

Nonlinear Optics on the Nanoscale: Solid State and Quantum Applications

by

PAUL MARIA DICHTL

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
and the Diploma of Imperial College London

Supervised by
Professor Stefan Maier and Professor Rupert Oulton

Experimental Solid State Group (EXSS)
Department of Physics
Imperial College London
London SW7 2AZ, U.K.

July 2021

To my parents

Declaration of Originality

I hereby declare that, to the best of my knowledge, the work presented in this thesis is my own and that I appropriately acknowledged and referenced the contributions of others.

Paul Dichtl

London, July 2021

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution 4.0 International Licence (CC BY).

Under this licence, you may copy and redistribute the material in any medium or format for both commercial and non-commercial purposes. You may also create and distribute modified versions of the work. This on the condition that you credit the author.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Nonlinear optical processes in solid-state systems are typically weak but the strength of the nonlinear interactions can be significantly increased in nanophotonic structures providing a local enhancement of the optical fields. This thesis deals with three experiments in this thematic area.

The Kerr nonlinearity of 2D Ruddlesden-Popper-phase (2D RPP) lead halide perovskite flakes is investigated by means of the Z-scan method, after which the flakes are combined with an array of Al nanoantennae to form a nonlinear metasurface. Both nonlinear absorption and refraction are strongly enhanced near the exciton ground state of the 2D RPP with peak values of $\beta_{\text{eff}} = -256 \text{ cm/MW}$ and $n_2 \sim \pm 10^{-13} \text{ m}^2/\text{W}$ comparable to top values in the literature. The combined metasurface leads to a further enhancement of the Kerr nonlinearity attributed to the strong near-fields of the nanoantennae, however, with a complicated saturation behaviour analysed separately. Moreover, a partial hybridisation of the exciton and plasmon modes is found.

Wavelength-sized cones etched into a GaN layer are studied for their improvement of frequency mixing efficiencies in the visible-NIR range compared to an unstructured GaN layer. As a preliminary result, the second-harmonic generation (SHG) efficiency is found to be increased by a factor of $\lesssim 10$. In future samples with smaller cones, pumped at the magnetic dipole resonance, a larger efficiency increase is expected.

Frequency mixing in a metasurface of Au nanoantennae is investigated with the aim of working towards sources of visible-NIR photon pairs with tailored spectral properties. Frequency

scans to map out the joint spectral density are conducted, which will allow for testing the spectral properties of improved future antenna designs. The expected spontaneous parametric down-conversion (SPDC) count rates in the GaN and Au nanoantenna samples are estimated via the stimulated emission tomography (SET) technique.

Contents

1	Introduction	1
1.1	Historical Background of Nanophotonics	1
1.2	Optical Fields on the Nanoscale	3
1.3	Applications in Nonlinear Optics	4
1.4	Quantum Applications	5
1.5	Scientific Aims and Thesis Overview	6
2	Nonlinear Optics	9
2.1	From Linear to Nonlinear Optics	9
2.2	Nonlinear Changes of the Refractive Index	12
2.3	Frequency Mixing	15
2.4	Phase-Matching	19
2.5	Photon Sources	22
2.6	Stimulated Emission Tomography	28
2.7	Conclusion	32
3	Plasmonics	33
3.1	Fundamentals	33
3.2	Plasmonic Resonances in Metal Nanoparticles	39
3.3	Plasmonic Nanoparticles as Optical Antennae	43
3.4	Conclusion	45
4	The Z-Scan Technique	47

4.1	Z-Scan Experimental Setup	47
4.2	Preliminary Remarks on Z-Scan Analysis	50
4.3	Off-Resonance Theory for a Weak $\chi^{(3)}$ Nonlinearity	53
4.3.1	Derivation of OA and CA Fitting Formulae	53
4.3.2	Linear Truncation of the OA and CA Expressions	58
4.4	Analysing Z-Scan Data with Saturable Absorption	60
4.4.1	Phenomenological Models of Saturable Absorption	61
4.4.2	A Physical Model for Saturable Absorption	66
4.5	Numerical Simulations of OA Z-Scan Traces	70
4.6	Conclusion	76
5	2D Ruddlesden-Popper Phase Perovskites	79
5.1	Chemical Composition and Crystal Structure	80
5.1.1	Lead-Halide Perovskites	80
5.1.2	Structure and Fabrication of 2D RPP Flakes	81
5.1.3	Mechanisms Affecting Exciton States and Optical Properties	83
5.2	Linear and Nonlinear Optical Properties	84
5.3	Perovskite Flakes Coupled to Al Nanoantennae	91
5.4	Analysing the Exciton Saturation Behaviour	98
5.5	Conclusion	109
6	Nonlinear Frequency Mixing in GaN	113
6.1	Motivation	114
6.2	Sample Structure	116
6.3	Coupling Input Fields and Active Tensor Elements	118
6.4	Frequency Mixing in the GaN Slab	121
6.4.1	Dual-Beam Study with Tilted Sample	122
6.4.2	SHG Wavelength Scan	124
6.4.3	Dual-Beam Frequency Mixing: Up- and Down-Conversion	128
6.5	Conversion Efficiency and SET Analysis	132

6.5.1	SET Assumptions	133
6.5.2	Analysis Framework and Overview of Results	135
6.5.3	Spontaneous Photon Pair Generation Rates	137
6.5.4	Frequency Mixing Efficiency	140
6.5.5	Reciprocity of Up- and Down-Conversion	141
6.6	Frequency Mixing in GaN Cones	143
6.6.1	Cone Fabrication	144
6.6.2	Simulations	147
6.6.3	Experimental Results	150
6.7	Future Measurements on New Cone Samples	156
6.8	Conclusion	159
7	Frequency Mixing in Gold Nanoantennae	163
7.1	Nonlinear Plasmonics	164
7.2	Origin of the Second-Order Nonlinearity of Gold	165
7.3	Sample and Simulations	169
7.4	Experimental Results	173
7.4.1	Output Polarisation Study	173
7.4.2	DFG Measurements	175
7.4.3	Suppression of DFWM from the Substrate	177
7.4.4	Spectrograms of Dual-Beam Signals	179
7.4.5	Joint Spectral Density (JSD) Maps of Dual-Beam Signals	181
7.5	Efficiency and SET Analysis	183
7.5.1	Spontaneous Photon Pair Generation Rates	185
7.5.2	Frequency Mixing Efficiency	187
7.5.3	Reciprocity of Up- and Down-Conversion	188
7.5.4	Frequency Mixing in a Single Antenna	189
7.6	New Triply-Resonant Antenna Design	191
7.7	Conclusion and Outlook	192

8 Conclusion	195
8.1 Summary	195
8.2 Outlook	197
Appendices	201
A Details of the Experimental Setups	203
A.1 Z-Scan	203
A.2 Frequency Mixing	204
B Fabry-Perot Model of the GaN Slab	209
C Details of the SET Analysis	213
C.1 Detection Efficiency Measurements	213
C.2 Beam Calculations	214
C.3 SET Details for the GaN Sample	216
C.4 SET Details for the Au Nanoantenna Sample	217
D Mie-Modes in Dielectric Spheres	221
Bibliography	223
Acknowledgements	249

Chapter 1

Introduction

1.1 Historical Background of Nanophotonics

Nanophotonics marks the intersection of optics and electronics [1] and is a consequence of the historical trajectories of both fields. On the optics side, until the 20th century, technological advances were mostly limited to magnifying devices aiding human vision, such as lenses and steadily improving versions of telescopes and microscopes [2]. Within this context it was discovered that light cannot resolve arbitrarily small structures, but that the minimum size is given by the diffraction limit $\sim \lambda/2$, described by Abbe in 1873, which cannot be overcome with macroscopic lens systems regardless of their design and quality [3]. By reciprocity, the diffraction limit also determines the minimum volume into which light can be squeezed [4].

Around the same time, the theoretical understanding of light was advanced by Maxwell, who identified it as electromagnetic waves [5,6]. The first experiments to explicitly generate electromagnetic waves with an antenna were conducted by Hertz [6,7], and soon wireless communication became a reality due to Marconi's pioneering work on improved antenna designs [6,8], however, radio-frequency waves were used for this purpose. The concept of optical antennae [9], marrying optics and electronics, was still a long way off and not yet imaginable at the time. With the ever-increasing need for fast remote communication during

the 20th century, radio technology, dealing with signal amplification and processing using electronic circuits, emerged as an active area of research in its own right [6]. It was this rapid development of electronics and the drive to miniaturise devices for a faster and more efficient performance that brought nanophotonics within reach as the logical consequence of this development [10].

One milestone on this journey was the introduction of semiconductor transistors allowing for energy-efficient signal amplification, first as point contact transistors [11] and bipolar junction transistors [12], and only a decade later as the much more scalable and now ubiquitous silicon-based field-effect transistor (MOSFET) [13,14]. Over the course of the following decades, integrated circuits with MOSFETs were built with an ever-increasing device density [15] for powerful computers, leading to enormous advances in fabrication technology. While the main development occurred in photolithography, allowing for large-scale manufacturing [16], alternative technologies, such as electron-beam lithography [17] and focussed-ion-beam lithography [18] were invented with the aim of producing structures of just a few nanometres in size, unhampered by the aforementioned diffraction limit of light[†]. These new nanofabrication capabilities have enabled researchers to transition from photonics [22], which is an umbrella term for the generation, modulation, and detection of optical signals, to nanophotonics [23].

This transition to sub-wavelength photonics is the logical consequence of the evolution of photonic communication systems driven by the demand for increasingly complex multiplexers and interconnects in order to handle the ever-growing data volume [24]. Optical circuits on silicon with a high density of components were constructed to implement such devices [25], however, given the diffraction limit of infrared light, the feature size of any component has been limited to several hundred nanometres despite the ability to fabricate much smaller structures. The last crucial ingredient for entering the realm of sub-100 nm optics with

[†]Of course, electron and ion beams also have an associated wavelength but the fabrication resolution is not limited by the diffraction limit but by other effects, such as unwanted resist exposure by secondary electrons in e-beam lithography [19]. Photolithography has seen enormous advances in shorter-wavelengths sources and optics in recent years. Modern extreme-UV systems [20] can easily compete with charged particle beams, reaching feature sizes of 6 nm half-pitch [21].

visible light was the availability of metal-dielectric surface modes, called plasmonic modes, which admit deep sub-wavelength nanoscale confinement of optical fields [23, 26, 27].

1.2 Optical Fields on the Nanoscale

Plasmonic modes, which occur as propagating surface plasmon polaritons (SPPs) or localised surface plasmon resonances (LSPRs), have two key properties making them scientifically and technologically interesting. The first feature is the aforementioned sub-wavelength confinement, while the second is the concomitant strong local field enhancement [28]. However, these benefits come at the cost of ohmic losses in the metal, such that a trade-off between confinement and loss needs to be considered in the design of any plasmonic device [28]. In fact, developing strategies to minimise or avoid plasmonic losses has claimed an increasing fraction of nanophotonics research activity in recent years, giving rise to the subfield of all-dielectric nanophotonics. In this sub-area at the low-loss low-confinement end of nanophotonics, Mie-resonances in high-index dielectric structures with a size of approximately half the wavelength are used to achieve a moderate local field enhancement [29].

In terms of applications, nanophotonics and particularly plasmonics has proven to be a horn of plenty, offering innovations in a diverse range of thematic areas [30]. Due to fast group velocities of SPPs close to the speed of light, they are interesting for fast interconnects in integrated circuits, promising to significantly reduce delay times compared to traditional electronic interconnects which suffer from increased wire resistivity on the nanoscale [31]. However, dissipative losses of SPPs restrict this application to short and intermediate-length interconnects [31]. On the other hand, plasmonic dissipation and heating may be used as a feature, such as in cancer therapy [32], where metal nanoparticles are accumulated in a tumour and heated up with a laser at a plasmonic resonance of the particle, destroying cancer cells by hyperthermia. Remaining with biology-related applications, plasmonic nanoparticles are also useful as highly sensitive biosensors [33]. Since the LSPR of a nanoparticle depends on the local refractive index close to the surface, functionalising it with receptors that will bind

to a specific molecule allows for the detection of even small concentrations of the latter [34]. Moreover, plasmonic particle dimers support extremely enhanced local fields allowing for fluorescence enhancement from a single molecule by three orders of magnitude [35] or the enhancement and re-shaping of a Raman spectrum from a single molecule [36].

1.3 Applications in Nonlinear Optics

The strong local field enhancement is also highly beneficial for nonlinear optics applications [37], which constitute the close scientific context of the research presented in this thesis. Nonlinear effects are weak perturbative corrections to the linear response of a material and hence are proportional to higher-order terms of the susceptibility [38]. Given this super-linear scaling, enhanced local fields may dramatically increase the efficiency of nonlinear processes [37]. Other key benefits are the down-scaling of nonlinear optical devices, since a strong nonlinear interaction is achievable in nanoscopic volumes rather than requiring accumulation in macroscopic structures, and the femtosecond response time of plasmonic excitations as well as the aforementioned refractive index sensitivity usable in conjunction with nonlinear refractive index changes [37]. An excellent overview of the wide range of plasmonic applications for nonlinear optical devices and fundamental phenomena at the intersection of plasmonics and nonlinear optics is provided in Ref. [37]. For the purposes of this introduction, it is sufficient to cover but a selection of recent research which is thematically closest to the results in the following chapters.

One topic of particular interest is the combination of a plasmonic metasurface, which is an array of plasmonic particles with subwavelength spacing [39], and an adjacent nonlinear medium with a strong Kerr nonlinearity. Since plasmonic particle resonances strongly depend on the local refractive index [40], a nonlinear refractive index shift will be accompanied by a shift of the plasmonic resonance and hence a strong variation of the extinction at a particular wavelength [41]. Of course, the Kerr nonlinearity itself is strengthened by the field enhancement provided by the plasmonic particles, which is interesting for low-power devices,

for example as a contribution to the field of metasurface optical switches [42–44]. In this thesis, a sample with cross-shaped Al nanoparticles and a 2D Ruddlesden-Popper-phase perovskite flake will be investigated in this context, with the results presented in Chapter 5.

The other interesting theme is the enhancement of nonlinear frequency mixing processes with plasmonic [45] and dielectric [46] nanophotonic structures, exploiting the aforementioned superlinear scaling. In this thesis, Chapters 6 and 7 showcase work in these areas, based on GaN cone resonators and plasmonic nanoparticle antennae, respectively. As we shall see particularly in Chapter 6, not only the local field enhancement, but also the alignment of the fields and the nonlinear susceptibility tensor components is important for a high frequency conversion efficiency [47]. Moreover, in the plasmonic antennae in Chapter 7, second-order processes benefit from an overall non-centrosymmetric arrangement of the constituent nanoparticles, which builds on previous work investigating this idea [48, 49].

1.4 Quantum Applications

Efficient nonlinear mixing in nanophotonics is not only interesting in itself but also for low-power photon sources. This is because certain nonlinear effects, such as spontaneous parametric down-conversion (SPDC), are characterised by a photon output where each photon has a partner photon generated at the same time, allowing for the construction of heralded single photon sources [50] and sources of entangled photon pairs [51]. A high mixing efficiency thus translates into lower power requirements at a given photon generation rate; moreover, the small required interaction volume allows for on-chip photon sources [52]. Furthermore, in the case of nonlinear metasurfaces consisting of plasmonic or dielectric nanoparticles, phase-matching constraints are relaxed and the properties of the photon pairs are controllable via the metasurface design [53]. One example for the far-reaching design and control possibilities is recent work on a metasurface directly generating photon pairs entangled in their orbital angular momentum states, with applications in quantum information processing [54].

The research presented in Chapter 7 of this thesis will focus on controlling the spectral properties of the photon output via the nanoparticle resonances, which could be interesting for applications in quantum imaging. In quantum imaging [55], signal and idler pairs are generated in two nonlinear interaction regions such that the idler modes are indistinguishable. This allows for an indirect imaging of the object in the idler beam via interference of the signal photons, which is technologically interesting for the case where the wavelength of the idler photons interacting with the object is different from that of the detected signal photons [55]. The plasmonic nanoparticles in Chapter 7 would be useful in this context, since the particle resonances can be designed such that photon pairs with a large frequency spacing are produced, potentially extending the spectral range of quantum imaging systems.

1.5 Scientific Aims and Thesis Overview

With the context presented above, the scientific aims of this thesis can be summarised as follows. The Kerr nonlinearity of a 2D perovskite flake is to be characterised with the Z-scan method and enhanced with a plasmonic metasurface. Moreover, the excitonic properties are to be studied in the context of a hybridisation of excitonic and plasmonic states, which goes beyond nonlinear optics and bridges the gap to solid-state physics. Next, dielectric and plasmonic nanostructures are to be studied for their frequency mixing properties and the resonant enhancement of the nonlinear material responses. Finally, the joint spectral density of the photon output of these nanostructures is to be characterised and the expected spontaneous photon pair generation rate is to be estimated using a technique called Stimulated Emission Tomography (SET) [56].

In Chapter 2, the fundamental concepts of nonlinear optics will be explained, with an emphasis on the key concepts needed to understand the experiments presented in later chapters. Aside from the fundamentals of frequency mixing processes, the concept of phase-matching and the SET technique will be discussed.

Plasmonics, which is responsible for the large near-field enhancements boosting the effective nonlinearity of nanostructures, is the topic of Chapter 3. After a streamlined derivation of plasmonic modes at a planar interface and a discussion of their key properties, sub-wavelength confinement and local field enhancement, plasmonic resonances in metal particles are discussed. The tuneability of the particle resonances is demonstrated for the simple case of spheroid particles. Finally, the antenna properties of plasmonic particles and the concept of a Fano resonance are explained, since these will be needed as a background for the antenna chapter.

Chapter 4 concludes the set of theory chapters with a discussion of the Z-scan method for determining the strength of the Kerr nonlinearity of thin-film samples. In contrast to the two preceding chapters, the Z-scan theory is about a specific technique, which is essentially already an application of nonlinear optics. After a derivation and explanation of the most commonly used formulae for analysing Z-scan data, their limitations and ways to extend the theory are discussed in great detail, since blindly applying the standard formulae is not sufficient to analyse the data presented in the following chapter. In particular, misconceptions in the literature are pointed out and a numerical framework for data analysis is introduced. Therefore, this chapter is not just a summary of existing theory, but contains significant original work on the analysis of Z-scan data from samples exhibiting strong saturable absorption.

Experimental work on 2D Ruddlesden-Popper-phase perovskite flakes (2D RPPs) is presented in Chapter 5. After an explanation of the sample structure and fabrication, the linear and nonlinear optical properties of the 2D RPPs are investigated. It is found that nonlinear absorption and refraction are strongly enhanced at the fundamental excitonic resonance of the 2D RPP flakes. Next, the flakes are combined with an array of plasmonic nanoparticles to form a nonlinear metasurface. Aside from the enhancement of the Kerr nonlinearity, a partial hybridisation of plasmonic and excitonic resonances is observed.

In Chapter 6, frequency mixing experiments with a GaN sample are presented. First, a GaN slab is examined as a control sample, with a study on the relevant tensor elements in second-

order interactions and a collection of various second- and third-order mixing processes. Then, GaN cones are introduced, which were etched directly into the slab. After a brief discussion of the fabrication technique, the frequency mixing experiments are repeated on the cone and compared to the slab data and supplemented with FDTD simulations. A moderate enhancement of up to one order of magnitude is found, which is attributed partly to a Mie-resonance of the cone, but mostly to an optical focussing effect. The expected photon pair generation rate in the quantum limit is estimated using the SET technique and discussed in the context of recent results in the literature. Moreover, the reciprocity of the up- and down-conversion frequency mixing efficiency is shown.

These results also serve as a reference for frequency mixing results with plasmonic nanoparticles, presented in Chapter 7. After a discussion on second-order nonlinear processes in Au, which exist despite the centrosymmetry of the lattice, the Au nanoparticle sample is introduced including simulation results describing the near-field coupling between the nanoparticles. Next, up- and down-conversion frequency mixing results are shown and the photon pair creation rates are estimated using the SET technique, as in the preceding chapter.

Finally, all results are summarised and discussed in Chapter 8. The thesis closes with an outlook on further work following the three main experimental paths shown in the preceding chapters, in particular the antenna-perovskite metasurface, novel GaN cones, and a triply resonant plasmonic antenna improving on the existing design for multi-beam nonlinear mixing experiments.

Chapter 2

Nonlinear Optics

In this chapter we shall see how nonlinear optical effects arise and why strong electrical fields are usually needed to bring their efficiency to an appreciable level. The effects covered are nonlinear refraction and absorption, which will be at the heart of Chapters 4 and 5, as well as frequency mixing, which is the central topic of the experimental work presented in Chapters 6 and 7. At the end of this chapter we shall see how frequency mixing is of use in quantum optics for single-photon and photon-pair generation and how the stimulated emission tomography (SET) technique allows for an efficient characterisation of the corresponding biphoton wavefunction using only classical nonlinear processes.

2.1 From Linear to Nonlinear Optics[†]

Light consists of electromagnetic fields obeying Maxwell's equations. Let us begin our discussion by considering the propagation of light in vacuum, where – due to the absence of charges and currents – Maxwell's equations are given by

$$\nabla \cdot \vec{E} = 0 \quad \nabla \cdot \vec{B} = 0 \quad \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad \nabla \times \vec{B} = \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \quad (2.1)$$

[†]This section and the following three sections closely follow selected arguments from Ref. [38] including similar re-drawn diagrams.

and upon taking the curl of the third equation and inserting it into the fourth equation we obtain the following wave equation:

$$\Delta \vec{E} - \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \quad (2.2)$$

with $c = (\mu_0 \varepsilon_0)^{-1/2}$. The well-known solutions are plane waves with perpendicular electric and magnetic fields propagating with the speed of light c . Note that due to the linearity of Eqs.(2.1) two solutions $(\vec{E}_1, \vec{B}_1)$ and $(\vec{E}_2, \vec{B}_2)$ can be added to form a third solution satisfying the equations. This fact implies that two light waves in vacuum cannot influence each other and – as long as we ignore nonlinear quantum electrodynamic effects – we are unable to study nonlinear optical phenomena.

This changes, however, as soon as we allow light to interact with matter. In the most general case, the material response to an external electromagnetic field consists of an oscillating polarisation and currents with a magnetic moment, which are included in the so-called macroscopic Maxwell's equations

$$\nabla \cdot \vec{D} = \rho \quad \nabla \cdot \vec{B} = 0 \quad \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad \nabla \times \vec{H} = \vec{j} + \frac{\partial \vec{D}}{\partial t} \quad (2.3)$$

with the electric displacement field $\vec{D} = \varepsilon_0 \vec{E} + \vec{P}$, where $\vec{P}$ is the induced polarisation given by the dipole moment per unit volume, and the magnetic field $\vec{H} = \frac{1}{\mu_0}(\vec{B} - \vec{M})$. As we are only interested in materials without a magnetic response, we will set $\vec{B} = \mu_0 \vec{H}$ for the remainder of this chapter. Moreover, unless stated otherwise we shall only consider interactions at a single optical frequency and isotropic materials to avoid unnecessary complexity and cluttered notation when explaining the key concepts.

The medium interacting with the light is said to have a linear response if $P = \varepsilon_0 \chi E \propto E$, meaning that the susceptibility χ must not depend on E . Experimentally, one finds that this is a good description of most materials for weak illumination, however, for sufficiently strong electric fields, this proportionality is significantly violated. Therefore, it is expedient

to express the susceptibility as a series expansion

$$\chi = \chi^{(1)} + \chi^{(2)}E + \chi^{(3)}E^2 + \dots \quad (2.4)$$

which captures the altered optical properties of the material for intense illumination via weak perturbative correction terms. It is clear that the nonlinear contributions to χ must decline rapidly with increasing order of the perturbation to ensure convergence of the series. Hence, even for strong optical fields, which are necessary to observe nonlinear effects of appreciable magnitude, $\chi^{(2)}$ and $\chi^{(3)}$ are sufficient for characterising the vast majority of practically relevant nonlinear optical effects. Of course, for a sufficiently strong illumination, the material may suffer irreversible damage, in which case Eq.(2.4) does not converge.

It is important to note that not all materials possess non-zero contributions of all orders in the expansion Eq.(2.4). The even-order nonlinear susceptibilities of inversion-symmetric media vanish, as can be made plausible with the following simple argument. Since the second-order polarisation $P = \varepsilon_0\chi^{(2)}E^2$ is subject to inversion symmetry, inverting the input field is accompanied by an inversion of the polarisation response, such that $-P = \varepsilon_0\chi^{(2)}(-E)^2 = \varepsilon_0\chi^{(2)}E^2$. However, this implies $P = -P$, which can only be satisfied for $\chi^{(2)} = 0$ except for the trivial case of a vanishing input field E . It is clear that this argument also holds for any even-order contribution to the susceptibility, making $\chi^{(3)}$ the leading nonlinear term in centrosymmetric media, which is of high practical relevance, as many commonly used materials in optical research fall into this category. Note that symmetry breaking at a surface makes it possible to obtain a non-zero second-order nonlinear response in centrosymmetric materials; this will be discussed in more detail for the example of gold in Chapter 7.

