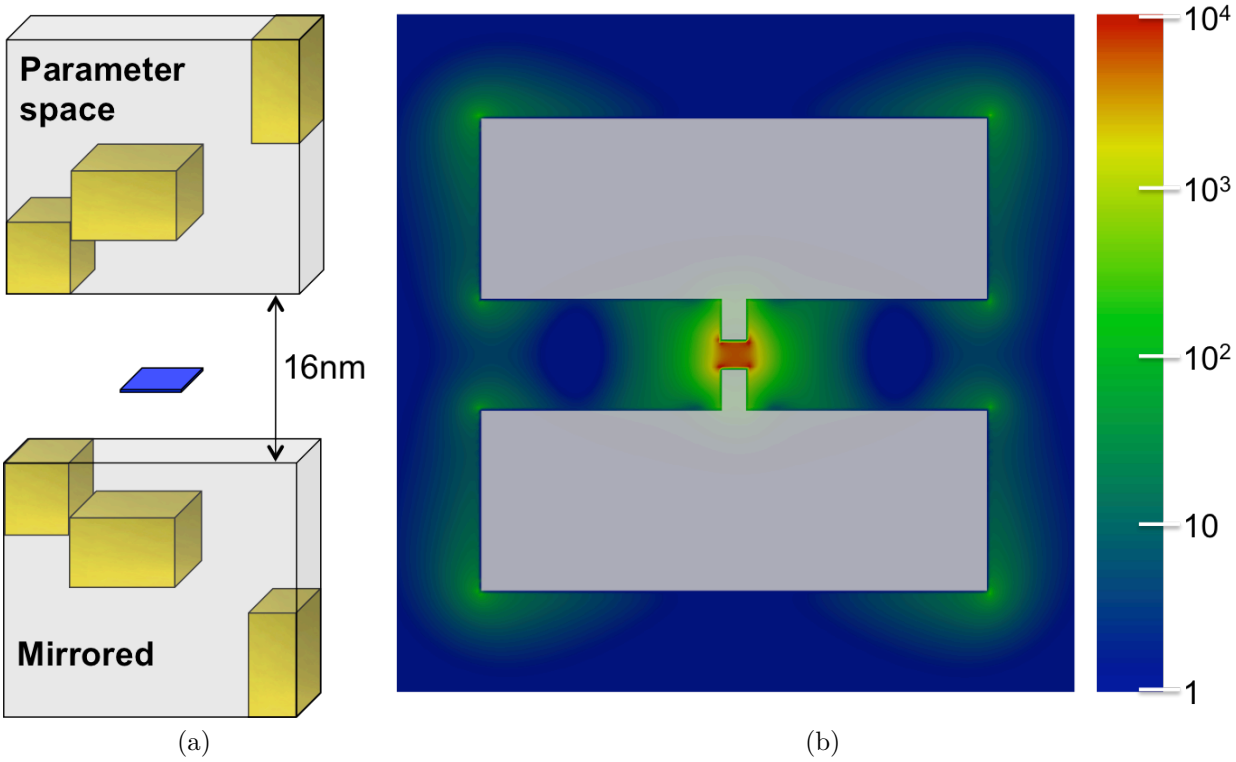


Fig. 4.25.: *a) Sketch of the optimization approach with the parameter space Ω indicated by the gray shaded box. b) Resulting structure of the optimization and the intensity enhancement for a resonant excitation in logarithmic scale, with the lowest value 1 (blue) and the highest value 10,000 (red).*

of the antenna is around 7000, which is almost twice as much as the best bow-tie with this periodicity as discussed in the previous section. As is known for other antennas the resonance frequency can be shifted by scaling the lengths of the two boxes.

Now the question arises, how the size of the supporting box influences the performance of the antenna. For that we changed the width of this box, recording the intensity enhancement in the middle of the antenna (c.p. Fig. 4.26). Here only a single antenna is calculated, which is surrounded by a PML in all directions, otherwise the periodicity of the array would have been changed as well and the results would not have been comparable. For this parameter sweep the simulation space is kept big enough to ensure, that the evanescent fields did not reach into the boundaries. One can see resonances with equidistant intervals of roughly 500 nm, which indicates some kind of standing wave inside of the supporting box.

To explain this we investigated the current density of these structures. In contrast to already known antennas, the optimized antennas current density has a large contri-

parameter	size
width of supp. box	$300nm$
length of supp. box	$108nm$
width of tip	$16nm$
length of tip	$24nm$
height	$28nm$
gap	$16nm$
periodicity Λ_x	$500nm$
periodicity Λ_y	$700nm$

(a) H-Antenna w/o array mode

parameter	size
width of supp. box	$253nm$
length of supp. box	$115nm$
width of tip	$16nm$
length of tip	$21nm$
height	$28nm$
gap	$16nm$
periodicity Λ_x	$760nm$
periodicity Λ_y	$785nm$

(b) H-Antenna with array mode

Tab. 4.2.: *Parameter set for an optimized box antenna. In a) the periodicity in x- and y-direction have been kept constant, whereas in b) they have been used as additional design variables.*

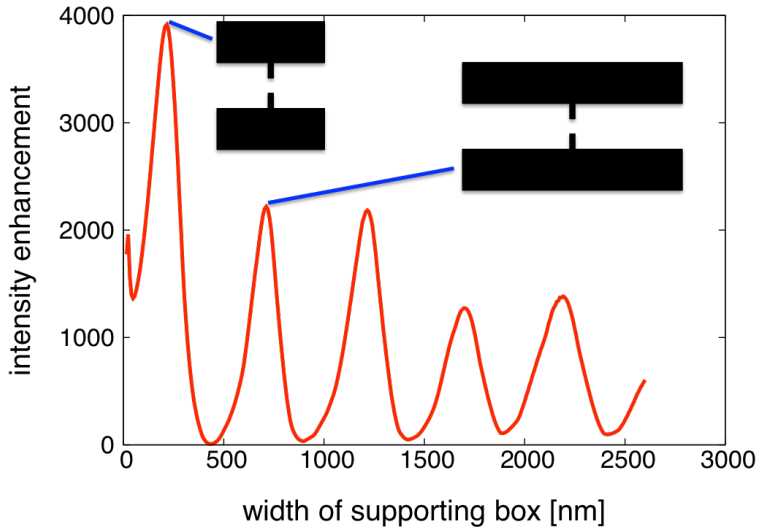


Fig. 4.26.: *Variation of the width of the supporting box of the optimized antenna, without changing any other parameters.*

bution in the direction perpendicular to the polarization of the incident light field. A line plot of the current density in the supporting box is shown in Fig. 4.27 a) for the

first two maxima and minima of Fig. 4.26. The second maximum (blue) looks like the first maximum (red), but has two half periods of an oscillation mode. This can be seen even better for the first two minima (green and pink). One can also see, that the closer one gets to the boundary of the supporting box, the more the oscillation becomes deformed. This is often attributed to a reactance at the ends of nanoantennas [16].

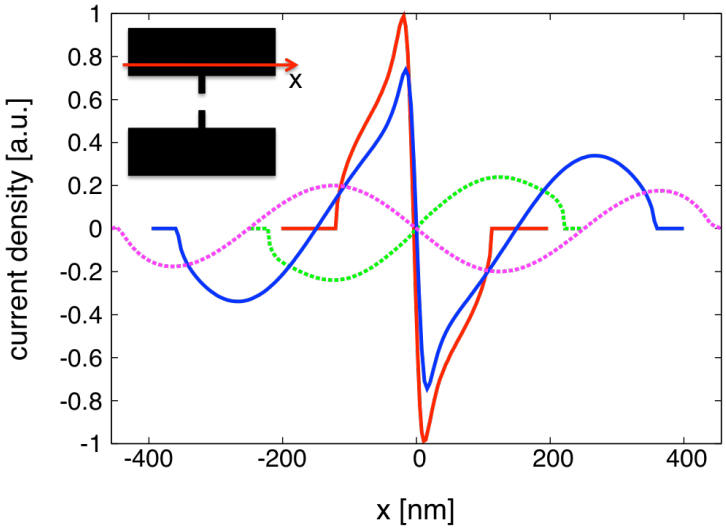
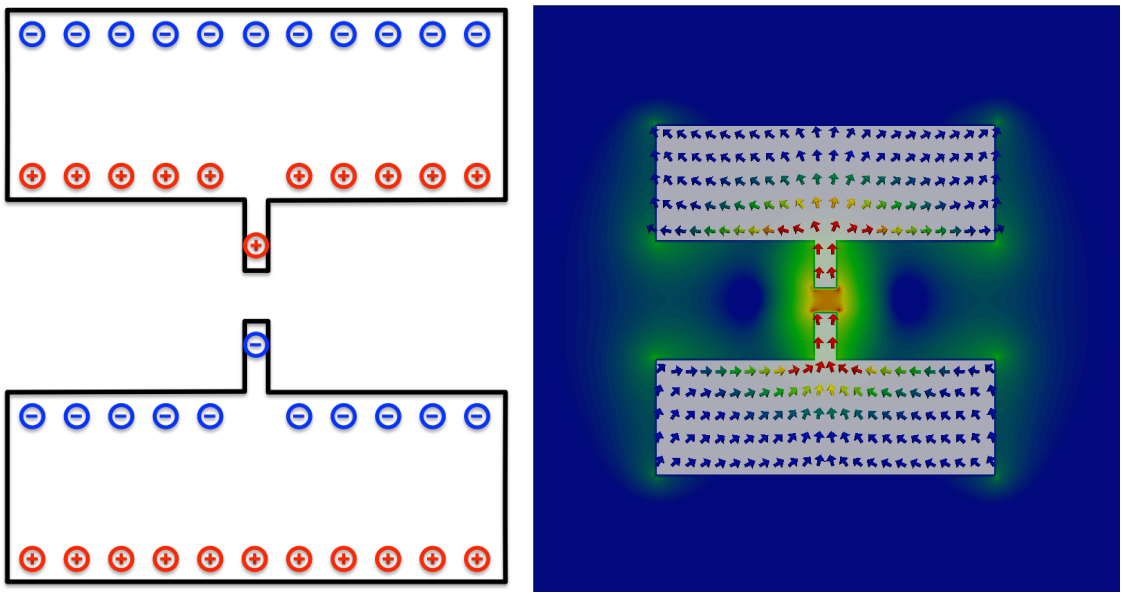


Fig. 4.27.: *a) Line plot of the current density parallel to the x-axis as indicated in the inset for different widths of the supporting box.*

The two minima exhibit a node in the middle of the box (where the tip is placed) with a small steepness, i.e. only a low charge carrier density is flowing to the tip. In contrast to that, the two maxima exhibit their maximum current density at that point, but have a change of the sign, which leads to a very high charge carrier density at the tip and with that a very high intensity enhancement within the gap of the antenna. Thus the width of the supporting box is crucial for a proper performance of this antenna.

In order to understand why such a high current density perpendicular to the polarization of the incident wave is generated, a sketch of the charge separation, when this antenna is excited resonantly, is shown in Fig. 4.28 a). If one looks to the electric field along the line, where the current density has been plotted, one would notice an inhomogeneity in particular for the E_y component, which increases when getting closer to the tip. Inserting this inhomogeneity into Maxwell's equations (2.2a) and (2.2b) results in an electric field pointing to the x-direction (or $-x$-direction respectively).

The induced field then causes the large current density parallel to the x-axis, with the charge carriers "flowing" towards the tip. This is more illustrated by Fig. 4.28, where the current density within the structure is pictured with arrows. The normalized absolute value is indicated by the color of the arrows (linearly scaled), reaching from 0 (dark blue) to 1 (red).



(a) (b)

Fig. 4.28.: *a) Sketch of the charge separation for the optimized antenna when excited resonantly. If one draws a line perpendicular to the antenna axis, it is apparent that the closer one gets to the tip, the higher an inhomogeneity of the resulting electric field becomes. Therefor a current towards the tip is induced, which further increase the charge at the tip, leading to a higher intensity enhancement. b) Current density inside of the antenna when excited resonantly. The color of the arrows indicate the absolute value (color scales linearly). Around the structure the electric field enhancement can be seen, with a logarithmic color scale.*

Now the question arises, if a collective array mode of this H-antenna[5] can be obtained. For that another optimization with the distance to neighboring antennas in x- and y-direction as additional design variables has been started. Since the stopping conditions have been finer as before (due to the reduction of design variables), the optimization converged after 1800 iterations. The found parameter set can be found in table 4.2 b). As discussed in section 4.3.2, a collective mode leads to a redshift of the resonance. Since the maximum intensity enhancement for the specific wavelength of $\lambda = 800nm$

[5]For a shorter name the optimized box antenna is from here on referred to as "H-antenna", because it looks like an H rotated by 90°.

has been optimized, the resulting antenna had to be slightly smaller in order to compensate the redshift. In Fig. 4.29 the spectrally resolved intensity enhancement for this antenna is shown (red line), which reaches at its maximum an intensity enhancement of around $I_{enh} \approx 35000$, which is almost 4 times as much as the best bow-tie achieved. The Q-factor is $Q = 18.1$, affirming that a collective array mode of these antennas has been found.

To confirm that this high enhancement, the same structure has been calculated with the frequency domain solver in CST Microwave Studio. There the grid is a tetrahedral, unstructured grid, where a staircase effect can be avoided. Also the problem of the epsilon averaging, where cells are partially filled with material, becomes not significant. Also the boundary conditions are implemented differently, and instead of an iterative time integration a large system of equations has to be solved. So this method, though it is still a FIT, works different in many critical points. As one can see, that in the results obtained by the frequency domain solver (green dots) are almost identical compared to the results before. The convergence in both simulations has been carefully checked, so one can safely assume, that this high intensity enhancement is not resulting from a numerical instability. To my knowledge this is by far the best performing antenna with a gap of $16nm$.

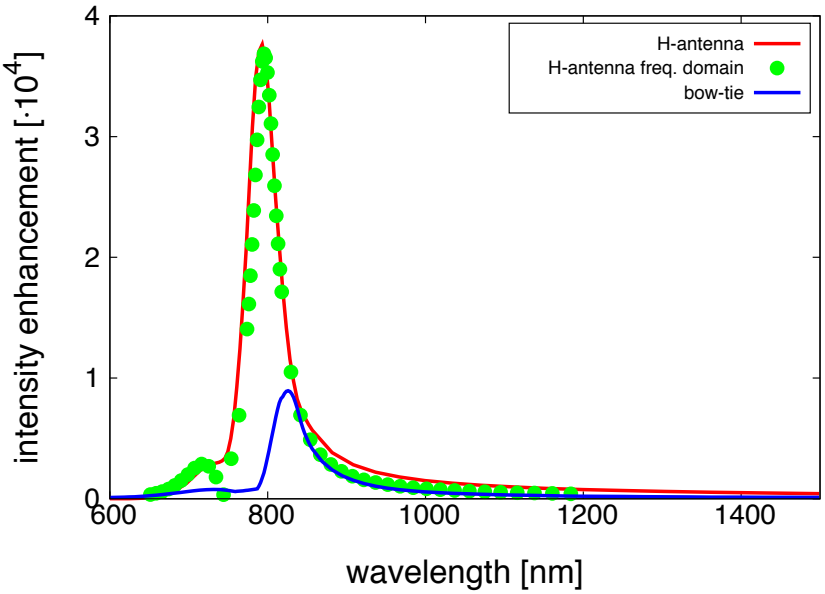


Fig. 4.29.: *Spectrally resolved intensity enhancement of the optimized H-Antenna calculated in time domain (red line) and in frequency domain (green dots) with the parameters given in table 4.2 b). For comparison the best obtained bow-tie is shown (blue).*

4.5. Fabrication related constraints

In this section the fabrication and experimental characterization of such nanoantennas is wrapped up. Since this is a theoretical work only, this section is kept brief, addressing only some restrictions or requirements, which have to be considered even when simulating nanoantennas.

As mentioned in the previous sections, the performance of nanoantennas is very sensitive with respect to the geometrical shape and its dimensions. Therefor the fabrication of such requires reliable and reproducible techniques with a typical resolution below $10nm$, which proves to be one of the main challenges in the realization of plasmonic devices. For that several approaches are possible, but in this section only the electron beam lithography (EBL), which is one of the most popular techniques, is discussed, based on [57].

The EBL (sketch in Fig. 4.30) is a top-down approach, usually starting with a substrate (e.g. glass) covered with a thin film of an optically transparent but electrically conducting material (often ITO), which is needed to avoid charging effects. On top of this a layer of an electron-sensitive resist (e.g. Poly(methyl methacrylate), also known as Plexiglas, (PMMA)) is deposited. With a focused electron beam 2D shapes can be written into the resist, which then can be selectively removed. A thin film of metal (e.g. gold) is then evaporated, covering both, the voids and the remaining resist. Finally within the so called lift-off the remaining resist including the covering metal is removed, leaving the metal structures in the voids unaffected.

Since the writing is done by a focused electron beam, the spatial resolution of the pattern is usually below $5nm$. However, the evaporated gold layer is usually polycrystalline, i.e. during the lift-off some small gold crystals are removed. This not only leads to a strongly increased surface roughness, but to a decreased structural resolution. Partially the surface roughness can be healed by tempering the sample though. In order to increase the stability of the structure, one often deposits titanium or chromium as a thin adhesion layer (below $5nm$). The drawback of such adhesion layers is often a significant increase of damping for surface plasmon polaritons.

It has been shown, that indeed single-crystalline gold can be grown in mesoscopic dimensions, where the structure can be cut out by focused ion-beam milling (FIB). This leads to a significant reduction of surface roughness, increased resolution and, unfortunately, much increased costs [78].

For the simulations some consequences arise for the design of nanoantennas. Firstly

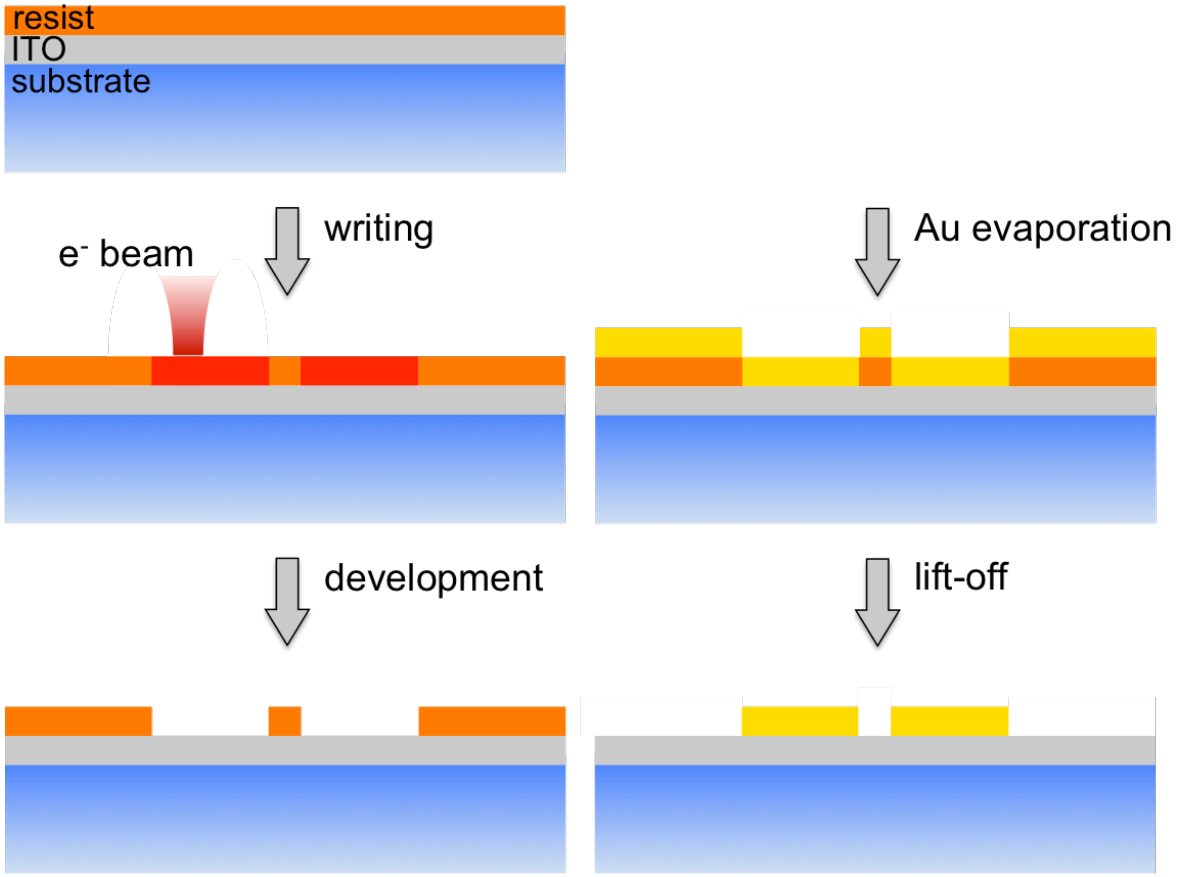
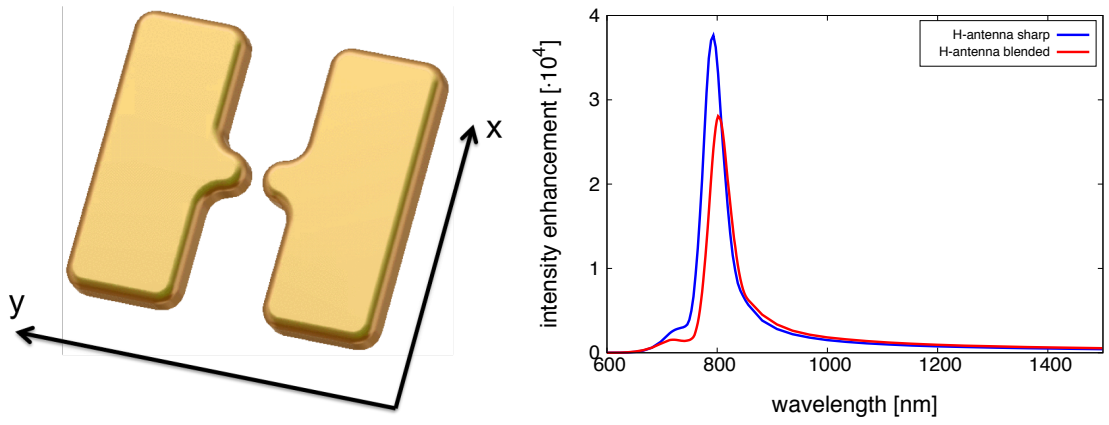


Fig. 4.30.: *Sketch of the most important steps in order to fabricate gold nanoantennas using electron beam lithography.*

one cannot fabricate sharp edges or infinitely thin structures. A minimum diameter of the resulting structure depends on the used machine, but is usually around $40nm$. One can still fabricate structures with smaller diameter, but with rapidly decreasing quality. Secondly during the lift-off a part of the gold is removed at the most unstable regions, which are the highest horizontal edges, i.e. they become rounded. This is essentially no problem, but since the antennas are very sensitive to the geometrical shape and its dimensions, one has to adjust the parameters when designing a nanoantenna.

In Fig. 4.31 a) this has been illustrated. After blending all horizontal and vertical edges, another optimization has been performed in order to compensate the different amount of material. Additionally the width of the tip has been enlarged from $16nm$ to $40nm$ (c.p. table 4.3). Please note, that the height and the gap in between of the two arms has been kept constant. In Fig. 4.31 b) one can see the spectrum of both, the H-antenna with sharp edges (blue), discussed in the previous section, and the obtained antenna with blended edges (red). Due to the broader tip the intensity enhancement is



(a) (b)

Fig. 4.31.: *a) Illustration and coordinate system of the optimized H-antenna with blended edges. b) Corresponding spectrally resolved intensity enhancement.*

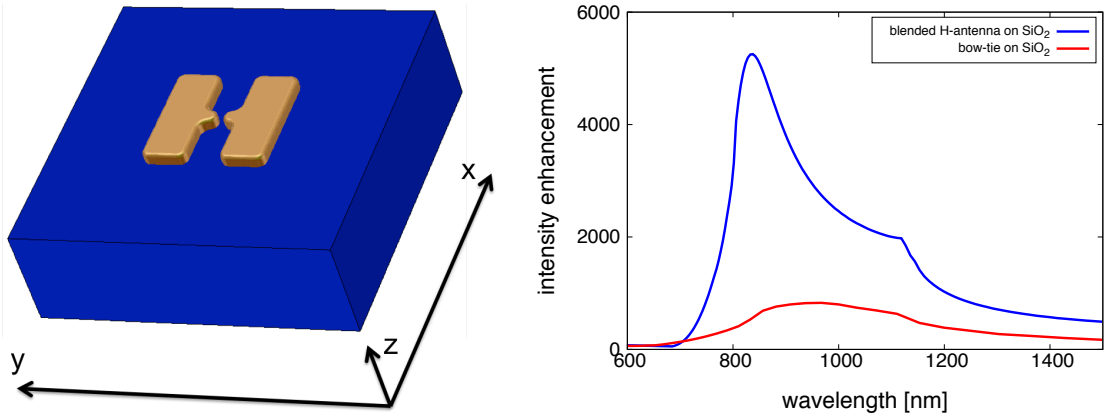
decreased by 20%, but not as much as expected. The resonance is still narrow, which indicates, that again an array mode has been excited.

parameter	size
width of supp. box	$272nm$
length of supp. box	$110nm$
width of tip	$40nm$
length of tip	$33nm$
height	$28nm$
gap	$16nm$
periodicity Λ_x	$770nm$
periodicity Λ_y	$786nm$

Tab. 4.3.: *Parameter set for the H-antenna. All horizontal edges have been blended with a radius of $r_{top} = 10nm$. All vertical edges have been blended with a radius of $r_{vert} = 20nm$.*

Now this blended antenna is placed on a SiO_2 substrate with a refractive index of $n = 1.46$ as illustrated in Fig. 4.32 a). As shown in chapter 4.3.3 it is expected, that the intensity enhancement decreases dramatically, while the resonance is slightly red shifted. The corresponding spectrally resolved intensity enhancement is shown in Fig.