Nonlinear processes can not only be categorised by their order in the driving field E , but also by whether they lead to the generation of frequencies distinct from the input fields or not. Let us consider the example of a material with a $\chi^{(3)}$ nonlinearity pumped with a

monochromatic input field $E = E_0 \cos(\omega t)$. The nonlinear polarisation is given by

$$P^{(3)} = \varepsilon_0 \chi^{(3)} E^3 = \frac{\varepsilon_0}{4} \chi^{(3)} E_0^3 \cos(3\omega t) + \frac{3\varepsilon_0}{4} \chi^{(3)} E_0^3 \cos(\omega t) \quad (2.5)$$

where the first term corresponds to third-harmonic generation (THG), which is a frequency mixing process, and the second term oscillates at the input frequency, which causes an intensity-dependent refractive index change as we shall see below.

In the next section, we shall take a closer look at the latter process, which is crucial for the Z-scan technique discussed in Chapters 4 and 5. Frequency mixing, the other main class of nonlinear processes, will be discussed in Section 2.3 and is the common theme of the results in Chapters 6 and 7.

2.2 Nonlinear Changes of the Refractive Index

In the previous section, we have seen that the third-order polarisation possesses a component oscillating at the incoming frequency and hence at the same frequency as the linear part of the response. Combining the relevant term in Eq.(2.5) with the linear polarisation response, we obtain

$$P_\omega = \varepsilon_0 \chi^{(1)} E_0 + \frac{3\varepsilon_0}{4} \chi^{(3)} E_0^3 = \varepsilon_0 \underbrace{\left(\chi^{(1)} + \frac{3}{4} \chi^{(3)} E_0^2 \right)}_{\chi_{\text{eff}}} E_0 \quad (2.6)$$

where P_ω is the polarisation component oscillating at the driving frequency ω . Using the relationships $n^2 = 1 + \chi_{\text{eff}}$, $n_0^2 = 1 + \chi^{(1)}$, and $n^2 - n_0^2 = (n + n_0)(n - n_0) \approx 2n_0(n - n_0)$ we find

$$2n_0(n - n_0) = \frac{3}{4} \chi^{(3)} E_0^2 = \frac{3}{2\varepsilon_0 c n_0} \chi^{(3)} I \quad (2.7)$$

and introducing the intensity-dependent refractive index $n = n_0 + n_2 I$ we can identify

$$n_2 = \frac{3}{4\varepsilon_0 c n_0^2} \chi^{(3)} \quad (2.8)$$

which demonstrates that the third-order susceptibility is indeed linked to an intensity-dependent refractive index change. The result above can be generalised to include absorption by introducing a complex refractive index $n \rightarrow n + i\kappa$ and a complex $\chi^{(3)}$. Matching the real and imaginary parts of Eq.(2.7) in that case leads to the approximate relationships

$$n_0 n_2 - \kappa_0 \kappa_2 = \frac{3}{4\varepsilon_0 c n_0} \Re(\chi^{(3)}) \quad \text{and} \quad n_0 \kappa_2 + \kappa_0 n_2 = \frac{3}{4\varepsilon_0 c n_0} \Im(\chi^{(3)}) \quad (2.9)$$

which imply $\Re(\chi^{(3)}) \propto n_2$ and $\Im(\chi^{(3)}) \propto \kappa_2$ for materials with weak linear absorption[†].

The arguments so far have been ‘macroscopic’ in the sense that $\chi^{(3)}$ only identifies the part of the material response proportional to the third order of the incident field strength, but does not provide details about the underlying microscopic physical processes. Let us now outline the characteristics of the processes linked to the real and imaginary part of $\chi^{(3)}$. The real part of $\chi^{(3)}$, and more generally the real part of the whole susceptibility tensor, arises due to so-called parametric processes, where only transitions between virtual levels occur, but no absorption takes place. Hence, the response time is usually very fast, as energy conservation may only be violated for small times $t \sim \Delta E/\hbar$, which is typically on the order of femtoseconds.

Figure 2.1(a) shows the possible third-order parametric processes causing a change of the real part of the refractive index, which can be decomposed into one-photon resonant processes and two-photon resonant processes. It can be shown that the former lead to a nonlinear response with the same handedness as the driving field, whereas the latter produce a nonlinear response of both the same and opposite handedness. As a consequence, non-trivial effects, such as the rotation of the polarisation ellipse occur even for the simple case of an isotropic medium with off-resonant electronic contributions to $\chi^{(3)}$, whereas linearly polarised light is not subject to a rotation of its polarisation state. The relative strength of one-photon resonant processes and two-photon resonant processes depends on the underlying physical process, such as molecular orientation.

[†]This particular derivation involving Eqs.(2.7)-(2.9) is adapted from Ref. [57].

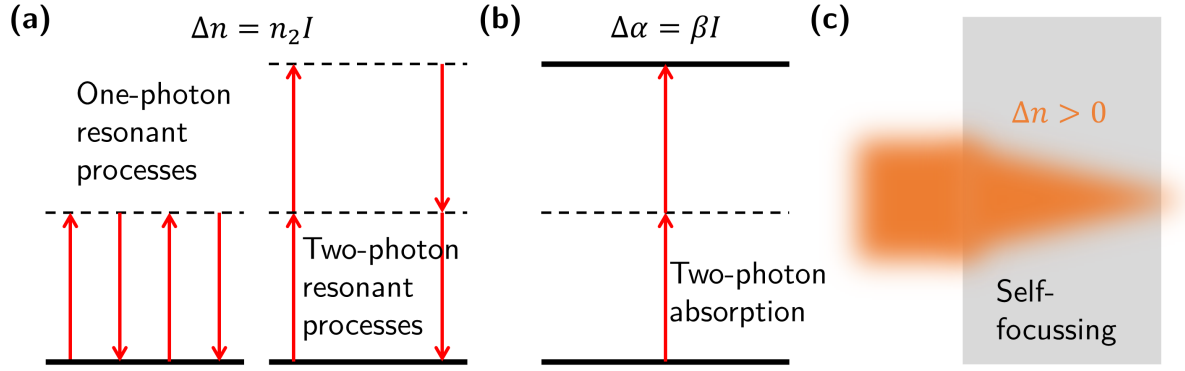


Figure 2.1 – Nonlinear changes of the complex refractive index. (a) Parametric processes via virtual levels change only the real part of the complex refractive index. The processes associated with the optical Kerr effect, and hence with the third-order susceptibility, can be decomposed into one-photon resonant processes and two-photon resonant processes. (b) The imaginary part of $\chi^{(3)}$ is determined by non-parametric processes, which involve the transfer of charges between real energy levels. An example of such a process is two-photon absorption. (c) Self-focussing is a well-known effect in nonlinear media, where an intense laser beam creates a positive refractive index change and focusses itself. Given the self-reinforcing nature of this process, it is dangerous for high-power optical applications, where it may lead to optical damage. All diagrams were re-drawn similar to Ref. [38]

Non-parametric processes, on the other hand, are linked to the imaginary part of χ and involve the promotion of electrons to a higher energy level, which is why the total energy of the propagating field is not conserved. One example of this is two-photon absorption (TPA), shown in Figure 2.1(b), which is often characterised with a parameter β_{TPA} such that [58]

$$\frac{dI}{dz} = -\tilde{\alpha}I = -\alpha I - \beta_{\text{TPA}}I^2 \quad (2.10)$$

for light propagating in the z -direction and with a linear absorption coefficient $\alpha = \frac{4\pi}{\lambda}\kappa_0$. Substituting the imaginary part of the total refractive index, $\kappa_0 + \kappa_2 I$, into the differential equation for linear absorption with a modified absorption coefficient $\tilde{\alpha}$

$$\frac{dI}{dz} = -\tilde{\alpha}I = -\frac{4\pi}{\lambda}\kappa_0 I - \frac{4\pi}{\lambda}\kappa_2 I^2 \quad (2.11)$$

shows that $\beta_{\text{TPA}} = \frac{4\pi}{\lambda}\kappa_2$, which links the TPA coefficient to the imaginary part of the nonlinear refractive index and hence to $\chi^{(3)}$.

It is also possible that the absorption of a material decreases with increasing intensity, which is called saturable absorption [59]. There are different physical processes contributing to this phenomenon, which will be discussed in detail in Chapters 4 and 5. In those chapters, the Z-scan technique for determining nonlinear parameters, such as n_2 and β_{TPA} , and its application to a highly nonlinear perovskite flake will be covered. We shall see that both TPA and saturable absorption play a crucial role near the fundamental exciton resonance of the perovskite flake.

Let us now return to parametric processes associated with n_2 and the real part of $\chi^{(3)}$. A convenient way of determining n_2 experimentally, which will be explained in Chapter 4, is via the effect of self-focussing, shown schematically in Figure 2.1(c). Assuming a positive n_2 , an intense laser beam will locally increase the refractive index at its wavelength. If the beam is in a fundamental Gaussian TEM_{00} mode, which is the case in many optical experiments, the refractive index change will be larger at the centre of the beam than in its outer region. The resulting refractive index profile is equivalent to that of a weak lens – one often speaks of a Kerr lens being created by the beam – such that the beam focusses itself. Of course, this leads to an increase of the intensity in the centre of the beam and hence an even larger nonlinear index change increasing the optical power of the Kerr lens. Therefore, the process is self-reinforcing, which makes it dangerous due to optical damage in high-power optical applications. If $n_2 < 0$, the Kerr lens has a negative optical power, leading to self-defocussing of the laser beam.

2.3 Frequency Mixing

After this brief discussion of nonlinear changes of the complex refractive index, let us now consider nonlinear processes generating frequency components distinct from those in the input fields. A simple way of understanding the generation of new frequencies is via a classical anharmonic oscillator model, which is shown schematically in Figure 2.2(a). Electrons in the medium are assumed to move independently and in anharmonic potentials. For small

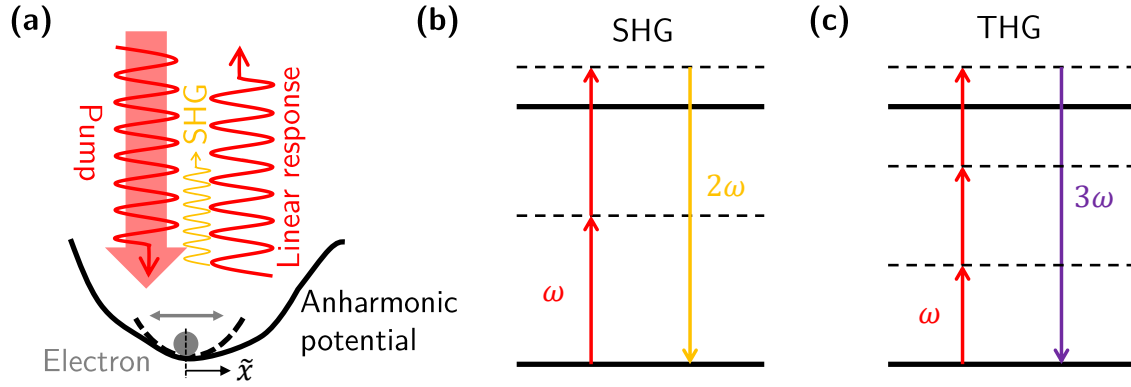


Figure 2.2 – Nonlinear frequency mixing. (a) The emergence of new frequency components in nonlinear mixing processes can be explained with a simple classical anharmonic oscillator model. (b) Energy diagram of a second-order process, SHG, where two pump photons are annihilated and one photon at twice the frequency is created in each elementary event. SHG involves virtual transitions (dashed lines) and does not require the presence of real energy levels, although there may be a resonant enhancement of the process. (c) Energy diagram of a third-order process, THG, where three pump photons are annihilated and one photon with three times the energy is created. Diagrams drawn similar to those in Ref. [38].

amplitudes, the oscillations are approximately harmonic with a resonance frequency ω_0 . This suggests that this model is only a good description of materials significantly below their lowest electronic resonance, where higher-energy resonances can be neglected. In general, a full quantum mechanical treatment is necessary to predict nonlinear interactions in systems with multiple resonances. However, since we would only like to briefly motivate the fundamentals of nonlinear optics, it is sufficient to explain the generation of new frequencies using the somewhat simplistic anharmonic oscillator model. Moreover, as we shall see in Chapter 6, the $\chi^{(2)}$ material of our dielectric sample, GaN, will only be used in its transparency window where this model is a good approximation.

Let us consider the example of a system without inversion symmetry, which possesses a leading-order anharmonic contribution to the potential proportional to x^3 . Such a system can be analysed analytically with a perturbative approach, where the driving field is multiplied by a control parameter λ and the solutions x are expanded in orders of the control parameter. The corresponding differential equation is

$$\ddot{x} + 2\gamma\dot{x} + \omega_0^2x + ax^2 = -\lambda eE/m \quad (2.12)$$

with $x = \lambda x_1 + \lambda^2 x_2 + \dots$ and where we assume a monochromatic driving field $E(t) = E_0 e^{-i\omega t} + c.c.$ for simplicity. Eq.(2.12) can be written as a system of differential equations, with one equation for each order in the control parameter λ . The first-order equation in λ is

$$\ddot{x}_1 + 2\gamma\dot{x}_1 + \omega_0^2 x_1 = -eE/m \quad (2.13)$$

which describes the well-known forced harmonic oscillator with the steady-state solution

$$x_1(t) = \underbrace{\frac{-eE_0}{m(\omega_0^2 - \omega^2 - 2i\gamma\omega)}}_{\tilde{x}_1} e^{-i\omega t} + c.c. \quad (2.14)$$

The terms proportional to λ^2 lead to the equation

$$\ddot{x}_2 + 2\gamma\dot{x}_2 + \omega_0^2 x_2 = -ax_1^2 = -a(\tilde{x}_1^2 e^{-2i\omega t} + c.c. + 2|\tilde{x}_1|^2) \quad (2.15)$$

which possesses a steady-state solution for x_2 at the second harmonic frequency

$$x_{2,\text{SHG}}(t) = \underbrace{\frac{-a\tilde{x}_1^2}{\omega_0^2 - (2\omega)^2 - 2i\gamma(2\omega)}}_{\tilde{x}_{2,\text{SHG}}} e^{-2i\omega t} + c.c. \quad (2.16)$$

The constant term on the right-hand side of Eq.(2.15) leads to another contribution to the second order solution x_2 corresponding to optical rectification, but that term is not of interest in this discussion. Returning to Eq.(2.16) one can now determine the strength of $\chi^{(2)}$ for second-harmonic generation (SHG) as expected from the anharmonic oscillator model. Using the amplitude term $\tilde{x}_{2,\text{SHG}}$ with the abbreviation $D(\omega') = \omega_0^2 - \omega'^2 - 2i\gamma\omega'$ and the relation $P(2\omega) = \varepsilon_0 \chi^{(2)}(2\omega, \omega, \omega) \underbrace{E^2(\omega)}_{E_0^2} = -Ne\tilde{x}_{2,\text{SHG}}$ one obtains

$$\chi^{(2)}(2\omega, \omega, \omega) = \frac{Ne^3 a}{\varepsilon_0 m^2 D^2(\omega) D(2\omega)}. \quad (2.17)$$

This result can be expressed in terms of first-order susceptibilities $\chi^{(1)}(\omega') = \frac{Ne^2}{\varepsilon_0 m D(\omega')}$ yielding

$$\chi^{(2)}(2\omega, \omega, \omega) = \frac{\varepsilon_0^2 m a}{N^2 e^3} \chi^{(1)}(2\omega) (\chi^{(1)}(\omega))^2. \quad (2.18)$$

Note that analogous calculations lead to expressions for the $\chi^{(2)}$ contribution corresponding to optical rectification. When considering two driving fields with distinct frequencies, one finds the analogous expressions for sum-frequency generation (SFG) and difference-frequency generation (DFG). Equation (2.18) is sometimes called Miller's rule, which states that the second-order response is directly proportional to the product of the linear response at the participating frequencies, since $\varepsilon_0^2 m a / N^2 e^3$ is almost constant for noncentrosymmetric crystals [38, 60]. While Miller's rule does work well for bulk crystals, it is worth noting that it was recently demonstrated to be unsuitable for predicting the SHG emitted from plasmonic nanoantenna arrays [61], where one should not use the simple relationship $I_{\text{SHG}} \propto I_{\text{scattered}}(2\omega) I_{\text{scattered}}^2(\omega)$ [62].

The interaction leading to SHG, characterised by $\chi^{(2)}(2\omega, \omega, \omega)$, can be visualised in a schematic energy level diagram, which is shown in Figure 2.2(b) as an example for second-order frequency mixing. The diagram illustrates an elementary interaction between two pump photons at frequency ω and one photon at the second harmonic frequency 2ω . As in the case of nonlinear changes of the real part of the refractive index described above, the process is parametric, meaning that it involves only virtual transitions. Therefore, SHG does not require the presence of real energy levels with the appropriate spacing but may always occur, as long as the material is non-centrosymmetric. However, even in centrosymmetric materials, there is a contribution to SHG near the surface due to symmetry breaking, which will be discussed in Chapter 7. Whenever a real level is close to one of the virtual levels involved in the interaction, the process is resonantly enhanced, however, at the cost of non-parametric processes and hence absorption losses.

Third-order frequency mixing, which is the leading-order nonlinear bulk response in centrosymmetric media, can be motivated in a similar way as above, however, the anharmonicity term

ax^2 in Eq.(2.12) needs to be replaced with ax^3 because the electronic potential has no terms proportional to x^3 , but the leading term is proportional to x^4 . A perturbative expansion in the control parameter λ yields oscillation amplitudes at the new frequencies, such as the third harmonic 3ω , from the equation containing all terms proportional to λ^3 . Figure 2.2(c) shows a schematic of third-harmonic generation (THG), where three photons at the pump frequency ω are converted into one photon at 3ω in a parametric process. Since THG does not require the breaking of inversion symmetry in the nonlinear medium, it may occur in all crystals. In general, for a third-order interaction of three distinct input frequencies, 19 new frequencies are generated.

2.4 Phase-Matching

So far, we have been concentrating on energetic considerations in frequency mixing processes. However, momentum, represented by the wavevector $\vec{k}$ of each participating field is a crucial aspect as well, particularly for optimising frequency conversion efficiencies in experiments. Let us consider the example of SHG to discuss the interplay of the momenta of the participating fields. Assuming an undepleted pump beam, meaning that only a very small fraction of the pump beam is converted into radiation at the second harmonic, one obtains the following expression for the SHG intensity after a propagation length L

$$I_{\text{SHG}} = \frac{2(\chi^{(2)})^2 \omega_{\text{SHG}}^2 I_{\text{pump}}^2}{n_{\text{pump}}^2 n_{\text{SHG}} \varepsilon_0 c^2} L^2 \text{sinc}^2 \left(\frac{\Delta k L}{2} \right). \quad (2.19)$$

The term $\text{sinc}^2(\Delta k L/2)$ with the momentum mismatch $\Delta k = 2k_{\text{pump}} - k_{\text{SHG}}$ contains the entire k -dependence of the output, implying that for vanishing Δk , the SHG output intensity grows quadratically with the interaction length L . For any non-zero value of Δk , however, the output can be greatly reduced and even be suppressed completely for $\Delta k L = 2\pi \cdot \nu$ with the integer $|\nu| \geq 1$. Hence, achieving $\Delta k = 0$, which is called phase-matching, is critical for ensuring the maximum efficiency of a frequency conversion process. Considering

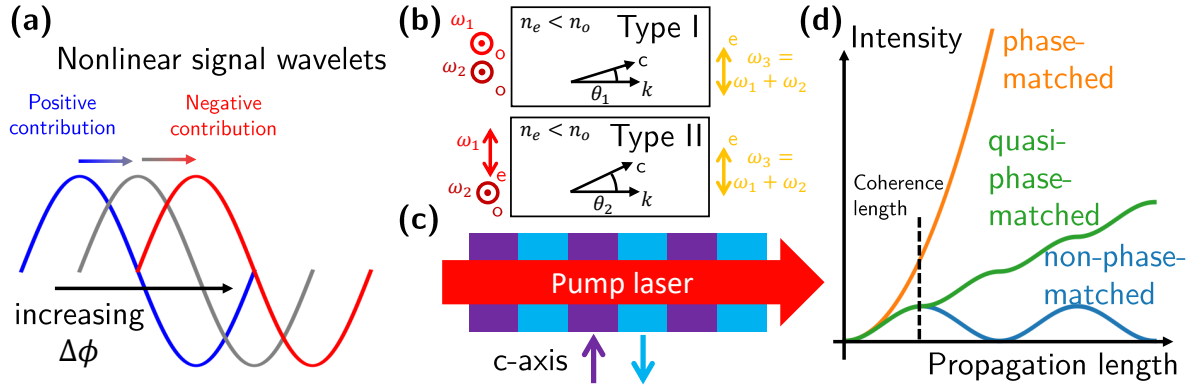


Figure 2.3 – Phase-matching. (a) Dispersion means that input fields and nonlinear output travel at different speeds, which is quantified by the wavevector mismatch Δk . Therefore, new wavelets after a propagation distance L are shifted by $\Delta\phi = \Delta k L/2$ which for $\pi/2 < \Delta\phi < \pi$ is a negative contribution to the signal. Thus, the energy transfer between input and output fields oscillates in a long crystal with $L \gg |\Delta k|^{-1}$. (b) Schematic of critical phase-matching, where birefringence is used to compensate for dispersive effects. In Type-I phase-matching, the lower two frequency inputs are of the same polarisation (both ordinary) while in Type-II phase-matching they are perpendicularly polarised (one ordinary, one extraordinary). Both diagrams visualise the example of sum-frequency generation (SFG) in a negative uniaxial crystal, where $n_e < n_o$. (c) Schematic of quasi-phase-matching where a nonlinear crystal is cut into slices of thickness $L_{\text{coh}} = \pi/|\Delta k|$ and the orientation of the c-axis is inverted in every slice yielding a stack of layers with alternating crystal orientation allowing for a continuous build-up of nonlinear signal. (d) Comparison of the intensity build-up as a function of the propagation length for critically phase-matched (orange), quasi-phase-matched (green), and non-phase-matched (blue) configurations for the exemplary case of SHG. (b)-(d) re-drawn similar to Ref. [38]

the phase-matching condition for SHG, we find that

$$2k_{\text{pump}} = k_{\text{SHG}} \Leftrightarrow 2n_{\omega} \frac{\omega}{c} = n_{2\omega} \frac{2\omega}{c} \Leftrightarrow n_{\omega} = n_{2\omega} \quad (2.20)$$

which means that dispersion is the reason for the momentum mismatch. A simple and intuitive illustration of the negative effect of dispersion on the nonlinear output intensity is given in Figure 2.3(a). As the fundamental wave travels through the nonlinear medium, wavelets of the nonlinear signal are created; however, since they travel at a different velocity than the fundamental, there is a phase difference of $\Delta\phi = \Delta k L/2$ between any two wavelets at the second harmonic created at a distance of L between each other. This means that after an initial accumulation phase where $\Delta\phi < \pi/2$ (blue), all new wavelets with $\pi/2 < \Delta\phi < \pi$ (red) reduce the nonlinear signal and energy is transferred back to the fundamental in that

regime. For even larger values of $\Delta\phi$ there is another accumulation phase and this oscillation of energy between the pump and nonlinear signal fields continues until it is interrupted by the end surface of the nonlinear crystal. Therefore, a large part of a long non-phase-matched nonlinear crystal, namely the largest integer multiple of the energy oscillation period, is redundant and only the small remaining length of the crystal contributes to the nonlinear output.

This shows that phase-matching, that is achieving $\Delta k = 0$, allows for an efficient use of the entire nonlinear crystal and is hence very desirable. One strategy called critical phase-matching is based on using birefringence, the property of some crystals to exhibit a polarisation-dependent refractive index, to counterbalance dispersion. Figure 2.3(b) illustrates two configurations for critical phase-matching using the example of sum-frequency generation (SFG) in a negative uniaxial crystal with an extraordinary index n_e smaller than the ordinary index n_o . The ordinary index is defined as the refractive index for light polarised perpendicular to the plane containing $\vec{k}$ and the optic axis, also called c-axis, of the crystal, while the extraordinary index is experienced by light polarised in the c- k -plane. In Type-I phase-matching, shown in the upper panel, the lower-frequency waves ω_1 and ω_2 have the same polarisation. Assuming normal dispersion, where the refractive index for low frequencies is lower than for high frequencies, ω_1 and ω_2 must be polarised perpendicular to the c- k -plane to be exposed to the high ordinary index n_o , while the SFG output must be polarised in the c- k -plane to be exposed to the low extraordinary index n_e counterbalancing its higher index due to dispersion. The value of n_e depends on the angle θ_1 and can hence be tuned via the crystal orientation. Type-II phase-matching is shown in the lower panel and is defined as the case where the two lower-frequency fields are orthogonally polarised. Similar arguments to above apply and given sufficiently strong birefringence, there is a tuning angle $\theta_2 \neq \theta_1$ for phase-matching with this polarisation configuration. Note that in crystals with a strong temperature dependence of the birefringence, temperature tuning can be used as an alternative to angle tuning with the terms Type-I and Type-II phase-matching retaining their meaning.