4.32 b). There the blended H-antenna (blue) is compared to the optimized bow-tie (red). One can see two overlying peaks, though this is not apparent for the bow-tie. In case of the H-antenna one resonance occurs at an incident wavelength of $\lambda_1 \approx 840nm$, the other at $\lambda_2 \approx 1120nm$. That first peak is still narrow, compared to the resonance of the bow-tie, which again indicates an array mode.



(a) (b)

Fig. 4.32.: *a) Illustration and coordinate system of the optimized H-antenna with blended edges placed on a* SiO_2 *substrate. b) Corresponding spectrally resolved intensity enhancement.*

There is an additional feature of the H-antenna, which has not been discussed yet. As could be seen for the optimized bow-tie, e.g. in Fig. 4.22, the length of each antenna arm has been independent on the opening angle of the underlying triangle. For an incident wavelength of $\lambda = 800nm$, the optimal length has been around $L_{BT} = 144nm$. Though the H-antenna has an optimal length of roughly $L_H = L_{tip} + L_{sb}$ between $130nm$ and $140nm$, the resonance can be shifted via the width of supporting box as well without a significant decrease of the performance. However, since such structures are calculated before fabrication, it is beneficial to optimize all design parameters instead of adjusting just one.

Before the chapter is finished, a short remark about the measurement of plasmonic modes (and therefore the intensity enhancement) is necessary. For the characterization of such plasmonic modes several techniques have been developed over the years, of which some should be mentioned here. The most outstanding techniques are certainly electron energy-loss spectroscopy (EELS) [79] and cathodoluminescence (CL) [80], with which the highest resolutions have been obtained. In these experiments a beam of fast electrons is focused onto the sample, where some electrons are able to transfer energy

into the nano structure via electromagnetic interaction [81]. In the case of EELS the corresponding loss of energy of the transmitted electrons is measured, whereas in the case of CL the emitted light is detected (mostly in reflective geometry). In Ref. [82], where both methods have been applied simultaneously, it is shown, that the CL is only able to measure bright modes, while the EELS also gives data about dark modes. Apart from these two techniques, other experimental setups like scanning near-field optical microscopy are also available, but the spatial resolution is usually much worse [83].

To summarize this chapter, many properties of dipole antennas have been discussed. The bow-tie antenna, surrounded by vacuum, has been systematically optimized and then set a reference. Then different ways of optimizing more general structures have been discussed, where it seems convenient to use simple geometries to form more complex shapes in order to reduce design variables and define proper constraints. It is evident, that the whole optimization process strongly depends on the definition of the objective function and the used constraints, particularly if several objective functions have to be combined (e.g. maximizing the field in area A and simultaneously minimizing the field in area B). In such cases one has to make sure, that the order of magnitude of both goal functions is similar. Since they often contradict each other, one has somehow reformulate the problem and weight the objective functions.

The optimization using bricks has been very successful, yielding a geometrical shape which generates an intensity enhancement four times higher compared to the optimized bow-tie. Even when blending the edges, the antenna performed 3.5 times better than the reference. Placing both, the optimized bow-tie and the H-antenna on glass substrate, resulted, as expected, in a dramatic decrease of the performance, but the H-antenna still performed much better than the bow-tie.

This shows, that there is still room for optimization, for not only a better performance has been achieved, but also a surprising geometrical shape found. Using automatic optimization routines leads to shapes, which might not have been thought of, and yields, while trying to understand an optimized structure, to more insight of the problem at hand and the involved processes.

5. Optical and NIR antennas with high directivity

For future light-on-chip integration the control of visible or infrared light is crucial on nanoscale dimensions [84]. Due to the diffraction limit, conventional optical devices like lenses are not functional in sub-wavelength scale, which is why antennas and metamaterials are often considered. Optical and infrared antennas provide a promising way to couple photons in and out of nanoscale objects [85, 64]. As counterpart of conventional radio antennas, they are able to increase optical fields in sub-wavelength volumes [63, 62], to enhance excitation and emission of quantum emitters, to direct light, emitted by quantum emitters [86, 87, 88, 89, 4] or to apply a color routing [90, 91]. So far, these functionalities have been mainly pursued by surface plasmon based devices [92], like nanolenses [93], nano-waveguides [94] or nanoantennas [95, 96, 97, 57]. Most of these devices become complicated due to coupling of several metallic particles and their sensitivity to the geometrical shape. However, there were many successes during the past years, but the main drawback of using plasmonic structures is their high intrinsic losses.

So far, in order to obtain a near infrared (NIR) directional antenna, usually the conventional Yagi-Uda antenna is scaled down and the lengths of the elements are adjusted to plasmonic resonances. The concept is theoretically and experimentally proven [61], but the directivity of such an antenna is not nearly as good as in the RF regime, which has several reasons, e.g. the coupling between plasmonic modes in different elements or the influence of the substrate.

As a promising alternative to plasmon based directional antennas, recently all-dielectric antennas are investigated. Tang et al. have changed the dielectric background using transformation optics [22], whereas in a more recent approach small silicon spheres are positioned similar to a Yagi-Uda antenna to direct the emitted light [84, 98]. In the experiment a plane wave is applied, since the positioning of the quantum emitter is crucial, but very difficult, for this kind of antenna.

First some aspects of the classical antenna theory are summarized. In the following sections an optimization of nanoscopic antennas in the near infrared regime starting from a metallic Yagi-Uda structure has been performed. The optimization is done via a particle-swarm algorithm, using full time domain finite integration simulations to obtain the characteristics of the investigated structure, also taking into account substrates. Furthermore dielectric antennas are shown, which perform even better, due to the lack of losses by an appropriate choice of the dielectric material. The section[1] 5.2.2 is based on Ref. [99].

5.1. Classical antenna theory

In this section some important quantities and properties of classical antennas are summarized. It is based on Ref. [100], in which a much more detailed discussion can be found.

Starting point is the radiation of electromagnetic waves by a general current (and charge) distribution. Typically, the current distribution is localized in a defined region, e.g. a wire antenna and it can generate electromagnetic fields, which can propagate to a far distance from the source position. In order to describe this radiation it proves convenient to use the electric and magnetic potentials instead of the electric and magnetic fields.

Since the divergence of the magnetic flux is zero (c.p. equation (2.2d)), one can define the so called vector potential, from which the magnetic flux can be calculated:

$$\boldsymbol{B}(\boldsymbol{r},t) = \nabla \times \boldsymbol{A}(\boldsymbol{r},t) \tag{5.1}$$

Inserting this into Faraday's law in equation 2.2a leads to the definition of the so called scalar potential

$$\boldsymbol{E}(\boldsymbol{r},t) = -\nabla\varphi(\boldsymbol{r},t) - \frac{\partial}{\partial t}\boldsymbol{A}(\boldsymbol{r},t). \tag{5.2}$$

These potentials are not uniquely defined, since they can be changed by adding constants to them. Additionally, Maxwell's equations are gauge invariant, with the Lorenz condition known as the most often used gauge:

[1] Publications for sections 5.3 and 5.4 are in preparation.

$$\nabla \cdot \boldsymbol{A}(\boldsymbol{r},t) + \frac{1}{c_0^2}\frac{\partial}{\partial t}\varphi(\boldsymbol{r},t) = 0. \tag{5.3}$$

After a short calculation, one can derive the wave equations for those potentials under the assumption, that ε and μ are constant with respect to $\boldsymbol{r}$.

$$\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\varphi(\boldsymbol{r},t) - \nabla^2\varphi(\boldsymbol{r},t) = \frac{1}{\varepsilon}\rho(\boldsymbol{r},t) \tag{5.4}$$

$$\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\boldsymbol{A}(\boldsymbol{r},t) - \nabla^2\boldsymbol{A}(\boldsymbol{r},t) = \mu\boldsymbol{J}(\boldsymbol{r},t) \tag{5.5}$$

So the charge density ρ and the current density $\boldsymbol{J}$ may be considered as sources, which generate the potentials φ and $\boldsymbol{A}$, from which the electric and magnetic fields can be calculated. The main (or causal) solution of the wave equations (5.4) and (5.5) for the case, that the source densities φ and $\boldsymbol{A}$ are known, is given by the so called retarded potentials in equations (5.6) and (5.7), where $R = |\boldsymbol{r} - \boldsymbol{r'}|$ is the distance from a point in the region of the source density to the field point.

$$\varphi(\boldsymbol{r},t) = \int_V \frac{\rho(\boldsymbol{r'}, t - \frac{R}{c_0})}{4\pi\varepsilon R} d^3\boldsymbol{r'} \tag{5.6}$$

$$\boldsymbol{A}(\boldsymbol{r},t) = \int_V \frac{\mu\boldsymbol{J}(\boldsymbol{r'}, t - \frac{R}{c_0})}{4\pi R} d^3\boldsymbol{r'} \tag{5.7}$$

So far the retarded potentials are general and apply to any current and charge distribution. For typical radiation problems several approximations have to be made, e.g. the far field approximation, which assumes that the investigated field point $\boldsymbol{r}$ is very far located from the current source. Additionally a first order expansion of the distance $R = |\boldsymbol{r} - \boldsymbol{r'}| \simeq |\boldsymbol{r}| - \hat{\boldsymbol{r}} \cdot \boldsymbol{r'} \simeq |\boldsymbol{r}|$ has been applied. This is indeed valid as long as the dimension of the source region l is much smaller than $r = |\boldsymbol{r}|$. Assuming a time harmonic dependence, transforming the potentials into frequency space and using spherical coordinates, one obtains the phasor part of the retarded potentials. With the definition of the wave vector $\boldsymbol{k} = k\hat{\boldsymbol{r}}$, where k is the magnitude, the retarded single-frequency potentials in the far-field approximation can be expressed as follows:

$$\varphi(\boldsymbol{r}) = \frac{e^{-ikr}}{4\pi\varepsilon r}\int_V \rho(\boldsymbol{r'}) e^{i\boldsymbol{k}\cdot\boldsymbol{r'}} d^3\boldsymbol{r'} \tag{5.8}$$

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu e^{-ikr}}{4\pi r}\int_V \boldsymbol{J}(\boldsymbol{r'}) e^{i\boldsymbol{k}\cdot\boldsymbol{r'}} d^3\boldsymbol{r'} \tag{5.9}$$

In these expressions the radial dependence has been separated from the angular de-

pendence. Since the integral factors are important for the directional properties of the radiated fields, the charge form-factor $Q(\boldsymbol{k})$ and the radiation vector $\boldsymbol{F}(\boldsymbol{k})$ are introduced. Basically, they are calculated by the three dimensional spatial Fourier transform of the charge and current density distribution. Since they are depending on the radial unit vector $\hat{\boldsymbol{r}}$, they can be written in dependence of the polar angle θ and the azimuthal angle ϕ.

$$Q(\theta,\phi) = \int_V \rho(\boldsymbol{r'}) e^{i\boldsymbol{k}\cdot\boldsymbol{r'}} d^3\boldsymbol{r'} \tag{5.10}$$

$$\boldsymbol{F}(\theta,\phi) = \int_V \boldsymbol{J}(\boldsymbol{r'}) e^{i\boldsymbol{k}\cdot\boldsymbol{r'}} d^3\boldsymbol{r'} \tag{5.11}$$

The retarded potentials are then described by equations (5.12) and (5.13).

$$\varphi(\boldsymbol{r}) = \frac{e^{-ikr}}{4\pi\varepsilon r} Q(\theta,\phi) \tag{5.12}$$

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu e^{-ikr}}{4\pi r} \boldsymbol{F}(\theta,\phi) \tag{5.13}$$

Neglecting all terms, which fall off faster than $1/r$, one obtains the expressions for the radiated electric and magnetic fields in equations (5.14) and (5.15), which are generated by a current source.

$$\boldsymbol{E} = ikZ_F \frac{e^{-ikr}}{4\pi r} (\hat{\boldsymbol{r}} \times \boldsymbol{F}) \times \hat{\boldsymbol{r}} \tag{5.14}$$

$$\boldsymbol{H} = -ik \frac{e^{-ikr}}{4\pi r} \hat{\boldsymbol{r}} \times \boldsymbol{F}, \tag{5.15}$$

with the impedance $Z_F = \sqrt{\mu/\varepsilon}$. If the radiation vector is decomposed into a longitudinal and transversal part $\boldsymbol{F} = \hat{\boldsymbol{r}} F_r + \boldsymbol{F}_\perp$, which can be done with any vector field according to Helmholtz's theorem, one can show, that neither the electric nor the magnetic field is depending on the radial component F_r, thus they may be expressed with the transversal component of the radiation vector only:

$$\boldsymbol{E} = -ikZ_F \frac{e^{-ikr}}{4\pi r} \boldsymbol{F}_\perp \tag{5.16}$$

$$\boldsymbol{H} = -ik \frac{e^{-ikr}}{4\pi r} \hat{\boldsymbol{r}} \times \boldsymbol{F}_\perp. \tag{5.17}$$

The flux vector of the electromagnetic energy can then be calculated by the time averaged Poynting vector:

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\Re(\boldsymbol{E} \times \boldsymbol{H}^*) = \hat{\boldsymbol{r}}\frac{Z_F k^2}{32\pi^2 r^2}|\boldsymbol{F}_\perp(\theta,\phi)|^2 \tag{5.18}$$

The power radiated into a solid angle $d\Omega$ will flow through the surface area dS, thus the power density is given by equation (5.19).

$$\frac{dP}{dS} = \frac{dP}{r^2 d\Omega} = \mathcal{P}_r = \frac{Z_F k^2}{32\pi^2 r^2}|\boldsymbol{F}_\perp(\theta,\phi)|^2 \tag{5.19}$$

Please note, that in many publications the power density is denoted by $p(\theta,\phi) = \mathcal{P}_r$. Often the radiation intensity is defined as in equation (5.20), with which the radiation pattern is characterized.

$$U(\theta,\phi) = \frac{dP}{d\Omega} = r^2\mathcal{P}_r = \frac{Z_F k^2}{32\pi^2}|\boldsymbol{F}_\perp(\theta,\phi)|^2 \tag{5.20}$$

The total radiated power then is obtained by integrating the radiation intensity over all solid angles $d\Omega = \sin\theta d\theta d\phi$:

$$P_{rad} = \int_0^\pi \int_0^{2\pi} U(\theta,\phi)d\Omega \tag{5.21}$$

In order to characterize antennas, a useful concept is the comparison of a radiation pattern with an isotropic radiator, whose radiated power is equally distributed over all solid angles. Thus the isotropic radiation intensity is

$$U_I = \frac{P_{rad}}{4\pi}. \tag{5.22}$$

To get insight into an antenna behavior, one often uses the so called directive gain, defined in equation (5.23), which is the radiation intensity of the investigated structure normalized with by the corresponding isotropic intensity.

$$D(\theta,\phi) = \frac{U(\theta,\phi)}{U_I} = \frac{4\pi\mathcal{P}_r}{P_{rad}} \tag{5.23}$$

It measures the ability of the structure to direct power towards a given direction. The maximum of the directive gain is called the directivity[2], so that

[2]In most publications concerning nanoantennas the directive gain is confused with the directivity. It is often referred to as directivity and the maximum directivity respectively, but by the context it becomes clear, what exactly is meant.

$$D_{max} = \frac{U_{max}}{U_I}. \tag{5.24}$$

Expressing the power density in terms of the directive gain gives

$$\mathcal{P}_r = \frac{P_{rad} D(\theta, \phi)}{4\pi r^2}, \tag{5.25}$$

which motivates the definition of the directive gain. Please note, that the directive gain is a relative quantity, i.e. that losses have not been considered so far, though they play a significant role within the design of an antenna. To include losses, the power gain is introduced in equation (5.26).

$$G(\theta, \phi) = \frac{4\pi U(\theta, \phi)}{P_{total}} = \frac{4\pi U(\theta, \phi)}{P_{rad} + P_{loss}} \tag{5.26}$$

By defining the radiation efficiency

$$\eta = \frac{P_{rad}}{P_{total}}, \tag{5.27}$$

one can rewrite the power gain to its most popular form:

$$G(\theta, \phi) = \eta D(\theta, \phi). \tag{5.28}$$

Another useful quantity is the half-power beam width, which is the angle between the points, where the directive gain (or the power gain) is equal to the half of its maximum value. Apparently the opening angles are defined separately for θ and ϕ and are denoted with $\Delta\theta_B$ and $\Delta\phi_B$ respectively.

To summarize this description, in order to describe an antenna, one usually needs the directive gain, radiation efficiency and the half-power beam width angles. It is easy to see, that with increasing directivity the beam width angles become smaller, so the radiation is directed more efficiently. Unfortunately most antennas exhibit side lobes, which means that a part of the radiation is directed into an entirely different direction compared to the main lobe with the maximum directive gain. For a proper design of antennas, the side lobes should be as small as possible, since some of the radiation is lost with them.

Another important quantity is the so called front-to-back ratio r_{fb}, for which many different definitions are used. The most common definition is the ratio of the directive

gain in direction of the main lobe and the directive gain in the opposite direction. Assuming the maximum directive gain is obtained for the angles θ_{max} and ϕ_{max}, it can then expressed by

$$r_{fb} = \frac{D(\theta_{max}, \phi_{max})}{D(\pi - \theta_{max}, \phi_{max} - \pi)}. \tag{5.29}$$

However, since this work is about nanoantennas, which usually lay on a substrate, it is expected, that the largest part of the radiation of a point source will penetrate through it. Thus another definition of the front-to-back ratio is convenient, where the polar angle is left constant:

$$R_{fb} = \frac{D(\theta_{max}, \phi_{max})}{D(\theta_{max}, \phi_{max} - \pi)} \tag{5.30}$$

5.1.1. Radiation pattern of a Hertzian Dipole

Now a Hertzian dipole will be discussed. It is actually the simplest linear antenna consisting of an infinitesimal thin wire with the length L, located at the origin and directed along the z-axis. The current density is the of the form $\boldsymbol{J}(\boldsymbol{r}) = \hat{\boldsymbol{z}} I(z)\delta(x)\delta(y)$, where $I(z)$, $-L/2 \leq z \leq L/2$, is the current distribution along the antenna element. A sketch of this linear antenna is shown in Fig. 5.1 a).

It is evident, that the radiation vector $\boldsymbol{F}$ from equation (5.11) will have only a z-component:

$$\boldsymbol{F} = \hat{\boldsymbol{z}} F_z = \hat{\boldsymbol{z}} \int_{-L/2}^{L/2} I(z') e^{ik_z z'} dz' \tag{5.31}$$

The wave vector $\boldsymbol{k}$ can be transformed to Cartesian coordinates, with $k_z = k\cos\theta$. Further it is assumed, that the antenna length is electrically short, i.e. $L \ll \lambda_0$, which leads to the current distribution of a Hertzian dipole $I(z) = IL\delta(z)$. With that the z-component of the radiation vector can be expressed as in eq. (5.32).

$$F_z = \int_{-L/2}^{L/2} I(z') L e^{ikz'\cos\theta} dz' = IL \tag{5.32}$$

For this simple antenna F_z is a constant and independent of θ. Transforming the unit vector $\hat{\boldsymbol{z}}$ to spherical coordinates yields the transversal part of the radiation vector in spherical coordinates:

$$\boldsymbol{F}_{\perp}(\theta) = \hat{\boldsymbol{\theta}} F_{\theta}(\theta) = -\hat{\boldsymbol{\theta}} F_z(\theta) \sin\theta \tag{5.33}$$

The calculation of the electric and magnetic radiation fields can then be calculated as in equations (5.34) and (5.35).

$$\boldsymbol{E} = \hat{\boldsymbol{\theta}} E_{\theta} = \hat{\boldsymbol{\theta}} ikZ_F \frac{e^{-ikr}}{4\pi r} IL \sin\theta \tag{5.34}$$

$$\boldsymbol{H} = \hat{\boldsymbol{\phi}} H_{\phi} = \hat{\boldsymbol{\phi}} ik \frac{e^{-ikr}}{4\pi r} IL \sin\theta \tag{5.35}$$

The fields are omnidirectional, that is, independent of the azimuthal angle ϕ. The radiation intensity from eq. (5.20) is then

$$U(\theta) = \frac{Z_F k^2}{32\pi^2} |IL|^2 \sin^2\theta \tag{5.36}$$

and the directive gain

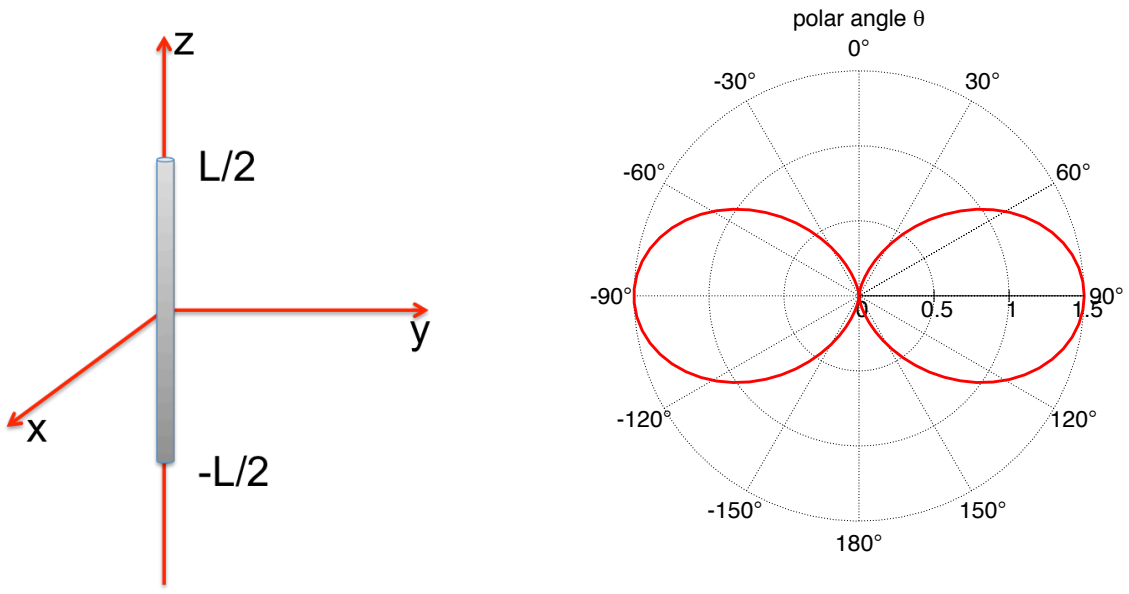
$$D(\theta) = \frac{4\pi U(\theta)}{P_{rad}} = \frac{4\pi U(\theta)}{\frac{Z_F k^2 |IL|^2}{12\pi}} = \frac{3}{2} \sin^2\theta. \tag{5.37}$$

The directive gain has been plotted in a polar plot in Fig. 5.1 b) for the polar angle θ. It is apparent that such a source won't radiate along the z-axis (i.e. for $\theta = 0$). Since the polarization of the radiation fields are perpendicular to the radial unit vector $\hat{\boldsymbol{r}}$, one obtains, depending on θ, electromagnetic waves with polarizations along all three coordinate axes (for Cartesian coordinates). The directivity of a Hertzian dipole is $D_{max} = 1.5$, which is close to an isotropic radiator.

In practice point sources are usually implemented in algorithms like FIT or FDTD by a short and thin antenna. In the extreme this antenna would cover only one cell, but, at least in the case of the used CST Microwave Studio, slightly longer wires significantly increase the stability of the simulations.

5.1.2. Yagi-Uda antenna in radio frequency regime

In this section a prominent directive antenna in the radio frequency regime, the Yagi-Uda antenna (or often called Yagi antenna), will be discussed. It is simple to construct and has a high gain, typically $G \geq 10dBi$. For most applications they are designed to work in HF or UHF bands, which means a range from about $3MHz$ to $3GHz$, though



(a) (b)

Fig. 5.1.: *a) Sketch of a linear antenna. For the case, that the length L is electrically small, it becomes the well known Hertzian dipole. b) Polar plot of the directive gain for a Hertzian dipole. Interesting is only the polar angle θ, since the directive gain does not depend on the azimuthal angle ϕ. A three dimensional representation of the directive gain would look like a donut.*

their bandwidth is rather small (typically a few percents of the center frequency). They were first published in Japan in 1926 by Shintaro Uda, but presented in English, and thus made available for engineers in the whole world, by his colleague Yagi.