Quasi-phase-matching, depicted schematically in Figure 2.3(c) is an alternative to critical phase-matching and sometimes the only viable phase-matching scheme, such as in the case of a vanishing birefringence. The basic idea is to cut the nonlinear crystal into slices of thickness $L = \pi/|\Delta k| = L_{\text{coh}}$ and invert the c-axis of each segment and hence the sign of $\chi^{(2)}$ in order to avoid the frequency mixing process operating in the reverse direction. In practice, periodic poling during the crystal growth process is used to achieve this effect.

Figure 2.3(d) shows a schematic comparison of the intensity build-up in SHG for the various phase-matching situations. In the case of critical phase-matching (orange), the intensity grows quadratically with L , as expected from Eq.(2.19). In the absence of phase-matching, the oscillatory behaviour of $I \propto \sin^2(\Delta k L/2)$, again following from Eq.(2.19), can be observed (blue). Finally, in quasi-phase-matching (green), all segments of the non-phase-matched curves with a negative slope are mirrored against the x -axis because of the periodic inversion of the sign of $\chi^{(2)}$, leading to a continuous, approximately linear growth of the intensity at the second harmonic. The points of inversion are at multiples of the coherence length, where the gradient of the built-up intensity just vanishes.

As a final remark on phase-matching, it is important to note that the suppression of mixing processes with $\Delta k \neq 0$ compared to phase-matched processes becomes progressively more extreme as the thickness of the nonlinear crystal spans more and more coherence lengths. In other words, an extremely thin crystal would not be selective in terms of the nonlinear output signal of a particular configuration, but the conversion efficiency would be very low due to the small interaction volume. We shall return to this problem – and a strategy for overcoming it – when discussing the GaN and Au nanostructures in Chapters 6 and 7, where efficient frequency mixing occurs in sub-wavelength volumes.

2.5 Photon Sources

As we have seen in the previous section, there is a wealth of frequency mixing processes associated with the nonlinear susceptibilities $\chi^{(2)}$ and $\chi^{(3)}$, many of which are used for

technological applications, such as laser systems with a tuneable frequency output [63]. In this thesis we are particularly interested in quantum optics applications, where specific photon states are produced with nonlinear optical processes. In that context, the generation of single photons has been at the centre of recent research efforts, given the surge of quantum technologies and the need for single photons as carriers of quantum information [64].

Single-photon sources can be subdivided into deterministic and non-deterministic sources [65]. Deterministic sources, often called on-demand sources, are based on a two-level system, for instance a suitable transition in a trapped ion, which is prepared in the excited state [66]. The transition back to the ground state will produce only one photon at a time if a single ion is excited. In order to improve the emission rate, the ion is typically placed in a cavity [67] to benefit from the enhanced light-matter interaction provided by the Purcell effect [68]. Ion traps allow for high-purity [69] and high-indistinguishability [70] photon production, however, with the drawback of a large and expensive experimental setup and the trapping time limiting the duration for which the photon source can operate continually. Alternative on-demand sources, such as semiconductor quantum dots [71, 72] do not suffer from those problems, but instead require the spectral selection of a single-exciton state to avoid multi-photon events and cooling with liquid helium to reduce interactions with phonons. The photon emission rate is limited by the life time of the excited state and the indistinguishability is limited by fluctuations in the energy of the exciton transition [73, 74]. All on-demand single photon sources have in common that they suppress the probability of multi-photon emission, meaning that the photon output is highly non-Poissonian, however, at the cost of isolating a single emitter from its surroundings and reliably preparing it in its excited state [65].

Non-deterministic sources on the other hand are based on the heralding principle [50], which is visualised in Figure 2.4(a) for the example of a nonlinear process creating photon pairs. A more detailed discussion of such processes will be given below. For an explanation of heralding it is sufficient to know that the pumped nonlinear medium has the property of emitting photon pairs at random times. While the arrival time of a photon in the given output mode is not known, it is possible to exploit the fact that both photons are created

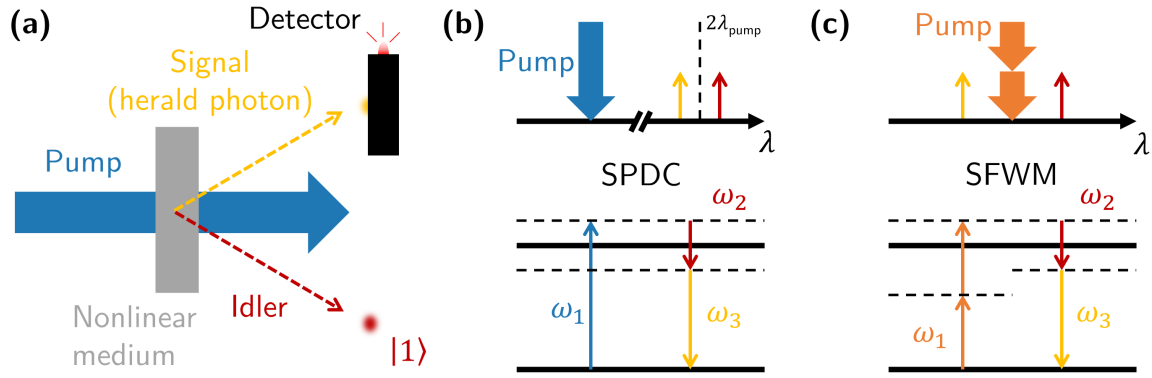


Figure 2.4 – Heralded single photon source. (a) The heralding scheme requires a nonlinear process generating two photons at the same time, such as SPDC or SFWM. The photon pairs are created at random times, but for every photon in the signal mode there must be a photon in the idler mode. Therefore, conditional on detection events in the signal arm, one obtains single photon states in the idler arm. (b) In spontaneous parametric down-conversion (SPDC), which is a second-order process, one pump photon is converted into a pair of a signal and an idler photon sharing the energy and momentum of the pump photon. The spectral distribution of the down-converted photons is hence centred around twice the pump wavelength, which means that there is a large energy difference between the pump photon and the generated pair. SPDC involves transitions between virtual energy levels, as is the case for other frequency mixing processes. (c) Spontaneous four-wave mixing (SFWM) is a third-order process similar to SPDC, except that there are two pump photons contributing the total energy and momentum of the down-converted signal and idler pair. The schematic spectrum shows the case of degenerate pump photons. In contrast with SPDC, all photons are spectrally close together.

at the same time [75]. Upon detecting a photon in a given output mode, the presence of a partner photon in the corresponding output mode, determined by energy and momentum conservation of the nonlinear process, is known [50]. Thus, postselection allows for the isolation of single photon states [50]. Since the detected photon heralds the presence of its partner, it is often called the herald photon [76]. If the detector is not number-resolving, a detection event may also herald the presence of a multi-photon state limiting the achievable purity of the photon output, as we shall see below. However, even for a number-resolving detector this problem is only reduced but not avoided completely [77].

Let us now take a closer look at nonlinear processes suitable for the production of photon pairs. The most popular process, spontaneous parametric down-conversion (SPDC), is a second-order interaction where a pump photon is converted into a signal and idler photon in each elementary event [78], as shown in the energy diagram in Figure 2.4(b). As no

absorption is involved in this parametric process via virtual levels, energy conservation holds for the participating photons, implying a symmetric distribution of the signal and idler photons around $\hbar\omega_p/2$ absent any stimulation of the signal and idler modes. Therefore, there is a large gap in the spectrum between the pump photons and the down-converted photons, as visualised in the schematic spectrum in Figure 2.4(b), which allows for an easy rejection of pump light after the nonlinear interaction [79].

A crucial difference between SPDC and other $\chi^{(2)}$ processes, such as SHG in the previous section, is that instead of two of the three participating photons, only one is provided by an input laser beam [78]. Hence, vacuum fluctuations play a crucial role in the frequency conversion process and a quantum mechanical description is warranted.

The following derivation is a summary of arguments presented by Gerry and Knight [80]. Let us start with the Hamiltonian

$$\hat{H} = -i\hbar\chi^{(2)}\hat{a}^\dagger\hat{b}^\dagger\hat{c} + h.c. + \hbar\omega_s\hat{a}^\dagger\hat{a} + \hbar\omega_i\hat{b}^\dagger\hat{b} + \hbar\omega_p\hat{c}^\dagger\hat{c} \quad (2.21)$$

where $\hat{a}$ and $\hat{b}$ are the annihilation operators of the signal and idler modes and $\hat{c}$ is the annihilation operator of the pump mode. Since the pump field is a strong classical beam, which is in very good approximation undepleted, it is possible to write the pump state as a coherent state $|\gamma e^{-i\omega_p t}\rangle$ such that

$$\hat{H} = -i\hbar\chi^{(2)}\hat{a}^\dagger\hat{b}^\dagger\gamma e^{-i\omega_p t} + h.c. + \hbar\omega_s\hat{a}^\dagger\hat{a} + \hbar\omega_i\hat{b}^\dagger\hat{b} \quad (2.22)$$

which is called the parametric approximation. In the interaction picture, the interaction Hamiltonian is given by

$$\hat{H}_I(t) = -i\hbar\chi^{(2)}\gamma\hat{a}^\dagger\hat{b}^\dagger e^{-i(\omega_p - \omega_s - \omega_i)t} + h.c. \quad (2.23)$$

which upon applying energy conservation $\omega_p = \omega_s + \omega_i$ reduces to a time-independent interaction Hamiltonian

$$\hat{H}_I = -i\hbar\chi^{(2)}\gamma\hat{a}^\dagger\hat{b}^\dagger + h.c. \quad (2.24)$$

The associated time evolution operator is

$$\hat{U} = e^{-i\hat{H}_I t/\hbar} = e^{\xi^*\hat{a}\hat{b} - \xi\hat{a}^\dagger\hat{b}^\dagger} = \hat{S}_2(\xi) \quad (2.25)$$

which is a two-mode squeezing operator $\hat{S}_2$ with the squeezing parameter $\xi = \chi^{(2)}\gamma t$. In an experiment, only the pump beam is provided and the signal and idler modes are in their vacuum states before the interaction. Hence the output signal and idler states are obtained by evaluating $\hat{S}_2|0\rangle_s|0\rangle_i$ which yields the two-mode squeezed vacuum state

$$\hat{S}_2|0\rangle_s|0\rangle_i = \frac{1}{\cosh r} \sum_{n=0}^{\infty} (-1)^n e^{in\phi} (\tanh r)^n |n\rangle_s |n\rangle_i \quad (2.26)$$

where $\xi = re^{i\phi}$. This result highlights that the SPDC output not only consists of vacuum and pairs of single photons in the signal and idler modes, but also pairs of multiphoton states, albeit with a strongly decreasing probability for higher photon number states. Experimentally, the squeezing parameter can be tuned via the pump intensity, since ξ is proportional to the pump amplitude γ , and the performance of the photon source is optimised for the maximum γ such that the sum of all number states with $n \geq 2$ is smaller than a pre-defined value expressing the purity requirements of the photon output. This is because the state in either mode, conditional on a detection event in the other mode, is not the pure Fock state $|1\rangle$, but an incoherent superposition of number states. In other words, one obtains a mixed state with $\hat{\rho} = \sum_{n=1}^{\infty} \lambda_n |n\rangle\langle n|$ with $\lambda_1 \lesssim 1$ where the purity is determined by $\sum_n |\lambda_n|^2$.

After this calculation, it is clear that in the above discussion of the heralding principle and the introduction of non-deterministic sources we considered the case of a sufficiently weak pump beam where multi-photon events can be neglected. Now, we are in a position to fairly compare deterministic sources to non-deterministic sources, concluding that the latter

are much easier to implement, as they only require a laser beam and a suitable nonlinear crystal, but at the cost of a significantly worse suppression of multiphoton states, limiting the achievable purity of the single-photon output. In fact, the multi-photon states are a built-in feature of nonlinear heralded sources, which places a very inconvenient intrinsic limitation on the achievable brightness for any acceptable maximum probability of multi-photon events. On the other hand, parametric photon sources have the built-in advantage of producing photon pairs, which is in itself interesting in quantum optics research and which constitutes a key advantage over single-emitter on-demand sources. A classic example of this is the production of polarisation-entangled photon pairs with a type-II phase-matched crystal [51] which can be used for tests of Bell's inequality or quantum teleportation [81].

An alternative to SPDC is degenerate four-wave mixing (DFWM), where only one pump beam is provided, but two pump photons are annihilated per elementary interaction [82], which is visualised in Figure 2.4(c). As can be seen in the energy diagram, DFWM is essentially the same process as SPDC, just with two pump photons at half the energy of the SPDC pump photon. The differences are that DFWM is a $\chi^{(3)}$ process and is hence supported even in centrosymmetric materials, and that the pump photons are spectrally close to the signal and idler photons, residing in the centre between them, which is shown in the schematic spectrum in Figure 2.4(c). DFWM is a sensible option when one wants to minimise dispersive effects and when the pump photons can be rejected by other means than spectral filtering, such as their polarisation, or if very narrow spectral filtering is possible. In terms of the theoretical treatment, the calculations above remain the same, except for the substitution $\chi^{(2)}\hat{c} \rightarrow \chi^{(3)}\hat{c}^2$.

From the discussion above, we can see that in order to maximise the efficiency of a non-deterministic photon source based on a nonlinear frequency mixing process – that is to achieve the maximum acceptable squeezing parameter ξ with a minimum of optical power – it is critical to work with high- $\chi^{(2)}$ or $-\chi^{(3)}$ materials and to optimise the strength of the optical fields at a given input power. In Chapters 6 and 7 we shall study dielectric and metallic nanostructures, respectively, which are designed according to this principle and

allow for efficient frequency mixing in sub-wavelength volumes thanks to a strong local field enhancement.

However, achieving high photon pair generation rates and efficiencies is not the only design aspect for parametric photon-pair sources. Controlling the spectral correlations between the output photons is crucial for applications in quantum information processing, where often uncorrelated photon pairs are needed [83]. The following section introduces a technique called stimulated emission tomography (SET) which allows for a convenient and efficient characterisation of those spectral correlations [56].

2.6 Stimulated Emission Tomography

When one wants to measure the spectral properties of the biphoton output of a parametric photon source, which is based on a spontaneous process such as SPDC, the pair generation rates are so low as to require the use of single-photon detectors. In order to achieve spectral resolution of the coincidences, it is common to work with either grating monochromators [83] or dispersive fibres [84]. In the latter case, the spectral information is mapped to the arrival time and hence the spectral resolution is limited by the timing jitter of the detectors. This is not the case for the former strategy, however, the low count rates in SPDC lead to a limitation due to dark counts, as the resolution becomes sufficiently fine for the signal-to-noise ratio to fall below an acceptable level. In any case, long integration times are needed due to the low count rates.

The SET technique [56], visualised schematically in Figure 2.5, attempts to overcome those problems by studying the output characteristics of the spontaneous nonlinear process in question indirectly via the corresponding stimulated process with the same phase-matching configuration. This discussion follows closely the arguments in the original SET paper [56], but using a simplified notation omitting details that are not relevant for the purposes of this thesis. Let us consider the example of SPDC, which is depicted in Figure 2.5(a), with a sufficiently weak pump beam (blue) such that multi-photon events are negligible and the output

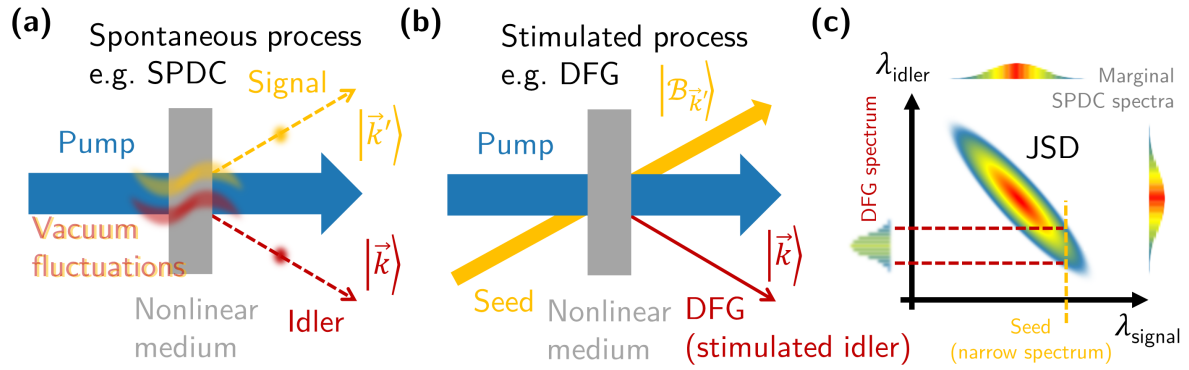


Figure 2.5 – Stimulated emission tomography. (a) The two-photon output of a spontaneous process, such as SPDC, is given by a joint distribution of wavevectors in the signal and idler modes. SPDC can be understood as being seeded by vacuum fluctuations. (b) In DFG, which is the stimulated process corresponding to SPDC, the signal mode is seeded with a second classical input beam. This stimulates the production of idler photons, which makes it possible to detect them with a usual spectrometer. (c) Joint spectral density (JSD) and marginal spectra. Since the spontaneous and stimulated process share the same pump and phase-matching configuration, it is possible to study the JSD of the SPDC biphoton wavefunction with a high resolution and good signal-to-noise ratio by scanning the seed beam through the parameter space and detecting the DFG signal. Re-drawn similar to Ref. [84].

consists of signal (yellow) and idler (red) photon pairs emitted at random times. In order to characterise the biphoton, one needs to consider pairs of output modes $|\vec{k}'\rangle$ and $|\vec{k}\rangle$, which can be selected with pinholes and bandpass filters, and record the coincidence rate for each combination. Note that in general the photons may have different polarisations, however, for our purposes it is sufficient to consider a Type-I phase-matched configuration where the down-converted photons are always in the same polarisation state. The key argument is that the biphoton amplitude and hence the distribution function of the coincidence rates as a function of $\vec{k}'$ and $\vec{k}$ is uniquely determined by the pump and phase-matching configuration. Expressing the general biphoton state as

$$|\psi\rangle = \frac{1}{\sqrt{2}} \int d\vec{k}_1 d\vec{k}_2 \phi(\vec{k}_1, \vec{k}_2) \hat{a}_{\vec{k}_1}^\dagger \hat{a}_{\vec{k}_2}^\dagger |\text{vac}\rangle \quad (2.27)$$

with an additional pair creation probability amplitude γ , one finds that the expected number of signal-idler pairs in a small parameter region around the configuration $(\vec{k}', \vec{k})$ is

$$\langle \hat{a}_{\vec{k}'}^\dagger \hat{a}_{\vec{k}'} \hat{a}_{\vec{k}}^\dagger \hat{a}_{\vec{k}} \rangle \delta\vec{k}' \delta\vec{k} = 2|\gamma|^2 |\phi(\vec{k}', \vec{k})|^2 \delta\vec{k}' \delta\vec{k}. \quad (2.28)$$

Note that the signal and idler modes are initially in their vacuum state, which can be expressed by saying that the modes $|\vec{k}'\rangle$ and $|\vec{k}\rangle$ are seeded by vacuum fluctuations.

If instead the mode $|\vec{k}'\rangle$ is seeded with a laser beam, described by the coherent state $|\mathcal{B}_{\vec{k}'}\rangle$, one obtains the situation shown in Figure 2.5(b) which is in fact a difference-frequency generation (DFG) setup. Assuming that the seed beam is narrowly distributed around $\vec{k}'$ and that the spontaneous emission into the idler mode can be neglected compared to the stimulated contribution, it can be shown that the expected number of photons in the idler mode in a small parameter region around $\vec{k}$ is

$$\langle \mathcal{B}_{\vec{k}'} | \hat{a}_{\vec{k}}^\dagger \hat{a}_{\vec{k}} | \mathcal{B}_{\vec{k}'} \rangle \delta \vec{k} \approx 2|\gamma|^2 |\phi(\vec{k}', \vec{k})|^2 \delta \vec{k}' \delta \vec{k} |\mathcal{B}_{\vec{k}'}|^2 = \langle \hat{a}_{\vec{k}'}^\dagger \hat{a}_{\vec{k}'} \hat{a}_{\vec{k}}^\dagger \hat{a}_{\vec{k}} \rangle \delta \vec{k}' \delta \vec{k} |\mathcal{B}_{\vec{k}'}|^2 \quad (2.29)$$

where $|\mathcal{B}_{\vec{k}'}|^2$ is the average number of photons in a stimulating pulse. Note that this result for the stimulated DFG experiment also contains the expression $|\gamma|^2 |\phi(\vec{k}', \vec{k})|^2$ since it shares the pump and phase-matching configuration with the SPDC experiment. It follows that there is a simple relationship between the expected number of pairs in the SPDC experiment and the expected photon number in the idler mode of a DFG experiment where the only change is that the signal mode $|\vec{k}'\rangle$ is seeded with a second classical input beam, namely that the ratio of the DFG idler photons to the SPDC pairs is given by the average number of photons in the seed beam:

$$\frac{\langle \mathcal{B}_{\vec{k}'} | \hat{a}_{\vec{k}}^\dagger \hat{a}_{\vec{k}} | \mathcal{B}_{\vec{k}'} \rangle \delta \vec{k}}{\langle \hat{a}_{\vec{k}'}^\dagger \hat{a}_{\vec{k}'} \hat{a}_{\vec{k}}^\dagger \hat{a}_{\vec{k}} \rangle \delta \vec{k}' \delta \vec{k}} \approx |\mathcal{B}_{\vec{k}'}|^2 \quad \Leftrightarrow \quad \frac{N_{\text{idler,DFG}}}{N_{\text{pairs,SPDC}}} \approx N_{\text{stimulating}}. \quad (2.30)$$

This result effectively allows for an amplification of SPDC coincidence measurements and hence an improved indirect characterisation of the biphoton spectral correlations [84], shown schematically in Figure 2.5(c). Assembling the spectrally resolved coincidences of an SPDC experiment in a two-dimensional histogram yields a map of the joint spectral density (JSD) [84]. The resolution of such a JSD map, however, is limited by detector dark counts and – if dispersive fibres are used to map the spectral information onto a distribution of photon arrival times – timing jitter [84], as explained above. However, if the same JSD map is

obtained by assembling the DFG spectra with the seed beam at the appropriate wavelengths, the resolution is limited by the linewidth of the seed beam [84]. In effect, one performs a high-resolution scan of the JSD map with the additional advantage of each constituent spectrum being taken with a classical spectrometer since the relevant photon counts are amplified by the seeding process [84]. The drastic improvement of the resolution and the signal-to-noise ratio was demonstrated in recent publications [84, 85], where fine spectral features that could not be resolved with SPDC coincidence measurements were made visible in great detail with DFG measurements, with the additional advantage of a significantly reduced acquisition time due to the much higher count rate.

In conclusion, the SET technique provides two main insights about a spontaneous frequency mixing process. Firstly, the expected photon generation rate can be estimated from the classical stimulated nonlinear mixing process using Eq.(2.30). Secondly, the JSD can be obtained conveniently, fast, and with a higher resolution than in the spontaneous experiment, allowing for a precise characterisation of the spectral correlations of the biphoton state. This information allows one to quantify the entanglement using the Schmidt decomposition [86–88] where the joint spectral amplitude $\mathcal{A}$ is written as [87]

$$\mathcal{A}(\omega_s, \omega_i) = \sum_n \sqrt{\lambda_n} \Psi_n(\omega_s) \Phi_n(\omega_i) \quad (2.31)$$

with $\sum_n |\lambda_n| = 1$. In the absence of correlations, $\mathcal{A}$ can be written as a product of a signal spectral amplitude $\Psi_0(\omega_s)$ and an idler spectral amplitude $\Phi_0(\omega_i)$ and $\lambda_0 = 1$, $\lambda_n = 0$ for $n \geq 1$ [87]. As soon as correlations occur, $\mathcal{A}$ cannot be factorised and $\lambda_n \neq 0$ for $n \geq 1$. The entanglement of such a state can be quantified with the entropy of entanglement $S = -\sum_n \lambda_n \log_2 \lambda_n$ which is only zero if there are no spectral correlations [87]. Experimentally, if a photon source with $S \neq 0$ was used as a heralded source, its output would be in a mixed state, requiring further spectral filtering to obtain a pure output, which reduces its efficiency [83]. Therefore, designing the phase-matching condition such that $\mathcal{A}$ is factorisable and $S = 0$ is highly desirable for high-brightness parametric photon sources [83].

In Chapters 6 and 7 we shall make use of the SET technique and in particular Eq.(2.30) to estimate the spontaneous photon pair creation rate in a wavelength-sized GaN structure and in Au nanoantennae, respectively. The analysis in those chapters will also contain a discussion of the concrete experimental conditions compared to the assumptions that were made in the derivation of Eq.(2.30) in the original SET paper [56]. Given the early stage of the project, the photon sources are not yet designed with optimised spectral properties of the biphoton wavefunction, which is why a quantitative entanglement analysis as outlined above will not be performed. However, this analysis will inform the design of future samples – an idea that we shall return to in the outlook in Chapter 8.

2.7 Conclusion

In this chapter, we have discussed the optical Kerr effect and nonlinear frequency mixing processes, which are essential for understanding the experimental results in Chapters 5-7. In particular, we have focussed on the link between nonlinear optics and quantum optics, given by parametric photon pair sources, which can be used as heralded single-photon sources. We have seen that the biphoton wavefunction of such sources can be studied indirectly via the corresponding stimulated process, an approach called the SET technique. A common theme of all the nonlinear processes considered in this chapter is that their efficiency is increased by an enhancement of the optical fields. In the following chapter, we shall see how plasmonic structures can be used to achieve this strong local field enhancement.

Chapter 3

Plasmonics

In this chapter we shall explore the physical origin of the local field enhancement of plasmonic nanostructures, which enables stronger nonlinear processes [37]. Given the wide scope of the field of plasmonics, we will limit ourselves to the discussion of a list of key concepts needed to better understand the role of the plasmonic antennae studied in Chapters 5 and 7. After reviewing the conditions for the existence of plasmonic surface modes with subwavelength confinement and the enhancement of the optical fields, and the properties of metals which make such modes possible, we will discuss plasmonic resonances in metal nanoparticles, where nanoscopic confinement is achieved in all three spatial dimensions. Next, we shall see how those plasmonic particle resonances are influenced by the particle shape and size and how particles can be used as optical antennae coupling electromagnetic energy from free space to nanoscopic volumes and vice versa. Finally, aspects of antenna design and the concept of near-field coupling are discussed in preparation for Chapter 7.

3.1 Fundamentals

Plasmonics deals with optical fields near metallic surfaces, which are locally strongly enhanced and confined below the diffraction limit [23]. Let us explore the conditions for the existence

of plasmonic modes for the most basic scenario of two half-spaces divided by an interface at $z = 0$, with different dielectric functions ε , following closely the derivation in Ref. [28]. The starting point of our calculation is the wave equation, which for a time-harmonic field assumes the form

$$\nabla^2 \vec{E}(x, y, z) + \underbrace{\frac{\omega^2}{c^2} \varepsilon}_{k_0^2} \vec{E}(x, y, z) = 0 \quad (3.1)$$

known as the Helmholtz equation. We are interested in finding a solution which is confined to the interface and can hence only propagate in the xy -plane. Without restriction of generality, the propagation direction may be chosen as the x -direction, such that the ansatz for the plasmonic field distribution is

$$\vec{E}(x, y, z) = \vec{E}(z) \cdot e^{ik_x x} \quad (3.2)$$

and the wave equation reduces to an ordinary differential equation

$$\frac{d^2}{dz^2} \vec{E}(z) + (k_0^2 \varepsilon - k_x^2) \vec{E}(z) = 0. \quad (3.3)$$

Using the result that only TM modes exist for this geometry [28] and hence all electromagnetic field components other than E_x , E_z , and H_y vanish, the relevant Maxwell's equations

$$\nabla \times \vec{E} = -\mu_0 \dot{\vec{H}} \quad \text{and} \quad \nabla \times \vec{H} = \varepsilon_0 \varepsilon \dot{\vec{E}} \quad (3.4)$$

reduce to the following relationships between the remaining components

$$\frac{dE_x}{dz} - ik_x E_z = i\omega \mu_0 H_y \quad \frac{dH_y}{dz} = i\omega \varepsilon_0 \varepsilon E_x \quad ik_x H_y = -i\omega \varepsilon_0 \varepsilon E_z \quad (3.5)$$

where we have also used the ansatz Eq.(3.2). Re-arranging Eq.(3.5) yields a wave equation for $H_y(z)$

$$\frac{d^2}{dz^2} H_y(z) + (k_0^2 \varepsilon - k_x^2) H_y(z) = 0 \quad (3.6)$$

for which we can use the ansatz

$$H_{y,z>0}(z) = H_{0,z>0} e^{-k_{z,z>0} z} \quad \text{and} \quad H_{y,z<0}(z) = H_{0,z<0} e^{k_{z,z<0} z} \quad (3.7)$$

where the amplitudes $H_{0,z>0}$ and $H_{0,z<0}$ are equal due to the continuity of $\vec{H}$ at the interface. Therefore, and due to the continuity of E_x , the middle expression in (3.5) evaluated at $z = 0$ yields the condition

$$-\frac{k_{z,z>0}}{k_{z,z<0}} = \frac{\varepsilon_{z>0}}{\varepsilon_{z<0}}. \quad (3.8)$$

Finally, inserting the ansatz (3.7) into Eq.(3.6), we obtain the conditions

$$k_{z,z>0}^2 + k_0^2 \varepsilon_{z>0} = k_x^2 \quad \text{and} \quad k_{z,z<0}^2 + k_0^2 \varepsilon_{z<0} = k_x^2 \quad (3.9)$$

which, combined with Eq.(3.8) yield the dispersion relation for the plasmonic interface mode

$$k_x = k_0 \sqrt{\frac{\varepsilon_{z>0} \cdot \varepsilon_{z<0}}{\varepsilon_{z>0} + \varepsilon_{z<0}}} \quad (3.10)$$

where the expression under the square root must be positive for a mode propagating along the interface. For an exponentially confined mode with $k_z > 0$, Eqs.(3.8) and (3.10) yield the conditions

$$\varepsilon_{z>0} \cdot \varepsilon_{z<0} < 0 \quad \text{and} \quad \varepsilon_{z>0} + \varepsilon_{z<0} < 0 \quad (3.11)$$

which must be satisfied for such a mode to exist. The mode we have just derived is called a surface plasmon polariton (SPP), whose field distribution is shown in the schematic diagram in Figure 3.1(a), and whose propagation properties have been investigated at length in the literature [90–93]. However, for the purposes of this thesis, we are only interested in SPPs as the most basic system where the key properties of plasmonic field enhancement and subwavelength confinement can be shown in a brief derivation from first principles. In the following section, we will see as a next step how spatial confinement in all three spatial dimensions creates stationary plasmonic resonances, which are a key component in the samples studied in Chapters 5 and 7.

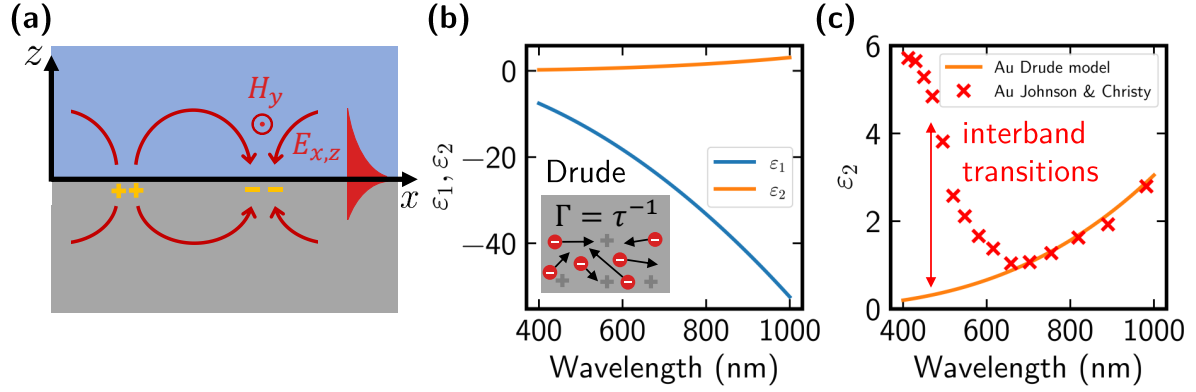


Figure 3.1 – Plasmonic mode and metal properties. (a) Surface plasmon polariton (SPP) at the interface between two half-spaces. The mode is exponentially confined to the interface and the near-fields are strongly enhanced. SPPs only exist in TM polarisation and require $\varepsilon_{z>0} \cdot \varepsilon_{z<0} < 0$ and $\varepsilon_{z>0} + \varepsilon_{z<0} < 0$. This is satisfied at the interface of a metal below its plasma frequency and a dielectric. Re-drawn similar to Ref. [57]. (b) Real (blue) and imaginary part (orange) of $\varepsilon = \varepsilon_1 + i\varepsilon_2$ for a metal according to the Drude model and using the parameters $\Gamma = 1.075 \cdot 10^{-15} \text{ s}^{-1}$ and $\omega_P = 13.8 \cdot 10^{-15} \text{ s}^{-1}$ for gold [89]. The inset is a schematic diagram of the Drude model with free electrons scattering off positive background charges on a characteristic timescale τ . (c) In the higher-energy half of the visible spectrum, gold deviates from the Drude model (orange) due to interband transitions. This is particularly well visible when comparing the imaginary part of ε with measured values (red crosses). Drawn with values measured by Johnson and Christy [89].

Returning to the conditions (3.11), it is clear that the SPP can only exist at the boundary between a medium with a strongly negative dielectric function ε and a medium with a positive ε that is smaller in magnitude. In the visible and NIR, these conditions are satisfied at the interface between a metal and a dielectric [28]. In the dielectric, bound charges are slightly shifted by the incident electric field, causing a polarisation aligned with the external field, which is described by a positive ε . In the metal, free electrons typically dominate the optical response, which is well described by the classical Drude model [2], where the electrons are driven by the external field and are assumed to scatter with positive background charges at the phenomenological collision rate Γ , leading to the differential equation

$$m_e \ddot{\vec{r}} + m_e \Gamma \dot{\vec{r}} = e \vec{E}_0 e^{-i\omega t} \quad (3.12)$$

with the effective electron mass m_e . This equation is solved by a harmonic oscillation $\vec{r}(t) = \vec{r}_0 e^{-i\omega t}$ of the electrons, with the amplitude $\vec{r}_0 = -e \vec{E}_0 / m \cdot (\omega^2 + i\Gamma\omega)^{-1}$, such that

with $\vec{P} = ne\vec{r}_0$ and $\vec{P} = \varepsilon_0\chi\vec{E}_0$ we obtain the dielectric function

$$\varepsilon_{\text{Drude}}(\omega) = 1 + \chi_{\text{Drude}}(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\Gamma\omega} = \underbrace{1 - \frac{\omega_p^2}{\omega^2 + \Gamma^2}}_{\varepsilon_1} + i \underbrace{\frac{\omega_p^2 \Gamma / \omega}{\omega^2 + \Gamma^2}}_{\varepsilon_2} \quad (3.13)$$

with the plasma frequency $\omega_p = (ne^2/m\varepsilon_0)^{-1/2}$ [2]. Therefore, it can be said that metals do indeed satisfy the requirement of a large negative ε_1 for $\omega < \omega_p$. The real and imaginary parts of the dielectric function are shown in Figure 3.1(b) for the example of gold, with the parameters $\Gamma = 1.075 \cdot 10^{-15} \text{ s}^{-1}$ and $\omega_p = 13.8 \cdot 10^{-15} \text{ s}^{-1}$ calculated from measurements by Johnson and Christy [89]. It is important to note that the optical properties of gold deviate from the predictions of the Drude model for $\lambda \lesssim 600 \text{ nm}$ due to interband transitions, which is shown in Figure 3.1(c) in a comparison of ε_2 predicted by the Drude model and the values measured by Johnson and Christy [89]. This additional strong absorption makes gold a poor material for most plasmonic applications in that wavelength range. In the experiments in Chapter 5, where plasmonic antennae were designed to match the excitonic resonance of a perovskite flake at $\lambda = 515 \text{ nm}$, aluminium was chosen instead of gold due to its much weaker absorption at that wavelength [94].

Let us conclude this section by explicitly demonstrating the plasmonic confinement and field enhancement for the example of a gold-air interface illuminated at $\lambda = 700 \text{ nm}$, just before the onset of the interband transitions. Since $\varepsilon \approx -26$ for gold, which can be seen in Figure 3.1(b), the confinement on the gold side must be about 26 times stronger than on the air side according to Eq.(3.8). The dispersion relation Eq.(3.10) yields $k_x^2 \approx 1.04 k_0^2$ and inserting this result in (3.9) shows that the confinement on the air side is characterised by a decay length of $\sim 5\lambda$, whereas it is $\sim \lambda/5$ on the gold side. For comparison, the hypothetical maximum achievable confinement is given by the maximum value of k_x , which is achieved for a vanishing real part of the denominator in Eq.(3.10) at $\lambda \approx 195 \text{ nm}$ corresponding to the surface plasmon frequency $\omega_{sp} = \omega_p/\sqrt{2}$. Due to the finite imaginary part of the dielectric function of gold, there is no divergence. Analogous calculations yield a decay length along the z -axis of $\sim 35 \text{ nm}$ both on the gold and on the air side, which is about an order of

magnitude smaller than the wavelength. Of course, the Drude model does not describe the optical properties of gold well in the UV due to transitions that are not taken into account, as mentioned earlier, so the latter estimate should be taken with a grain of salt. However, other metals, such as silver may be excited at ω_{sp} and a similarly strong confinement may be achieved; this was shown [95] with a silver slab used as a superlens with a resolution of $\lambda/6$. This shows that plasmonic modes can in principle be confined far below the wavelength, even in the absence of any nanostructuring of the metal surface.

The field enhancement close to the interface is easily estimated via the electromagnetic energy density. Evaluating the expression $u = \frac{1}{2}(\vec{E} \cdot \vec{D} + \vec{B} \cdot \vec{H})$ [96] for the SPP fields above yields

$$u(z) = \frac{1}{2}(\varepsilon_0 E^2 + \mu_0 H^2) = \frac{1}{2} \underbrace{\mu_0 H_0^2}_{\text{pl.wave}} \frac{2\varepsilon_{\text{Drude}}}{1 + \varepsilon_{\text{Drude}}} e^{-2k_z z} \quad \text{for } z > 0 \quad (3.14)$$

which, in the long-wavelength limit, is identical to the plane-wave expression in vacuum since $\varepsilon_{\text{Drude}}$ becomes large and negative with a dominating real part. However, for shorter wavelengths, as the real part of $\varepsilon_{\text{Drude}}$ approaches -1 , high local energy densities are achieved. The local field enhancement can then be estimated as $\sqrt{u/u_0}$, due to the quadratic field dependency of the energy density, with the plane-wave-in-vacuum expression $u_0 = \mu_0 H_0^2 = \varepsilon_0 E_0^2$. For the above example of a gold-air interface at $\lambda = 700$ nm, $\varepsilon/(1 + \varepsilon) \approx 1.04$, so the field enhancement just above the gold layer is only $\sim 2\%$.

This result seems to suggest that gold does not offer any real benefit as a plasmonic material, as the field confinement and enhancement in the region without interband transitions is negligible, given that the plasma frequency is in the deep UV. On the other hand, interband transitions and hence losses are high in the UV, where a substantial plasmonic field enhancement is expected. The solution to this dilemma is the use of nanofabrication to confine and resonantly enhance the plasmonic mode, which allows for controlling the resonance via the particle size and shape, as we shall see in the following section.

3.2 Plasmonic Resonances in Metal Nanoparticles

Fuelled by recent improvements in wet chemistry processes and lithography for the generation of designer particles with specific optical properties [97], research interest in metal nanoparticles has been strong for many years due to their numerous practical applications, such as light harvesting [98] or biomedical sensing and imaging [99]. In this section, we shall introduce the concept of a plasmonic particle resonance using the example of a sub-wavelength sphere, followed by a discussion of more general nanoparticles and their resonant properties, as well as their suitability as optical antennae, which plays a major role in the samples investigated in Chapters 5 and 7.

Plasmonic resonances in particles are localised surface plasmon resonances (LSPR), in contrast to SPPs, which form the other major category of plasmonic modes [28]. The simplest model system, in which the resonance condition for a LSPR can be derived is a metallic sphere with a radius a much smaller than the wavelength of the excitation beam, where the plasmonic mode is confined in all three spatial dimensions rather than just in one dimension perpendicular to the interface of a metal surface. The following derivation is a summary of the discussion in Ref. [28].

Let us assume a metal particle with radius $a \ll \lambda$ located at the origin and an incident plane wave polarised along the z -axis, which is drawn schematically in Figure 3.2(a). The key element of this derivation is the assumption that there are no phase differences of the fields inside and around the particle, which is a good approximation as long as the particle is much smaller than the wavelength. Under this so-called quasi-static approximation, the problem reduces to an electrostatic one, where it is only necessary to solve the Laplace equation

$$\Delta\Phi = 0 \tag{3.15}$$

to obtain the electrostatic potential Φ and the electric field distribution $\vec{E} = -\nabla\Phi$. For the time-dependent solution, this field distribution is simply multiplied by $e^{-i\omega t}$. In spherical

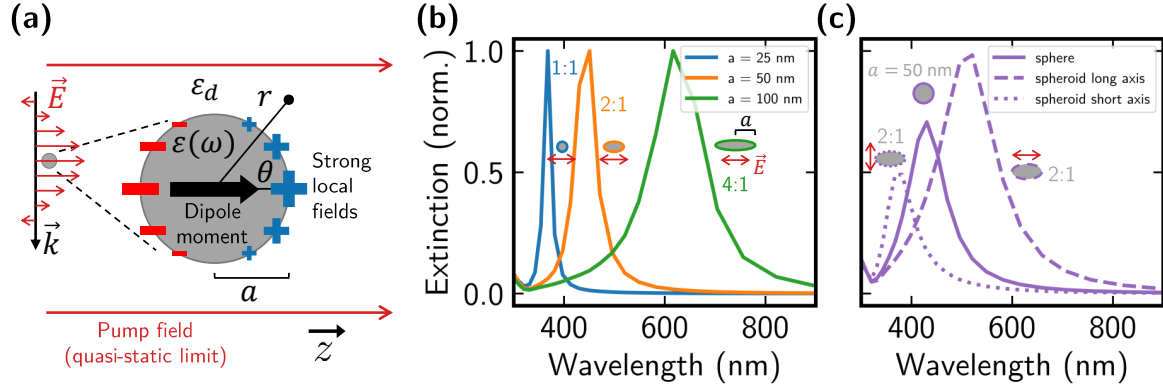


Figure 3.2 – Plasmonic resonances in metal nanoparticles. (a) Schematic of a sub-wavelength metal sphere with a radius a and a dielectric function $\varepsilon(\omega)$ in an external optical field and surrounded by a medium with a dielectric constant ε_d . In the quasi-static limit, the whole particle is assumed to react to the external field in phase, reducing the problem to an electrostatic one, which is a good approximation when $a \ll \lambda$. (b) For larger particles, where $a \lesssim \lambda$, the modified long-wavelength approximation (MLWA) leads to good predictions of the dipolar particle resonance. For a prolate spheroid with growing aspect ratio, a red-shift and a broadening of the resonance occurs, which is due to depolarisation fields and radiative damping, included to leading order in the MLWA. All spectra are normalised by their respective maxima. (c) Transforming a spherical particle into a prolate spheroid at constant particle volume splits the resonance of the spherical particle (solid line) into a blue-shifted short-axis resonance (dotted line) and a red-shifted long-axis resonance (dashed line). All spectra were generated for silver particles using refractive index data measured by Johnson and Christy [89].

coordinates, the solution of Eq.(3.15) is given by the series expansion

$$\Phi(r, \theta) = \sum_{l=0}^{\infty} (A_l r^l + B_l r^{-(l+1)}) P_l(\cos \theta) \quad (3.16)$$

with $P_l(\cos \theta)$ denoting the Legendre polynomials of order l . After applying the boundary conditions that the potential stay finite at the origin, converge towards $-E_0 z = -E_0 r \cos \theta$ for $r \rightarrow \infty$ and that $\vec{E}$ and $\vec{D}$ satisfy the usual continuity conditions at $r = a$, the solution takes the form

$$\Phi_{r \leq a} = \frac{-3\varepsilon_d}{\varepsilon(\omega) + 2\varepsilon_d} E_0 r \cos \theta \quad \text{and} \quad \Phi_{r \geq a} = -E_0 r \cos \theta + \frac{\vec{p} \cdot \vec{r}}{4\pi\varepsilon_0\varepsilon_d r^3} \quad (3.17)$$

with the induced dipole moment

$$\vec{p} = 4\pi\varepsilon_0\varepsilon_d a^3 \frac{\varepsilon(\omega) - \varepsilon_d}{\varepsilon(\omega) + 2\varepsilon_d} \vec{E}_0 = \varepsilon_0\varepsilon_d \alpha \vec{E}_0 \quad (3.18)$$

where α is the polarisability of the particle. The resonance condition for maximising the polarisability and hence the local fields of this LSPR is given by the Fröhlich condition

$$\Re(\varepsilon(\omega)) = -2\varepsilon_d \quad (3.19)$$

as long as $\Im(\varepsilon(\omega))$ is small or constant. This shows that the particle resonance depends on the refractive index of the surrounding dielectric, which is important for sensing applications [100].

For a small gold particle in air, the resonance is shifted towards the visible range compared to the surface plasmon frequency $\omega_{sp} = \omega_p/\sqrt{2}$, since Eq.(3.19) is satisfied for $\omega = \omega_p/\sqrt{3}$ corresponding to $\lambda \sim 240$ nm. Of course, this is not sufficient to avoid the strong absorption caused by the interband transitions; indeed, plasmonic resonances at $\lambda \gtrsim 600$ nm are needed for high-performance gold nanostructures. When allowing for larger particles, which however are still smaller than the wavelength, finite-size effects start to affect the particle resonance, enabling one to tune the particle resonance in the visible and NIR part of the spectrum for designer plasmonic nanostructures [97]. The two leading effects, retardation due to depolarisation fields and radiative damping can be introduced by means of the so-called modified long-wavelength approximation (MLWA), where one substitutes

$$\alpha = 4\pi a^3 \frac{\varepsilon(\omega) - \varepsilon_d}{\varepsilon(\omega) + 2\varepsilon_d} \rightarrow \alpha \cdot \left(1 - \frac{2}{3}ik^3\alpha - \frac{k^2}{a}\alpha\right)^{-1} \quad (3.20)$$

with the radiative damping term $\frac{2}{3}ik^3\alpha$ and the term accounting for the depolarisation across the particle surface $\frac{k^2}{a}\alpha$ [97]. The latter term is responsible for the red-shift of the resonance with growing particle size, while the first term leads to a broadening of the resonance.

Let us study these effects for the example of prolate (cigar-shaped) spheroids with increasing aspect ratios, to not only show the tuneability throughout the visible part of the spectrum but also to bridge the gap to the use of metal nanoparticles as optical antennae, which will be discussed below. Figure 3.2(b) shows theoretical extinction spectra for prolate silver spheroids using the Johnson and Christy [89] dataset for the complex refractive index of silver,

which is relatively coarse and responsible for the curves not being smooth. Silver was chosen as an example, because of its lack of interband transitions in the visible part of the spectrum, which is better suited for this demonstration than gold with its additional absorption due to interband transitions for $\lambda \lesssim 600$ nm. The radius of the particles perpendicular to the polarisation of the incoming light was fixed at 25 nm and the semi-axis parallel to the polarisation was set to 25 nm (blue), 50 nm (orange), and 100 nm (green) leading to an aspect ratio of 1:1, 2:1, and 4:1, respectively. The extinction was evaluated using [101]

$$C_{\text{ext}} = C_{\text{abs}} + C_{\text{sca}} = k \Im(\alpha) + \frac{k^2}{6\pi} |\alpha|^2 \quad (3.21)$$

and inserting for α the electrostatic polarisability of a prolate spheroid [101] with the long semi-axis a

$$\alpha = 4\pi ab^2 \frac{\varepsilon(\omega) - \varepsilon_d}{3\varepsilon_d + 3L(\varepsilon(\omega) - \varepsilon_d)} \quad (3.22)$$

modified according to Eq.(3.20), and where the shape parameter L is given by

$$L = \frac{1 - e^2}{e^2} \left(-1 + \frac{1}{2e} \ln \left(\frac{1 + e}{1 - e} \right) \right) \quad (3.23)$$

with the eccentricity parameter $e = \sqrt{1 - b^2/a^2}$. The extinction spectra in Figure 3.2(b), each normalised by its respective maximum value, do indeed demonstrate that controlling the length of a cigar-shaped nanoparticle unlocks the ability to tune the dipole resonance of the particle throughout the visible range of the spectrum, which, however, is accompanied by a broadened lineshape due to radiative damping. Note that a prolate spheroid does not only possess a dipolar resonance along its long axis, but also along its short axis, which is shown in Figure 3.2(c) for the example of a silver spheroid with the same volume as a sphere with radius 50 nm, but an aspect ratio of 2. The extinction spectrum of the spheroid excited parallel to its long axis (dashed line) is red-shifted compared to the extinction spectrum of the equivalent sphere (solid line), while it is blue-shifted for a polarisation perpendicular to the long axis (dotted line). This behaviour is intuitively expected due to the longer resonant wavelengths of oscillations in longer potential wells. To conclude the discussion

of this figure, it is important to note that the extinction spectra were generated with an approximate analytical formula and that higher-order resonances, as well as effects beyond order k^3 were neglected, which leads to inaccurate predicted spectra when the particle size approaches and exceeds the wavelength of the incident light. In general, particularly for more complicated shapes, a numerical simulation of the extinction spectra is necessary, which is at the heart of the design of nanoplasmonic antennae.