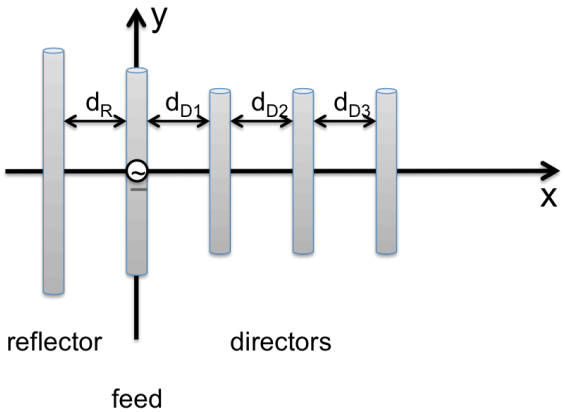


Fig. 5.2.: *Sketch of a classical Yagi-Uda antenna in radio frequency regime. The feed element is connected to a coaxial cable and therefore acting as a dipolar source. The reflector, typically larger than the feed, is effectively reflecting radiation backwards towards the feed, while the directors further decrease the light beam width.*

The basic geometry is shown in Fig. 5.2. It consists of a single feed element, which is usually a dipole antenna. This feed is the only element that is externally excited, typically via a connected coaxial cable. The rest of the elements are parasitic, i.e. they only help to reflect or further direct radiation emitted by the feed into the desired direction. Almost always only one reflector exists, so the feed is the second element to the end. It has to be resonant to the excitation frequency, which means a length L_F of the feed slightly smaller than half of the wavelength, in order to compensate the presence of the other parasitic dipoles.

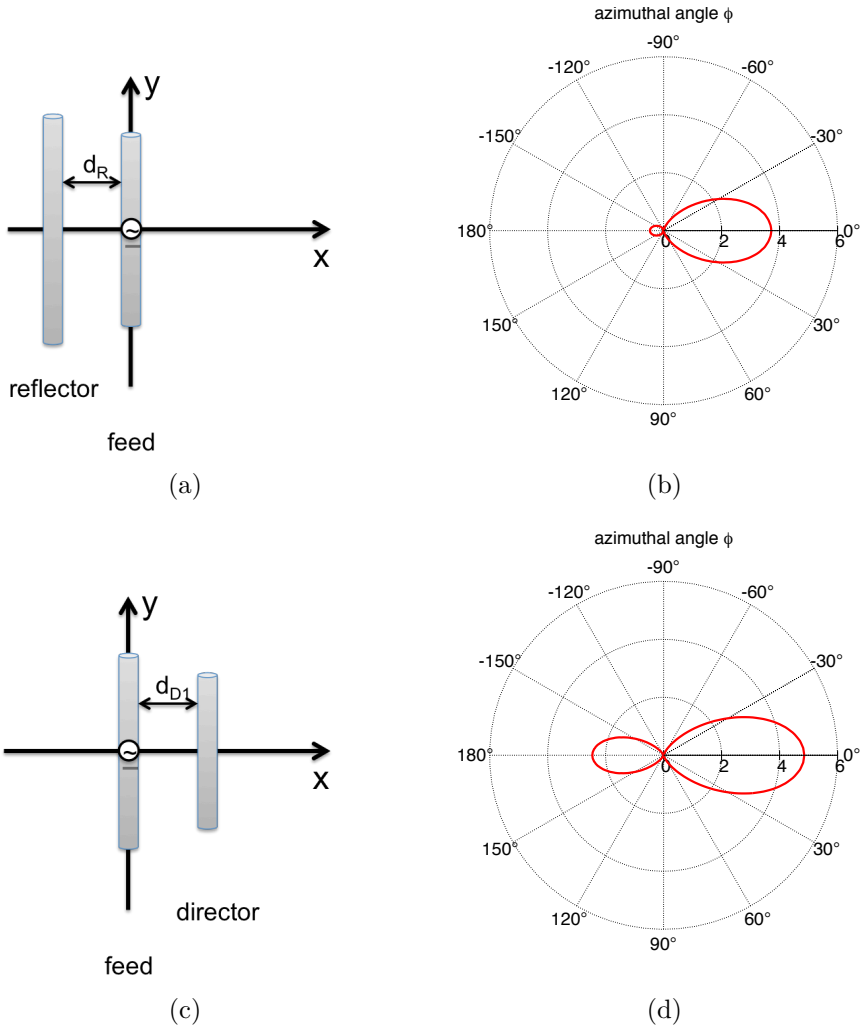


Fig. 5.3.: *Sketch and Directivity for both, a reflector (a) - b)) and a director (c) - d)) next to the feed. The polar plot show the azimuthal angle, since for this structure the radiation in plane of the antenna axis (x-axis) and the polarization of the feed (y-axis) is interesting. The frequency of the excitation signal has been* $f = 1GHz$.

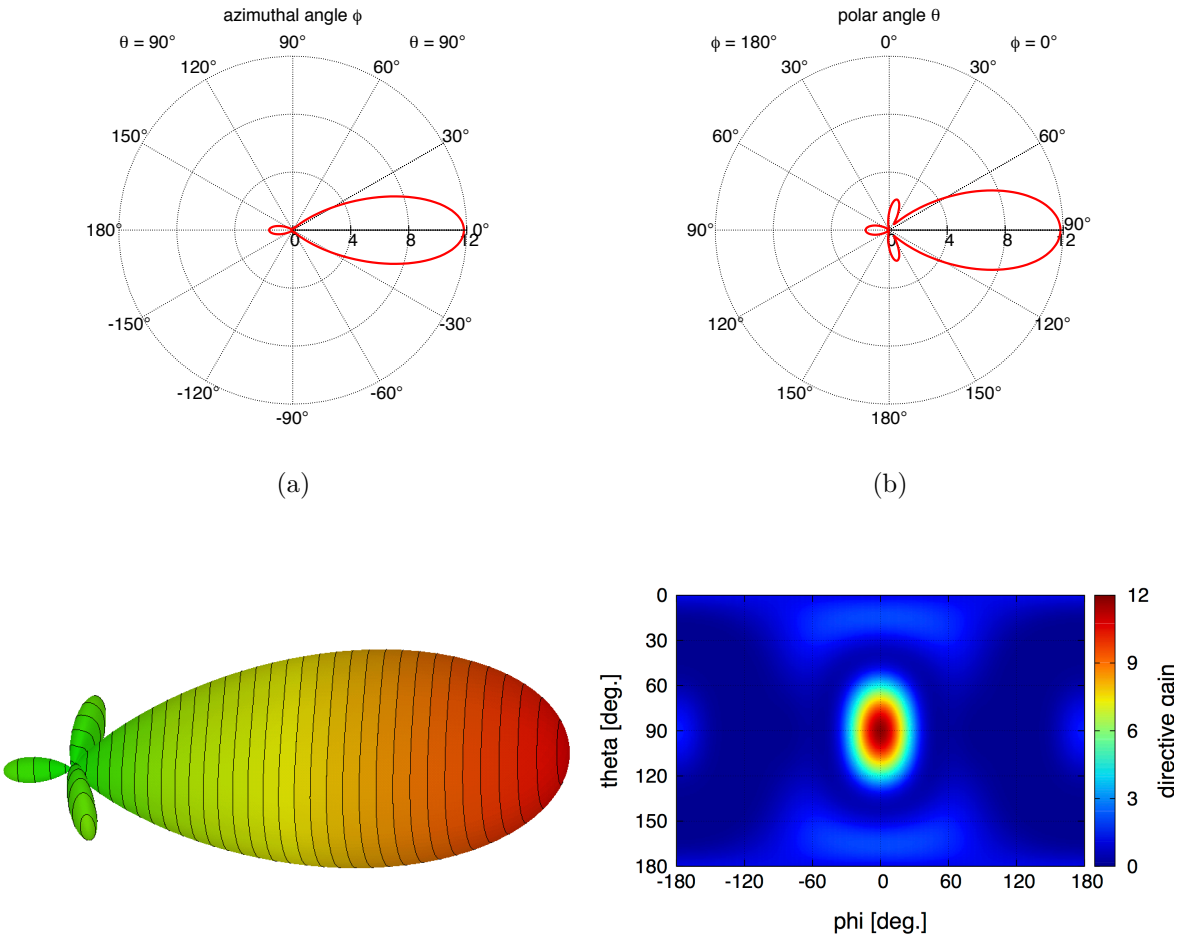
On one side of the feed the reflector is placed, which is slightly larger than the feed.

With a length slightly larger than the resonance length, the current on the reflector lags the induced voltage. Thus the impedance of the reflector becomes inductive. The opposite is the case for the directors, which are placed on the other side of the feed. They are smaller than the resonance length, so their impedance becomes capacitive. This arrangement causes a phase distribution to occur across the elements, simulating the phase progression of a plane wave across the array of elements. This leads to the array being designated as a so called traveling wave antenna. By choosing the lengths in this manner, the Yagi-Uda antenna becomes an end-fire array. Within the RF-regime the electromagnetic fields don't really penetrate into the metals, i.e. one can assume, that the currents are flowing on the surface of the wires, which simplifies all calculations significantly.

In Fig. 5.3 the two most simple Yagi-Uda arrays are shown. In a) the feed with a reflector and no director is illustrated, with the simulated directive gain in b). The same has been done for the feed with a director and no reflector present. While the reflector essentially reflects the radiation in backwards direction, but does not qualitatively change the radiation pattern in forward direction, the director clearly collimates the light beam, with only a slight decrease of the radiation in backward direction.

For a larger array, i.e. reflector, feed, and several directors, the lengths and distances have to be adjusted. Even in the RF regime, a closed expression for the design has not been found yet, so it is a typical optimization problem, dependent on the application, the frequency band, the used materials, the required space, etc. Thus over the years a vast amount of lookup tables have been published, where these parameters have been calculated or experimentally determined. One of the first is given in Ref. [101], which has been determined by extensive experimental studies.

The Yagi-Uda antenna sketched in Fig. 5.2 consists of a reflector, the feed and three directors. For parameters given in Tab. 5.1 the radiation pattern of the Yagi-Uda antenna looks like in Fig. 5.4. In a) and b) polar plots of the directive gain for the azimuthal angle ϕ and the polar angle θ in the far field are shown. One can see a nicely directed radiation towards the x-axis, with three small side lobes. The front-to-back ratio $r_{fb} = R_{fb} = 7.4$. In Fig. 5.4 c) a 3D radiation pattern without axes is shown, which is essentially the combination of the polar plot of the azimuthal angle for many polar angles. They often are shown to give an impression of the directivity. In d) a color plot of the directive gain over both angles is shown. Though not common, it is the preferred illustration of radiation patterns in this work, since it contains the most information. It might appear unintuitive, but one just has to keep in mind, that



(a) (b)

(c) (d)

Fig. 5.4.: *Directive gain of a Yagi-Uda antenna with three directors in polar plots for a) the azimuthal angle ϕ with constant polar angle $\theta = 90°$, which is actually the xy-plane, and b) the polar angle θ with constant azimuthal angle $\phi = 0°$, which corresponds to the zx-plane. In order to get a qualitative impression of the behavior, one might also plot the three dimensional radiation pattern in spherical coordinates as has been done in c), where the axes have been neglected. However, such an illustration is not that practical, if one is interested in the absolute values of the directive gain. A better illustration is shown in d), where a color map over both angles is given. Since such a plot contains more information than the former ones, it is preferred in this work.*

the angle $(\theta = 90°, \phi = 0°)$ points into the direction of the x-axis, whereas the angle $(\theta = 90°, \phi = 90°)$ points into the direction of the y-axis.

As stated before, the directivity can be increased with the number of directors, as is shown in Fig. 5.5. The red dots represent the directivities of antennas, which have been simulated with CST Microwave Studio. They consist of copper, with parameters taken

parameter		size in wavelengths
length of reflector	L_R	0.482
length of feed	L_F	0.45
length of director 1	L_{D1}	0.428
length of director 2	L_{D2}	0.424
length of director 3	L_{D3}	0.428
spacing between elements	d_i	0.2
radius of all wires	R	0.004

Tab. 5.1.: *Parameter set for a Yagi-Uda antenna with three directors taken from [101].*

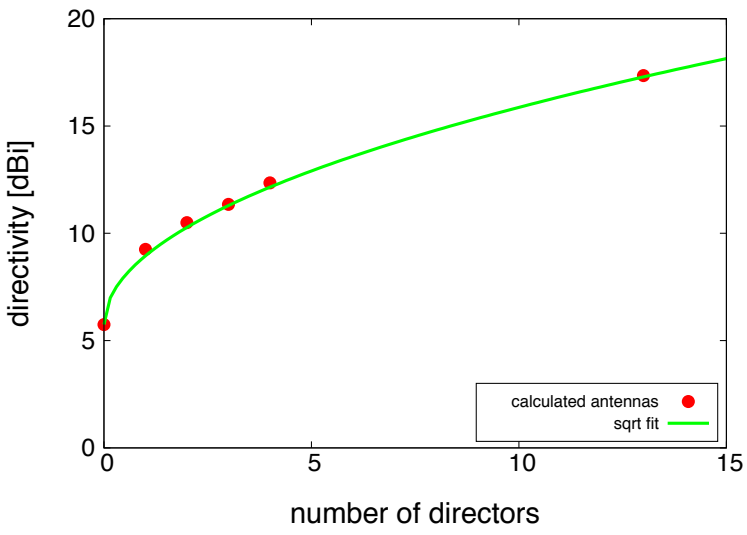


Fig. 5.5.: *Directivity in dBi of a Yagi-Uda antenna with varying number of director elements. A fit of a square root has been plotted, so one can can see, how the directivity behaves with increasing number of directors. Also increasing are the losses for real metals, so that a maximum power gain is achieved at about* $20dBi$.

from [101], which probably can be further optimized. One can see, that with increasing number of directors the directivity increases. Please note, that here the directivity is expressed in dBi, which is common within the antenna community. Additionally, a square root function has been fitted to the directivities of the different antennas. Since

in real metals losses cannot be avoided, the radiation efficiency decreases with increasing number of directors. So the power gain exhibits a maximum when using between 15 and 20 directors, depending on the material system and the detailed geometry used. Thus with this simple geometry a maximum power gain of about $G = 20dBi$ (or linearly scaled $G = 100$) has been achieved. Additionally, some other geometries have been applied, leading to even higher directivities.

To summarize this section, the principle of the Yagi-Uda antenna design has been discussed, together with a reference in the radio frequency regime. In the next section this concept is applied to near infrared antennas, where all dimensions are scaled down and adjusted. There are two major differences. Firstly, the structures become so small, that, in contrast to the RF regime, the electric field penetrates through the antenna and thus might excite plasmons. Secondly, the antennas are not free-standing anymore, but laying on a substrate, which probably causes a completely different radiation pattern.

5.2. Optical Yagi-Uda antennas

5.2.1. Optical Yagi-Uda antenna in vacuum

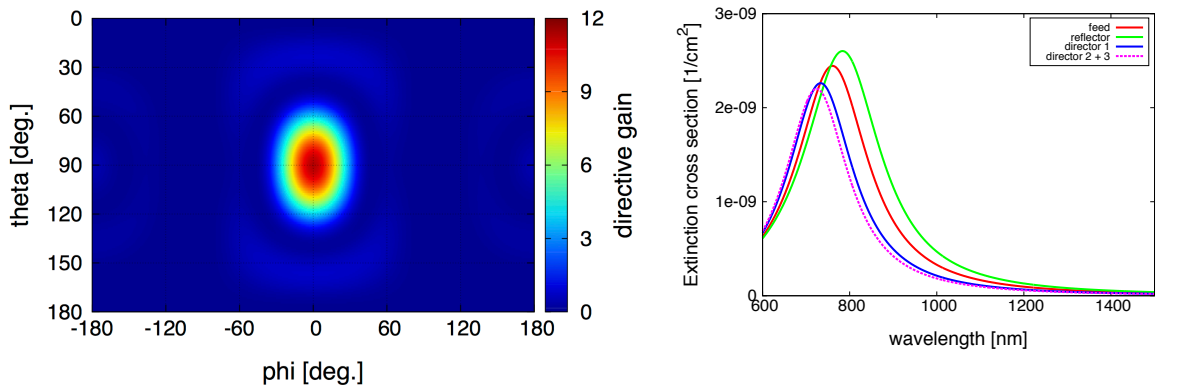
In chapter 5.1.2 the Yagi-Uda antenna in the radio frequency regime has been discussed. In this section the entire concept is transferred to plasmonic antennas, which operate in the near infrared. As mentioned before, the material behaves differently and such antennas cannot be fabricated free-standing, i.e. they usually lay on a substrate, which again changes the material behavior and modifies the radiation pattern of a dipole source.

It consists basically of a feed, which is resonant to the desired frequency, a reflector, which is slightly larger than the feed, and some directors, which are smaller than the feed. While the feed is driven, the other elements are parasitic, i.e. they are not externally excited. Responsible for high directive gains is the interference of the dipolar radiation of all elements. This is a hand-waving argument, though reasonable, but has to my knowledge not been proven so far, since analytical expressions as for the Hertzian dipole are not available due to the complexity of the antenna. Following this argument however, means, that one can create a Yagi-Uda like antenna by a collection of any particles, which have appropriate dipole resonances, which will be proven in this section.

Starting point is a five element Yagi-Uda like antenna, consisting of prolate spheroids, whose dipolar radiation has been calculated via the coupled dipole approximation, where all higher order moments have been neglected (c.p. chapter 2.5). Only the feed element has been excited with a vacuum wavelength of $\lambda_0 = 800nm$, thus the other 4 elements are parasitic. The material of the elements has been chosen to be gold, with the same Drude parameters as in previous chapters (see also in table A.1).

The solved equations for 5 coupled dipoles are given in appendix B, equation (B.1). They are placed along the x-axis, with their major axis parallel to the y-axis. The excitation light field has been polarized along the y-axis as well. Then the size parameters have been optimized, where the directive gain along the x-axis ($\theta = 90°$ and $\phi = 0°$) has been maximized (see table 5.2). According to Ref. [41] one can easily calculate the extinction cross section by equation (5.38).

$$C_{ext} = \frac{4\pi k}{|\boldsymbol{E}_0|^2} \sum_{j=1}^{N_p} \Im(\boldsymbol{E}^*_{incident,j} \cdot \boldsymbol{p}_j) \tag{5.38}$$



(a) (b)

Fig. 5.6.: *a) Color plot for the directive gain in spherical coordinates, with the polar angle θ and the azimuthal angle ϕ. b) Extinction cross section for each element (isolated) for the 5 element Yagi-Uda antenna.*

The resulting directive gain is shown in Fig. 5.6 a) in spherical coordinates. As one can see, this antenna works similarly good as the RF counterpart in the previous chapter, which is actually surprising. In order to get more insight to the antenna, in 5.6 b) the extinction cross sections of all elements (isolated) has been calculated. There one can see, that the resonance of the feed is slightly below the incident wavelength due to the compensation of the parasitic elements. As expected, the resonance of the reflector

is above of the incident wavelength, whereas the resonance of the directors is smaller compared to the feed. If one looks to the parameter set, one can actually see, that the directors, though almost identical in length, are separated differently. That the optimization process indeed converged, can be seen in Fig. 5.7, where the extinction cross section[3] of the whole antenna is plotted, showing a narrow peak at exactly the incident wavelength.

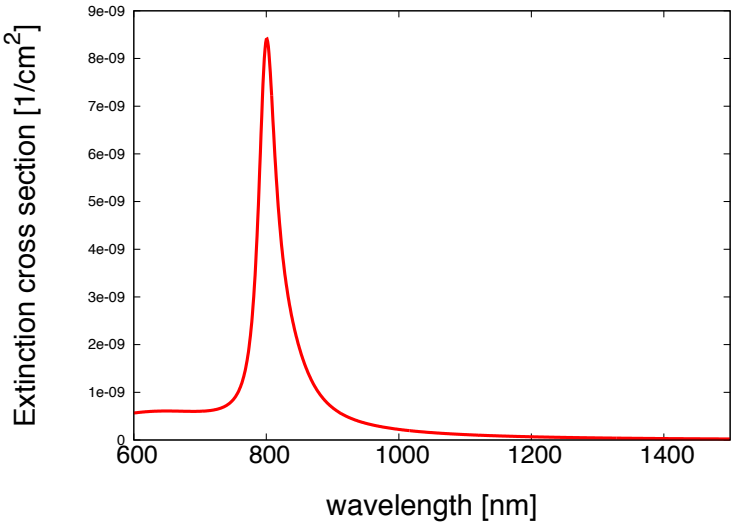


Fig. 5.7.: *Extinction cross section of the whole Yagi-Uda like antenna, consisting of 5 prolate spheroids.*

It is apparent then, that also with other particles, which exhibit a sufficient scattering cross section, Yagi-Uda like antennas can be built, even if the particles don't consist of metal, which will be shown later.

5.2.2. Plasmonic Yagi-Uda antennas on substrate

As a reference antenna we chose the one proposed by Curto et al. [61], because it was theoretically and experimentally well investigated. In Fig. 5.8 a) the structure of this antenna is illustrated, where the sphere indicates the location of the point source. In this publication the point source has been a core-shell quantum dot, which has been

[3]Usually the standard extinction cross section is calculated with the whole structure excited by a plane wave. In this case only one particle (the feed) has been excited, so that the extinction cross section should be treated carefully.

element	r_y [nm]	$r_{x,z}$ [nm]	separation [nm]
reflector	95	30	177
feed	90	30	
director 1	84	30	193
director 2	82	30	350
director 3	82	30	308

Tab. 5.2.: *Parameter set for the optimized spheroids building a Yagi-Uda like antenna. The radius of the major axis is r_y, and the radii of the minor axes are r_x and r_z, respectively.*

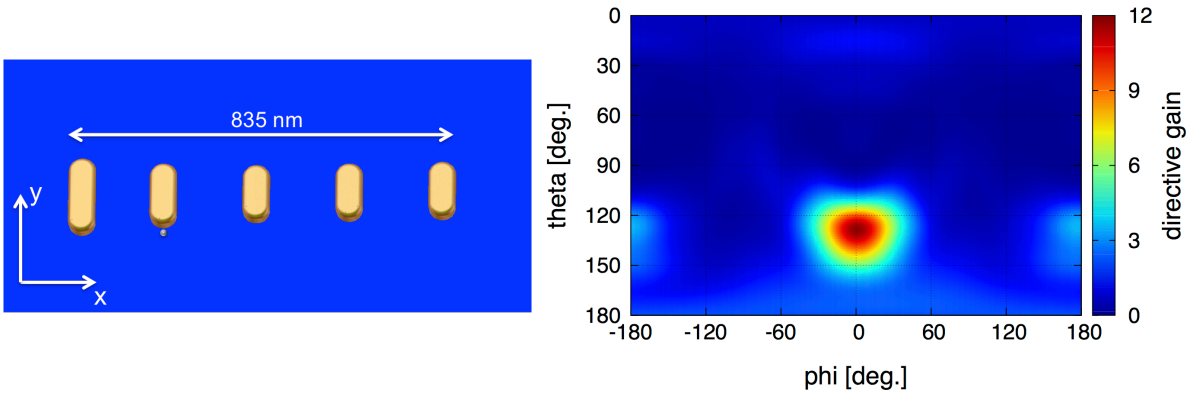
excited by a circularly polarized laser beam with a wavelength of $\lambda = 633nm$. After some relaxation the quantum dots emit light between $720nm$ and $880nm$.

All simulations are done with the commercial software CST Microwave Studio, using the time-domain solver (see chapter 2.2). In order to get the desired antenna properties, a near-field to far-field transformation is applied, taking into account the substrate, which is glass with $\varepsilon_r = 2.25$.

To model the material gold, a second order fit to the experimental data [38] is used, which perfectly matches the standard Drude model with appropriate parameters. In general the hexahedral mesh had 15 cells per wavelength, but for accurate results a local mesh refinement is applied, i.e. a non-uniform grid is used, where the cells in or near to the gold structure had a maximum size of $5nm \times 5nm \times 5nm$. Furthermore the adaptive mesh refinement has been used, where the mesh is locally refined at certain hot spots, until (according to defined criteria) the simulation is converged.