3.3 Plasmonic Nanoparticles as Optical Antennae

An optical antenna is a device which couples free-space optical modes to localised near-field modes smaller than or on the order of the wavelength [9], which is schematically illustrated in Figure 3.3(a). By reciprocity, the ability to efficiently transform electromagnetic energy in free-space modes to energy bound in sub-wavelength modes implies a high efficiency of the reverse process [9], where energy in a near-field mode radiates into the far-field. Figure 3.3(b) schematically compares dipole emission and quadrupole emission, which are the two leading terms in a multipole expansion of the nanoparticle. While the dipole couples well to the far-field, the quadrupole does not, which is why the latter is often called a dark mode and the former is called a bright mode in the literature [103]. The bad emission properties of a quadrupole can be rationalised intuitively by decomposing it into two dipoles of opposite orientation. This is equivalent to a relative phase shift of π leading to destructive interference in the forward direction. Due to the finite separation of the dipoles, the mode is not completely dark, but it is possible to weakly couple to it at high angles [103].

An alternative method for exciting dark modes is via near-field coupling, which can be achieved in multiparticle plasmonic nanostructures [104]. Figure 3.3(c) shows the schematic extinction spectrum of such a structure and a diagram of its relevant modes. The incident light couples to the broad dipole mode of the small particle, which in turn excites the narrow quadrupole mode of the large particle via near-field coupling, creating a so-called Fano resonance. In a Fano resonance, a spectrally narrow mode couples to a much broader mode

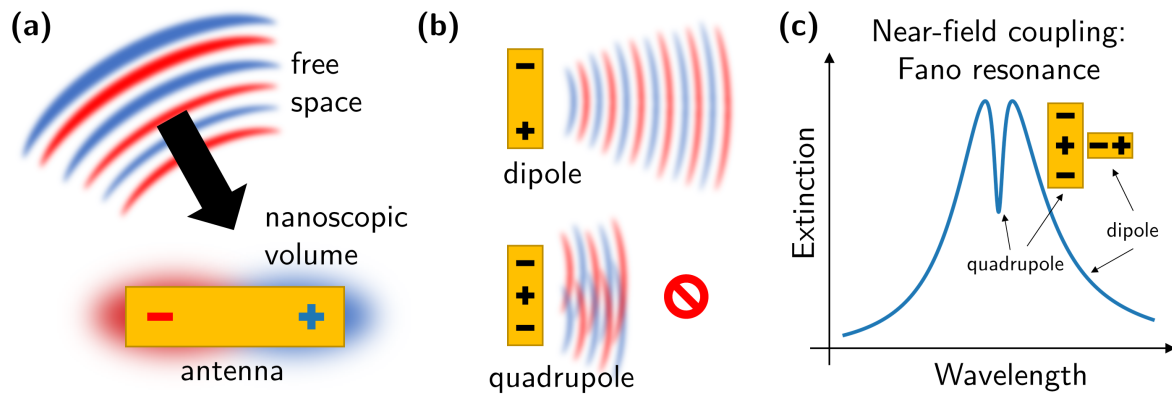


Figure 3.3 – Plasmonic antennae and near-field coupling. (a) Schematic of a nanoparticle acting as an optical antenna with the purpose of coupling free-space optical modes to tightly confined nanoscopic antenna near-field modes and vice versa as efficiently as possible [9]. (b) The dipole mode of a plasmonic bar antenna radiates well to the far-field, whereas the quadrupole mode does not. Heuristically, this can be understood by two small dipoles radiating out of phase and hence cancelling each other out in the far-field. Due to reciprocity, the quadrupole mode cannot be excited from the far-field at normal incidence and is hence called a dark mode, while the dipole mode can be excited well and is hence called a bright mode. (c) A dark quadrupolar mode can be indirectly excited from the far-field in a plasmonic structure via a dipole mode, which couples to the quadrupole in the near-field. The extinction spectrum of such a structure, such as a nanoparticle dimer, exhibits a sharp dip due to the narrow quadrupole resonance reducing the population of the broad dipole resonance. This is an example of a so-called Fano resonance, where a spectrally narrow mode is coupled to a spectrally broad mode or a continuum [102].

or a continuum, with a lineshape that in general may be asymmetric [102]. In the plasmonic nanoparticle system considered here, the different width of the modes is due to the high radiative loss rate of the dipole compared to the quadrupole, which translates into a much lower lifetime and hence a much broader linewidth [102].

All aspects discussed above, in particular the plasmonic near-field enhancement, the tuneability of the plasmonic particle resonance, the radiative properties of dipole and quadrupole modes of particle antennae, and near-field coupling play a crucial role in the nanoparticle ensemble antenna presented in Chapter 7 designed for efficient SHG on the nanoscale. The individual nanoparticles of the antenna are optimised for an efficient radiative coupling between the near-field and the far-field at the fundamental and second harmonic wavelengths, which was achieved by appropriately sized particles with dipolar resonances at those wavelengths. The high SHG efficiency was due to a good overlap of the strongly enhanced dipolar near-field at the fundamental and the quadrupolar near-field at the second harmonic wavelength;

the latter was designed to couple to two nearby particles with a dipole resonance at the second harmonic wavelength for an efficient radiation of the SHG signal into the far-field. This system, which is described in detail in an earlier publication [49] and in Chapter 7, demonstrates the importance of plasmonic concepts for efficient frequency mixers on the nanoscale.

3.4 Conclusion

In this chapter, we have covered selected fundamental aspects of plasmonics, which are crucial for understanding the plasmonic antennae in the hybrid perovskite sample in Chapter 5 and in the frequency mixing arrays in Chapter 7. After deriving the most fundamental plasmonic surface mode, a surface plasmon polariton at a metal-dielectric interface, and demonstrating its sub-wavelength confinement and enhancement of the local fields, we investigated localised surface plasmon resonances in metal nanoparticles which are tuneable throughout the visible and near-infrared range of the spectrum via their size and shape. We have seen that these particles allow for the design of optical antennae, since dipolar resonances, through which near- and far-fields are coupled efficiently, can be controlled directly via the particle morphology. Finally, we have covered the near-field coupling between plasmonic nanoparticles, which can be observed experimentally through the characteristic Fano lineshape of the extinction spectrum and which is a key design element for the frequency mixing arrays in Chapter 7.

Chapter 4

The Z-Scan Technique

Many technological applications of nonlinear optics, such as Kerr-lens mode-locking [105], pulse compression based on self-phase-modulation [106], and photon pair generation with degenerate four-wave-mixing [107] rely on materials with a large Kerr-type nonlinearity. A convenient and popular way to identify such materials is the Z-scan technique, which was developed by Sheik-Bahae and colleagues in the late 1980's [58]. In this chapter, the Z-scan setup, the working principle, and the key theoretical aspects of the data analysis will be discussed in detail. In particular, a major part of the chapter is devoted to advanced aspects of the analysis of saturable absorption data, which will prove essential for the nonlinear characterisation of 2D perovskite flakes in the following chapter. Being mostly theoretical, this chapter thus bridges the gap between the general nonlinear optics theory in Chapter 2 and the experimental results in Chapter 5.

4.1 Z-Scan Experimental Setup

Let us begin our discussion with the description of a typical Z-scan setup as introduced by Sheik-Bahae and co-workers [58], which is depicted schematically in Figure 4.1(a). Note that this section only deals with the working principle of a generic Z-scan setup; details on

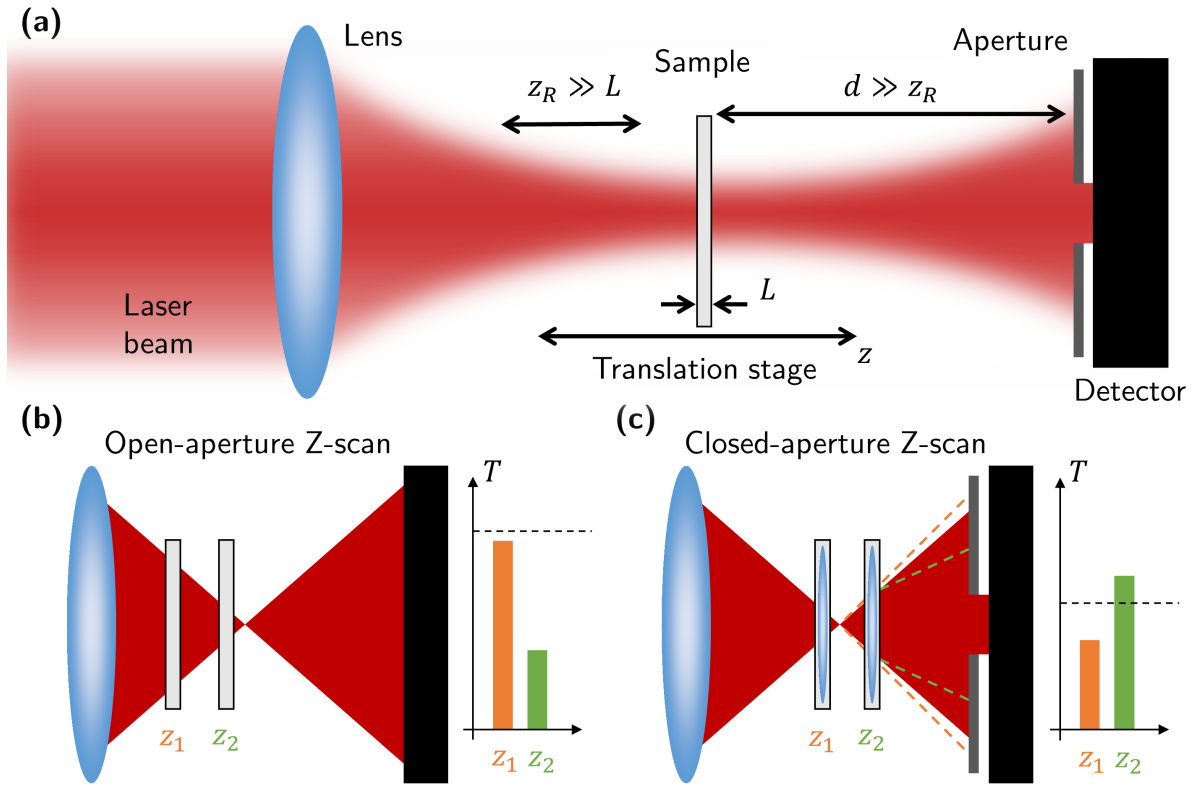


Figure 4.1 – Z-scan setup and working principle. (a) The sample of thickness L is moved through a focused laser beam with a Rayleigh range z_R , assumed to be much larger than L in the following discussion, and the transmitted intensity recorded as a function of the z -position of the sample. The detector is assumed to be far away at a distance $d \gg z_R$ from the sample. Due to the varying intensity, the transmittance changes as a function of the z -position. An aperture can be introduced to restrict the detection to light close to the optical axis which makes the technique sensitive to nonlinear refraction. (b) Working principle of the open-aperture (OA) Z-scan. The aperture is removed such that the total transmittance is measured as a function of the sample z -position. Far away from the focus at position z_1 , nonlinear absorption is weak, while at position z_2 close to the focus its impact is increased. Thus, analysis of the transmittance curve as a function of the z -position gives access to the nonlinear absorption coefficient. (c) Working principle of the closed-aperture (CA) Z-scan for which an aperture selecting light near the optical axis is introduced. The nonlinear refractive index change inside the sample generates a Gaussian index profile acting as a weak lens. For a sample position before the focus, z_1 , the beam divergence behind the focus is slightly increased, leading to a reduced transmission through the aperture. When the sample is behind the focus, at position z_2 , the beam is slightly re-focused, increasing the transmission through the aperture. Therefore, the nonlinear refractive index change can be determined by analysing the transmittance in CA configuration as a function of the z -position of the sample.

the setup used in the experiments are included in Appendix A. A laser beam is focused with a lens, and a detector is placed several Rayleigh lengths z_R behind the focus such that it captures the whole beam spot. The detector can be fitted with an aperture to restrict the detection to a small area at the centre of the beam. Next, the sample to be investigated –

typically a thin film – is inserted perpendicularly to the beam axis near the focus. In terms of our coordinate system, we choose the sample plane as the xy -plane and the beam axis as the z -axis. The sample is mounted on a stage such that it can be shifted along the beam axis, allowing one to control the intensity inside the sample as a function of the z -position of the stage, which gives the Z-scan technique its name. For the theory presented in the following sections it is assumed that the sample of thickness L is thin compared to the Rayleigh length z_R , the necessity of which will be explained in the next section.

Let us now explore how the nonlinear parameters associated with the $\chi^{(3)}$ nonlinearity can be determined with this setup. In the open-aperture (OA) Z-scan, which is visualised in Figure 4.1(b), the aperture is removed such that the whole transmitted power is detected as a function of the z -position of the sample, which allows one to quantify the nonlinear absorption. At position z_1 far away from the focus, the sample experiences a low intensity and the nonlinear absorption is weak, leading to only a slight reduction of the transmittance compared to the limit of vanishing input intensity. However, closer to the focus, at position z_2 , the intensity is much higher, which increases the amount of nonlinear absorption and hence reduces the transmittance even further relative to the zero-intensity limit. The resulting pairs of transmittance values and z -positions can be plotted as a function of the z -position, which we will be calling an OA Z-scan trace, the analysis of which yields a nonlinear absorption coefficient β . The discussion of the meaning and limitations of such a nonlinear absorption coefficient β and possible alternative characteristic parameters will occupy a large part of this chapter.

In order to determine the nonlinear refractive index change, the detector aperture is inserted, such that only light very close to the centre of the beam is detected. The working principle of this configuration, called the closed-aperture (CA) Z-scan, is illustrated in Figure 4.1(c). Due to the AC Kerr effect, the Gaussian beam profile gives rise to a Gaussian refractive index profile in the sample, which in the case of a positive refractive nonlinearity – also called self-focussing – acts as a weak convex lens. At position z_1 before the focus, this Kerr lens will focus the beam more strongly and thus increase the divergence of the beam behind

the focus. The total power in the detector plane is unchanged, but the on-axis intensity is reduced, which is measured because of the aperture. When the sample is placed close to but behind the focus at position z_2 , the Kerr lens reduces the beam divergence, increasing the on-axis intensity in the detector plane and hence the transmission through the aperture. At the focus, the Kerr lens reaches its maximum optical power, but the marginal ray height is zero, which means that the transmission through the aperture is not affected. Therefore, we expect the transmission as a function of the z -position to exhibit a dip before the focus, a peak behind the focus and a zero-crossing at the focus. Of course, in the case of a negative nonlinear refractive index change, which corresponds to self-defocussing rather than self-focussing, the order of the peak and the dip is reversed. Analysing such CA Z-scan traces with suitable formulae, which will be derived in the next section, allows one to determine the nonlinear refractive index n_2 .

For the sake of completeness it is worth mentioning that there are several extensions of the Z-scan technique described elsewhere, such as eclipsing Z-scan [108], two-colour Z-scan [109, 110], which also allows for time-resolved measurements [111], single-beam time-resolved Z-scan for slow nonlinearities [112], and reflection Z-scan [113]. However, since none of these adaptations of the Z-scan method were used in the measurements for this thesis and since they would require a separate theoretical treatment, they will not be discussed further.

4.2 Preliminary Remarks on Z-Scan Analysis

Before we start our derivations, let us briefly discuss the assumptions allowing one to link the nonlinear refraction and absorption in the sample to the experimental transmission data as a function of the z -position by means of a closed-form analytical expression. The simplest and most commonly used analytical formulae were derived in the original Z-scan paper by Sheik-Bahae and co-workers [58] for the case of weak two-photon absorption (TPA) and a weak nonlinear refractive index change. In their calculations, a TEM_{00} Gaussian profile was assumed, requiring a high-quality pump beam. This is easily satisfied in most Z-scan

experiments given the high quality of the commercially available lasers today. Moreover, the sample thickness L was assumed to be much less than the Rayleigh range z_R of the focused pump beam in order to be able to treat the intensity propagation through the sample as an effectively one-dimensional problem. If the sample thickness is equal to or larger than the Rayleigh range, the theory in Refs. [114,115] should be used instead. Note that by choosing a lens with a sufficiently long focal length, one can in principle always satisfy $L \ll z_R$; however, this may often be impracticable due to the need for a long translation stage and a higher input power to reach the desired peak intensity at the focus. As a final remark on the sample length L , one should also be aware that Fabry-Perot interference in the thin film may lead to artefacts in the data if $L \gtrsim \lambda$ – of course, the extent of this problem depends on the refractive index contrast at the surfaces and the surface roughness according to standard Fabry-Perot theory [116]. These étalon effects can be avoided by, for example, using thick vapour cells with an anti-reflection coating [115].

Under the conditions above, experimental Z-scan traces can be analysed with the theory by Sheik-Bahae and colleagues [58] to extract the value of the nonlinear refraction coefficient, n_2 and two-photon absorption coefficient, β_{TPA} . However, in some cases, particularly near a resonance, samples may exhibit a complicated saturation behaviour, which goes beyond the scope of that theory. Therefore, if proper care is not taken, one may obtain wrong or meaningless values of the nonlinear parameters even if the experiments are conducted at weak intensities to achieve a sufficiently weak nonlinear response. Another issue is the morphology of metamaterial or metasurface samples distorting the transmitted beam profile, such that the assumption of a TEM_{00} Gaussian profile does not hold true any more, leading to CA data that cannot be meaningfully analysed with these formulae. Both problems will play a major role in the data analysis of the perovskite samples in the next chapter. Despite those limitations, the theory by Sheik-Bahae and colleagues [58] is a good starting point for the analysis of most Z-scan data.

Therefore, the remainder of this chapter will be organised as follows: In Section 4.3, we shall present a streamlined derivation of the most commonly used formulae for Z-scan analysis

as laid out by Sheik-Bahae and co-workers. The reason for including this derivation in the main text rather than in an appendix is that the problems related to an incorrect use of these formulae can only be understood if the key steps of the derivation are explained first. Moreover, this simplifies referring back to certain aspects of the derivation when discussing the analysis problems in later sections. In our presentation of the derivation, we shall, for reasons of convenience, first show the results for continuous-wave illumination (CW) and then adapt the formulae for pulsed excitation, which will be needed since our experiments in the following chapter were conducted with a pulsed laser. The resulting formulae are truncated series, and the truncation error as well as a so-called division method for the CA trace will be discussed separately in Subsection 4.3.2.

Once the theoretical foundations are thus established, we shall in Section 4.4 extensively discuss the analysis of Z-scans with saturable absorption since it is present in most of the data shown in the next chapter. In particular, we shall examine under which circumstances such data can and cannot be fitted with adapted versions of the Sheik-Bahae formulae, and which pitfalls are to be avoided. This is necessary, since some workers in the field have incorrectly used or adapted these formulae, and uncritically adopting their approach would have had detrimental effects on the data analysis in the next chapter. At the end, in Subsection 4.4.2, a two-level saturation model will be discussed and compared to phenomenological approaches for analysing Z-scan data with saturable absorption.

In Section 4.5 we shall introduce a numerical analysis framework allowing for the simulation of OA Z-scan data for arbitrary propagation models, as long as they can be stated in terms of a differential equation for an effectively one-dimensional propagation of the fields. This allows, for example, for a direct comparison of saturation models regardless of their analytical complexity, since the numerical framework does not require any analytical approximations. We shall use this tool to analyse a complicated OA Z-scan trace and show explicitly that two alternative analysis approaches are incorrect, justifying again the development of the numerical analysis code and the rejection of some approaches in the literature. The chapter closes with concluding remarks on the key problems of Z-scan analysis.

4.3 Off-Resonance Theory for a Weak $\chi^{(3)}$ Nonlinearity

As mentioned above, this section is based on the results by Sheik-Bahae and colleagues [58]. First, we shall follow their derivation to obtain the most common formulae for fitting Z-scan data. After that, we shall discuss the caveats when working with the linear truncations of those formulae.

4.3.1 Derivation of OA and CA Fitting Formulae

As mentioned above, we shall first cover the cases of nonlinear refraction and absorption under CW excitation, which is followed by a discussion of pulsed excitation of instantaneous third-order effects. In the thin-sample approximation, $L \ll z_R$, a slowly varying envelope of the laser beam is assumed and the propagation through the sample becomes effectively one-dimensional with the equations

$$\frac{d\Delta\phi(z')}{dz'} = kn_2 I(z') \quad \text{and} \quad \frac{dI(z')}{dz'} = -\alpha I(z') - \beta_{\text{TPA}} I^2(z') \quad (4.1)$$

governing the intensity distribution, $I(z')$, and nonlinear phase shift, $\Delta\phi(z')$, in the sample of physical length L , where z' is the z -coordinate inside the sample. Solving Eq.(4.1)b yields

$$I(z') = \frac{I_{\text{in}} e^{-\alpha z'}}{1 + I_{\text{in}} \beta_{\text{TPA}} (1 - e^{-\alpha z'}) / \alpha} \stackrel{z'=L}{=} \frac{I_{\text{in}} e^{-\alpha L}}{1 + I_{\text{in}} \beta_{\text{TPA}} L_{\text{eff}}} \quad (4.2)$$

with an effective length $L_{\text{eff}} = (1 - e^{-\alpha L}) / \alpha$ and $I_{\text{in}}(r, z)$ being the intensity inserted at the front surface of the sample after subtraction of Fresnel reflection losses $R = |(1-n)/(1+n)|^2$ at the interface. With this result, Eq.(4.1)a can be solved

$$e^{i\Delta\phi(L)} = \underbrace{\left(1 + I_{\text{in}} \beta_{\text{TPA}} L_{\text{eff}}\right)}_{q(r,z)}^{ikn_2/\beta_{\text{TPA}}} \xrightarrow{\beta_{\text{TPA}} \rightarrow 0} e^{ikn_2 I_{\text{in}} L_{\text{eff}}} \quad (4.3)$$

where the nonlinear absorption parameter $q(r, z) = \beta_{\text{TPA}} L_{\text{eff}} I_{\text{in}}(r, z)$ was defined and the expression in the limiting case of only nonlinear refraction is consistent with the intuitively expected total phase shift of $\Delta\phi(z' = L) = kn_2 I_{\text{in}} L_{\text{eff}}$.

Let us next consider the evolution of the xy beam profile $E(r, z', z)$ in the sample. For a thin sample with two-photon absorption as in Eq.(4.1)b and an intensity-dependent refractive index $n(I) = n_0 + n_2 I = n_0 + \Delta n$, the electric field, $E(r, z' = L, z)$, of the laser beam at the rear surface of the sample at a scan position z must be

$$E(r, z' = L, z) = E_{\text{in}}(r, z) e^{-\alpha L/2} (1 + q(r, z))^{-\frac{1}{2} + ikn_2/\beta_{\text{TPA}}} \xrightarrow{\beta_{\text{TPA}} \rightarrow 0} E_{\text{in}}(r, z) e^{-\alpha L/2} e^{ikn_2 I_{\text{in}} L_{\text{eff}}} \quad (4.4)$$

where we have assumed negligible material birefringence enabling a scalar description. The incident electric field, $E_{\text{in}}(r, z)$ at a scan position z is the assumed TEM₀₀ Gaussian mode of the pump beam,

$$E_{\text{in}}(r, z) = E_0 \frac{w_0}{w(z)} \exp\left(-\frac{r^2}{w^2(z)} - \frac{ikr^2}{2R(z)}\right) e^{-i\phi(z)} \quad (4.5)$$

with the corresponding intensity distribution

$$I_{\text{in}}(r, z) = I_0 \frac{1}{1 + \tilde{z}^2} \exp\left(-\frac{2r^2}{w^2(z)}\right) \quad (4.6)$$

where $\tilde{z} = z/z_R$, $z_R = kw_0^2/2$ is the Rayleigh range, $R(z) = z(1 + \tilde{z}^2)$ is the radius of curvature, and $w(z) = w_0\sqrt{1 + \tilde{z}^2}$ is the $1/e^2$ beam width.