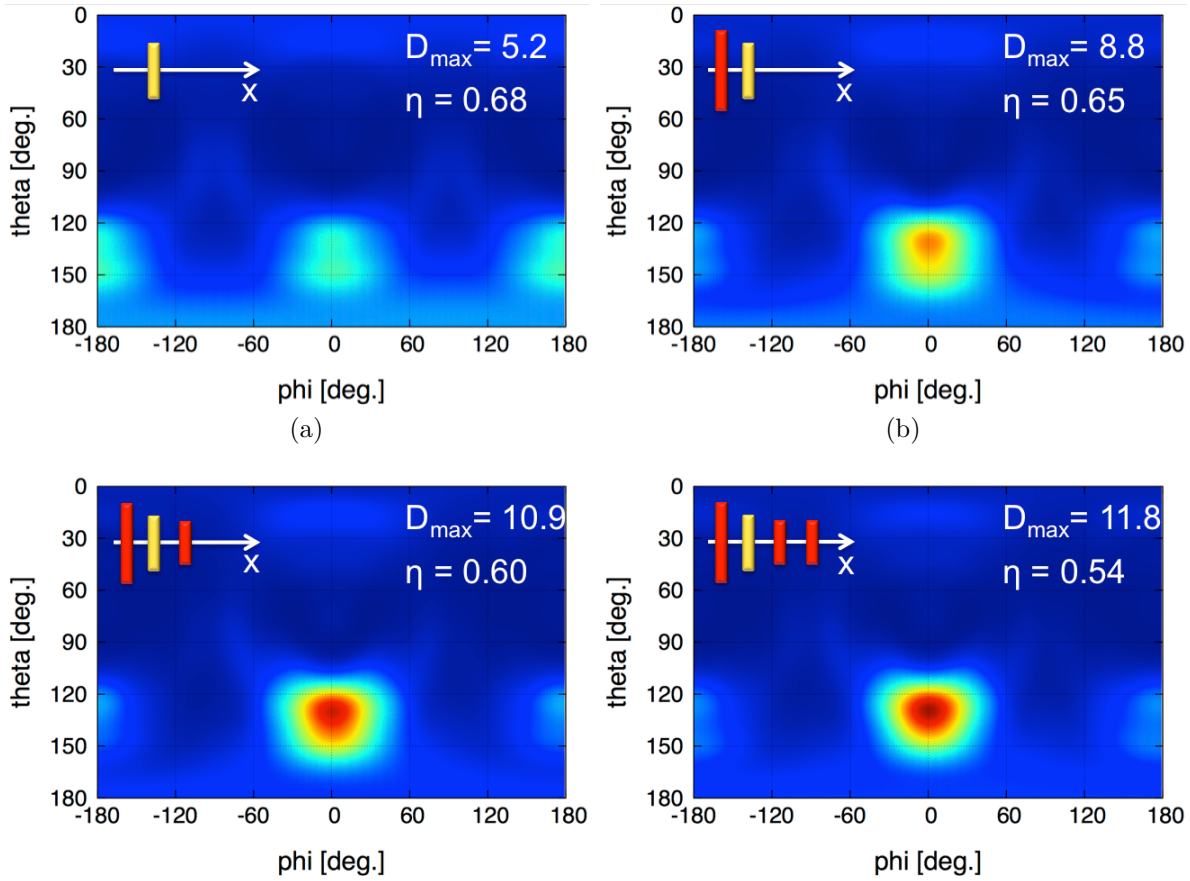
In Fig. 5.8 b) the calculated radiation pattern is shown. The maximum directivity of this antenna was $D = 12.1$ and the radiation efficiency was approximately $\eta \approx 0.5$ (own calculations) for an incident wavelength of roughly $\lambda = 900nm$. The front-to-back ratio, defined in equation 5.30 is then $R_{FB} \approx 6dBi$. The half power beam width angles are $\Delta\theta = 25°$ and $\Delta\phi = 74°$, which agrees with the publication.

In order to get more insight into the difference between Yagi-Uda antennas in optical and in RF regime, in Fig. 5.9 the radiation patterns of the building blocks of the antenna are shown. In a) only the feed remains, excited by a point source. The major difference to the RF regime is, that the radiation splits into two bulbs, in positive and negative x-direction. Also the radiation almost completely penetrates into the substrate, making it difficult to build a sending-receiving antenna system. In b) the reflector is added. That it indeed reflects the electromagnetic field and does not absorb



(a) (b)

Fig. 5.8.: *a) Illustration of the investigated antenna in Ref. [61]. The sphere indicates the position of the point dipole. b) Color plot for the directive gain in spherical coordinates, with the polar angle θ and the azimuthal angle ϕ.*



(a) (b) (c) (d)

Fig. 5.9.: *Directive gain in spherical coordinates with the polar angle θ and the azimuthal angle ϕ. The calculated antennas consist of a) only feed, b) a reflector and feed, c) a reflector, feed and one director, and d) a reflector, feed and two directors. The color scale is the same as in Fig. 5.8 b).*

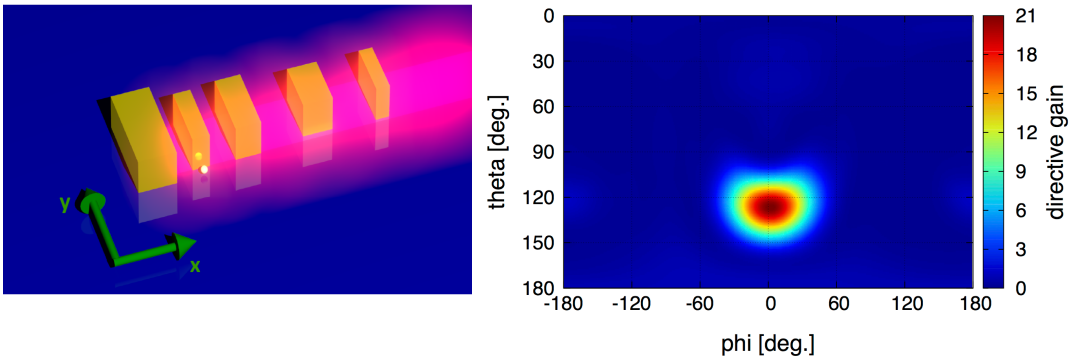
one half, can be seen when comparing the radiation efficiency and the directivity of both configurations. While the radiation efficiency slightly decreases, i.e. that a bit light is absorbed by the reflector, which is expected, the directivity strongly increases.

In c) and d) one additional director has been added to the system. One can see, that with each director the increase of the directivity decreases, while the radiation efficiency decreases as expected. One reason for this is, that the radiation penetrates into the substrate below the directors, i.e. that each added director becomes less relevant for the system. The second reason is, that this system has not been optimized. As discussed before, the presence of parasitic dipoles changes their resonance frequency, so at least the lengths have to be adjusted. This becomes clear, when the antenna with two directors in d) is compared with one, which has three directors (5.8 b)). While the directivity only slightly increases by adding the third director, the radiation efficiency decreases almost as much as was the case with less directors, which means, that the third director still interacts with the radiated field.

Though the directivity is comparable to RF antennas, the radiation efficiency is quite low, i.e. about the half of the emitted light is absorbed by the antenna. To emphasize this, if one looks to a sending-receiving antenna system, 75% of the emitted power would be lost within the antennas themselves, with additional losses neglected, which arise, when the beam width becomes larger then the dimensions of the receiving antenna. For strongly coupled quantum systems the energy transfer is of utmost importance, if they are spatially separated. This leads to the question, if such an plasmonic antenna can be improved, thus this structure has been optimized.

For the optimization process the *Particle Swarm Algorithm* is chosen, which is close to a genetic one. The algorithm is purely heuristic and is a so called global optimization algorithm. It is very robust, which means, that it is still working well, when the objective function has many local maxima and minima. Unfortunately it needs many function evaluations for convergence, but for our cause the robustness is far more important (see chapter 3.4).

We started with a 5-element Yagi-Uda antenna, where all parameters, determining the geometry, were optimized. For simplicity all elements had the same height. Several constrains were introduced, e.g. a minimum width of each element of $w = 35nm$ for an easier fabrication and a minimum distance between two elements to avoid numerical instabilities. The system was excited by a single electric point source, which was placed at the end of the feed element and 10 nm above the substrate. The electric field was y-polarized, which means along each element and perpendicular to the antenna axis.



(a) (b)

Fig. 5.10.: *(a) Optimization of a five element Yagi-Uda antenna consisting of gold leads to the shown antenna geometry. The quantum emitter is indicated by a small sphere above the substrate. The reflector is left, the feed element next to it. The three directors (right) don't have the same size and separation any more. (b) Directivity of the optimized gold antenna in spherical coordinates. The angle ϕ is in plane of the antenna axis, θ is orthogonal.*

Using 15 design parameters, the optimization process converged after approximately 800 iterations.

element	length [nm]	width [nm]	separation [nm]
reflector	232	102	150
feed	163	45	
director 1	174	71	146
director 2	131	79	208
director 3	136	35	153

Tab. 5.3.: *Parameters of the optimized Yagi-Uda antenna consisting of gold. The elements are listed from left to right, with the second element acting as feed element. The height of each element is 100nm.*

The resulting structure is illustrated in Fig. 5.10 a), the corresponding parameters are enlisted in Tab. 5.3. As expected, the reflector (left) is bigger than the feed element (next to the reflector), but the directors have different sizes and separation distances, which might be explained by the coupling of plasmonic modes in these elements. The directive gain of this antenna is shown in Fig. 5.10 b), where the directivity is roughly $D = 21$, which is almost twice as good, compared to the reference antenna. The radiation efficiency is around $\eta = 0.4$, so the losses are higher than these of the reference antenna, which is caused by the usage of more lossy material. Taking the power gain into account, the performance of this optimized antenna is still 40% higher than of the

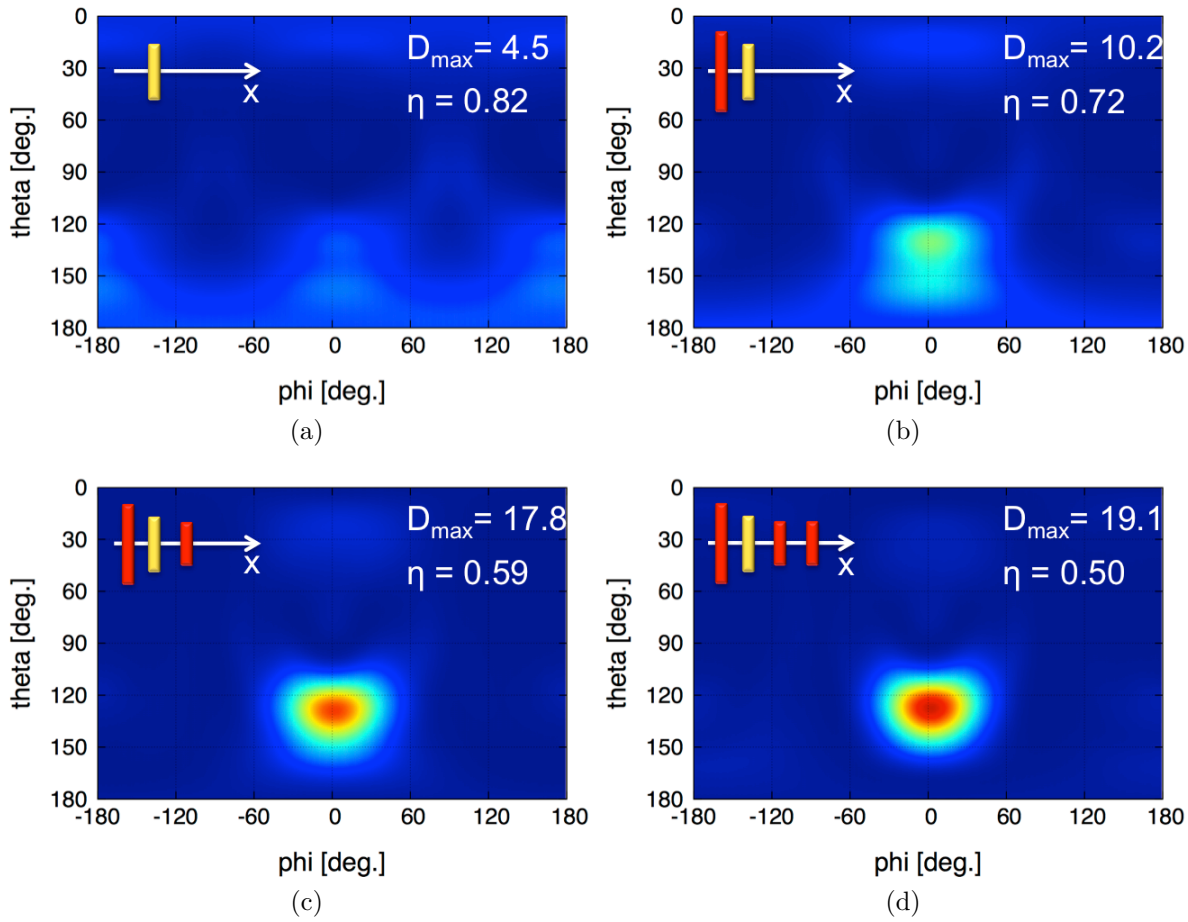


Fig. 5.11.: *Directive gain in spherical coordinates with the polar angle θ and the azimuthal angle ϕ. The calculated antennas consist of a) only feed, b) a reflector and feed, c) a reflector, feed and one director, and d) a reflector, feed and two directors. The color scale is the same as in Fig. 5.10 b).*

reference antenna.

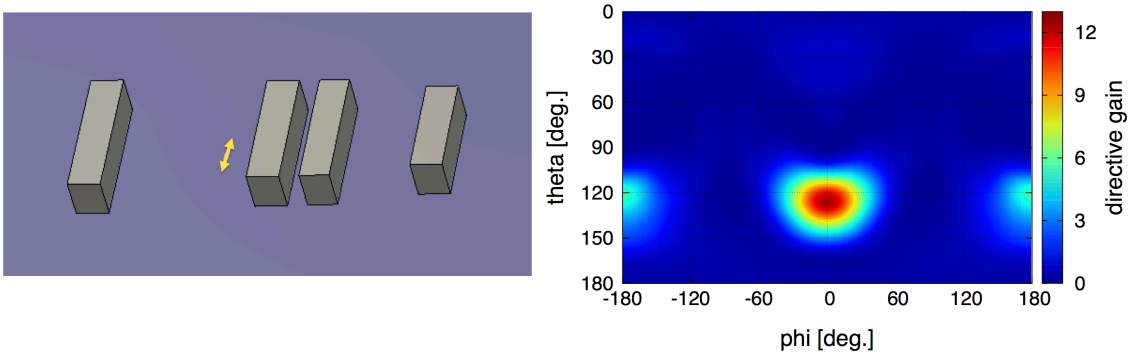
In Fig. 5.11 the radiation patterns of the cumulatively built antenna are shown (with the same color scale as the corresponding directive gain of the full antenna shown in Fig. 5.10 b)). In a) only the feed remains, exhibiting a typical, non-directed dipole radiation, where the dipole lays on the substrate. In b) one can see, that the left most elements indeed reflects the light and not just absorbs the light emitted in backwards direction. The radiation along the $-z$-axis decreases significantly due to the interference of the dipoles. By adding the directors in c) and d) the main bulb of the directive gain becomes more distinguished, but simultaneously the radiation efficiency decreases. Dependent on the application and the corresponding requirements one would probably use the three element antenna from c), because its directivity is already higher compared to

the reference antenna, but exhibits a higher radiation efficiency.

The directivity of the full optimized antenna is about $D = 21$ with a radiation efficiency of about $\eta = 0.4$. The half power beam width angles are $\Delta\theta = 30.3°$ and $\Delta\phi = 55.2°$. The front-to-back ratio $R_{FB} = 21.4dBi$, which is approximately three times as much compared to the reference.

5.2.3. Optical Yagi-Uda antennas using dielectrics

As shown in chapter 5.2.1, such antennas can be build with any particles, which exhibit a sufficient dipole resonance. To emphasize this point, in this section a Yagi-Uda-like antenna consisting of silicon is shown. Since this is not a plasmonic antenna, the point dipole does not couple to the feed element like before to get the Hertzian Dipole. For this reason the feed element has been replaced by the point dipole. Again the geometrical parameters have been optimized, leading to the structure in Fig. 5.12 a). The corresponding radiation pattern is shown in Fig. 5.12 b), with a directivity of $D = 12.5$. This seems to be not better than the reference antenna, but the radiation efficiency of this antenna is $\eta \approx 0.98$, which is nearly lossless. The parameters of this antenna can be found in table A.6.



(a) (b)

Fig. 5.12.: *(a) Optimization of a four element Yagi-Uda antenna consisting of silicon leads to the shown antenna geometry. The dipole moment of the point source is indicated by a yellow arrow. The reflector is left, the feed element next to it. The three directors (right) don't have the same size and separation any more. (b) Directivity of the optimized silicon antenna in spherical coordinates. The angle ϕ is in plane of the antenna axis, θ is orthogonal.*

In Fig. 5.13 the antenna is again shown decomposed, with the insets showing which elements have been simulated. The reflector does not significantly increase the direc-

tivity, which is not surprising, because of its large distance to the dipole source. The directors on the other hand significantly increase the directivity, without decreasing the radiation efficiency. However, they also increase the radiation in backwards direction, leading to a lower front-to-back ratio. Thus one can see a reflector for this antenna is sorely needed.

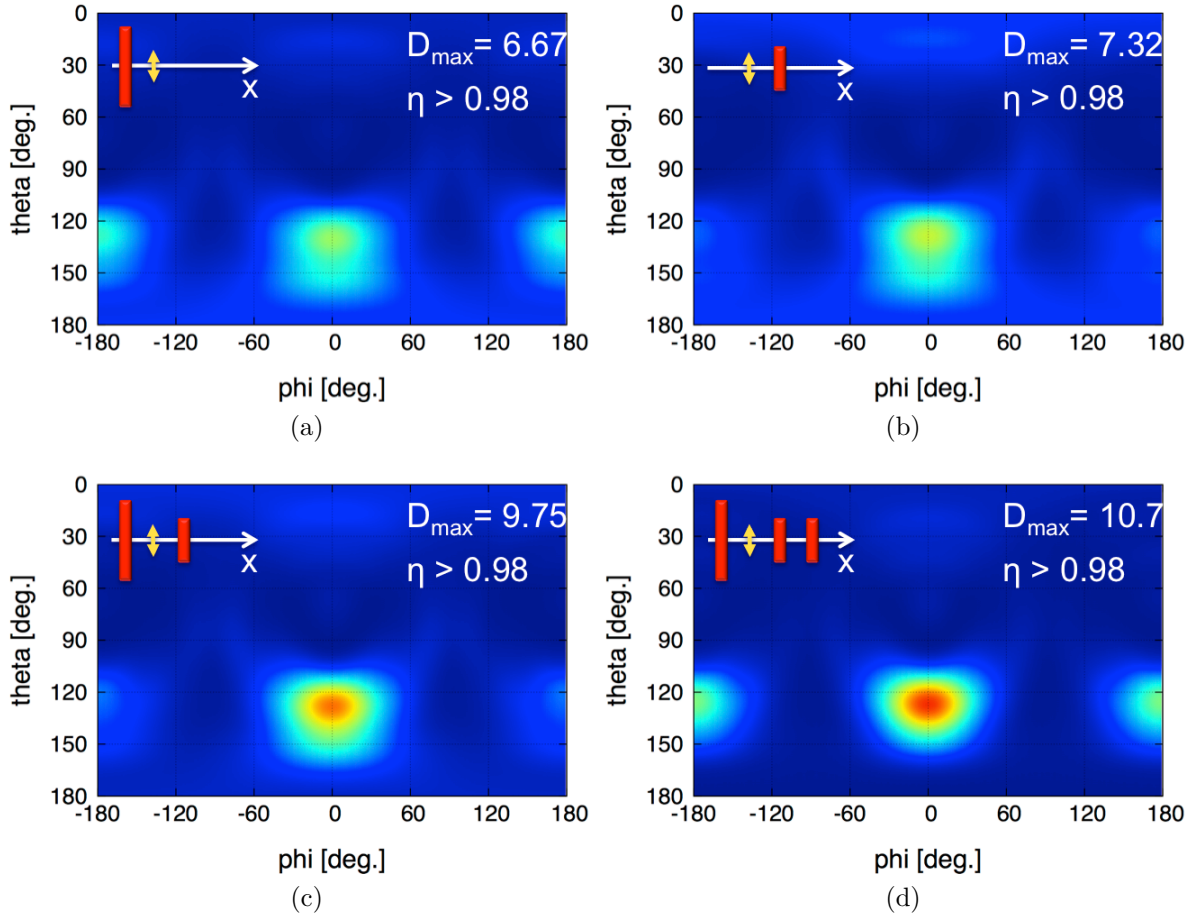


Fig. 5.13.: *Directive gain in spherical coordinates with the polar angle θ and the azimuthal angle ϕ. The calculated antennas consist of a) only reflector, b) only director, c) reflector and director, and d) reflector and two directors. The color scale is the same as in Fig. 5.12 b). Note, that the feed has been replaced by the point source, whose dipole moment is indicated by a yellow arrow.*

Please note, that the optimization might have found only a local optimum, so this antenna can probably further optimized. One can also replace the silicon reflector with a golden one, leading to hybrid antennas. Advantage of such is a better performance, disadvantage is the significantly increased difficulty in the fabrication of such structures.

5.3. Highly directive, all-dielectric optical antennas

5.3.1. A large, but high performance, all-dielectric antenna

As discussed in the previous chapter, Yagi-Uda antennas can be built by particles with fitting dipole resonances. In [84] a directional antenna was made by several silicon spheres of different sizes. Since it is hard to place a quantum emitter in such an antenna, we tried to design an antenna, which is similar to the plasmonic one, but made of another material. For that we chose *Hafniumoxide* (HfO_2), which is already established in the semiconductor industry, because it can be structured on small scales, is transparent for visible light and the dielectric function is almost constant in the visible ($\varepsilon \approx 3.57$).

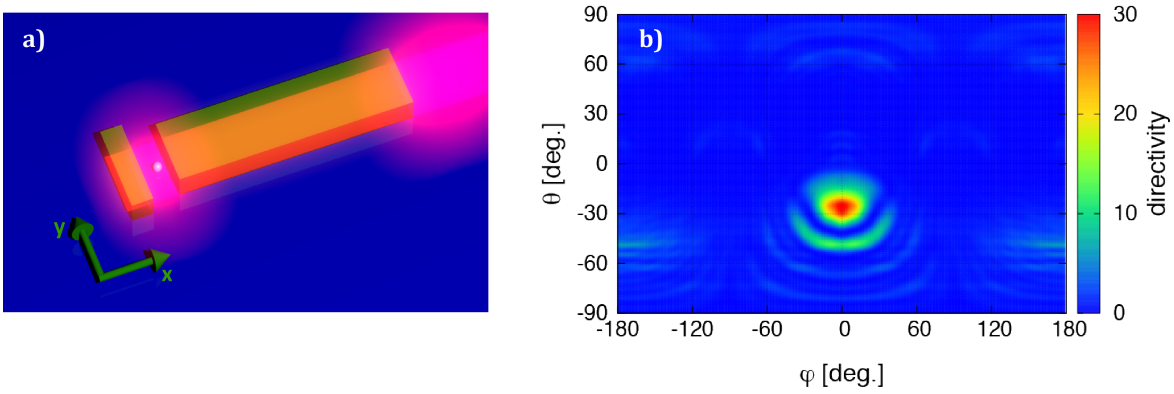


Fig. 5.14.: *(a) Optimization and simplification of a dielectric antenna leads to the shown structure, where the left element is the reflector and the right element acts as a director. The quantum emitter is indicated by a small sphere above the substrate. (b) Corresponding directivity of the optimized antenna in spherical coordinates. The angle ϕ is in plane of the antenna axis, θ is orthogonal.*

The setup for the optimization process including the design parameters were the same as before, only the material of the antenna was changed. Also the point source was placed between the two left elements, replacing the feed element of the plasmonic antenna (again $10nm$ above the substrate). After approximately 900 iterations the algorithm converged to a geometry, where all four elements on the right of the source had roughly the same size, with the minimum distance between them. So the geometry was further simplified to the one shown in Fig. 5.14 a), without loosing any performance.