First, we consider the case of only nonlinear refraction, where $\beta_{\text{TPA}} = 0$, and after inserting Eq.(4.6) in Eq.(4.4), the electric field of the laser beam just behind the sample at position z becomes

$$E(r, z' = L, z) = E_{\text{in}}(r, z) e^{-\alpha L/2} e^{ikn_2 I_{\text{in}} L_{\text{eff}}} = \mathcal{E} \sum_m \frac{(ikL_{\text{eff}} n_2 I_0)^m}{m! (1 + \tilde{z}^2)^m} e^{-\frac{(2m+1)r^2}{w^2(z)} - \frac{ikr^2}{2R(z)}} \quad (4.7)$$

with $\mathcal{E} = E_0 e^{-\alpha L/2} (1 + \tilde{z}^2)^{-1/2} e^{-i\phi(z)}$. The series expansion of the nonlinear phase term leaves us with a sum of Gaussians that can all be propagated from the sample to the detector plane at distance d individually – a technique called the Gaussian decomposition method [117]. This is not the only way to obtain the electric field distribution in the detector plane, but merely a convenient alternative to a zeroth-order Hankel transform. When carrying out the propagation, it is important to note that at the position of the sample each constituent beam of the decomposition has a different width, but the radius of curvature $R(z) = z(1 + \tilde{z}^{-2})$ is the same for all partial beams. Introducing the complex beam parameter [118] $q_{\text{beam}}(z)$ with the properties

$$q_{\text{beam}}(z) = q_{\text{beam}}(0) + z \quad \text{and} \quad \frac{1}{q_{\text{beam}}(z)} = \frac{1}{R(z)} + i \frac{\lambda}{\pi w^2(z)}, \quad (4.8)$$

the m -th partial beam can conveniently be propagated to the detector plane at a distance d by finding the real and imaginary parts of

$$q_{\text{beam}}(z + d) = q_{\text{beam}}(z) + d = \frac{1}{\frac{1}{R(z)} + i \frac{\lambda}{\pi w_m^2(z)}} + d \quad (4.9)$$

and identifying R and w using Eq.(4.8)b. Note that the complex beam parameter is usually just called q , but is denoted by q_{beam} here to distinguish it from the nonlinear absorption parameter q . Knowing the real and imaginary parts of $q_{\text{beam}}(z + d)$ also allows one to calculate the Gouy phase θ_m accumulated by the m -th partial beam between the sample and the detector. The initial Gouy phase just behind the sample is subtracted using the identity $\arctan(x) - \arctan(y) = \frac{x-y}{1+xy}$. As a result of this propagation step, the field distribution at the detector is found to be

$$E_{\text{detector}} = \mathcal{E} \sum_m \frac{(i\Delta\phi_0)^m}{m! (1 + \tilde{z}^2)^m} \left(g^2 + \left(\frac{d}{d_m} \right)^2 \right)^{-\frac{1}{2}} \exp \left(-\frac{r^2}{w_m^2} - \frac{ikr^2}{2R_m} + i\theta_m \right) \quad (4.10)$$

with $w_m^2 = \frac{w^2(z)}{(2m+1)}$, $R_m = d \left(1 - \frac{g}{g^2 + (\frac{d}{d_m})^2} \right)^{-1}$, $d_m = kw_m^2/2$, $g = 1 + \frac{d}{R(z)}$, $\theta_m = \tan^{-1} \left(\frac{d/d_m}{g} \right)$, and the peak on-axis nonlinear phase shift $\Delta\phi_0 = kL_{\text{eff}}n_2I_{\text{in}}(r = 0, z = 0)$.

For an aperture in the far field ($z_0 \ll d$) the closed-aperture Z-scan signal is given by the on-axis ($r = 0$) intensity. It is customary to normalise this intensity by the 0th-order term corresponding to an absence of nonlinear effects. This means that under all assumptions stated above, the relative intensity signal in a closed-aperture (CA) Z-scan is

$$T_{CA} = 1 + \Delta\phi_0 \frac{4\tilde{z}}{(\tilde{z}^2 + 1)(\tilde{z}^2 + 9)} + \Delta\phi_0^2 \frac{4(3\tilde{z}^2 - 5)}{(\tilde{z}^2 + 1)(\tilde{z}^2 + 9)(\tilde{z}^2 + 25)} + \mathcal{O}(\Delta\phi_0^3). \quad (4.11)$$

For sufficiently weak nonlinear effects, it is acceptable to fit the data with a truncated version of this equation up to first order in $\Delta\phi_0$.

In order to consider the simultaneous impact of two-photon absorption and nonlinear refraction on the CA Z-scan we must evaluate Eq.(4.4) for non-vanishing β_{TPA} . Using

$$(1 + x)^\alpha = \sum_m \binom{\alpha}{m} x^m = \sum_m x^m \prod_{n=1}^m \frac{\alpha - n + 1}{n} = \sum_m \frac{x^m}{m!} \prod_{n=1}^m (\alpha - n + 1) \quad (4.12)$$

we find

$$E(r, z)_{\text{after}} = \mathcal{E} \sum_m \frac{(ikL_{\text{eff}}n_2I_0)^m}{m!(1 + \tilde{z}^2)^m} \delta_m e^{-\frac{(2m+1)r^2}{w^2(z)} - \frac{ikr^2}{2R(z)}} \quad (4.13)$$

with $\delta_m = \prod_{n=1}^m \left(1 + i(2n - 1)\frac{\beta_{TPA}}{2kn_2}\right)$ just being a constant factor for each partial beam in the Gaussian decomposition. Propagation to the detector and calculation of the on-axis intensity are done in the same way as above and result in a first-order expression for the CA trace of

$$T_{CA} = 1 + \Delta\phi_0 \frac{4\tilde{z}}{(\tilde{z}^2 + 1)(\tilde{z}^2 + 9)} - q_0 \frac{\tilde{z}^2 + 3}{(\tilde{z}^2 + 1)(\tilde{z}^2 + 9)} + \mathcal{O}(\Delta\phi_0 q_0, \Delta\phi_0^2, q_0^2) \quad (4.14)$$

consistent with Eq.(4.11), and using the simplified notation $q(r = 0, z = 0) = q_0$ for the nonlinear absorption parameter.

For the open-aperture (OA) scan, the entire transmitted intensity is detected in an experiment. Hence, the nonlinear refraction does not have an impact any more and it suffices to integrate the intensity distributions given by Eq.(4.2) and Eq.(4.6). This leads to the

expression

$$P_{\text{transmitted}} = \int_0^{2\pi} d\varphi \int_0^\infty dr r \frac{I_0(z) e^{-\alpha L} e^{-2r^2/w^2(z)}}{1 + \underbrace{\beta_{\text{TPA}} L_{\text{eff}} I_0(z)}_{q(r=0,z)=q_0(z)} e^{-2r^2/w^2(z)}} = P_0(z) e^{-\alpha L} \frac{\ln(1 + q_0(z))}{q_0(z)} \quad (4.15)$$

with $P_0(z) = \frac{\pi w_0^2}{2} I_0(z)$ and $I_0(z) = I_{\text{in}}(r = 0, z)$. Again, in an experiment it is customary to work with a normalised transmission signal, such that the OA trace is given by

$$T_{\text{OA}} = \frac{P_{\text{transmitted}}}{P_{\text{transmitted}, q=0}} = \frac{\ln(1 + q_0(z))}{q_0(z)} = \sum_{m=0}^{\infty} \frac{(-q_0(z))^m}{m+1} \quad (4.16)$$

The theory considered so far is based on the assumption of the incident electric field being constant in time and hence only applies to experiments with a CW pump beam. As pulsed lasers with their high peak intensities at low average powers are often used to increase the nonlinearity while minimising thermal effects [119], adapting the formulae above to the case of pulsed excitation is of high practical relevance. In many experiments, the pump pulses are Gaussian in time. For the OA formula, after replacing $I_0(z) \rightarrow I_0(z) e^{-t/\tau_p}$ in Eq.(4.15) and taking the average over the pulse, where the integration limits can be extended to infinity as the pulse duration τ_p is much smaller than the repetition period between pulses, we obtain

$$T_{\text{OA}} = \sum_{m=0}^{\infty} \frac{(-q_0(z))^m}{(m+1)^{\frac{3}{2}}} = 1 - \frac{1}{2\sqrt{2}} \frac{q_0(0)}{1 + \tilde{z}^2} + \mathcal{O}(q_0^2) \quad (4.17)$$

with a correction factor $1/\sqrt{2}$ in the linear term as given by Sheik-Bahae and colleagues [58]. Incorrectly using the CW formula in Eq.(4.16) for an OA Z-scan with a pulsed laser would lead to an underestimation of β_{TPA} by $\sim 30\%$ up to first order in q_0 .

For the CA scan, we must make the same substitution $I_0(z) \rightarrow I_0(z) e^{-t/\tau_p}$ in Eq.(4.13) and take the time-average. Following this approach, the same $1/\sqrt{2}$ correction factor is found in the first-order result for the CA Z-scan trace [120]

$$T_{\text{CA}} \approx 1 + \Delta\phi_0 \frac{4\tilde{z}}{\sqrt{2}(\tilde{z}^2 + 1)(\tilde{z}^2 + 9)} - q_0 \frac{\tilde{z}^2 + 3}{\sqrt{2}(\tilde{z}^2 + 1)(\tilde{z}^2 + 9)} \quad (4.18)$$

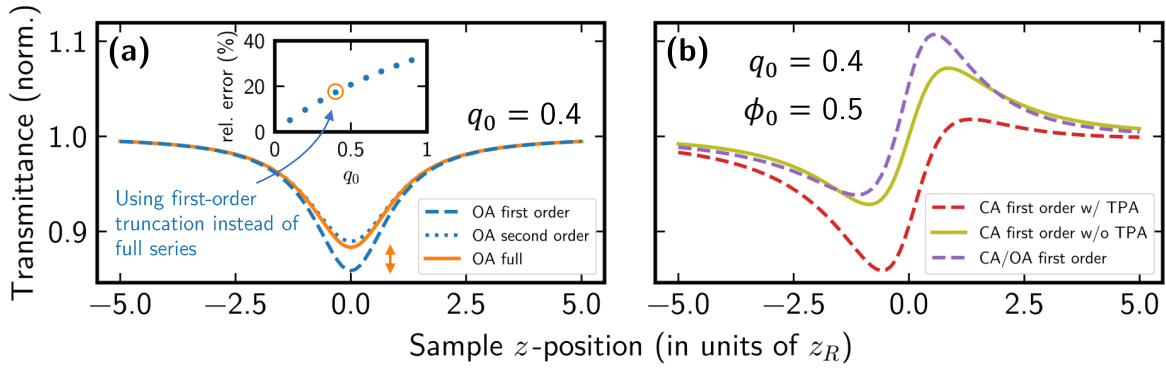


Figure 4.2 – OA and CA Z-scan fit curves in first-order approximation. (a) Comparison of OA Z-scan traces in first- and second-order approximation with the full series expansion for an intermediate strength of two-photon absorption (TPA) with $q_0 = \beta_{\text{TPA}} I_0 L_{\text{eff}} = 0.4$. The first-order approximation leads to an underestimation of β_{TPA} by $\sim 20\%$. (b) CA Z-scan traces with the same nonlinear absorption and a nonlinear phase shift of $\phi_0 = 0.5$ rad. The dashed red line shows a CA Z-scan trace that would be measured in an experiment for those parameters. For reference, the scan with a purely refractive nonlinearity is shown in yellow. It can be approximated by dividing the CA by the OA trace (dashed purple) to within a 10% error in ϕ_0 if $\Re(\chi^{(3)}) > \Im(\chi^{(3)})$ [58].

again with the simplified notation $q_0(z = 0) = q_0$, such that now $t = 0$ and $r = 0$ are implied and the subscript 0 means $z = 0$. For the remainder of this thesis, this convention will be adopted, meaning in particular $I(z)$ denoting the peak intensity as a function of z with implied $t = 0$ and $r = 0$ and the abbreviation I_0 for $I(z = 0)$. The same applies to the nonlinear absorption parameter $q(z)$, where $q(z = 0)$ will be abbreviated as q_0 .

4.3.2 Linear Truncation of the OA and CA Expressions

Let us now take a moment to discuss the outcome of the above derivations. Given the assumptions described above, we have found that Z-scan data in pulsed experiments can be fitted with Eqs.(4.17) and (4.18). In practical terms, the linear truncation of those expressions is most often used by workers in the field, which makes it interesting to take a closer look at the limitations of this approach.

Fitting OA data with the first-order truncation of (4.17) leads to a systematic error in the estimate of β_{TPA} , which is visualised in Figure 4.2(a). The full OA expression Eq.(4.17) (orange) is plotted with the first-order (blue dashed) and second-order (blue dotted) truncations

for the example of $q_0 = \beta_{\text{TPA}} I_0 L_{\text{eff}} = 0.4$. Comparing the curves reveals that the first-order truncation will lead to a systematic underestimation of β_{TPA} , since the parameter q_0 of the first-order curve needs to be reduced in order to match the orange curve representing the full expression which indicates what the Z-scan data would really be in an experimental situation with a peak intensity and sample properties such that $q_0 = 0.4$. Adding the quadratic term significantly improves the agreement, which is unsurprising given the convergence radius of $|q_0| = 1$ for Eq.(4.17). The inset visualises the relative error in the estimation of β_{TPA} as a function of q_0 in a given experimental situation when blindly using the first-order truncation rather than the full expression Eq.(4.17) for fitting OA data. For small q_0 values up to 0.2, the truncation error is below 10% and hence likely acceptable, given that there are many other and often dominant sources of error, such as intensity fluctuations or an inaccurate estimation of the Rayleigh range z_R . However, for larger but still realistic q_0 values of ~ 0.8 it can be seen that – all else being equal – by ignoring the truncation error one unknowingly underestimates β_{TPA} by about one third.

For the CA data in the next chapter, the key practical problem was not a too strong nonlinear phase shift requiring higher-order terms but rather the separation of nonlinear refraction and nonlinear absorption. Therefore we shall discuss the method introduced by Sheik-Bahae and colleagues [58] where the normalised CA data is divided by the normalised OA data and then fitted according to Eq.(4.18) with $q_0 = 0$. Figure 4.2(b) shows a CA curve according to Eq.(4.18) with $q_0 = 0.4$ and $\phi_0 = 0.5$ rad (dashed red) together with a hypothetical purely refractive case where $q_0 = 0$ (yellow) and the curve obtained by dividing the CA curve by the OA curve with $q_0 = 0.4$ in an attempt to eliminate the nonlinear absorption from the data (dashed purple). The agreement between the yellow and purple curves is reasonable and Sheik-Bahae and co-workers showed that an error of below 10% is introduced with this method as long as $\Re(\chi^{(3)}) > \Im(\chi^{(3)})$ and $q_0 < 1$ [58].

It is important to note that so far, we have only considered quasi-instantaneous off-resonant nonlinear processes. However, the Z-scan technique is sensitive to all intensity-dependent absorptive and refractive effects [121] and the theory summarised in this section with its

nonlinear parameters β_{TPA} and n_2 is often not suitable for a good interpretation of more complicated data, particularly when the assumptions of a weak nonlinear response, or of a material response time much shorter than the laser pulse duration are violated. In the limiting case of the nonlinearity being much slower than the laser pulse duration, it was shown that a correction factor of $\frac{1}{2}$ independent of the pulse shape must be used [58]. Note that in this case q_0 and $\Delta\phi_0$ are not linked to instantaneous electronic nonlinearities any more, but for instance to excited-state cross-sections when the process involves a long-lived excited intermediate state [121]. However, there are other cases, such as the nonlinear behaviour of perovskite flakes on plasmonic antennae in the following chapter, where strong saturable absorption occurs near an excitonic resonance and the time-dependence and non-local effects are too complicated to be accounted for by a simple model or an extension of the formulae derived above. As we shall see there, the most reasonable analysis strategy will be to take a step back and phenomenologically describe the results rather than attempting to apply a complicated model that is not supported by information from complementary experiments and does not lead to genuine insights.

Of course, the nonlinear optical response in most experiments is not prohibitively complicated and can indeed be modelled reasonably well. This is the case for saturable absorption, on which we shall focus next because it is both very common in the literature and critical for the analysis of the data in the following chapter. Unfortunately, the analysis of Z-scan data with saturable absorption is not as straightforward as that with two-photon absorption and the treatment in the literature is sometimes problematic, which warrants an extensive discussion in the following section.

4.4 Analysing Z-Scan Data with Saturable Absorption

It is a very common experimental situation in Z-scan for the OA trace to exhibit a positive transmittance change, meaning that the sample becomes more transparent as the pump intensity increases, which is called saturable absorption [122]. Of course, the two-photon

absorption model and fitting formulae developed above cannot be directly applied to analyse such data and the question arises how to best perform the analysis and quantify the nonlinear response of the sample. In this section we shall discuss several approaches, some of which are flawed but nonetheless popular. At the end, we shall introduce a simple two-level saturation model, which requires a numerical solution to produce an OA trace. This goes beyond the scope of analytical fitting and will serve as a link to the next section, which deals with the numerical simulation of OA traces for arbitrary models with ordinary differential equations governing the intensity propagation through the sample.

4.4.1 Phenomenological Models of Saturable Absorption

Let us first consider the simplest and most popular approach, where negative values of q_0 are introduced to be able to use the theoretical framework laid out in the previous section and to fit the data with the very common linear truncation of Eq.(4.17). The rationale behind this method is that the data often resembles a mirror image of OA traces with two-photon absorption, which gives little reason to suspect potential problems. Further contributing factors are the convenience and ubiquitousness of Eq.(4.17). Since the intensity and effective length are always positive quantities, this approach implies a negative β_{TPA} , which is then called α_{NL} [123] or just β [124]. However, if proper care about the actual meaning of a negative β is not taken, problematic consequences about the physical interpretation of such an analysis may arise. In this subsection we will find that under some circumstances a negative β can be meaningful, but not when it is interpreted as a negative two-photon absorption coefficient β_{TPA} . In order to understand why a negative β_{TPA} has problematic implications, it is helpful to consider Figure 4.3. A comparison of the full Eq.(4.17) (solid) and its linear truncation (dashed) for various negative values of q_0 is shown in Figure 4.3(a). While the agreement is good for q_0 values close to 0, the discrepancy between both curves grows until the full expression Eq.(4.17) diverges for $q_0 < -1$. This is consistent with the convergence radius $|q_0| = 1$ of Eq.(4.17). The maximum achievable finite transmittance value with the full expression is $T_{\text{OA}}(q_0 = -1) = \text{Li}_{3/2}(1) \approx 2.61$, while there is no such

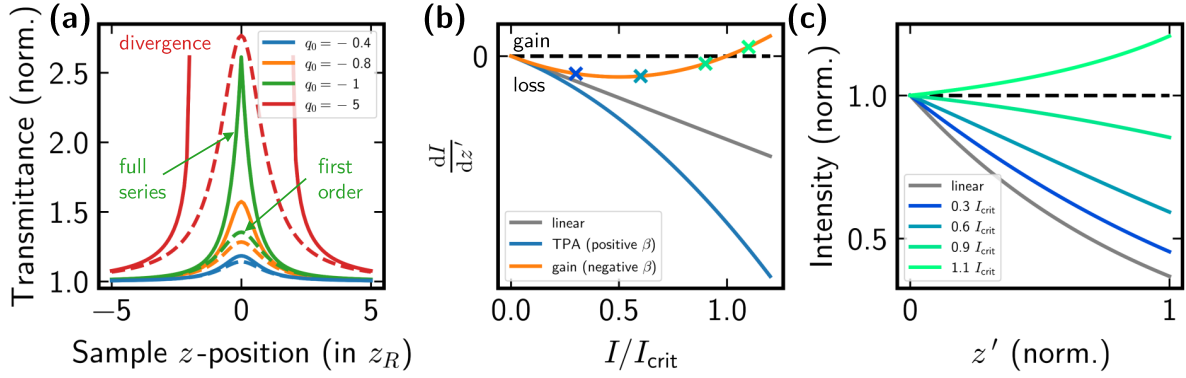


Figure 4.3 – Divergence problem with a negative β_{TPA} . (a) Hypothetical OA Z-scan traces for increasingly negative values of $q_0 = \beta_{TPA} I_0 L_{eff}$. The solid lines represent the full series expression, the dashed lines indicate the first-order truncations. The discrepancy between the dashed and solid lines increases as q_0 tends towards -1 and the full series diverges for $q_0 < -1$. (b) Comparison of $\frac{dI}{dz'}$ as a function of the intensity for β_{TPA} positive (blue), negative (orange), and zero (grey). At a critical intensity, $\beta_{TPA} \cdot I$ is large enough to compensate the linear absorption, beyond which an unphysical gain regime emerges, where the laser beam reinforces itself. (c) Visualisation of the normalised intensity distributions inside the sample for various input powers. Starting from the low-intensity limit of an exponential decay according to the linear absorption coefficient α , the $\beta_{TPA} \cdot I$ increases with intensity and leads to gain above the critical intensity. The intensities for the various curves are marked with crosses on the orange gain curve in (b).

limitation for the first-order truncation. This behaviour indicates that blindly substituting negative values for q_0 into the framework presented in the last subsection is not a meaningful way of doing the analysis. Even though a good agreement with the data may be achieved with negative q_0 values in the linear truncation of Eq.(4.17), this is not a valid application of the Z-scan theory above, but merely a truncation of a diverging series if $q_0 < -1$.

Looking at this problem from another angle demonstrates that this is not only a mathematical issue, but also one of the underlying physics. Figure 4.3(b) shows the intensity gradient $\frac{dI}{dz'}$ inside the sample as a function of the intensity. In the case of linear absorption (grey), the intensity gradient along the sample coordinate z' is proportional to the current intensity value, which implies an exponential decay. For positive values of β_{TPA} (blue), the decay is faster than exponential, while for negative β_{TPA} values the decay is slower than exponential (orange). However, in the latter case the decay turns into gain above a critical intensity $I_{crit} = -\alpha/\beta_{TPA}$, which corresponds to the nonlinear term in Eq.(4.1)b exceeding the linear loss term. This demonstrates that the negative β_{TPA} approach is not a real saturation model,

but merely uses a ‘two-photon gain’ to achieve an intensity-dependent increase in transmittance. The unphysical runaway self-amplification above the critical intensity demonstrates how unsuitable this model is. Note, that I_{crit} is close to but not equivalent to the convergence radius condition $q_0 = -1$, which would imply $I'_{\text{crit}} = -1/(L_{\text{eff}}\beta_{\text{TPA}}) = I_{\text{crit}}/(1 - e^{-\alpha L}) > I_{\text{crit}}$. Hence, excluding the region with unphysical gain is the stronger restriction.

Finally, Figure 4.3(c) shows the transition from loss to gain for increasing incident intensity and a fixed negative β_{TPA} , which is visualised by the normalised intensity distributions inside the sample. The corresponding initial intensities and gradients are marked by crosses on the orange curve in Figure 4.3(b). In the low-intensity limit, the intensity decays exponentially, with the decay decreasing and turning into runaway gain as the incident intensity increases and eventually exceeds I_{crit} . It is interesting to note that just below the critical intensity, the intensity distribution is concave instead of convex, which is not the case for a genuine saturation model, as we shall see at the end of this subsection.

After this demonstration that the negative β_{TPA} approach, if understood in the context of the two-photon absorption Z-scan theory above, leads to a meaningless analysis, this fact needs to be reconciled with the good agreement of many OA traces with the first-order truncation of Eq.(4.17) and a negative q_0 . Figure 4.4(a) shows hypothetical OA Z-scan data points generated with the formula

$$T_{\text{OA}} = 1 - \frac{\beta_{\text{NL}} L_{\text{eff}}}{2\sqrt{2}} \frac{I_0}{1 + \tilde{z}^2} \quad \text{where} \quad \beta_{\text{NL}} < 0 \quad (4.19)$$

to mimic a realistic dataset for a saturable absorber under weak irradiation. Of course, Eq.(4.19) is formally equivalent to the first-order truncation of Eq.(4.17) with a negative q_0 , but it should be viewed as a formula in its own right, rather than being connected to the two-photon absorption theory presented above. Note that β_{NL} is similar to but distinct from β_{TPA} except in the limit $q_0 \rightarrow 0$, which is why it is denoted with a separate name. Indeed, Eq.(4.19) does have a sensible interpretation when viewed on its own, as can be seen in Figure 4.4(b), where the dataset from Figure 4.4(a) was plotted as a function of the intensity

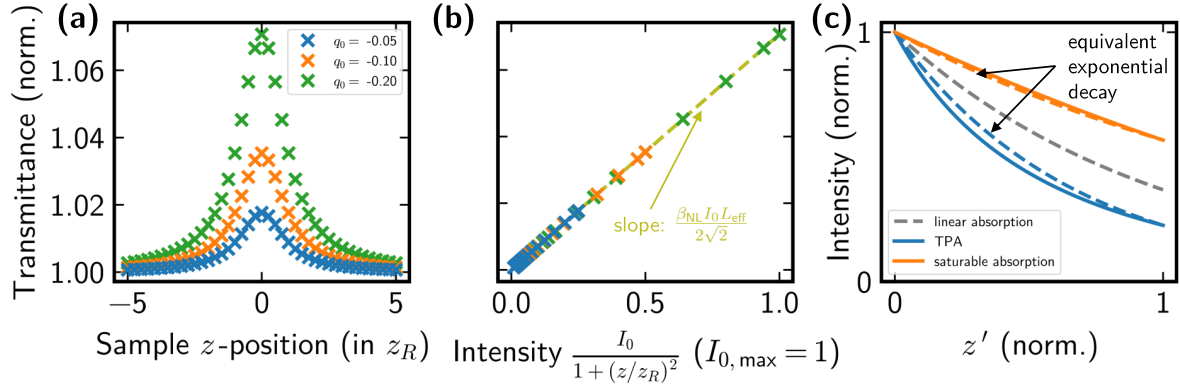


Figure 4.4 – Negative β_{NL} as a slope in intensity space. (a) Hypothetical OA data for the case of weak saturable absorption. The points were generated with the first-order OA expression, but with a phenomenological negative β_{NL} instead of β_{TPA} . (b) The meaning of the negative β_{NL} becomes clear when the same points are plotted against intensity. It is not implying a gain mechanism, but purely a constant of proportionality for a case where the relative transmittance increases linearly with intensity, emancipating the first-order truncation to an expression in its own right. (c) An alternative approach [125] is to consider the relative transmittance change behind the sample and to then find the necessary change in the absorption coefficient α to match that outcome. The approach is applicable to both saturable absorption (orange) and two-photon absorption (blue), but is only phenomenological in the sense that there is no implied prediction of the intensity distribution inside the sample.

for each data point according to the transformation $z \rightarrow I(z) = \frac{I_0}{1+z^2}$ and normalised by the intensity of the highest-intensity data point. All data points lie on a line with slope $\frac{\beta_{NL} I_0 L_{eff}}{2\sqrt{2}}$, which shows that Eq.(4.19) establishes β_{NL} as a coefficient of proportionality of the nonlinear absorption, if the effect is growing approximately linearly with intensity. This has nothing to do with the theory presented above and is a merely phenomenological statement of linear growth of the nonlinear effect with intensity. It is possible to interpret β_{TPA} in the linear truncation of Eq.(4.17) in the same way. Viewed from that perspective, the off-resonant theory by Sheik-Bahae [58] is an addition which identifies Eq.(4.19) as a low-intensity approximation of Eq.(4.17), which in turn is linked to the two-photon absorption parameter β_{TPA} in the physical model Eq.(4.1). Eq.(4.19) works well for weak nonlinearities, because even complicated nonlinear behaviour almost always starts with an effect strength proportional to the incident intensity. At the end of this section, we shall show this explicitly for a two-level saturation model.

There is another valid phenomenological way to characterise the strength of the nonlinear absorption, which is discussed here because it was used in a recent publication with a high impact [125]. In it, the nonlinear absorption is characterised via the parameter $\beta_{\text{eff}} = \frac{\Delta\alpha}{I}$. Figure 4.4(c) visualises the working principle of this approach. A nonlinear process, such as two-photon absorption (blue) or saturable absorption (orange) alters the transmittance through the sample compared to the case of just linear absorption (grey). In order to quantify the strength of the effect, one then determines the attenuation coefficient α_{eff} that would be necessary to reach the observed transmittance after the propagation over the entire sample length. In other words, one finds the equivalent exponential decays (dashed lines) consistent with the measured transmittance, which of course is agnostic about the physical process bringing about the transmittance change, unlike a physical model such as Eq.(4.1). The intrinsic nonlinearity of the material, expressed as the parameter β_{eff} is then found by taking the difference in attenuation coefficients $\Delta\alpha = \alpha_{\text{eff}} - \alpha_0$ and dividing it by the peak intensity of the excitation beam.

It is important to note that β_{eff} determined in that way is distinct from β_{NL} in Eq.(4.19), which can be shown as a simple example for the case of weak two-photon absorption, where $\beta_{\text{NL}} \approx \beta_{\text{TPA}}$. Assuming a 1% reduction in transmittance due to two-photon absorption, one obtains

$$\frac{-\beta_{\text{TPA}} I_0 L_{\text{eff}}}{2\sqrt{2}} = -0.01 \quad (4.20)$$

with the linear truncation of Eq.(4.17) and

$$e^{-\alpha_{\text{eff}} L} = (1 - 0.01) e^{-\alpha_0 L} \quad (4.21)$$

with the method above, leading to the relation

$$\beta_{\text{eff}} \approx \frac{0.01}{L I_0} = \beta_{\text{TPA}} \frac{L_{\text{eff}}}{2\sqrt{2} L} < \beta_{\text{TPA}} \quad (4.22)$$

implying that for a weakly absorbing sample where $L_{\text{eff}} \approx L$, the different conventions for the nonlinear absorption parameter β_{TPA} and β_{eff} lead to a discrepancy of a factor of $2\sqrt{2} \approx 3$

for the same experiment. This shows how important discussing different analysis methods is to avoid mistakes when comparing literature values rather than assuming that the nonlinear absorption parameters amount to the same quantity just because their name is similar.

After this discussion of popular phenomenological descriptions of OA Z-scan data with saturable absorption, it is interesting to see if a genuine saturation model can be used to describe the data and extract saturation parameters. Below, we shall demonstrate that this is possible for the example of a simple two-level saturation model.

4.4.2 A Physical Model for Saturable Absorption

The most basic physical system allowing for saturation consists of a transition between a ground state and an excited state, the population dynamics of which can be described with a rate equation model. This is a special case of a three-level system analysed for its saturation properties by Hercher [126], on which our derivation is based. Assuming an excitation rate $W = \sigma I$, where σ is the absorption cross-section, and a decay rate from the excited state back to the ground state of $A = \tau^{-1}$, with τ being the decay time, the system is described by the following equations

$$\dot{n}_1 = -W(n_1 - n_2) + n_2 A \quad \dot{n}_2 = -\dot{n}_1 \quad n_1 + n_2 = 1 \quad (4.23)$$

with the relative populations $n_{1,2} = N_{1,2}/N_{\text{total}}$. The steady state solution for $n \equiv n_1 - n_2$ is

$$n = \frac{1}{1 + \frac{2W}{A}} = \frac{1}{1 + 2\sigma I\tau} = \frac{1}{1 + \frac{I}{I_{\text{sat}}}} \quad (4.24)$$

with the saturation intensity $I_{\text{sat}} = (2\sigma\tau)^{-1}$. Using the fact that $\alpha(I) = \alpha_0 \cdot n$, we arrive at the simple two-level saturation model

$$\frac{dI}{dz'} = -\frac{\alpha_0}{1 + \frac{I}{I_{\text{sat}}}} I \quad (4.25)$$

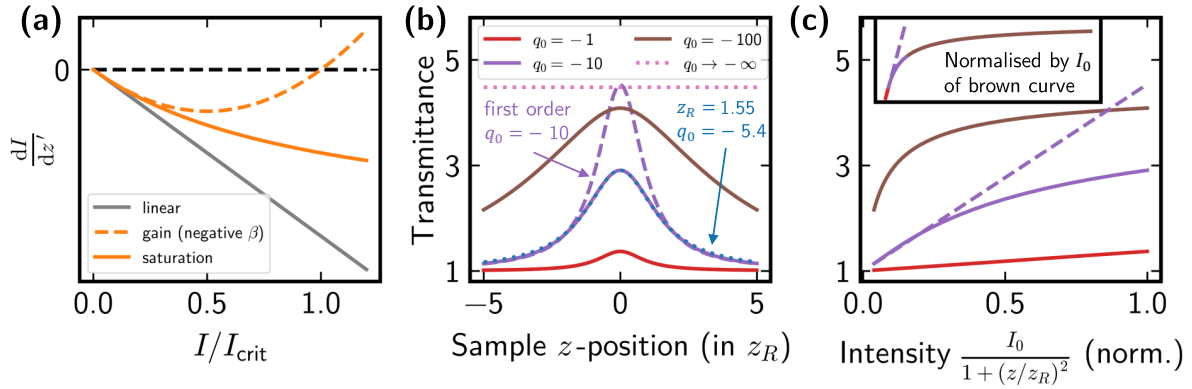


Figure 4.5 – Two-level saturation model for saturable absorption. (a) Comparison of $\frac{dI}{dz}$ as a function of the intensity for the two-level saturation model (solid orange) to the negative β_{TPA} approach (dashed orange) and linear absorption (grey). At low intensities, the negative β_{TPA} produces the same intensity decay as the two-level saturation model; however, the discrepancies grow as the intensity tends towards $I_{\text{crit}} = \alpha/\beta_{\text{TPA}}$. (b) Numerical simulations of the OA Z-scan traces corresponding to the two-level saturation model at varying intensities and otherwise constant parameters (solid lines). The traces broaden and approach the maximum achievable increase corresponding to the full saturation of the linear absorption in the limit of infinite intensity (dotted pink). As expected from (a), the first-order OA expression (dashed purple) agrees with the numerical simulation in the low-intensity regime away from the focus, but predicts too large increases of the transmittance at the focus. Using the first-order OA expression with z_R as a free parameter (dotted blue) seems to be in good agreement with the simulation results, but leads to a significant error in the estimation of q_0 . (c) The same data plotted as a function of the intensity, normalised to the maximum intensity of each curve, demonstrates the increasing saturation of the saturable absorption curves. It also shows that the first-order OA expression is a useful approximation in the low-intensity regime, where the transmittance change is approximately linear with intensity. In the inset, all curves are normalised by the peak intensity of the brown curve with $q_0 = -100$ to demonstrate that all curves coincide when plotted against an absolute intensity scale.

with a low-intensity absorption coefficient α_0 . Note that a three-level model would have led to the same differential equation, just with a more complicated I_{sat} depending on all three decay times instead of just one for the two-level system [126].

Figure 4.5(a) visualises the intensity gradient inside the sample along the propagation direction for the two-level saturation model (solid orange) compared to linear absorption (grey) and the negative β_{TPA} approach (dashed orange) discussed above. The saturation model possesses realistic properties, such as an absence of an unphysical gain regime as well as a monotonously falling intensity gradient as the input intensity is increased. This means that with increasing input intensity there is more and more bleaching, but never a situation

where the loss rate increases during propagation through the sample, which is one of the many problematic implications of the negative β_{TPA} approach. It is interesting to note that both curves are equivalent at low intensities, which is the same as saying that the bleaching according to the two-level saturation model at low intensities is equivalent to a hypothetical two-photon gain mechanism. The reason for this is that the first-order Taylor expansion of Eq.(4.25) is equal to the two-photon absorption model Eq.(4.1)b with the substitution $\beta_{\text{TPA}} \rightarrow -\alpha/I_{\text{sat}} < 0$, which explains why many OA Z-scan traces of saturable absorbers at low irradiances look like mirrored OA traces with two-photon absorption.

Example OA Z-scan traces of a sample described by the two-level saturation model are shown for increasing input intensities in Figure 4.5(b). For the sake of easy comparability with the figures above, each curve is characterised by the parameter $q_0 = -L_{\text{eff}} \alpha I_0 / I_{\text{sat}}$. For the parameters chosen in the simulation, $q_0 = -1$ (red) corresponds to $I_0 / I_{\text{sat}} \approx 1.29 \gtrsim 1$, meaning that the red curve marks the case where the exhaustion of the saturation capabilities of the sample starts to become relevant, which is also the characteristic intensity where the linear Taylor expansion of Eq.(4.25) ceases to be a good approximation of the saturation behaviour. The impact of this becomes clear for the case $q_0 = -10$ (purple), where the input intensity was increased by an order of magnitude. The OA trace is broadened, which is expected, as in a saturation process the reduction in α for a given increase in the input intensity should gradually reduce at higher input intensities – possibly with a residual non-saturable component due to excited-state absorption [126]. In other words, the transmittance increase due to the saturation process cannot continue at the same pace for arbitrarily high intensities, as the maximum achievable increase in transmittance $T_{\text{max}} = e^{\alpha L}$ is dictated by the linear absorptivity and the sample thickness. For ever increasing intensities, the OA Z-scan trace must keep broadening, as can be seen for the case $q_0 = -100$ (brown) until it reaches a constant at T_{max} in the limit $I \rightarrow \infty$ (dotted pink) where the sample has become transparent due to complete saturation of the transition at the pump wavelength. Of course, this limit is never reached in practice, as the sample would disintegrate due to excessive energy deposition much earlier.

Returning to the case $q_0 = -10$ (purple), the magnitude of the broadening can be appreciated in comparison with Eq.(4.19) for the same q_0 (dashed purple), corresponding to a first-order approximation in intensity space, as we have seen earlier. As the linear approximation in intensity space implies a constant growth rate of the transmittance with intensity it overestimates the strength of the effect. This discrepancy is strongest near the focus at $z = 0$ where the intensity is highest. For the parameters chosen in this example, the maximum transmittance value of the linear approximation for $q_0 = -10$ even slightly exceeds the maximum possible transmittance $T_{\max}$ indicated by the dotted pink line.

Finally, it is important to discuss the dangerous option of allowing for larger z_R values for fitting broadened OA data with Eq.(4.19), which is shown for the same example $q_0 = -10$ as the dotted blue curve. Of course, such a fit would be physically wrong as z_R is fixed by the focal length of the lens and the beam width, even though a very good agreement between the curves can be achieved. In our example the extracted q_0 and hence the estimate of I_{sat} is wrong by a factor of ~ 2 . This highlights the importance of correctly determining the Rayleigh range z_R and fixing that parameter for any fitting procedure. A concrete example for strongly broadened OA Z-scan data where this point applies can be found in Figure 5.8 in the next chapter.

Figure 4.5(c) shows the curves in Figure 4.5(b) as a function of the intensity, using the transformation $z \rightarrow I(z) = \frac{I_0}{1+\tilde{z}^2}$ with $\tilde{z} = z/z_R$, to support the arguments in the discussion above from another perspective. For low intensities (red), the increase in transmittance is almost linear as a function of the intensity; however, the transmittance curve starts levelling off more and more for the higher-intensity cases $q_0 = -10$ (purple) and $q_0 = -100$ (brown). Indeed, the linear approximation of $q_0 = -10$ (dashed purple) works well for the lower-intensity part of the curve corresponding to large values of $|z|$, but less so for high intensities corresponding to $z \approx 0$. If one normalises all intensities by the maximum intensity I_0 of the $q_0 = -100$ curve (brown) instead of normalising by I_0 of each individual curve, all datasets coincide on the same saturation curve, which can be seen in the inset of Figure 4.5(c). The linear approximation (dashed purple) does in fact have the same slope for all curves when

plotted on this absolute intensity scale and allows for the extraction of I_{sat} . Performing fits with z_R as a free parameter would lead to incorrect values in the transformation $z \rightarrow I(z)$, changing the curves that ought to be bent in intensity space into straight lines. Thus, an unsuspecting scientist pursuing this approach would be able to fit the data well, but would extract a seriously wrong value for I_{sat} and miss the crucial saturation of the transmittance curve in intensity space.

In this section we have investigated the analytical fitting of Z-scan data in detail, covering phenomenological approaches and others based on a genuine physical model. The OA traces for the two-level saturation model mark the end of this section, as only the first-order approximation curves according to Eq.(4.19) were analytical but the OA traces resulting from the differential equation Eq.(4.25) were simulated numerically. In the next section we shall discuss the inner workings of these simulations and the extended possibilities of data analysis they offer.

4.5 Numerical Simulations of OA Z-Scan Traces

One big limitation of the analytical approaches in the previous section is that they require simple differential equations governing the nonlinear dynamics of the intensity and phase throughout the sample in order to yield useful formulae for analysing Z-scan data. Already a model as simple as the two-level saturation model in Eq.(4.25) does not result in a simple closed-form expression for the OA trace and is best treated with a numerical simulation. In this section we will be focussing on the simulation of OA Z-scan traces, since these are needed for the data analysis in the following chapter, unlike simulations of CA traces which will be briefly discussed at the end of this section.

Figure 4.6 gives an overview of the implementation and a test of the numerical simulation of an OA Z-scan trace. The laser spot at the front surface of the sample $z' = 0$ needs to be discretised by sampling the TEM_{00} Gaussian profile in a regular two-dimensional grid, as shown in Figure 4.6(a). It is important to set the limits of the grid such that the intensity

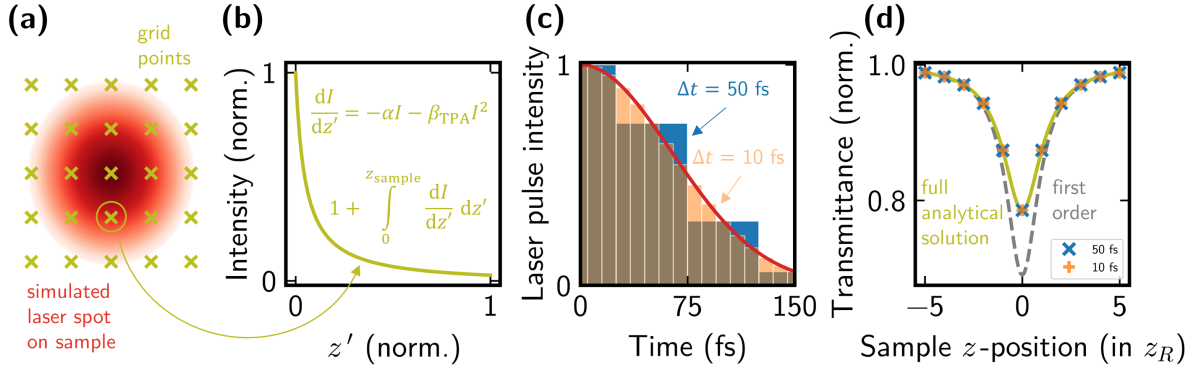


Figure 4.6 – Implementation of the numerical OA Z-scan simulations. (a) The Gaussian beam profile on the front surface of the sample is discretised as a two-dimensional grid of intensity values. (b) At each grid point, the differential equation governing the intensity propagation through the sample is solved, shown for the example of two-photon absorption. (c) In order to correctly model the nonlinear response of the sample in time, the input pulses are sampled with a finite time step. (d) The OA Z-scan simulations are tested for the case of two-photon absorption, for which the analytical solution was derived above. Even a relatively large time step of 50 fs (blue) leads to a very good agreement with the analytical curve (yellow) and a further simulation with a 10 fs time step (orange) confirms that the simulation has converged.

distribution is essentially not clipped for the largest spot size for the maximum value of $|z|$ in the simulation. A good value for the spacing between the grid points can then be found in a convergence test. The differential equation used for modelling the intensity distribution inside the sample has to be solved at each grid point, which is visualised in Figure 4.6(b) for the example of two-photon absorption.

Summing up the output at all grid points for each value of z and dividing the resulting curve by one obtained in a simulation where $\frac{dI}{dz} = -\alpha_0 I$, is all that is needed for simulating OA Z-scan traces under CW illumination conditions. However, for pulsed illumination, the correct time average needs to be implemented. Figure 4.6(c) shows how a pulse with a Gaussian time dependence and a FWHM of 150 fs, which is the case for the experiments in the following chapter, is sampled in time with timesteps of 50 fs (blue) and 10 fs. The time average is taken at each sample position z by averaging the output behind the sample over all input intensity values $I(t_i)$.

A test of the simulations can be seen in Figure 4.6(d). Two-photon absorption was chosen because in that case, the OA Z-scan problem with temporally Gaussian laser pulses possesses

a known analytical solution, Eq.(4.17), which we have re-derived in Section 4.3. Comparing the simulated data points to the full analytical solution (yellow) shows that even a step size of 50 fs for sampling the pulse in time produces a very accurate prediction of the OA data (blue crosses). Repeating the simulation with a step size of 10 fs leads to no appreciable change, which indicates that for this model and pulse duration, 50 fs is a sufficiently fine step size and the simulation has converged with respect to the time step.

Having demonstrated that the simulation does work as expected greatly expands our analysis capabilities, as we are now in a position to generate OA Z-scan traces for any nonlinear absorption behaviour that can be well described by a one-dimensional differential equation. As an example of a complicated nonlinear absorption behaviour, which we would not have been able to analyse without the numerical simulation code, let us discuss an OA dataset from a 2D perovskite flake about 160 meV above its fundamental excitonic resonance. The sample and the physics behind its nonlinear properties will be discussed in detail in the following chapter. As can be seen in Figure 4.7, the OA trace in question exhibits a positive transmittance change in the high-intensity region around the focus, but a negative transmittance change in the low-intensity region away from it. Therefore, the differential equation for the intensity propagation must be such that the optical loss is lower than for the case of just linear absorption at high intensities, but higher at low intensities. One possibility to achieve this is visualised in Figure 4.7(a), where the differential equation for the two-level saturation model Eq.(4.25) was combined with two-photon absorption. The two-photon absorption term was assumed to be saturable, with a saturation behaviour that was previously reported in the literature [128], leading to the combined model

$$\frac{dI}{dz'} = -\frac{\alpha_0}{1 + \frac{I}{I_{\text{sat}}}}I - \frac{\beta_{\text{TPA}}}{1 + \frac{I}{I_{\text{sat},\beta}}}I^2 \quad (4.26)$$

with three free parameters β_{TPA} , I_{sat} , and $I_{\text{sat},\beta}$.

The ability to numerically simulate OA Z-scan traces corresponding to this model allows for an analysis of the data in Figure 4.7(b) and moreover for a direct test of alternative analysis

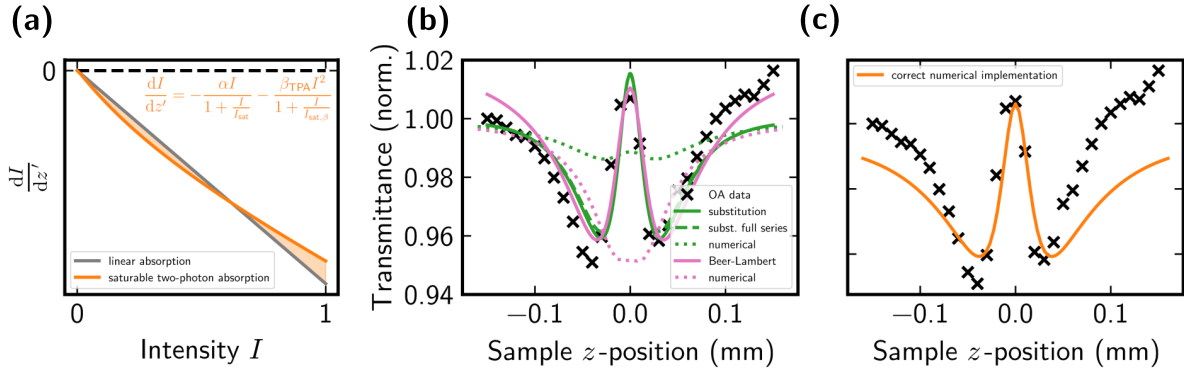


Figure 4.7 – Application of the numerical simulation to complicated OA data. (a) Analysing complicated OA Z-scan data, where the transmittance is increased near the focus, but decreased away from the focus, requires a more complex model than the two-level absorption model presented above. One option is to introduce an additional saturation parameter, such that the two-photon absorption and linear absorption saturate independently. This makes it possible to describe situations where, compared to linear absorption (grey), the absorption is reduced at high intensities, but increased at low intensities (orange). (b) Example of OA Z-scan data (black), where this extended model can be applied. A simple substitution approach [127] with the first-order truncation (solid green) and full series (dashed green) OA formula Eq.(4.17) seems to agree well with the data, but produces unrealistic parameter values and does not agree with a numerical simulation with the same model and parameters (dotted green). The same is the case for an approach with Beer-Lambert's law (pink). (c) A correct implementation of the model with the numerical simulation framework described above allows for the extraction of realistic parameter values. The discrepancies between the data and the model away from the focus suggest that the saturation behaviour is even more complicated than the model predicts.

methods for similarly complicated OA data with both increased and reduced transmittance in the same dataset, which we shall examine first. The first method is a seemingly straightforward substitution approach used by Lee and co-workers [127], where the saturation model is directly inserted into Eq.(4.16). Adapting this to our model and using the pulsed OA formula Eq.(4.17), given that the data was collected in a pulsed experiment, we obtain

$$T_{\text{OA}} = 1 - \frac{\beta_{\text{TPA}} I_0 L_{\text{eff}}}{2\sqrt{2}}, \quad \beta_{\text{TPA}} \rightarrow \alpha(I) - \alpha_0 = \frac{\alpha_0}{1 + \frac{I}{I_{\text{sat}}}} + \frac{\beta_{\text{TPA}}}{1 + \frac{I}{I_{\text{sat},\beta}}} I - \alpha_0 \quad (4.27)$$

which can be fitted to the data well (solid green), however, this good agreement should not distract one from the fact that this equation is not dimensionally correct, resulting in a meaningless parameter value for L_{eff} . In fact, many seemingly good fits to the data could be achieved for a broad range of values for L_{eff} and we arbitrarily chose $L_{\text{eff}} = 1 \cdot 10^{-20}$ near

the centre of that range. Note that we did not add a unit to this value because it clearly should not be interpreted as an effective length L_{eff} on the order of 10^{-20} m. As we have seen in Section 4.3, L_{eff} has a precise meaning and is uniquely determined by the physical length and the linear absorption of the sample; in other words, it should not be treated as an arbitrary fitting parameter. In the case of the data in Figure 4.7(b), $L_{\text{eff}} = 47$ nm, but fixing this parameter in the fit led to artefacts due to numerical cancellation, which is why a different value was chosen. Note, however, that the parameters of interest, namely the saturation parameters $I_{\text{sat}} = 28$ TW/m² and $I_{\text{sat},\beta} = 24$ TW/m², and the two-photon absorption coefficient $\beta_{\text{TPA}} = 35.8$ cm/MW appear reasonable. It might seem unnecessary to include this incorrect approach into our discussion; however, because of the problematic combination of a good fit to the data, reasonable saturation parameters and the popularity of this approach in the literature, it is in fact worthwhile to explicitly show that it is not a valid way of applying the model in question. Performing a numerical simulation with the obtained parameter values (dotted green) demonstrates that the solid green curve does not in fact represent the model Eq.(4.26) for the obtained parameter values. Note that performing the substitution in Eq.(4.27) with the full series expression Eq.(4.17), indicated by the dashed green line in Figure 4.7(b) does not alter the situation.