The parameters are enlisted in tab. 5.4 b). As can be seen, the size of this dielectric antenna is more than twice the size of the optimized gold antenna, but the performance

is dramatically increased as well. The directive gain of this antenna is shown in Fig. 5.14 b), with a directivity of $D = 30.3$ and a radiation efficiency of $\eta \approx 1.0$.

element	length [nm]	width [nm]	separation [nm]
reflector	786	180	220
director	600	2200	40

Tab. 5.4.: *Parameters of the optimized, all-dielectric antenna. The first element is placed to the left of the source, the second element to the right. The height of the dielectric antenna is 180nm. The polarization of the source has been parallel to the y-axis, which is perpendicular to the antenna axis.*

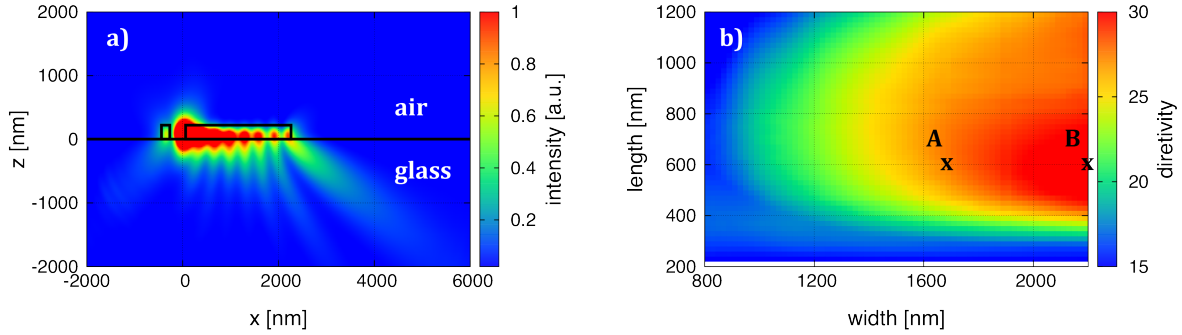


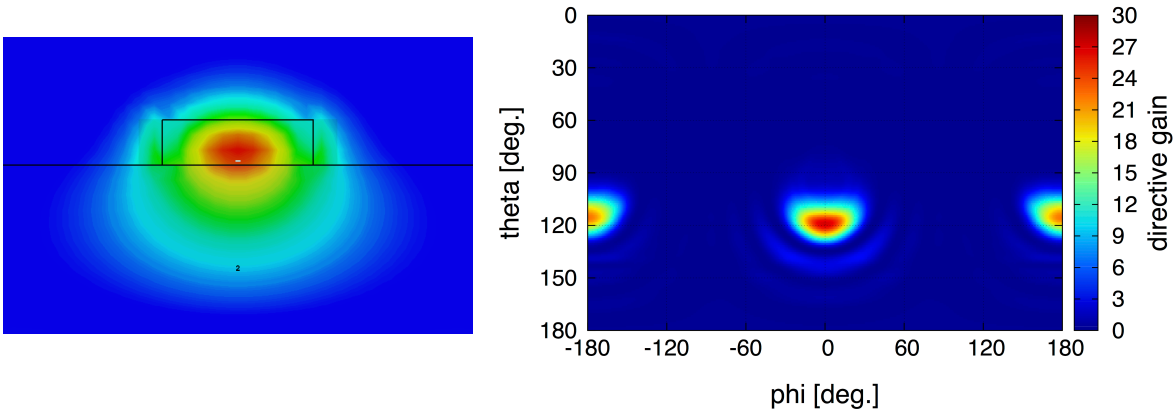
Fig. 5.15.: *In a) a snapshot of the absolute value of the electric field is pictured. One can see, that most of the light is traveling along the interface of substrate and antenna element, before it is scattered at the end of the antenna. The appearance of the rings in Fig. 5.14 b) around the main lobe can be seen as well. In b) the directivity of the dielectric antenna is shown in dependence of length and width of the element to the right of the source.*

In order to understand this antenna, the absolute value of the electric field at a wavelength of $\lambda = 900nm$ is plotted in Fig. 5.15 a). This is obtained via a Fourier transform in every point of the simulation grid, where the desired frequency component is picked. In this plot one can see, that the small element indeed acts as a reflector, whereas the huge element is more like a lossy waveguide. The electric field propagates along the interface between substrate and antenna element, until it is finally scattered into the substrate. While traveling along this interface one can see, that some light is coupled out in equidistant periods, corresponding to the half of the wavelength of the light inside of the material. Each of these spots is directed in a slightly different angle, which explains the half-rings in the directivity of this antenna (Fig. 5.14 b)). However, though clearly visible in this plot, these spots are small compared to the main lobe scattered at the end of the element.

In Fig. 5.15 b) the maximum directivity is plotted in dependence of length and width of the big element. Increasing the width of this element leads to a smooth increase of the directivity. One consequence of this is, that the antenna is robust against fabrication tolerances. Another one is, that the size of this antenna can be reduced without loosing much of the performance as well.

The spots of the waveguide, where some energy is coupled into the substrate can be seen as new dipole sources, which interfere with each other. Thus this leaky waveguide works quite similar to a classical Yagi-Uda antenna, where several dipole sources (caused by the polarizability of the corresponding particles) in different positions - and thus also with different phase - interfere with each other.

To prove this point, the HfO_2 waveguide has been replaced by a number of point dipoles, achieving a comparable directive gain. The number and positions of those point dipoles can be read out from Fig. 5.15 a). The necessary phase difference is then $\varphi = \beta x_i$, where x_i denotes the distance of each source to the primary one and β is the propagation constant for the leaky waveguide mode. To obtain the propagation constant a 2D eigenvalue solver has been applied, additionally calculating the mode profile, which is shown in Fig. 5.16 a). The resulting parameters of the dipoles are listed in Tab. A.7, with a calculated propagation constant $\beta = 1.0239420 \cdot 10^7 1/m$.



(a) (b)

Fig. 5.16.: *a) Electric field strength for the mode profile of calculated leaky mode in* HfO_2 *waveguide (linear scale), leading to propagation constant* β*. b) Resulting directive gain, where the* HfO_2 *waveguide has been replaced by 7 coherent dipoles.*

In Fig. 5.16 b) the resulting directive gain is shown in spherical coordinates. The main lobe as well as the side lobes (or rings) look even quantitatively like the radiation pattern of the hafnium oxide antenna, though the radiation in backward direction here

is much stronger. Please note, that this model is at best an estimation, since the positions and the phase differences are roughly estimated. Still one can deduce, that this leaky waveguide is not so different to a classical Yagi-Uda antenna, since it consists of a number of interfering dipoles.

Apart from that a major problem for optical antennas has been overcome here. Such small antennas cannot be designed as free standing Yagi-Uda antennas as is possible in the RF regime, but have to placed on a substrate. The interfering dipoles then emit light primarily into the substrate - not directly to the other dipole elements. As a consequence each added director will become less relevant for the antenna, since there is not much radiation left, which can excite this added director. That means, that for plasmonic antennas one gets no increased performance after placing three or four directors. However, using this leaky waveguide, the main energy is following the interface between glass and hafnium oxide, so adding some additional length increases the directivity as much, as is the case for RF antennas. Additionally, this antenna exhibit no losses at all, i.e. the power gain is more than 5 times better compared to the reference antenna. Also this antenna is less sensitive regarding fabrication tolerances, since the propagation constant does not change dramatically with the variation of the geometrical parameters.

In Fig. 5.17 the maximum directivity over the wavelength for three different antennas can be seen. The optimized plasmonic antenna (blue) works best for the wavelength of $\lambda = 900nm$, which is a good sign, that the optimization algorithm has found the global maximum. The FWHM of this antenna is around $\Delta\lambda_{FWHM} = 150nm$. The other two lines correspond to the dielectric antenna with two different lengths of the big element, indicated by the marks A and B in Fig. 5.14 b), both working best for the wavelength $\lambda = 830nm$. The smaller variant of this dielectric antenna (red) has a much broader peak ($\Delta\lambda_{FWHM} = 550nm$), than the optimized plasmonic antenna. The bigger variant (green) is working for even a broader variety of wavelengths, there doesn't seem to be a peak for smaller wavelengths any more, which makes it a perfect broadband antenna.

5.3.2. Experimental verification of the HfO_2 antenna

The fabrication of the antenna with the parameters in 5.4 has been realized by a two-step lithography by cooperation partners. First[4] the HfO_2 structure has been deposited

[4]This step has been done by Christian Schlickriede, who works in the group of Prof. Dr. Thomas Zentgraf, Paderborn University.

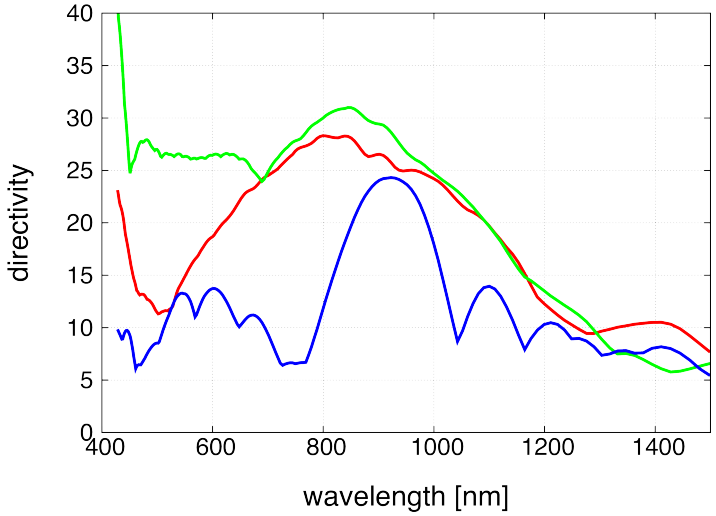


Fig. 5.17.: *The spectra of the optimized plasmonic (blue) and two realizations of the optimized dielectric antenna, one with a shorter (red) and one with a longer (green) second element, are pictured.*

on glass as described in chapter 4.5. In the second step[5] a layer of PMMA has been placed on the sample. Then the area, where quantum dots should be deposited, has been defined by the exposure of a focused electron beam, and the PMMA layer there has been removed. After functionalizing this area, CdSe core-shell quantum dots have been deposited, building a link to the substrate via the functionalizing, so that they stay at their respective position, when the PMMA layer has been removed by a lift-off process. A more detailed description of this fabrication process can be found in [102]. The resulting structure is shown in Fig. 5.18, where the quantum dots are not resolved, but the area where they are deposited is indicated by the red circle.

The quantum dots are then pumped with blue light and, after some relaxation, emit light around $\lambda = 780nm$, with a radiation pattern comparable to a point dipole. The emitted light has been focused via an objective, which has a numerical aperture of $NA = 1.49$, and detected by a standard ccd-camera. As a result of this setup, only radiation propagating into the substrate ($\theta > 90°$) could be detected. In Fig. 5.19 one can see the measured fluorescence intensity in spherical coordinates. One can clearly see a directive behavior, where much more energy is directed along the positive x-direction ($\phi = 0°$) than along negative x-direction ($\phi = \pm 180°$). The narrow main bulb is prominent, as are the first two rings below it. Also a relatively bright ring for higher polar angles can be seen, which is caused by quantum dots, which do not couple

[5]This second step as well as the measurement itself has been done by Manuel Peter, who works in the group of Prof. Dr. Stefan Linden, Bonn University.

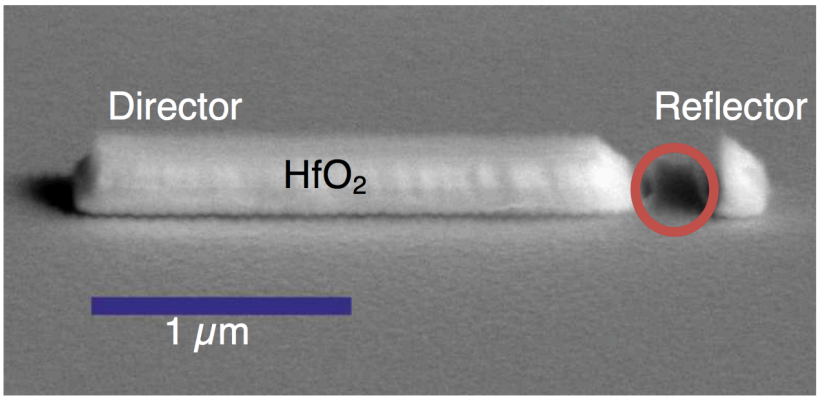


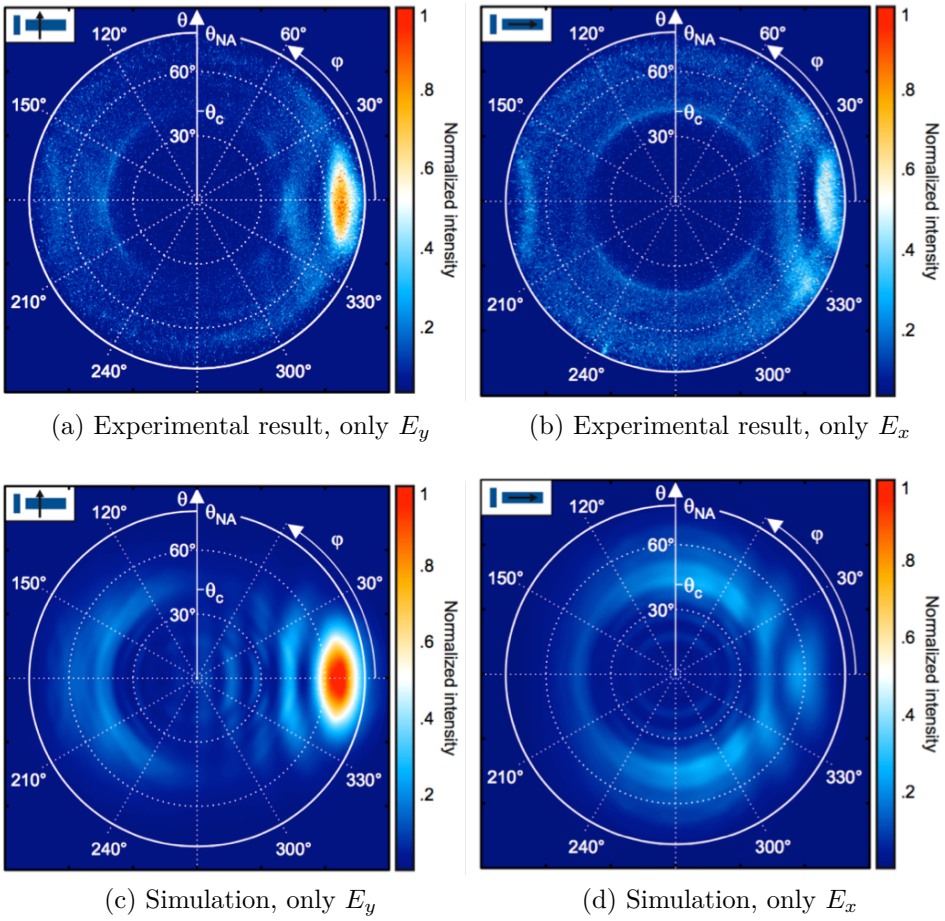
Fig. 5.18.: *Electron micrograph of hafnium oxide antenna by Manuel Peter, Bonn University. The red circle indicates the spot, where some quantum dots are deposited.*

to the leaky wave guide mode. Those quantum dots have been deposited far from the desired position or exhibit a dipole moment parallel to the antenna axis.

It seems, that the ratio of the intensity of the main bulb and the first ring is rather small compared to the theoretical results. This is caused by the fact, that only a part of the main lobe is actually detected. For smaller polar angles the intensity is definitely higher, but the critical angle is reached, which means, that instead of being detected by the camera, this part of the main lobe is subject to total internal reflection. The ideal critical angle $\theta_c = 111°$ is indicated by the white circle, which is determined by product specifications of the objective.

The main difference to former simulations is the fact, that colloidal quantum dots have been used, which means that they have equal distributions of dipole moments pointing in x-, y-, and z-direction. Thus the antenna, which has been designed for a point dipole pointing in y-direction, loses some of its performance, but the unidirectional behavior is still prominent. Also apparent is the good agreement between simulation and experiment.

That this antenna is indeed robust against fabrication tolerances can be seen in Fig. 5.20, where the same procedure has been used to fabricate a structure, where the geometrical parameters (except the height) are scaled by a factor of a) 0.8 and b) 1.4. Though the performance decreases significantly, one can still see all features of this structure, including the directional properties.



(a) Experimental result, only E_y

(b) Experimental result, only E_x

(c) Simulation, only E_y

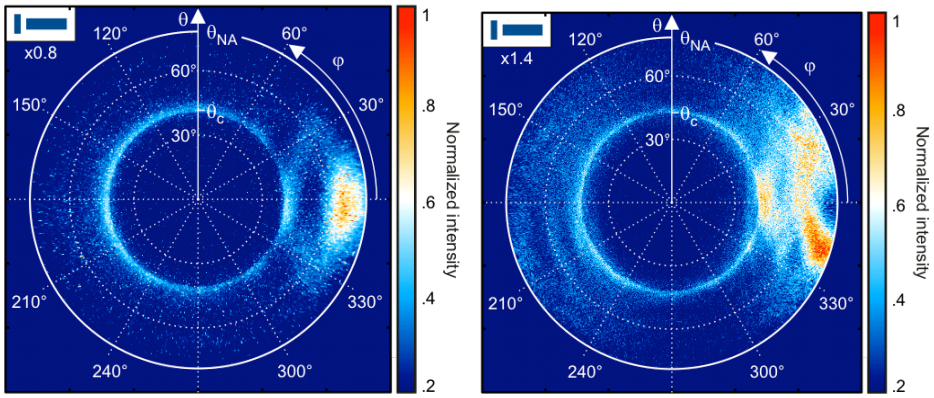
(d) Simulation, only E_x

Fig. 5.19.: *Angular distribution of the normalized quantum dot fluorescence guided by the antenna. The angular $\theta\prime$ is defined in spherical coordinates by $\theta\prime = 180° - \theta$, which is zero in the origin. In a) and b) the normalized intensity of the E_y and E_x components of the experimental measurements are shown. For comparison the numerical simulations are shown in c) and d). The white circle indicates the ideal numerical aperture. For larger values of $\theta\prime$ the light is reflected back into the substrate and cannot be measured.*

5.3.3. A one-material, all-dielectric directional antenna

In the studies of the previous section, only glass was considered as a substrate, but using different substrates is of interest, since the choice of the substrate depends on the growing of an appropriate point source. In our theoretical investigations we are not restricted to use certain materials, so we placed this antenna on a substrate with varying relative permittivity, without changing the antenna itself.

In Fig. 5.21 the maximum directivity over the permittivity of the substrate (red) and the corresponding vertical angle of the main lobe of the radiation (blue) are plotted.



(a) Dimensions scaled by factor of 0.8, experimental result

(b) Dimensions scaled by factor of 1.4, experimental result

(c) Dimensions scaled by factor of 0.8, simulation

(d) Dimensions scaled by factor of 1.4, simulation

Fig. 5.20.: *Angular distribution of the normalized quantum dot fluorescence guided by the antenna. Essentially the same as in Fig. 5.19. The lateral dimensions of the antenna have been scaled by the factor of 0.8 in a) and c), and by the factor of 1.4 in b) and d).*

One can see, that there appear some minor resonances, but at least for $\varepsilon_r > 2$ the directivity never drops below a value of $D = 19$. As expected, with increasing permittivity the main lobe angle of the radiation decreases. At one point ($\varepsilon_r = 3.57$) the substrate and the antenna have the same relative permittivity, but still a strong directivity ($D = 20$) can be seen. That means, that such an antenna, where the antenna is structured into the substrate, might work for all substrates, with the right geometry parameters.

For comparison we again chose the plasmonic antenna proposed in Ref. [61], designed for a wavelength of $\lambda = 900nm$, as a reference, as it was theoretically and experimentally well investigated. The maximum directivity of this antenna was $D = 12.1$ and

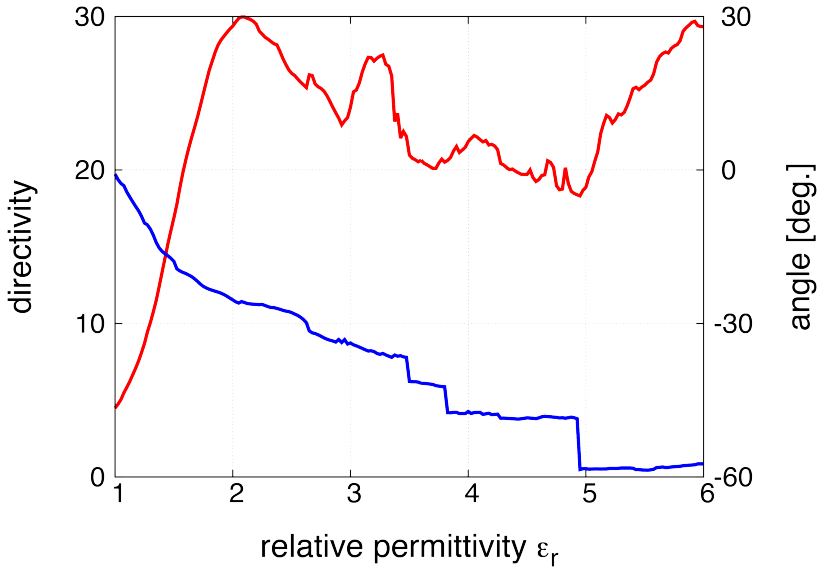


Fig. 5.21.: *HfO2 antenna from Fig. 5.14 on a substrate with varying relative permittivity. One can see that the directivity (red) never drops below $D = 19$ for a permittivity $\varepsilon_{substrate} \geq 1.5$. Also with increasing permittivity the main lobe angle decreases as expected, i.e. the more the substrate becomes optically dense, the steeper the radiation penetrates into the substrate.*

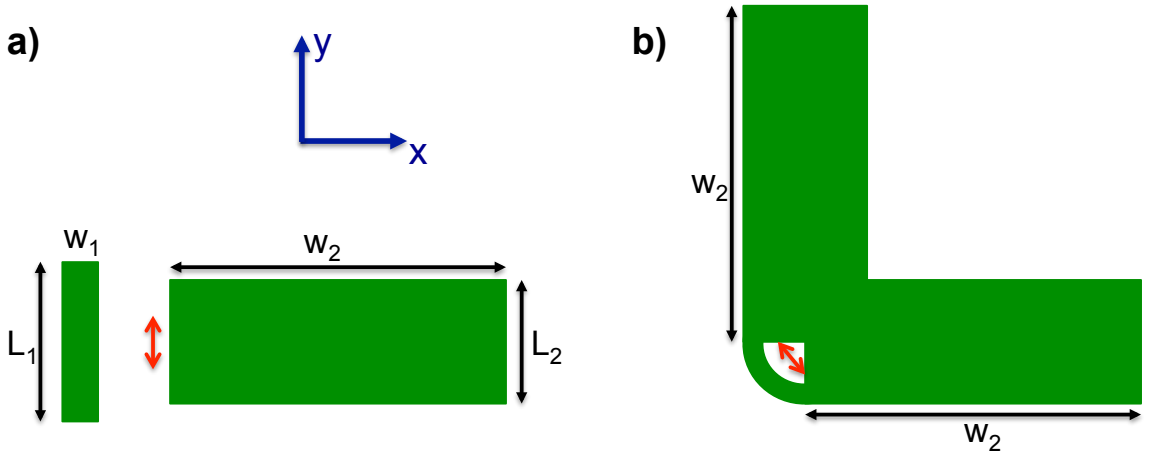


Fig. 5.22.: *Sketch of the top view of investigated GaAs structures placed on GaAs substrate. a) Antenna consisting of a small brick, acting as a reflector, and a large brick, adding to the directivity. The point source, indicated by a red arrow, is polarized parallel to the y-axis. b) Antenna using two orthogonal, large bricks as in a) to direct emitted light in x- and y-direction. The reflector consists then of a hollow, quarter cylinder. The point source, again indicated by a red arrow, might point anywhere in the x-y-plane.*

the radiation efficiency was approximately $\eta = 0.5$ (own calculations). In Fig. 5.22

a) a sketch of the structure (top view) is shown with all necessary parameters for the optimization process.

In contrast to the reference GaAs was chosen for both, substrate and antenna material ($\varepsilon_r = 12.9$), and the parameters were optimized for a wavelength of $\lambda = 900nm$ (which corresponds to the smallest emission wavelength of InGaAs quantum dots). For the optimization process again the *Particle Swarm Algorithm* [55] is chosen, which is similar to a genetic algorithm. The method is purely heuristic and is a so called global optimization algorithm. It is very robust, which means, that it is still working well, even if the objective function has many local maxima and minima. It needs many function evaluations for convergence, but in this case the robustness is far more important. Due to constraints the parameter space becomes strongly non-convex, comparable to Swiss Cheese, thus most simplex methods like the CRSLM would have difficulties. The resulting parameters are enlisted in Tab. 5.5. Note, that this parameter set might correspond to a local optimum and maybe can be further optimized.

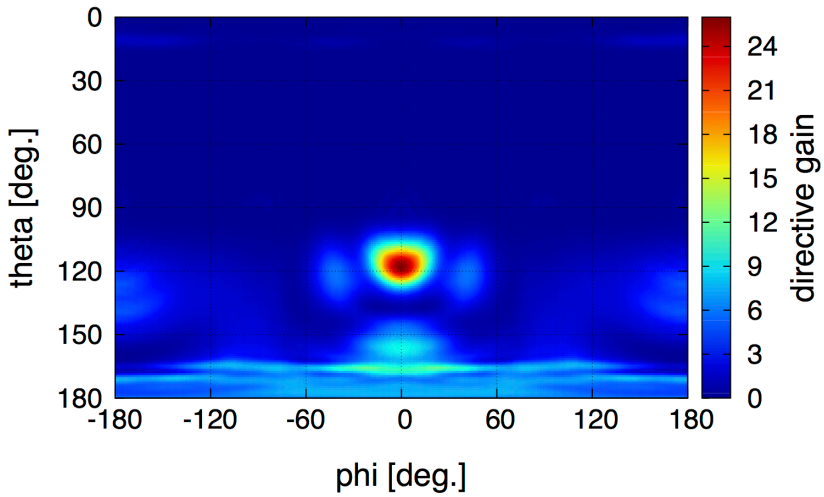


Fig. 5.23.: *Angular resolved directive gain (far field) of structured substrate, as sketched in Fig. 5.22 a), in spherical coordinates, where ϕ is the azimuthal and θ the polar angle.*

In Fig. 5.23 one can see the resulting directivity in spherical coordinates in the far field. The main lobe points into the x-direction with a polar angle of $\theta = -27$ degree (down into the substrate). This angle is surprisingly small, since the main lobe angle of a simple dipole source on a GaAs substrate would be around $\theta = -70$ degree. A part of this behavior still can be seen, for the directivity has a side lobe for $\theta = -75$ degree. Still the main lobe exhibits a directivity of $D = 25.7$ with an angular width of $\Delta\phi = 29$ degree of the azimuthal and $\Delta\theta = 17$ degree of the polar angle, which is

element	length [nm]	width [nm]	separation [nm]
reflector	408	210	90
director	530	1400	20

Tab. 5.5.: *Parameters of the optimized, GaAs antenna. The first element is placed to the left of the source, the second element to the right. The height of the dielectric antenna is* 140*nm. The polarization of the source has been parallel to the y-axis, which is perpendicular to the antenna axis.*

very narrow compared to known plasmonic Yagi-Uda like antennas, such as discussed in [61].

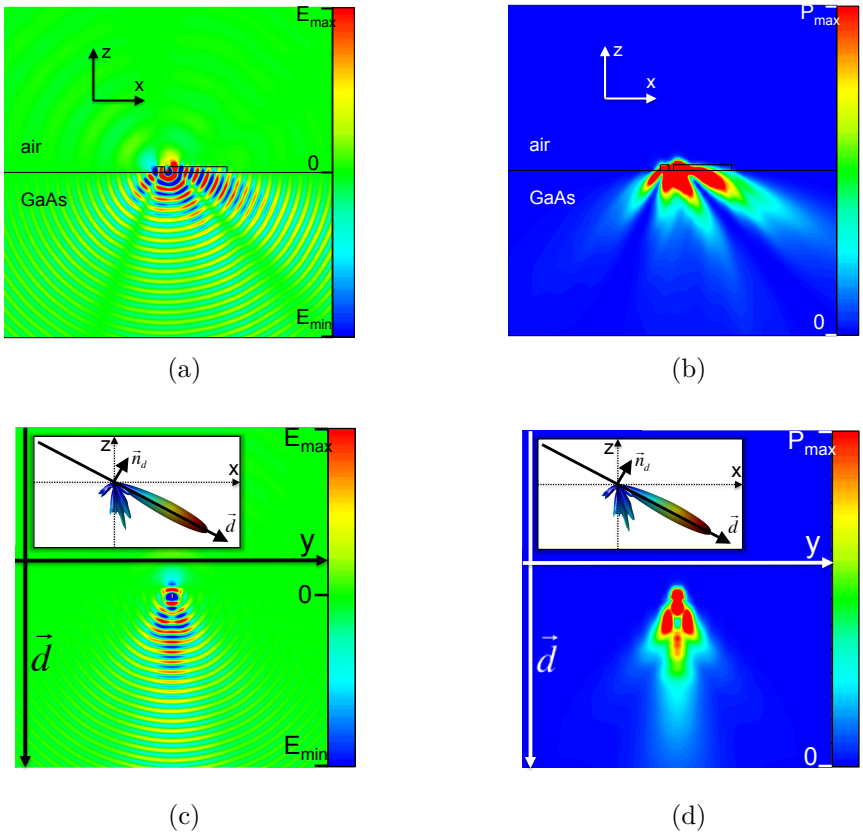


Fig. 5.24.: *2D cross sections of the simulation space showing a) the electric field component E_y and b) the power flow $|\vec{S}|$ in the x-z-plane, and same quantities in a plane parallel to the main lobe in c) and d).*

In order to explain this radiation pattern both the electric field component E_y and the absolute value of the Poynting vector has been plotted in Fig. 5.24. In a) and b) the x-z-plane is shown. Since the system is excited by a point source, the power flow reaches the saturation close to the origin. There one can see four paths, where almost

the whole amount of energy is directed to. Two of them point into the $-z$-direction, whose interference leads to the noisy appearance of the directive gain for large polar angles ($\theta > 160°$). Another path is directed basically in negative x-direction, but with larger polar angle ($\theta \approx 130°$). The main lobe, however, contains the strongest power flow. The high directivity of this antenna corresponds to a very narrow opening angle both in θ and in ϕ.

In order to have a more detailed look, again the electric field component E_y and the absolute value of the Poynting vector are plotted in Fig. 5.24 c) and d), but this time in a plane parallel to the main lobe, with the plane normal $\vec{n}$. The insets show the 3D radiation pattern of the structure and the corresponding coordinate system. There one can see, that the light is directed in two directions ($\phi \approx \pm 20°$), leading to an interference (akin to a double slit experiment), where the effective radiation is directed at $\phi = 0°$. However, small residues of both directions can also be seen in the radiation pattern (Fig. 5.23). It is safe to assume at this point, that no resonances of either the materials or the geometries have been used. Instead this antenna works by using simple interference and the habit of light to use the paths with the optical densest materials. A consequence of this is, that such an antenna works for any material system at any wavelength, given the right geometrical parameters.

Since the far field results consist of two simulations, where in one simulation the far field for polar angles $\theta > 90°$ and in the other one the far field for $\theta \leq 90°$ have been calculated, the directivity has been checked by obtaining all quantities using the near field power flow. Both approaches yielded the same results.

Also interesting is the spectral performance of such a structure. In Fig. 5.25 the maximum directive gain, also called directivity, in dependence of the incident wavelength is shown for this structure (red) and the plasmonic reference antenna (green). One can see, that the optimization algorithm indeed did not find the global maximum, since this structure has the highest directivity for a wavelength of $\lambda = 750\text{nm}$ with a directivity of $D = 29$. Please note, that losses of the material have not been considered in Fig. 5.25, which would increase significantly for smaller wavelengths due to the band gap of GaAs. The point here is, since the Maxwell equations are scalable with the wavelength, that the directivity of $D = 29$ should be easily obtainable in the NIR region as well. The maximum directivity never drops below $D = 19$ in the shown region without a significant variation of the main lobe direction, which is necessary for a proper broadband antenna. Further numerical calculations show that simulated fabrication tolerances like rounded edges do not change the performance significantly.

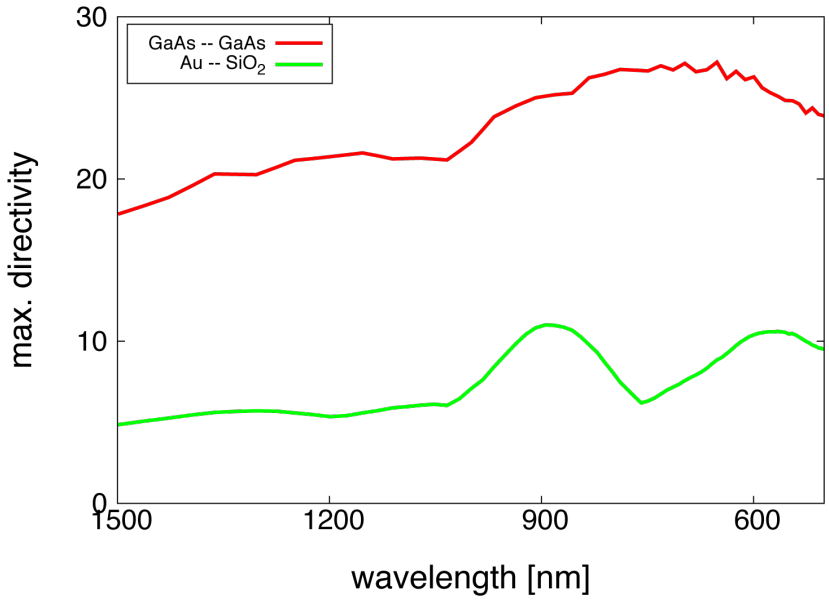
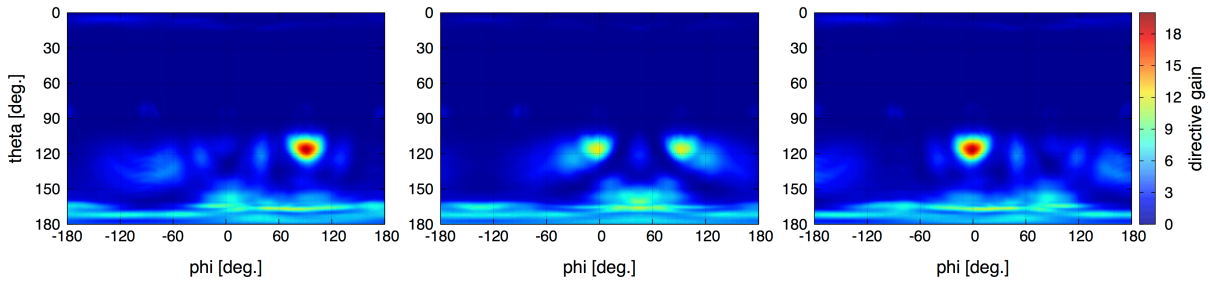


Fig. 5.25.: *Dependence of the maximum directivity on the radiation wavelength for the plasmonic Yagi-Uda antenna on glass (green), and the structured substrate (red).*

A logical continuation of this approach is the combination of two such antennas, which are orthogonally aligned, such as in Fig. 5.22 b). Since the reflectors would have crossed and therefore drastically decreased the performance, they have been replaced by a quarter of a hollow cylinder, with an outer radius of 265nm and an inner radius of 95nm. Any polarization in the x-y-plane can be decomposed in a x- and y-polarized wave, each directed into the orthogonal direction by the corresponding antenna arm. In Fig. 5.26 the directive gain for three different incident polarization angles are shown. In a) the electric field is x-polarized, leading to a directivity of $D \approx 19$ into the y-direction. The opposite occurs for a y-polarized incident field. Exciting this antenna with a polarization vector along the angle of $\phi = 45$ deg. (exactly between x- and y-axis), the radiated power gets symmetrically split up and directed into both directions.

Unfortunately, the maximum directive gain drops to slightly above $D = 19$, which is mostly caused by the compromise to obtain one reflector for all polarizations. This also results in the increased (unwanted) emission towards high polar angles (so basically straight down into the substrate).

Summing up it has been shown, that one can direct light emitted by point sources, like quantum dots, efficiently by dielectric antennas. Such an antenna can consist of the same material as the substrate and reached a directivity of $D = 25.7$ with almost no losses for GaAs in the NIR regime. This enables a variety of possibilities for



(a) x-polarized (b) xy-polarized (c) y-polarized

Fig. 5.26.: *Angular resolved directive gain (far field) of structured substrate, as sketched in Fig. 5.22 b), in spherical coordinates, where ϕ is the azimuthal and θ the polar angle. The electric field of the source is a) x-polarized, c) y-polarized, and b) in x- and y-polarized.*

designing optical antennas on nearly all material systems. Dependent on the properties of the used materials, the antenna becomes considerably larger than the plasmonic counterpart, but the directivity can be much higher, and the radiation efficiency can be close to 1 (dependent on material and wavelength). Additionally, these antennas are relatively easy to fabricate and much more robust to fabrication tolerances than plasmonic antennas. Still the main challenge might be the placing of the nano emitter within the gap of the antenna.

Adding another antenna arm to the system and modifying the reflector, one effectively obtains a broadband, polarization splitter for small dipole sources, leading to many possible applications, e.g. in spectroscopy.

5.4. One-particle, super small, all-dielectric optical antenna

As mentioned above, recently very small antennas have been studied, where a point source has been placed inside of a silicon sphere. It can be shown, that such a point source excites an electric and a magnetic dipole moment in a perfect sphere [103]. For materials with high permittivity these two, orthogonal dipoles exhibit similar strengths in the visible regime, if the size of the sphere is properly adjusted. These two dipoles then interfere with each other, which leads to a directed emission. In non-perfect spheres also higher order moments can be excited, which can lead to even more complex interference patterns. In order to increase the directivity of such a small particle, a

notch has been cut out of the sphere, in which the point source has been placed. Thus the excitation of higher order modes became the main effect responsible for high directivities. This has been theoretically and (partially) experimentally investigated in [98, 104, 105, 106, 84], though without any point sources.

There is one main drawback in this approach: Such an antenna is very sensitive to the positioning of the source. This source was actually "floating" within the notch around $50nm$ above the substrate. Since such an approach is unpractical, the content of this chapter is about a similar concept based on silicon nanodisks. Some properties, like the equivalence of the spectrum by excitation of a plane wave and a point-like dipole source, are already studied by J. van de Groep [107]. This chapter is also the result of a cooperation with BSc. David Hähnel.

Starting point is a silicon cylinder in vacuum, excited by a point source, which is placed $5nm$ next to the cylinder and polarized along the y-axis. For the beginning no substrate is considered, in order to characterize the cylinder itself, so the point source is "floating" at the middle of the height of the cylinder. The material has been modeled via a 10th order fit to the experimental data taken from Ref. [38], though in the shown frequency regime the complex permittivity is relatively smooth.

For this system the far field pattern has been calculated and the spectrally resolved directivity is shown in Fig. 5.27 for three different dimensions of the cylinder. The top curve shows a maximum of the directivity for $\lambda = 480nm$. More important though are the two minima at $\lambda_e = 428nm$ and $\lambda_m = 535nm$. The insets show the radiation patterns for these two wavelengths, which closely resemble the radiation patterns of dipoles, which is further affirmed by the directivity of about $D_{max} \simeq 1.5$. Interestingly, these dipoles are orthogonal to each other, so that they can be identified as the electric dipole at $\lambda_e = 428nm$ and the magnetic dipole at $\lambda_m = 535nm$. The maximum between these two points is then the interference of those dipoles, which is the definition of a Huygen's source.

In contrast to perfect spheres, the nanodisk offers two degrees of freedom for the design of antennas. In the middle curve the directivity of the same cylinder with larger height is shown, as well as a cylinder with the same height but larger diameter in the bottom curve. There one can see, that with both parameters the maximum can be shifted to higher wavelengths without decreasing the directivity. Particularly the center plot is of interest here. One can see that for smaller wavelengths additional maxima appear, which is the result of contributions of excited higher order multipole moments. The maximum at a wavelength of $\lambda = 445nm$ is caused by the interference of the electric

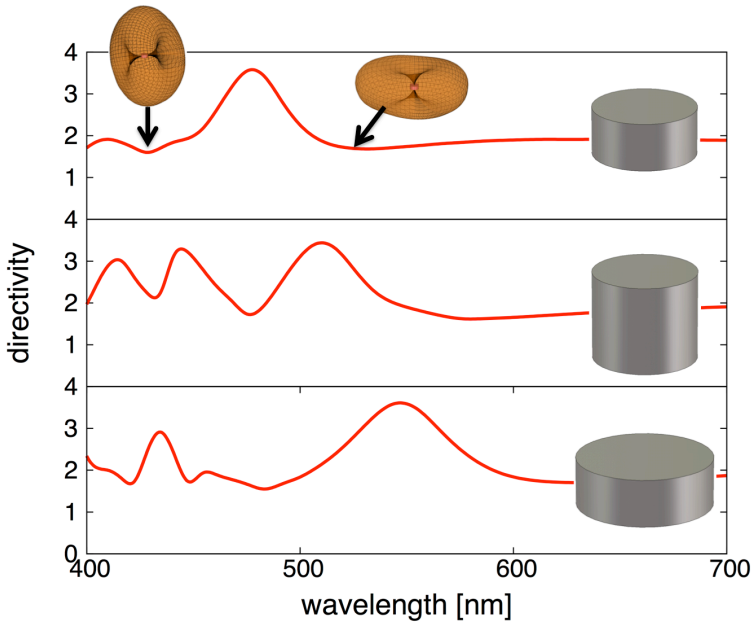


Fig. 5.27.: *Directivity depending on the wavelength for three different sizes of the nanodisk: the diameter and height are in a) 130nm and 60nm, b) 130nm and 110nm, and c) 170nm and 60nm. The two insets are calculated far field radiation patterns, showing basically two orthogonal dipoles, with the electric at 428nm and the magnetic at 535nm.*

dipole ($\lambda_{ed} = 475nm$) and the magnetic quadrupole moment ($\lambda_{mq} = 425nm$). In principle this holds true for the top and bottom case as well, but here the diameter and the height are close to each other, so that the higher order moments can be formed easier. In the following first the interference of the electric and magnetic dipole moments will be exploited, then the interference of higher order moments. In both cases the excitation wavelength is $\lambda = 545nm$, which corresponds the maximum directivity in the bottom curve.

In order to get more insight into the structure, it is useful to have a closer look to the position of the source. For that the source has been shifted along the x-axis from the center to a point of $35nm$ outside of the disk. In Fig. 5.28 the directive gain in forward ($\theta = 90°$, $\phi = 0°$, blue) and backward ($\theta = 90°$, $\phi = 180°$, red) direction is shown. The dashed black line indicates the end of the disk. For a source position of $x = 31nm$ one obtains a remarkably high front-to-back ratio of $R_{FB}^{31nm} = 16.2dBi$, which is several times as much as has been found using perfect spheres. For a source position of $x = 86nm$ a similar behavior can be seen, only in backwards direction. The reverse of the front-to-back ratio is then $(R_{FB}^{86nm})^{-1} = 12dBi$.

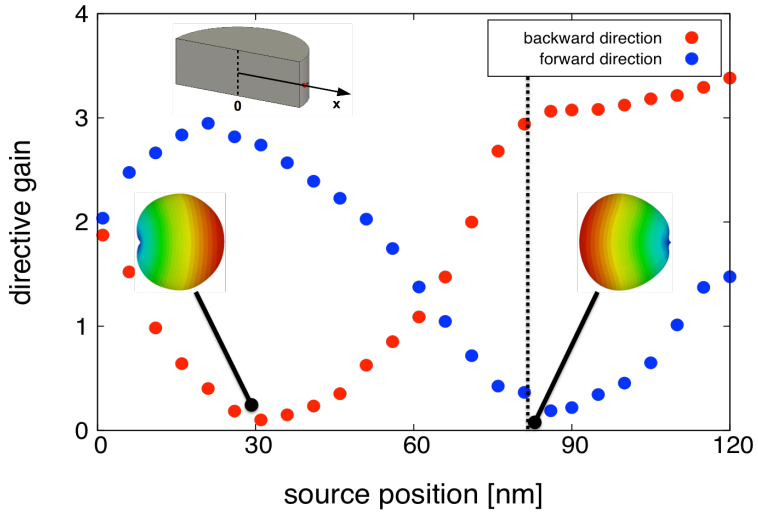


Fig. 5.28.: *Directive gain in forward and backward direction depending on the position of the point source for a nanodisk with diameter 170nm and height 60nm. The insets show the radiation pattern for the points with the highest front-to-back ratio.*

In contrast to any structure with a "floating" dipole, it is possible to grow a disk with an embedded quantum dot, acting as a dipole source [108]. However, using the interference of the magnetic and electric dipole can reach high front-to-back ratios, but not high directivities, as discussed in [98]. In order to increase the directivity of the system, now the interference of the magnetic hexapole and the electric dipole is exploited.

As mentioned earlier, the dimensions of the cylinder have to be appropriate for a hexagonal mode to form. Consider a standard hexagon with all six side lengths a. The height (distance from flat to opposite flat side) is then $h = \sqrt{3}a$ and the width (tip to opposite tip) is $w = 2a$. For a hexagon to form within the cylinder, a good start for the dimensions would be a ratio of diameter of the cylinder and its height $d/h = \sqrt{4/3}$.

The resulting disk had a diameter of $d = 228nm$ and a height of $h = 200nm$. The optimal source position has been $x = -73nm$ from the center, but well within the disk. In Fig. 5.29 the corresponding directive gain in spherical coordinates is shown. There one can see a clear hot spot in positive x-direction achieving a directivity of about $D \simeq 9$. Unfortunately the front-to-back ratio decreased to $R_{FB} \simeq 5dBi$.

In order to get more insight into the structure, in Fig. 5.30 a) the directivity is plotted in a polar plot for the polar angle θ. Also in b) the cross section (x-z-plane) of the

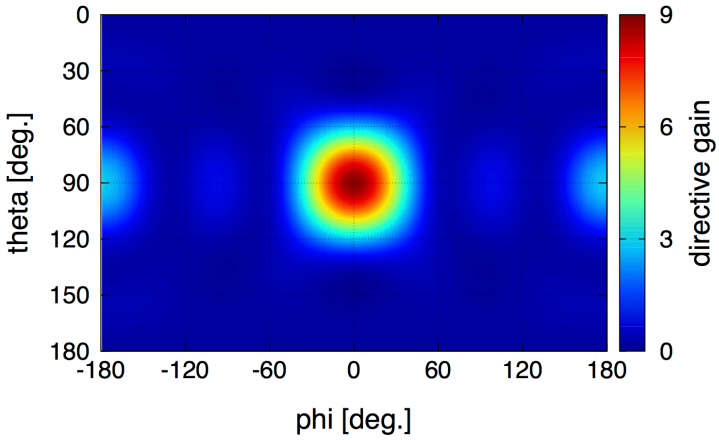


Fig. 5.29.: *Directive gain in spherical coordinates with the polar angle θ and the azimuthal angle ϕ for a silicon cylinder with diameter 228nm and height 200nm, where the source is positioned 73nm from the center.*

normalized, absolute value of the electric field is shown in logarithmical color scale. The pattern of the electric field corresponds to the interference of an electric dipole and a magnetic hexapole, leading to an asymmetric radiation, with the hexagon perfectly fitting into the nanodisk. With increasing (or decreasing) y, the width becomes much smaller than the height, which results in the concentration of the energy towards the center (y = 0nm). Thus the beam width angular is small not only for the polar angle, but the azimuthal angle as well. Also the side lobes are well suppressed, though the front-to-back ratio is small com-pared to a Huygens source.

Placing this cylinder on a SiO_2 substrate (n=1.46), does not change the mechanism itself (Fig. 5.30 c) and d)), only the angle of radiation is changed due to the change of the dielectric environment. Otherwise the field pattern remains the same, leading to a directivity of almost D=16.

The next step is the manipulation of the radiation pattern by cutting a cylinder out of the silicon nanodisk. The dipole source is placed in the middle of the cut out, 5nm above the SiO_2 substrate. Also the dimensions are changed (diameter $d = 295nm$ and height $h = 185nm$), because the air gap within the cylinder leads to distortions of the hexagon, i.e. the previous estimation is not applicable in this case. A demonstration of the fabrication of such hollow cylinders and an investigation of their dipole resonances can be found in Ref. [109].

In Fig. 5.31 a) the maximum directive gain into the air and into the substrate are plotted versus the radius of the cut-out. If no cut-out is present, the nanodisk resembles the one of the previous section, only with the source at the bottom at $x = 0nm$. In

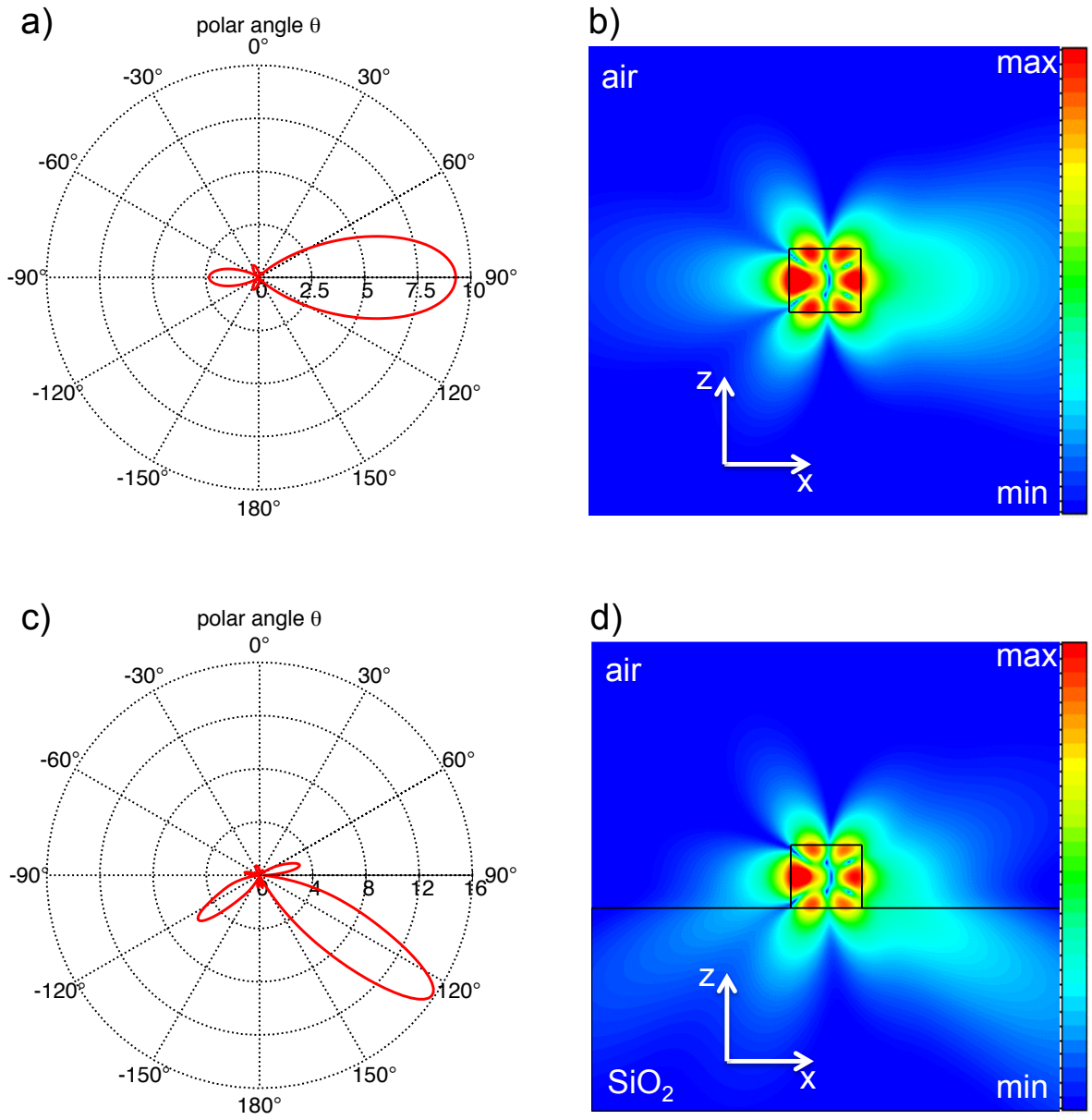


Fig. 5.30.: *a) Polar plot of the directive gain for the polar angle θ (constant $\phi = 0$). b) Corresponding normalized, absolute value of the excited electric field in dB scale, reaching from -40 dB (blue) to 0 dB (red). The shown cross section is normal to the dipole source and results from an excited magnetic hexapole moment. c) and d) Same plots as in a) and b) but for a structure placed on glass substrate.*

this case the hexapole is rotated due to the different position of the source. Since the nanodisk radiates to the direction opposite of the source, in this configuration over 95% of the power is radiated straight into the air ($\theta = 0°$), as can be seen in the first inset. One could argue, that the silicon nanodisk, which has the highest index of refraction in

this system, is placed on top of the point source, so that the light is just propagating through the optically densest medium. But then the size of the nanodisk, if sufficiently large, would not change much of the radiation pattern, which it actually does. Apart from that one would get a typical dipole radiation pattern, which cannot be seen here. This radiation pattern might be advantageous for spectroscopic applications, since the total internal reflection becomes insignificant then.

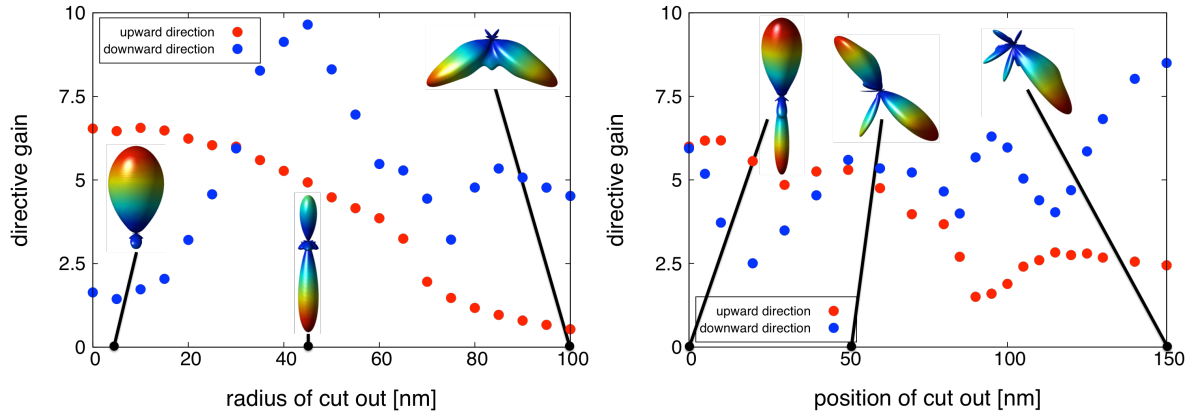


Fig. 5.31.: *a) Polar plot of the directive gain for the polar angle θ (constant $\phi = 0$). b) Corresponding normalized, absolute value of the excited electric field in dB scale, reaching from -40 dB (blue) to 0 dB (red). The shown cross section is normal to the dipole source and results from an excited magnetic hexapole moment. c) and d) Same plots as in a) and b) but for a structure placed on glass substrate.*

For large radii the radiation into the air becomes negligible, and the radiation pattern becomes similar to a single dipole on the substrate alone. This actually makes sense, since the radius of the cutout is so large then, that the small silicon ring does not influence the radiation pattern of a small source. In the intermediate region one can see a resonance, where roughly 80% of the energy is propagating directly into the substrate ($\theta = 180°$) in a narrow beam with a directivity of D = 9.8. Thus by changing the size of the cut-out one can choose if the power is radiated directly into the air or into the substrate.

Also interesting is the further utilization of the magnetic hexapole. This can be obtained by shifting the cut-out together with the dipole source (here in x-direction). For that a radius of the cutout of $r = 30nm$ has been chosen, since it is, to my knowledge, possible to fabricate and still the hexapole can be seen in our simulations. Of course one can observe distortions of the field profile due to the missing material. In Fig. 5.31 b) the maximum directive gain into the air as well as into the substrate region

is shown, depending on the shift of the cutout starting from the center. There one can see, that starting from the straight up / down radiation pattern (left inset), the behavior evolves to a unidirectional radiation pattern with a directivity of D = 8 (for a radius of $r = 150nm$).

To sum up this chapter, it has been shown, that with using a simple silicon nanodisk one can achieve a Huygens source as well as unidirectional antennas, just by changing the geometrical parameters of the disk (to adjust the electric and magnetic multipoles) and the source position. However, this only works, if the source is embedded in the middle of the height of the cylinder. Since the parasitic impedance of the substrate makes it difficult to excite the higher order poles of the disk, if the source is located at its bottom, at least the dimensions of the nanodisk have to be adjusted.

If the source cannot be embedded in the silicon disk, one can also etch a cylindrical cutout into the disk, place the source directly onto the substrate, and either get the radiation directed straight up into the air or down into the substrate, just by changing the radius of the cut out. By changing the position of the cut-out one may get directed emission also to the side, if desired. All in all, one can get almost any radiation pattern by changing the geometrical parameters of the disk and the source position. A consequence of this though is, that the system is sensitive to these parameters, so the fabrication might be challenging. Please note, that in this chapter almost no optimization has been done, so there is room for improvement, e.g. using an elliptical cutout instead of a circular one.

6. Conclusion and Outlook

6.1. Conclusions

In this work two different types of antennas are investigated and optimized, receiving antennas, yielding high intensity enhancement, and sending antennas, creating well defined radiation patterns from a small dipole source.

The receiving antennas, consisting of two particles, and their most important properties are discussed. Since the material of all considered receiving antennas has been gold, they are affected by high losses. But since they couple efficiently to the far field these losses become negligible compared to the (re-) emitted radiation from the antennas. It could be confirmed, that such antennas achieve a higher intensity enhancement in an array compared to isolated antennas. While with varying periodicity along the antenna axis the change of the maximum intensity enhancement was rather low, it dramatically changed with the periodicity perpendicular to the antenna axis. If the periodicity was around the wavelength of the incident light, one could achieve constructive interference between the incident light and the emitted light from neighboring antennas, resulting in a collective mode of the whole array. As a result the intensity enhancement becomes much higher and radiation losses significantly smaller. This effect is known for other resonators like quantum dots, but has not been reported for antennas so far, and still needs investigation, because such modes appear stronger for some structures like bow-ties, or weaker for others like rod antennas.

Apart from the optimization of an array of bow-ties and rod antennas, a nearly free geometry optimization, using a number of bricks with varying size, has been investigated. The result is a surprisingly simple antenna roughly shaped like an H, utilizing the inhomogeneity at its tip, which leads to a current perpendicular to the polarization of the incident light field, effectively collecting the charge carriers close to the gap, which leads to an intensity enhancement almost 5 times higher than has been observed for bow-ties. Although the intensity enhancement is decreased by placing the

optimized antenna on glass substrate, its intensity enhancement is more than 5 times higher compared to the optimized bow-tie, still showing signs of the collective array mode, which has vanished for bow-ties.

A free form optimization process has been briefly shown with an example of a polygon, where additional filter would have been necessary to obtain useful structures, requiring much additional computation power. Another approach could be to define a grid of small squares, where each square is filled with material or not. Since the parameter space would become binary (apart from being huge), this approach would render the strengths of optimization algorithms useless, resulting in a tremendous increase of required computation power. It turned out, that an optimization process, where structures are built by combination of simple geometries like bricks or triangles, is much more efficient than a free form optimization, since constraints are much easier implemented and processed by the corresponding algorithm, and almost arbitrary shapes can be generated.

Also the Yagi-Uda antennas, sending antennas generating required radiation patterns, are discussed. It has been shown, that common design rules, which have been developed for the radio frequency regime, only roughly apply in the optical regime, due to plasmon resonances and the parasitic impedance of the substrate. A Yagi-Uda antenna, which has been designed by such rules and thoroughly investigated, could be optimized, where the directivity increased from roughly $D = 12$ (reference) to $D = 21$ (optimized), but with the prize of a slightly decreased radiation efficiency.

More important, it could be shown using the coupled dipole approximation that indeed the directive properties are caused by the interference of appropriate dipole sources. A consequence of this is, that for such antennas any materials and geometrical shapes may be utilized, if they exhibit sufficient dipole resonances, which has been shown exemplarily for silicon bricks. An apparent drawback of plasmonic antennas is their high loss, caused by the used metal. Apart from that another drawback of such Yagi-Uda like antennas is, that most of the radiation penetrates into the substrate, so that the performance of such antennas cannot be increased by addition of further elements.

This can be overcome by utilization of a leaky waveguide, for which HfO_2 has been used in this work, since it exhibits almost no losses in the optical regime. The point source couples to the leaky waveguide mode, where all those points, where some of the radiation couples to substrate modes can be considered as new point sources. These point sources again interfere and thus generate the desired radiation pattern, in this case an extremely narrow main lobe with a directivity of $D = 30$. The phase difference of

those sources is manipulated by the geometric parameters of the HfO_2, which changes the propagation constant of the waveguide mode. Advantages of such an antenna are an extremely high directivity, caused by an increased number of participating, interfering sources, the lack of losses, and the robustness of the structure, since the propagation constant is not that sensitive to fabrication tolerances. Disadvantage is the significantly increased dimension (almost a factor of 2) compared to the plasmonic (reference) antenna.

One can also decrease the size of such an antenna by interference of higher order poles with the point source. This has been shown with silicon nanodisks, where the point dipole couples to the magnetic hexapole mode of the disk, leading to a relatively high directivity ($D = 16$) for a silicon nano cylinder on SiO_2 substrate. By changing the position of the source this magnetic hexapole gets rotated, leading to radiation in other directions, e.g. straight up into the air, which might be useful for spectroscopic measurements. One can also cut a smaller cylinder out of the nanodisk, letting a point dipole like colloidal quantum dots fall into the cutout, and obtain, depending on the size and position of the cutout, different radiation patterns, e.g. straight down into the substrate, or sidewards with high directivity, similar to Yagi-Uda antennas.

6.2. Outlook

In principle antennas, with the highest intensity enhancement couple best to the far field, resulting in only short-lived plasmonic modes and huge radiation losses. But this can be somewhat overcome, if a collective array mode can be excited, which has been shown in chapter 4.3.2. As mentioned, publications dealing with such modes are sparse, though such modes might significantly increase the performance. Also the strength of such modes is depending on the used geometry, i.e. there is potential for additional optimization, especially if a better understanding of such modes leads to design rules for such structures.

There are also some fundamental aspects of the optimization process itself, which are worth to look at, e.g. finding better algorithms for this type of problem or maybe develop a gradient-based optimization process. Finally one might use such antennas not only to increase the field in an area, but also to give that area a specific form, using interference or polarization effects.

For the sending antenna much more work is necessary, if they shall be used for quantum

information processing or computing. The Yagi-Uda antennas can further be miniaturized by choosing high index materials. It would be interesting, if the directive gain could further be increased by the usage of hybrid-material antennas. Also worth of investigation are other geometries like disk antennas. The silicon nanodisk can obviously improved further by usage of other (lossless) materials. Also interesting would be the utilization of other poles, which interfere with the dipole source.

A. Parameters

A.1. Nanoantennas for intensity enhancement

In order to model the behavior of noble metals like silver or gold, the standard Drude model with the following parameters has been used.

parameter	value
ε_∞	11.1
ω_p	$1.431 \cdot 10^{16} \frac{rad.}{s}$
γ_p	$1.357 \cdot 10^{14} \frac{rad.}{s}$

Tab. A.1.: *Used parameters to simulate the behavior of gold for a wavelength of* $500nm \leq \lambda \leq 1500nm$*. In this region only one pole is necessary for accurate results.*

parameter	value
L_x	40nm
L_y	160nm
L_z	40nm
r_{vert}	20nm
r_{top}	10nm

Tab. A.2.: *Geometrical parameters for a single rod with the lengths* L_i *of the corresponding coordinates and the radii of the blending of the edges.*

parameter	value
L_x	40nm
L_y	130nm
L_z	28nm
r_{vert}	20nm
r_{top}	10nm

(a) rod dimer

parameter	value
φ	45°
L_y	145nm
L_z	28nm
r_{vert}	20nm
r_{top}	10nm

(b) bow-tie

Tab. A.3.: *Geometrical parameters for a golden rod dimer with lengths L_i and a bow-tie with the opening angle φ and the lengths L_i. All horizontal edges have been blended with a radius of r_{top} and the vertical edges with r_{vert}. The structure has been excited via a plane wave propagating in z-direction.*

parameter	size
width of supp. box	$272nm$
length of supp. box	$110nm$
width of tip	$40nm$
length of tip	$33nm$
height	$28nm$
gap	$16nm$
periodicity Λ_x	$770nm$
periodicity Λ_y	$786nm$

Tab. A.4.: *Parameter set for the H-antenna. All horizontal edges have been blended with a radius of $r_{top} = 10nm$. All vertical edges have been blended with a radius of $r_{vert} = 20nm$.*

A.2. Nanoantennas for high directivity

element	length [nm]	width [nm]	separation [nm]
reflector	232	102	150
feed	163	45	
director 1	174	71	146
director 2	131	79	208
director 3	136	35	153

Tab. A.5.: *Parameters of the optimized Yagi-Uda antenna consisting of gold (see chapter 5.2.2). The elements are listed from left to right, with the second element acting as feed element. The height of each element is 100nm.*

element	length [nm]	width [nm]	separation [nm]
reflector	319	77	289
director 1	294	76	127
director 2	292	71	120
director 3	240	75	256

Tab. A.6.: *Parameters of the optimized Yagi-Uda antenna consisting of silicon (see chapter 5.2.3). The elements are listed from left to right, with the feed element replaced by the point source. The height of each element is 98nm.*

No. of dipole	distance to primary source [nm]	Phase difference [π]
1	386	0.069
2	679	0.121
3	985	0.176
4	1297	0.232
5	1602	0.286
6	1914	0.342
7	2224	0.397

Tab. A.7.: *Parameters of the dipoles, with which the hafnium oxide waveguide has been replaced, to obtain a similar radiation pattern.*

B. CDA equations for 5 elements

Starting from equation (2.81), the solution for 5 particles can be analytically derived. In order to make it short, some definitions are needed.

$$\underline{\mathbf{B}}_1 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{11} - \underline{\mathbf{A}}_{22}^{-1}\underline{\mathbf{A}}_{21} \qquad \underline{\mathbf{B}}_2 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{13} - \underline{\mathbf{A}}_{22}^{-1}\underline{\mathbf{A}}_{23}$$
$$\underline{\mathbf{B}}_3 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{14} - \underline{\mathbf{A}}_{22}^{-1}\underline{\mathbf{A}}_{24} \qquad \underline{\mathbf{B}}_4 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{15} - \underline{\mathbf{A}}_{22}^{-1}\underline{\mathbf{A}}_{25}$$
$$\underline{\mathbf{B}}_5 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{11} - \underline{\mathbf{A}}_{32}^{-1}\underline{\mathbf{A}}_{31} \qquad \underline{\mathbf{B}}_6 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{13} - \underline{\mathbf{A}}_{32}^{-1}\underline{\mathbf{A}}_{33}$$
$$\underline{\mathbf{B}}_7 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{14} - \underline{\mathbf{A}}_{32}^{-1}\underline{\mathbf{A}}_{34} \qquad \underline{\mathbf{B}}_8 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{15} - \underline{\mathbf{A}}_{32}^{-1}\underline{\mathbf{A}}_{35}$$
$$\underline{\mathbf{B}}_9 := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{11} - \underline{\mathbf{A}}_{42}^{-1}\underline{\mathbf{A}}_{41} \qquad \underline{\mathbf{B}}_{10} := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{13} - \underline{\mathbf{A}}_{42}^{-1}\underline{\mathbf{A}}_{43}$$
$$\underline{\mathbf{B}}_{11} := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{14} - \underline{\mathbf{A}}_{42}^{-1}\underline{\mathbf{A}}_{44} \qquad \underline{\mathbf{B}}_{12} := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{15} - \underline{\mathbf{A}}_{42}^{-1}\underline{\mathbf{A}}_{45}$$
$$\underline{\mathbf{B}}_{13} := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{11} - \underline{\mathbf{A}}_{52}^{-1}\underline{\mathbf{A}}_{51} \qquad \underline{\mathbf{B}}_{14} := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{13} - \underline{\mathbf{A}}_{52}^{-1}\underline{\mathbf{A}}_{53}$$
$$\underline{\mathbf{B}}_{15} := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{14} - \underline{\mathbf{A}}_{52}^{-1}\underline{\mathbf{A}}_{54} \qquad \underline{\mathbf{B}}_{16} := \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{15} - \underline{\mathbf{A}}_{52}^{-1}\underline{\mathbf{A}}_{55}$$

$$\underline{\mathbf{C}}_1 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_1 - \underline{\mathbf{B}}_6^{-1}\underline{\mathbf{B}}_5 \qquad \underline{\mathbf{C}}_2 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_3 - \underline{\mathbf{B}}_6^{-1}\underline{\mathbf{B}}_7 \qquad \underline{\mathbf{C}}_3 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_4 - \underline{\mathbf{B}}_6^{-1}\underline{\mathbf{B}}_8$$
$$\underline{\mathbf{C}}_4 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_1 - \underline{\mathbf{B}}_{10}^{-1}\underline{\mathbf{B}}_9 \qquad \underline{\mathbf{C}}_5 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_3 - \underline{\mathbf{B}}_{10}^{-1}\underline{\mathbf{B}}_{11} \qquad \underline{\mathbf{C}}_6 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_4 - \underline{\mathbf{B}}_{10}^{-1}\underline{\mathbf{B}}_{12}$$
$$\underline{\mathbf{C}}_7 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_1 - \underline{\mathbf{B}}_{14}^{-1}\underline{\mathbf{B}}_{13} \qquad \underline{\mathbf{C}}_8 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_3 - \underline{\mathbf{B}}_{14}^{-1}\underline{\mathbf{B}}_{15} \qquad \underline{\mathbf{C}}_9 := \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_4 - \underline{\mathbf{B}}_{14}^{-1}\underline{\mathbf{B}}_{16}$$

$$\underline{\mathbf{D}}_1 := \underline{\mathbf{C}}_2^{-1}\underline{\mathbf{C}}_1 - \underline{\mathbf{C}}_5^{-1}\underline{\mathbf{C}}_4 \qquad \underline{\mathbf{D}}_2 := \underline{\mathbf{C}}_2^{-1}\underline{\mathbf{C}}_3 - \underline{\mathbf{C}}_5^{-1}\underline{\mathbf{C}}_6$$
$$\underline{\mathbf{D}}_3 := \underline{\mathbf{C}}_2^{-1}\underline{\mathbf{C}}_1 - \underline{\mathbf{C}}_8^{-1}\underline{\mathbf{C}}_7 \qquad \underline{\mathbf{D}}_4 := \underline{\mathbf{C}}_2^{-1}\underline{\mathbf{C}}_3 - \underline{\mathbf{C}}_8^{-1}\underline{\mathbf{C}}_9$$

$$\tilde{\boldsymbol{E}}_2 := \underline{\mathbf{B}}_2^{-1}(\underline{\mathbf{A}}_{12}^{-1}\boldsymbol{E_1} - \underline{\mathbf{A}}_{22}^{-1}\boldsymbol{E_2}) \qquad \tilde{\boldsymbol{E}}_3 := \underline{\mathbf{B}}_6^{-1}(\underline{\mathbf{A}}_{12}^{-1}\boldsymbol{E_1} - \underline{\mathbf{A}}_{32}^{-1}\boldsymbol{E_3})$$
$$\tilde{\boldsymbol{E}}_4 := \underline{\mathbf{B}}_{10}^{-1}(\underline{\mathbf{A}}_{12}^{-1}\boldsymbol{E_1} - \underline{\mathbf{A}}_{42}^{-1}\boldsymbol{E_4}) \qquad \tilde{\boldsymbol{E}}_5 := \underline{\mathbf{B}}_{14}^{-1}(\underline{\mathbf{A}}_{12}^{-1}\boldsymbol{E_1} - \underline{\mathbf{A}}_{52}^{-1}\boldsymbol{E_5})$$

$$\hat{\boldsymbol{E}}_3 := \underline{\mathbf{C}}_2^{-1}(\tilde{\boldsymbol{E}}_2 - \tilde{\boldsymbol{E}}_3) \qquad \hat{\boldsymbol{E}}_4 := \underline{\mathbf{C}}_5^{-1}(\tilde{\boldsymbol{E}}_2 - \tilde{\boldsymbol{E}}_4)$$
$$\hat{\boldsymbol{E}}_5 := \underline{\mathbf{C}}_8^{-1}(\tilde{\boldsymbol{E}}_2 - \tilde{\boldsymbol{E}}_5)$$

Using these definitions, one can shortly write the expressions for the dipole moments as in equation (B.1), where the local excited, electric field is simply written as $\boldsymbol{E}_i$.

$$\boldsymbol{p}_1 = (\underline{\mathbf{D}}_2^{-1}\underline{\mathbf{D}}_1 - \underline{\mathbf{D}}_4^{-1}\underline{\mathbf{D}}_3)(\underline{\mathbf{D}}_2^{-1}(\hat{\boldsymbol{E}}_3 - \hat{\boldsymbol{E}}_4) - \underline{\mathbf{D}}_4^{-1}(\hat{\boldsymbol{E}}_3 - \hat{\boldsymbol{E}}_5)) \tag{B.1a}$$
$$\boldsymbol{p}_5 = \underline{\mathbf{D}}_2^{-1}(\hat{\boldsymbol{E}}_3 - \hat{\boldsymbol{E}}_4) - \underline{\mathbf{D}}_2^{-1}\underline{\mathbf{D}}_1\boldsymbol{p}_1 \tag{B.1b}$$
$$\boldsymbol{p}_4 = \underline{\mathbf{C}}_2^{-1}(\tilde{\boldsymbol{E}}_2 - \tilde{\boldsymbol{E}}_3) - \underline{\mathbf{C}}_2^{-1}\underline{\mathbf{C}}_1\boldsymbol{p}_1 - \underline{\mathbf{C}}_2^{-1}\underline{\mathbf{C}}_3\boldsymbol{p}_5 \tag{B.1c}$$
$$\boldsymbol{p}_3 = \underline{\mathbf{B}}_2^{-1}(\underline{\mathbf{A}}_{12}^{-1}\boldsymbol{E}_1 - \underline{\mathbf{A}}_{22}^{-1}\boldsymbol{E}_2) - \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_1\boldsymbol{p}_1 - \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_3\boldsymbol{p}_4 - \underline{\mathbf{B}}_2^{-1}\underline{\mathbf{B}}_4\boldsymbol{p}_5 \tag{B.1d}$$
$$\boldsymbol{p}_2 = \underline{\mathbf{A}}_{12}^{-1}\boldsymbol{E}_1 - \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{11}\boldsymbol{p}_1 - \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{13}\boldsymbol{p}_3 - \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{14}\boldsymbol{p}_4 - \underline{\mathbf{A}}_{12}^{-1}\underline{\mathbf{A}}_{15}\boldsymbol{p}_5 \tag{B.1e}$$

C. Notations and symbols

C.1. Acronyms

ADOL-C automatic differentiation by overloading in C++

BEM boundary element method

CDA coupled dipole approximation

CFL Courant-Friedrichs-Lewy

CL cathodoluminescence

COBYLA constrained optimization by linear approximation

CRS controlled random search

CRSLM controlled random search with local mutation

DOF degrees of freedom

EBL electron beam lithography

EELS electron energy-loss spectroscopy

FDTD finite difference time domain

FEM finite element method

FIB focused ion-beam milling

FIT finite integration technique

ITO indium tin oxide

NIR near-infrared

NTFF near to far field

PBC periodic boundary conditions

PEC perfect electric conductor

PML perfectly matched layers

PMMA Poly(methyl methacrylate), also known as Plexiglas,

PSA particle swarm algorithm

qubit quantum bit

QD quantum dot

RF radio frequency

SHG second harmonic generation

SPP surface plasmon polariton

TE transversal electric

THG third harmonic generation

TM transversal magnetic

Bibliography

[1] Xiaohong Ren, Chenghua Fu, Tianwen Wang, and Shuxiang Jia. Can bus network design based on bluetooth technology. In *Electrical and Control Engineering (ICECE), 2010 International Conference on*, pages 560–564. IEEE, 2010.

[2] Jinwook Kim, Hyeon-Chang Son, Do-Hyeon Kim, and Young-Jin Park. Optimal design of a wireless power transfer system with multiple self-resonators for an led tv. *IEEE Transactions on Consumer Electronics*, 58(3):775–780, 2012.

[3] Naomi J Halas. Plasmonics: an emerging field fostered by nano letters. *Nano letters*, 10(10):3816–3822, 2010.

[4] Jens Dorfmuller, Daniel Dregely, Moritz Esslinger, Worawut Khunsin, Ralf Vogelgesang, Klaus Kern, and Harald Giessen. Near-field dynamics of optical yagi-uda nanoantennas. *Nano letters*, 11(7):2819–2824, 2011.

[5] Pascal Anger, Palash Bharadwaj, and Lukas Novotny. Enhancement and quenching of single-molecule fluorescence. *Physical review letters*, 96(11):113002, 2006.

[6] Tim H Taminiau, Fernando D Stefani, and Niek F van Hulst. Enhanced directional excitation and emission of single emitters by a nano-optical yagi-uda antenna. *Optics express*, 16(14):10858–10866, 2008.

[7] Shuzhou Li, María L Pedano, Shih-Hui Chang, Chad A Mirkin, and George C Schatz. Gap structure effects on surface-enhanced raman scattering intensities for gold gapped rods. *Nano letters*, 10(5):1722–1727, 2010.

[8] Thorsten Schumacher, Kai Kratzer, David Molnar, Mario Hentschel, Harald Giessen, and Markus Lippitz. Nanoantenna-enhanced ultrafast nonlinear spectroscopy of a single gold nanoparticle. *Nature Communications*, 2:333, 2011.

[9] Guoxing Zheng, Holger Mühlenbernd, Mitchell Kenney, Guixin Li, Thomas Zent-

graf, and Shuang Zhang. Metasurface holograms reaching 80% efficiency. *Nature nanotechnology*, 10(4):308–312, 2015.

[10] Franziska Zeuner, Mulda Muldarisnur, Andre Hildebrandt, Jens Forstner, and Thomas Zentgraf. Coupling mediated coherent control of localized surface plasmon polaritons. *Nano letters*, 15(6):4189–4193, 2015.

[11] Andrei Andryieuski, Radu Malureanu, Giulio Biagi, Tobias Holmgaard, and Andrei Lavrinenko. Compact dipole nanoantenna coupler to plasmonic slot waveguide. *Optics letters*, 37(6):1124–1126, 2012.

[12] Y. M. Wu, L. W. Li, and B. Liu. Gold bow-tie shaped aperture nanoantenna: Wide band near-field resonance and far-field radiation. *IEEE Transactions on Magnetics*, 46(6):1918–1921, June 2010.

[13] Andrea Alù and Nader Engheta. Hertzian plasmonic nanodimer as an efficient optical nanoantenna. *Phys. Rev. B*, 78:195111, Nov 2008.

[14] Huanjun Chen, Lei Shao, Qian Li, and Jianfang Wang. Gold nanorods and their plasmonic properties. *Chemical Society Reviews*, 42(7):2679–2724, 2013.

[15] Benito Rodríguez-González, Farah Attouchi, M. Fernanda Cardinal, Viktor Myroshnychenko, Odile Stéphan, F. Javier García de Abajo, Luis M. Liz-Marzán, and Mathieu Kociak. Surface plasmon mapping of dumbbell-shaped gold nanorods: The effect of silver coating. *Langmuir*, 28(24):9063–9070, 2012.

[16] Lukas Novotny. Effective Wavelength Scaling for Optical Antennas. *Physical Review Letters*, 98(26):266802, June 2007.

[17] Thaddeus D. Ladd, Fedor Jelezko, Raymond Laflamme, Yasunobu Nakamura, Christopher Monroe, and Jeremy L OBrien. Quantum computers. *Nature*, 464(7285):45–53, 2010.

[18] Julien Claudon, Joël Bleuse, Nitin Singh Malik, Maela Bazin, Périne Jaffrennou, Niels Gregersen, Christophe Sauvan, Philippe Lalanne, and Jean-Michel Gérard. A highly efficient single-photon source based on a quantum dot in a photonic nanowire. *Nature Photonics*, 4(3):174–177, 2010.

[19] Mathieu Munsch, Nitin S Malik, Emmanuel Dupuy, Adrien Delga, Joël Bleuse, Jean-Michel Gérard, Julien Claudon, Niels Gregersen, and Jesper Mørk. Dielec-

tric gaas antenna ensuring an efficient broadband coupling between an inas quantum dot and a gaussian optical beam. *Physical review letters*, 110(17):177402, 2013.

[20] Sean D Barrett and Pieter Kok. Efficient high-fidelity quantum computation using matter qubits and linear optics. *Physical Review A*, 71(6):060310, 2005.

[21] Toon Coenen, Ernst Jan R Vesseur, Albert Polman, and A Femius Koenderink. Directional emission from plasmonic yagi–uda antennas probed by angle-resolved cathodoluminescence spectroscopy. *Nano letters*, 11(9):3779–3784, 2011.

[22] W. Tang, C. Argyropoulos, E. Kallos, W. Song, and Y. Hao. Discrete coordinate transformation for designing all-dielectric flat antennas. *IEEE Trans. Antennas Propag.*, 58:3795–3804, 2010.

[23] Allen Taflove and Susan C. Hagness. *Computational Electrodynamics*. Artech House, 3rd edition, 2005.

[24] T. Weiland. A discretization method for solution of maxwells equations for 6-component fields. *Electronics and Communications*, 31:116–120, 1977.

[25] CST AG. Cst studio suite 2014.

[26] T. Weiland and R. Schuhmann. Vorlesungsskript elektromagnetische feldsimulation, 2009.

[27] U. van Rienen and T. Weiland. Triangular discretization method for the evaluation of rf-fields in cylindrically symmetric cavities. *IEEE Transactions on Magnetics*, MAG-21:2317–2320, 1985.

[28] R. Schuhmann and T. Weiland. A stable interpolation technique for fdtd on nonorthogonal grids. *International Journal on Numerical Modeling*, 11:299–306, 1998.

[29] M. Clemens and T. Weiland. Discrete electromagnetism with the finite integration technique. *Progress In Electromagnetics Research*, 32:65–87, 2001.

[30] T. Weiland. Modes in infinitely repeating structures of cylindrical symmetry. *Proceedings of the International Linear Accelerator Conference (LINAC)*, pages 292–295, 1986.

[31] J.-P. Bérenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200, 1994.

[32] F. Teixeira. Complex space approach to perfectly matched layers: A review and some new developments. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 13:441–455, 2000.

[33] P. Hammes. *Numerische Simulation von Streumatrizen im Hochfrequenzbereich.* PhD thesis, Technische Universität Darmstadt, 1999.

[34] L. Zhao and A. Cangellaris. Gt-pml: Generalized theory of perfectly matched layers and its application to the reflectionless truncation of finite-difference time-domain grids. *IEEE Transactions on Microwave Theory and Techniques*, 44:2555–2563, 1996.

[35] K. S. Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Trans. Antennas Propagat.*, 14:302–307, 1966.

[36] C. A. Balanis. *Advanced Engineering Electromagnetics.* Wiley, 1989.

[37] Ilker R. Capoglu and Glenn S. Smith. A direct time-domain fdtd near-field-to-far-field transform in the presence of an infinite grounded dielectric slab. *Antennas and Propagation, IEEE Transactions on*, 54(12):3805–3814, 2006.

[38] E. D. Palik. *Handbook of Optical Constants of Solids.* Academic Press, 1991.

[39] John David Jackson. *Classical Electrodynamics.* John Wiley and Sons, 3rd edition, 1998.

[40] Michael P. Marder. *Condensed Matter Physics.* John Wiley and Sons, 2nd edition, 2010.

[41] Bruce T. Draine and Piotr J. Flatau. Discrete-dipole approximation for scattering calculations. *J. Opt. Soc. Am. A*, 11(4):1491–1499, Apr 1994.

[42] Vladimir D. Miljkovi, Tavakol Pakizeh, Borja Sepulveda, Peter Johansson, and Mikael Käll. Optical forces in plasmonic nanoparticle dimers. *The Journal of Physical Chemistry C*, 114(16):7472–7479, 2010.

[43] C. F. Bohren and D. R. Huffman. *Absorption and Scattering of Light by Small Particles*. John Wiley and Sons, New York, 1998.

[44] M. Meier and A. Wokaun. Enhanced fields on large metal particles: dynamic depolarization. *Opt. Lett.*, 8(11):581–583, Nov 1983.

[45] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2nd edition, 2006.

[46] Carl Geiger and Kristian Kanzow. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer Verlag, 1999.

[47] D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Applications*, 79:157–181, 1993.

[48] A. Dekkers and E. Aarts. Global optimization and simmulated annealing. *Mathematical Programming*, 50:367–393, 1991.

[49] R. Storn and K. Price. Differential evolution - a simple and efficient heuristics for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[50] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer Verlag, 1996.

[51] W. L. Price. Global optimization by controlled random search. *Journal of Optimization Theory and Applications*, 40:333–348, 1983.

[52] M. J. D. Powell. Direct search algorithms for optimization calculations. *Acta Numerica*, 7:287–336, 1998.

[53] P. Kaelo and M. M. Ali. Some variants of the controlled random search algorithm for global optimization. *Journal of Optimization Theory and Applications*, 130:253–264, 2006.

[54] Lawrence Davis. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991.

[55] Russ C. Eberhart and James Kennedy. A new optimizer using particle swarm theory. In *Proceedings of the sixth international symposium on micro machine and human science*, volume 1, pages 39–43. New York, NY, 1995.

[56] A. Walther and A. Griewank. *Getting started with ADOL-C.* In U. Naumann und O. Schenk, Combinatorial Scientific Computing, Chapman-Hall CRC Computational Science, 2012.

[57] Paolo Biagioni, Jer-Shing Huang, and Bert Hecht. Nanoantennas for visible and infrared radiation. *Reports on Progress in Physics*, 75(2):024402, 2012.

[58] Harry a Atwater and Albert Polman. Plasmonics for improved photovoltaic devices. *Nature materials*, 9(3):205–13, 2010.

[59] Anika Kinkhabwala, Zongfu Yu, Shanhui Fan, Yuri Avlasevich, Klaus Müllen, and W. E. Moerner. Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nature Photonics*, 3:654–657, 2009.

[60] Jan Becker, Andreas Trügler, and Arpad Jakab. The Optimal Aspect Ratio of Gold Nanorods for Plasmonic Bio-sensing. *Plasmonics*, 5:161–167, 2010.

[61] Alberto G Curto, Giorgio Volpe, Tim H Taminiau, Mark P Kreuzer, Romain Quidant, and Niek F van Hulst. Unidirectional emission of a quantum dot coupled to a nanoantenna. *Science (New York, N.Y.)*, 329(5994):930–3, 2010.

[62] Matthias Reichelt, Andrea Walther, and Torsten Meier. Tailoring the high-harmonic emission in two-level systems and semiconductors by pulse shaping. *J. Opt. Soc. Am. B*, 29:A36–A42, 2012.

[63] Matthias Reichelt, Andre Hildebrandt, Andrea Walther, Jens Förstner, and Torsten Meier. Engineering high harmonic generation in semiconductors via pulse shaping. *Proc. SPIE*, 8260:82601L, 2012.

[64] Lukas Novotny and Niek van Hulst. Antennas for light. *nature photonics*, 5:83–90, 2011.

[65] Seungchul Kim, Jonghan Jin, Young-Jin Kim, In-Yong Park, Yunseok Kim, and Seung-Woo Kim. High-harmonic generation by resonant plasmon field enhancement. *Nature*, 453(7196):757–60, 2008.

[66] Wei Ding, Renaud Bachelot, Sergei Kostcheev, Pascal Royer, and Roch Espiau de Lamaestre. Surface plasmon resonances in silver Bowtie nanoantennas with varied bow angles. *Journal of Applied Physics*, 108(12):124314, 2010.

[67] Garnett W. Bryant, F. Javier Garcia de Abajo, and Javier Aizpurua. Mapping

the Plasmon Resonances of Metallic Nanoantennas. *Nano letters*, 8(2):631–636, 2008.

[68] Carlo Forestiere, Alyssa J Pasquale, Antonio Capretti, Giovanni Miano, Antonello Tamburrino, Sylvanus Y Lee, Björn M Reinhard, and Luca Dal Negro. Genetically engineered plasmonic nanoarrays. *Nano letters*, 12(4):2037–44, 2012.

[69] Thorsten Feichtner, Oleg Selig, Markus Kiunke, and Bert Hecht. Evolutionary optimization of optical antennas. pages 1–5, 2012.

[70] Andre Hildebrandt, Matthias Reichelt, Torsten Meier, and Jens Förstner. Optimization of the intensity enhancement in plasmonic nanoantennas. In *THE FIFTH INTERNATIONAL WORKSHOP ON THEORETICAL AND COMPUTATIONAL NANO-PHOTONICS: TaCoNa-Photonics 2012*, volume 1475, pages 59–61. AIP Publishing, 2012.

[71] Stefan A. Maier. *Plasmonics: Fundamentals and applications*. Springer, 2007.

[72] Lukas Novotny. Strong coupling, energy splitting, and level crossings: A classical perspective. *American Journal of Physics*, 78(11):1199–1202, 2010.

[73] Jer-Shing Huang, Johannes Kern, Peter Geisler, Pia Weinmann, Martin Kamp, Alfred Forchel, Paolo Biagioni, and Bert Hecht. Mode imaging and selection in strongly coupled nanoantennas. *Nano letters*, 10(6):2105–2110, 2010.

[74] Kevin J Savage, Matthew M Hawkeye, Ruben Esteban, Andrei G Borisov, Javier Aizpurua, and Jeremy J Baumberg. Revealing the quantum regime in tunnelling plasmonics. *Nature*, 491(7425):574–577, November 2012.

[75] Jens Förstner. *Light Propagation and Many-Particle Effects in Semiconductor Nanostructures*. PhD thesis, Technische Universität Berlin, 2005.

[76] Bernd Metzger, Mario Hentschel, Thorsten Schumacher, Markus Lippitz, Xingchen Ye, Christopher B. Murray, Bastian Knabe, Karsten Buse, and Harald Giessen. Doubling the efficiency of third harmonic generation by positioning ito nanocrystals into the hot-spot of plasmonic gap-antennas. *Nano Letters*, 14(5):2867–2872, 2014.

[77] S.G. Johnson. The nlopt nonlinear-optimization package, http://ab-initio.mit.edu/nlopt.

[78] Padmanabhan Pramod and K. George Thomas. Plasmon coupling in dimers of au nanorods. *Advanced Materials*, 20(22):4300–4305, 2008.

[79] Guillaume Boudarham and Mathieu Kociak. Modal decompositions of the local electromagnetic density of states and spatially resolved electron energy loss probability in terms of geometric modes. *Physical Review B*, 85(24):245447, 2012.

[80] Martin Kuttge, Ernst Jan R Vesseur, AF Koenderink, HJ Lezec, HA Atwater, FJ García de Abajo, and Albert Polman. Local density of states, spectrum, and far-field interference of surface plasmon polaritons probed by cathodoluminescence. *Physical Review B*, 79(11):113405, 2009.

[81] FJ García De Abajo. Optical excitations in electron microscopy. *Reviews of modern physics*, 82(1):209, 2010.

[82] Arthur Losquin, Luiz F Zagonel, Viktor Myroshnychenko, Benito Rodríguez-Gonzalez, Marcel Tence, Leonardo Scarabelli, Jens Forstner, Luis M Liz-Marzan, F Javier Garcia de Abajo, Odile Stephan, et al. Unveiling nanometer scale extinction and scattering phenomena through combined electron energy loss spectroscopy and cathodoluminescence measurements. *Nano letters*, 15(2):1229–1237, 2015.

[83] Alain Dereux, Christian Girard, and Jean-Claude Weeber. Theoretical principles of near-field optical microscopies and spectroscopies. *The Journal of Chemical Physics*, 112(18):7775–7789, 2000.

[84] Y. Hsing Fu, A. I. Kuznetsov, A. E. Miroshnichenko, Y. Feng Yu, and B. Luk'yanchuk. Directional visible light scattering by silicon nanoparticles. *Nature Communications*, 4, 2013.

[85] Palash Bharadwaj, Bradley Deutsch, and Lukas Novotny. Optical Antennas. *Advances in Optics and Photonics*, 1(3):438, 2009.

[86] H. F. Hofmann, T. Kosako, and Y. Kadoya. Design parameters for a nano-optical yagi-uda antenna. *New J. Phys.*, 9, 2007.

[87] T. H. Taminiau, F. D. Stefani, F.B. Segerink, and N. F. van Hulst. Optical antennas direct single-molecule emission. *Nature Photonics*, 2:234–237, 2008.

[88] H. Giessen and M. Lippitz. Directing light emission from quantum dots. *Science*, 329:910–911, 2010.

[89] T. Kosako, Y. Kadoya, and F. Hofmann. Directional control of light by a nano-optical yagi-uda antenna. *Nature Photonics*, 4:312–315, 2010.

[90] Timur Shegai, Si Chen, Vladimir D Miljković, Gülis Zengin, Peter Johansson, and Mikael Käll. A bimetallic nanoantenna for directional colour routing. *Nature communications*, 2, 2011.

[91] Heykel Aouani, Oussama Mahboub, Eloïse Devaux, Hervé Rigneault, Thomas W Ebbesen, and Jérôme Wenger. Plasmonic antennas for directional sorting of fluorescence emission. *Nano letters*, 11(6):2400–2406, 2011.

[92] L. Novotny and B. Hecht. *Principles of Nano-Optics*. Cambridge University Press, 2006.

[93] K. Li, M. Stockman, and D. J. Bergman. Self-similar chain of metal nanospheres as an efficient nanolens. *Phys. Rev. B*, 91, 2003.

[94] S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koel, and A. A. G. Requicha. Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides. *Nature Materials*, 2:229–232, 2003.

[95] T. Pakizeh and M. Käll. Unidirectional ultracompact optical nanoantennas. *Nano Letters*, 9:2343–2349, 2009.

[96] D. Dregely, R. Taubert, J. Dorfmüller, R. Vogelsang, K. Kern, and H. Giessen. 3d optical yagi-uda nanoantenna array. *Nature Communications*, 2, 2011.

[97] N. Livneh, A. Strauss, I. Schwarz, I. Rosenberg, A. Zimran, S. Yochelis, G. Chen, U. Banin, Y. Paltiel, and R. Rapaport. Highly directional emission and photon beaming from nanocrystal quantum dots embedded in metallic nanoslit arrays. *Nano Letters*, 11:1630–1635, 2011.

[98] A. E. Krasnok, A. E. Miroshnichenko, P. A. Belov, and Y. S. Kivshar. All-dielectric optical nanoantennas. *Optics Express*, 20:20599–20604, 2012.

[99] Andre Hildebrandt, Matthias Reichelt, Torsten Meier, and Jens Förstner. En-

gineering plasmonic and dielectric directional nanoantennas. In *SPIE OPTO*, pages 89841G–89841G. International Society for Optics and Photonics, 2014.

[100] Sophokles J. Orfanidis. *Electromagnetic Waves and Antennas*. Rutgers University, 2014.

[101] P. Viezbicke. Yagi antenna design. *National Bureau of Standards Technical Note*, 688, 1968.

[102] M. Peter, C. Friesen, and S. Linden. Fabrication and characterization of active plasmonic antennas. In *2015 9th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics (METAMATERIALS)*, pages 256–258, 2015.

[103] Olivier Merchiers, F Moreno, F Gonzalez, and JM Saiz. Light scattering by an ensemble of interacting dipolar particles with both electric and magnetic polarizabilities. *Physical Review A*, 76(4):043834, 2007.

[104] Alexander E Krasnok, Constantin R Simovski, Pavel A Belov, and Yuri S Kivshar. Superdirective dielectric nanoantennas. *Nanoscale*, 6(13):7354–7361, 2014.

[105] Alexander E Krasnok, Dmitry S Filonov, Constantin R Simovski, Yuri S Kivshar, and Pavel A Belov. Experimental demonstration of superdirective dielectric antenna. *Applied Physics Letters*, 104(13):133502, 2014.

[106] PA Belov, AE Krasnok, DS Filonov, CR Simovski, and Yu S Kivshar. Superdirective all-dielectric nanoantennas: theory and experiment. In *IOP Conference Series: Materials Science and Engineering*, volume 67, page 012008. IOP Publishing, 2014.

[107] J Van de Groep and A Polman. Designing dielectric resonators on substrates: Combining magnetic and electric resonances. *Optics express*, 21(22):26285–26302, 2013.

[108] Mark J. Holmes, Kihyun Choi, Satoshi Kako, Munetaka Arita, and Yasuhiko Arakawa. Room-temperature triggered single photon emission from a iii-nitride site-controlled nanowire quantum dot. *Nano Letters*, 14(2):982–986, 2014.

[109] Marie Anne van de Haar, Jorik van de Groep, Benjamin JM Brenny, and Al-

bert Polman. Controlling magnetic and electric dipole modes in hollow silicon nanocylinders. *Optics express*, 24(3):2047–2064, 2016.

Own Publications

[1] M. Reichelt, A. Hildebrandt, A. Walther, J. Förstner, and T. Meier, "Engineering high harmonic generation in semiconductors via pulse shaping", Proc. SPIE **8260**, 82601L (2012)

[2] A. Hildebrandt, M. Reichelt, T. Meier, and J. Förstner, "Optimization of the intensity enhancement in plasmonic nanoantennas", in *Proceedings of the 5th International Workshop on Theoretical and Computational Nano-Photonics*, AIP Conference Proceedings **1475**: 59-61 (2012)

[3] A. Hildebrandt, M. Reichelt, T. Meier, and J. Förstner, "Engineering plasmonic and dielectric directional nanoantennas", Proc. SPIE **8984**, 89841G (2014)

[4] M. Hammer, A. Hildebrandt, and J. Förstner, "How planar optical waves can be made to climb dielectric steps", Optics Letters **40**(16): 3711-3714 (2015)

[5] F. Zeuner, M. Muldarisnur, A. Hildebrandt, J. Förstner, and T. Zentgraf, "Coupling Mediated Coherent Control of Localized Surface Polaritons", Nano Letters **15**(6): 4189-4193 (2015)

[6] M. Hammer, A. Hildebrandt, and J. Förstner, "Full Resonant Transmission of Semiguided Planar Waves Through Slab Waveguide Steps at Oblique Incidence", Journal of Lightwave Technology **34**: 997-1005 (2016)

[7] A. Hildebrandt, S. Alhaddad, M. Hammer, and J. Förstner, "Oblique incidence of semi-guided waves on step-like folds in planar dielectric slabs: Lossless vertical interconnects in 3D integrated photonic circuits", Proc. SPIE **9750**, 97501B (2016)

[8] S. Blumenthal, M. Bürger, A. Hildebrandt, J. Förstner, N. Weber, C. Meier, D. Reuter, and D. As, "Fabrication and characterization of two-dimensional cubic AlN photonic crystal membranes containing zincblende GaN quantum dots", Physica Status Solidi C **13**(5-6), 292-296 (2016)

List of Figures

Danksagung

Zuerst möchte ich mich bei meinem Doktorvater Prof. Dr. Jens Förstner für die freundliche Aufnahme in seiner Arbeitsgruppe danken. Als Gruppenleiter hat er immer für einen guten Ton und einer angenehmen Arbeitsatmosphäre gesorgt. Zudem hatte er immer ein offenes Ohr, wenn Probleme jeglicher Art anstanden und stand mir mit Rat und Tat zur Seite. Gedankt sei auch Prof. Dr. Torsten Meier, der nicht nur als Zweitgutachter viel zu lesen hatte, sondern mir immer mit seiner Expertise zur Seite stand und immer eine Antwort auf Fragen hatte.

Bedanken möchte ich mich auch bei Dr. Matthias Reichelt, der die unglaubliche Fähigkeit besitzt, die richtigen Fragen zu stellen, sodass man immer einen Ansatz findet Probleme zu verstehen. Auch allen anderen Mitgliedern der Gruppen von Prof. Dr. Meier, Prof. Dr. Schumacher und Prof. Dr. Förstner gilt mein Dank für die unzähligen, fruchtvollen Diskussionen beim Kaffee, und auch weil einige wirklich gute Freunde sind.

Peter Kölling sei dafür gedankt, dass er diese Arbeit bereitwillig korrektur-gelesen hat.

Bedanken möchte ich mich auch ganz herzlich bei meiner Frau Susanne, ohne deren Nachdruck und der Fähigkeit zuzuhören, auch wenn es sie nicht interessiert, ich längst verzweifelt wäre. Sie hat mich immer unterstützt und auch häufige Abwesenheit z.B. durch Workshops oder Konferenzen geduldig ertragen. Auch meinen Freunden möchte ich danken, denn sie waren immer da, wenn ich sie brauchte.

Weiterhin möchte ich der Forschungsgemeinschaft DFG danken, die mich im Rahmen des Graduiertenkollegs GRK 1464 *Micro- and nanostructures in optoelectronics and photonics* finanziell unterstützt hat.

Zuletzt will ich mich bei unserem Druckwasserwirkstoffextraktor bedanken, der immer fehlerlos arbeitete und dafür sorgte, dass ich auch bei langen Arbeitstagen unter Druck noch effizient arbeiten konnte.