Similarly, we can now show that an alternative strategy using Beer-Lambert's law (pink) leads to incorrect parameter values as well. This approach is included in the discussion to demonstrate that even if one tried to save the substitution approach by avoiding the dimensionally incorrect formula Eq.(4.27) above and instead would attempt to modify Beer-Lambert's law, which comes to mind as a seemingly logical alternative, the correct model parameters could still not be obtained. Essentially, we would like to justify our claim that there is no simple analytical formula available and that seemingly reasonable simple approaches attempting this are incorrect. The only valid way known to the author of avoiding a numerical simulation in these circumstances is based on an Adomian decomposition method, but has not enjoyed much popularity in the literature [122]. Returning to the Beer-Lambert approach, using the

equation

$$T_{OA} = \frac{\exp\left(-\left(\frac{\alpha_0}{1+\frac{I(z)}{I_{sat}}} + \frac{\beta_{TPA} I(z)}{1+\frac{I(z)}{I_{sat,\beta}}}\right)L\right)}{\exp(-\alpha_0 L)} \quad (4.28)$$

yields the saturation parameters $I_{sat} = 199 \text{ TW/m}^2$ and $I_{sat,\beta} = 61 \text{ TW/m}^2$, and the two-photon absorption coefficient $\beta_{TPA} = 10.4 \text{ cm/MW}$. A numerical simulation with the model Eq.(4.26) and the obtained parameters (dotted pink) leads to a different OA trace and does not reproduce the central peak in the data. The problems with the Beer-Lambert approach are that it does not take into account the spatio-temporal averaging over the Gaussian intensity distribution and that it assumes an exponential intensity decay inside the sample, even though the model implies a more complicated intensity loss as a function of z' . However, for an experiment with a CW top-hat beam and an intensity decay inside the sample, which does not significantly deviate from an exponential function, this approach may be a good approximation.

Finally, a correct application of Eq.(4.26) using the numerical simulation is shown in Figure 4.7(c). The obtained parameters $I_{sat} = 180 \text{ TW/m}^2$ and $I_{sat,\beta} = 11 \text{ TW/m}^2$, and $\beta_{TPA} = 25 \text{ cm/MW}$ are not far away from those found with the incorrect methods above; however, in other situations, the discrepancies might have been larger. It was important to perform this test, as it is not a priori clear by how much the values obtained with the incorrect methods would deviate from those estimated with a correct application of the model. Here, the main difference lies in the value of the saturation intensity $I_{sat,\beta}$ which is at least two times smaller than the values obtained with the incorrect methods. This explains why the central peak in the data, which is due to a strong saturation of the two-photon absorption – associated with a low corresponding saturation intensity – cannot be reproduced with the incorrect parameter values in the numerical simulations (dotted lines in Figure 4.7(b)). Of course, the simulated OA trace in Figure 4.7(c) does not perfectly agree with the data either, because it predicts too large an effect for the low-intensity data points away from the focus, but the model is correctly implemented and the orange curve in Figure 4.7(c) does actually correspond to the model Eq.(4.26) and the parameter values stated above. The

reason for the imperfect agreement might be that the saturation of two-photon absorption should actually increase much faster as a function of the intensity, which could be achieved by substituting $\frac{I(z)}{I_{\text{sat},\beta}} \rightarrow \left(\frac{I(z)}{I_{\text{sat},\beta}}\right)^2$ in Eq.(4.26) as in the article by Wang and colleagues [129]. On the other hand, in another study [128] it was found that two-photon absorption in all samples investigated saturated linearly, as in Eq.(4.26) and not quadratically, as it would after the substitution. We chose to go with the linear saturation behaviour that was observed in a well-conducted and correctly implemented study [128] rather than applying an ad-hoc modification without further justification other than a potential better agreement with the data. The model Eq.(4.26) may not be perfect for describing the physics inside the sample; however, it can explain the key features of the data and is preferable to using a modified version, unless further experimental evidence and first-principles theoretical modelling yield compelling reasons to do so.

At the end of this section, let us briefly comment on numerical simulations of CA Z-scan traces. Given that the CA data presented in the next chapter could be analysed well with the division method explained in Section 4.3, there was no need to implement such a simulation. Moreover, the implementation is much more complex, as the nonlinear phase shift inside the sample needs to be taken into account and the resulting phase and intensity distribution at the back surface of the sample needs to be propagated to the detector plane. If such an analysis tool is needed for future work, one approach is to implement the Fresnel-Kirchhoff diffraction integral with a FFT-based beam propagation method, which was successfully applied in an article by Alam and colleagues [41].

4.6 Conclusion

In this chapter, we have explained the Z-scan technique and discussed the aspects of the data analysis needed for the measurements presented in the following chapter. We particularly focussed on the correct interpretation of OA Z-scan traces, particularly in the case of saturable absorption. The situation in the literature is complicated and many methods and naming

conventions exist, which can easily lead to confusion and avoidable mistakes in the analysis. At the end, we presented a numerical framework to simulate OA Z-scan traces for arbitrary differential equations governing the intensity evolution inside the sample, which allows for straightforward testing of various models and analysis methods in the literature.

Z-scan experiments are extremely easy to conduct and require little experimental skill and resources, which is in stark contrast to the complexity of the theory and the many pitfalls in the data analysis. Therefore, it is crucial not to blindly follow seemingly simple analysis methods in the literature, as quite a number of them are at least partially flawed and do not really explain the data with the stated models. In other words, such approaches are misleading because on a surface level, they imply a good understanding of the physics in a given sample, when, in fact, this is not the case. Instead, a deeper understanding should be developed first, as we have attempted in this chapter, in order to achieve a meaningful analysis.

Chapter 5

2D Ruddlesden-Popper Phase Perovskites

This chapter explores the nonlinear optical characterisation of 2D Ruddlesden-Popper phase perovskites (2D RPPs) and a hybrid structure consisting of a 2D RPP on Al nanoantennae. The 2D RPPs under study belong to the family of lead halide perovskites, which, due to exceptional light harvesting properties, have been extensively investigated in the context of solar cell research [130–133]. However, 2D RPPs are also of interest in nonlinear optics because of their multilayer quantum well (MQW) structure, where strongly bound excitonic states give rise to a resonantly enhanced optical nonlinearity [134,135]. After a more detailed explanation of the chemical composition and structure of 2D RPPs, their linear and nonlinear optical properties will be discussed, followed by a comparison to the aforementioned hybrid perovskite-antenna system. Note that sections 5.1 and particularly 5.2 are based on previously published work [135]. This project was a collaboration with Dr Ibrahim Abdelwahab from the group of Prof Dr Kian Ping Loh at the National University of Singapore. All individual contributions are acknowledged with footnotes at the corresponding figures.

5.1 Chemical Composition and Crystal Structure

5.1.1 Lead-Halide Perovskites

Crystals are said to be of Perovskite structure when their composition is of the form ABX_3 , where A is a large cation, B is a small cation, X is an anion, and the spatial arrangement is as depicted in Figure 5.1. Each of the large cations A is at the centre of a cubic structure, with cations B in the corners and anions X in the centres of each edge, such that each cation B is surrounded by 6 anions X in a corner-sharing octahedral configuration [133]. In the case of the lead halide perovskites that revolutionised solar cell research and which are at the heart of this chapter, the large cation A is methylammonium $CH_3NH_3^+$, cation B is lead Pb^{2+} and the anion X is one of the halides Br^- or I^- [130]. Note that perovskites of such a composition are often called hybrid organic-inorganic perovskites [136] due to the organic methylammonium (MA) and the inorganic lead and halide ions; however, this is only of secondary importance for the investigation of this chapter and the word 'hybrid' will only be used to refer to samples where perovskites are combined with Al nanoantennae to avoid confusion.

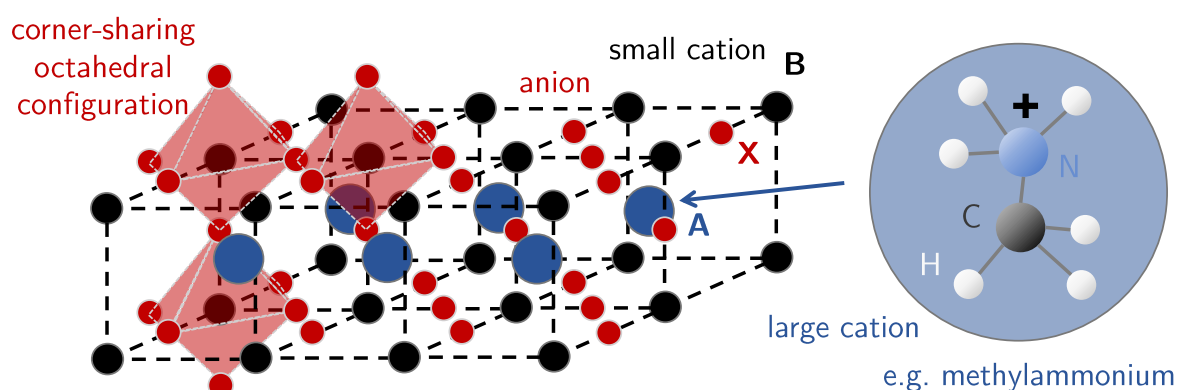


Figure 5.1 – Generic structure of a Perovskite crystal. Large cations A and small cations B are arranged in interdigitating cubic lattices, such that each cation of one species is in the centre of the cube formed by the other species. Anions X are centred on the edges of the cubic lattice of species B, such that they assume a corner-sharing octahedral configuration [133]. For lead-halide perovskites, which are investigated in this chapter, there is a Pb^{2+} cation on the B sites, an I^- anion on the X sites, and methylammonium (MA) is the large cation on the A sites [130].

A key problem of MAPbX_3 perovskites is their humidity instability [137]. In order to address this problem, which prevents any serious scalable use in optical devices, dimensionality reduction from 3D MAPbX_3 crystals to 2D layered structures has been explored [137, 138], based on the idea of increasing hydrophobicity by separating the perovskite layers with longer organic cations [139]. For a given 3D perovskite crystal, there are many possible ways of obtaining 2D layered structures [140], differing in the crystal plane of the slicing, the 2D layer thickness, and the choice of spacer cation between the negatively charged perovskite layers [141]. This allows one to tune electrical and optical properties with much greater flexibility than in 3D crystals, where for instance the choice of A cations is severely limited [142] due to size requirements [143] for the stability of the crystal structure. Arguably, for fundamental research purposes this drastically increased tuneability [144] is an even greater advantage of the dimensionality reduction than the increased moisture stability.

5.1.2 Structure and Fabrication of 2D RPP Flakes

One family of 2D structures derived from MAPbX_3 crystals is that of Ruddlesden-Popper-phase perovskites (2D RPPs) with butylammonium (BA) serving as organic spacer cations [138, 145]. A schematic overview of this homologous series of 2D RPPs with the chemical formula $(\text{BA})_2(\text{MA})_{n-1}\text{Pb}_n\text{I}_{3n+1}$, $n = 1, \dots, 4$ abbreviated as $\text{I}_{n=1/2/3/4}$ is depicted in Figure 5.2(a). Characteristically for this structure family, organic BA bilayers alternate with organic-inorganic 2D perovskite sheets of varying thickness, depending on the concentration of the initial reagents [145]. For the 2D RPP samples investigated in this chapter, the simple and inexpensive wet chemistry synthesis process was adapted from Ref. [145] with slight alterations [134]. The appropriate stoichiometric amounts of PbO_2 , BAI , and MAI for the desired crystal were made to react with HI and H_3PO_2 , which is visualised schematically in Figure 5.2(b). The choice of BA for our 2D RPPs is due to the fact that the fabrication process is established and reliable [134], but spacer molecules with longer alkyl groups would have been equally suitable for separating the perovskite layers. For the experiment described in Section 5.2, thin flakes of around 100 nm thickness were exfoliated from the 2D RPPs

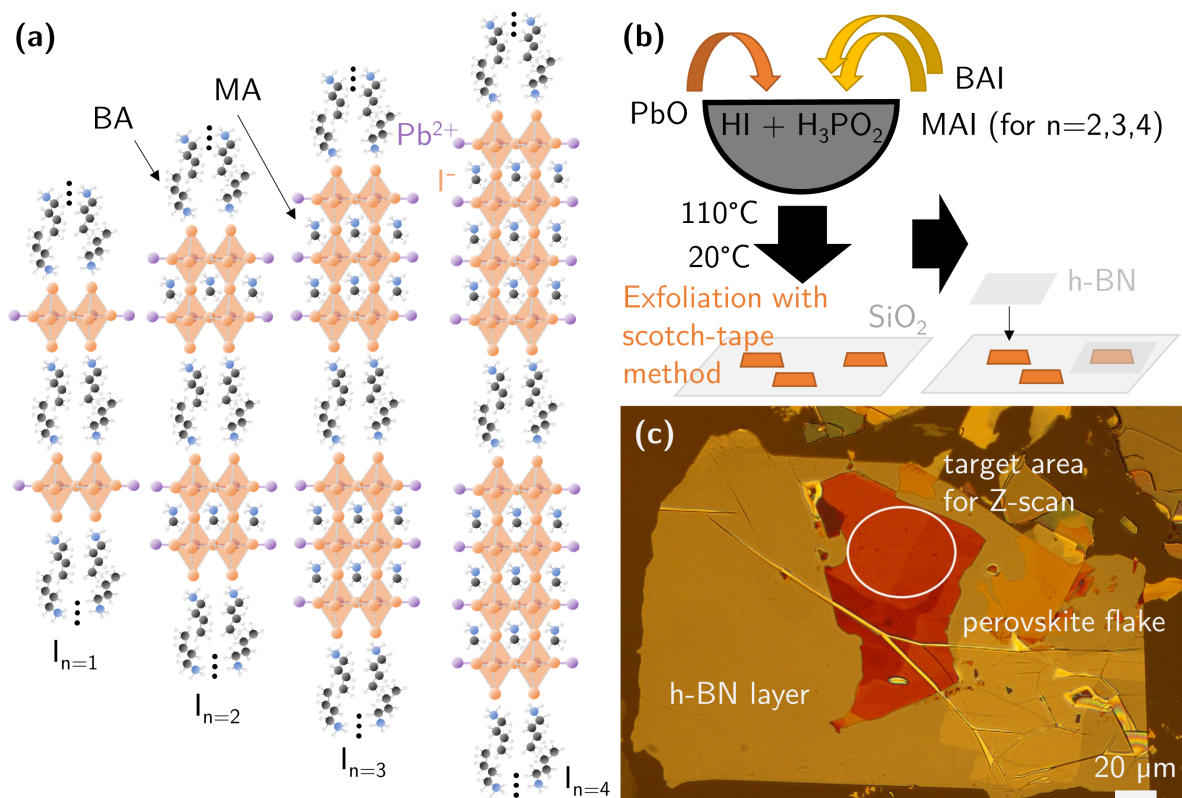


Figure 5.2 – Structure, fabrication, and image of perovskite flakes.[†] (a) Visualisation of the 2D RPP structure in the homologous series $(\text{BA})_2(\text{MA})_{n-1}\text{Pb}_n\text{X}_{3n+1}$, $n = 1, \dots, 4$. Re-drawn similar to Ref. [134]. (b) Fabrication diagram of the 2D RPP flakes. The correct stoichiometric amounts of PbO, BAI, and for $n=2,3,4$ MAI, are made to react with HI and H_3PO_2 and heated up to 110°C . After cooling down to room temperature, thin flakes of about 100 nm thickness are exfoliated from the resulting crystals and transferred onto a SiO_2 substrate. In order to achieve a stable performance over several months, the highest-quality flakes on our samples are covered with an h-BN layer using a stamp transfer method [134]. (c) Microscope image of a perovskite flake on the sample covered with a thin h-BN layer. A large homogeneous region, as indicated by the white circle, allows for Z-scan measurements with a good data quality.

with a scotch tape method and encapsulated with an hexagonal boron nitride (h-BN) layer of several nm thickness to prevent degradation of the flakes when in contact with air [134]. Despite the improved moisture stability mentioned above, this step was necessary to ensure stability over several months. A typical picture of such encapsulated flakes with a diameter of about $100\ \mu\text{m}$ is shown in Figure 5.2(c). A homogeneous region of several tens of μm in diameter, as indicated by the white circle, is necessary for a high data quality in the following Z-scan measurements. In terms of the sample stability it was also necessary to work well

[†]The author acknowledges Dr Ibrahim Abdelwahab for the fabrication of the samples and for taking the optical microscopy image in Figure 5.2(c).

below the laser damage threshold of the flakes to avoid degradation due to local heating by the laser beam.

5.1.3 Mechanisms Affecting Exciton States and Optical Properties

In preparation for the results in the following section, let us now focus on the physical properties that will be needed to explain the linear and nonlinear optical response of the sample. The organic spacer layers in the 2D RPPs have a much smaller dielectric constant and a larger bandgap than the perovskite layers [146], leading to a self-assembled multiple-quantum-well (MQW) structure [147]. This allows for extremely strongly bound room-temperature-stable excitons with binding energies of several $100 \text{ meV} \gg k_B T$, in contrast to excitons in epitaxially grown semiconductor heterostructures of a similar size, where excitonic absorption is still visible at room temperature but the binding energy is at least an order of magnitude smaller [148]. The reason for this is that the latter excitons only benefit from quantum confinement, which increases their binding energy in 2D quantum wells by up to a factor of ~ 4 , while the 2D RPP layer stack with its large refractive index contrast increases the binding energies much more strongly due to the so-called dielectric confinement effect [149–151]. This effect is a consequence of the electron and hole wavefunctions extending into the low-index regions outside the well, where there is less screening of their Coulomb interaction, leading to a more strongly bound exciton [147]. One implication of this is that the binding energy is expected to decrease with increasing perovskite layer thickness, which has been observed in other studies [134, 151] and is also the case for the samples used in this work. Not only the exciton binding energy, which is the difference in energy between the excitonic ground state and the continuum edge, but also the continuum edge itself depends on the perovskite layer thickness [151], leading to an absorption edge varying from the UV to the red part of the spectrum. One explanation of this behaviour is the dimensionality reduction from 3D to 2D, where the electronic states become more and more localised normal to the perovskite layer, leading to a band structure with narrower bands and a larger bandgap [152]. As an extension of this dimensionality reduction argument, it was pointed out [153] that the

halide anions at the boundary of each inorganic layer interact much more strongly with a lead cation on the inside than with the BA cations on the organic outside. In this picture, the larger difference in electronegativity between the halide anions and the BA cations gives rise to a larger bandgap [153]. While this relatively simple argument explains the increasing bandgap for 2D RPPs well, this is not true for general low-dimensional perovskite structures. For an accurate prediction of bandgap energies in those cases, more subtle effects such as the coordination angles and distances in the lead halide octahedra and in- and out-of-plane distortions have to be taken into account [144].

5.2 Linear and Nonlinear Optical Properties

Let us now return to the $I_{n=1/2/3/4}$ RPP samples under study, an example of which was shown in Figure 5.2(c). Before conducting the Z-scan measurements, a linear optical characterisation of the flakes was carried out. Figure 5.3(a),(b) shows the real and imaginary part of the complex dielectric function ε , which was calculated according to $\Re(\varepsilon) = n^2 - k^2$ and $\Im(\varepsilon) = 2nk$ from complex refractive index data $n + ik$ from ellipsometry measurements. Note that the flakes exhibit a layered structure, as we have seen above, and are hence anisotropic, but here an effective isotropic index model was chosen, which allows for a straightforward identification of the excitonic resonances. Each of the RPPs exhibits a strong resonant feature with the characteristic trough-peak signature in the real part and a Lorentzian peak in the imaginary part. The resonance wavelength is lowest for the $I_{n=1}$ flake at $\lambda = 515$ nm and then gradually shifts to longer wavelengths as the thickness of the perovskite layers is increased.

As explained in the previous section, the main reasons for this behaviour are the strong dielectric confinement of the excitons in the perovskite layer, which increases their binding energy, and the shift of the continuum edge due to dimensionality reduction [151–153]. Note that these effects work in different directions, as the increasing bandgap blue-shifts the excitonic resonance, while a stronger exciton binding energy results in a red-shift. Overall,

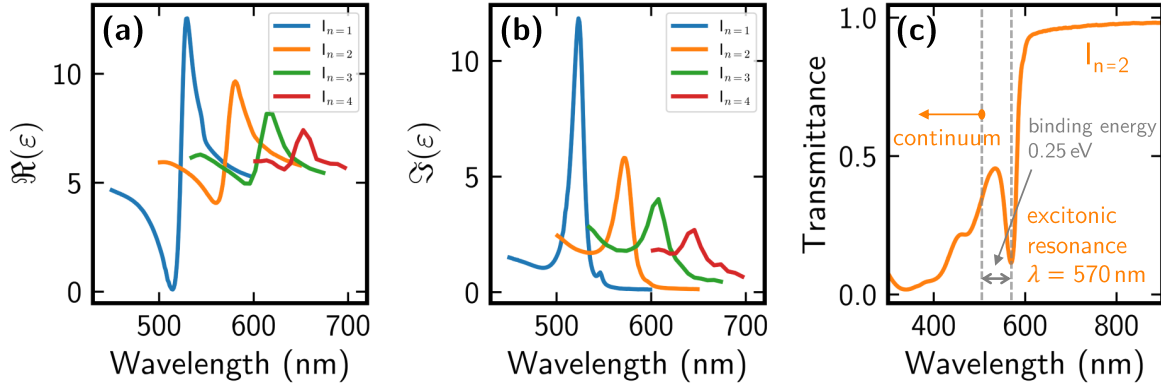


Figure 5.3 – Linear optical characterisation of 2D RPPs.[†] (a) Real part of the dielectric function ϵ for the 2D RPPs in the homologous series $(\text{BA})_2(\text{MA})_{n-1}\text{Pb}_n\text{X}_{3n+1}$, $n = 1, \dots, 4$. (b) Imaginary part of the dielectric function for the same 2D RPPs. The curves were obtained from ellipsometry data of the complex refractive index $n + ik$ via $\Re(\epsilon) = n^2 - k^2$ and $\Im(\epsilon) = 2nk$. (a),(b) re-drawn from Ref. [135]. (c) Transmittance of the 2D RPP flake with $I_{n=2}$, which will be discussed in more detail in this section. The strong dip at $\lambda = 570$ nm is attributed to the excitonic ground state, which has a binding energy of 0.25 eV [151]. The transmittance is low for energies above the continuum edge due to the excitation of electrons into the conduction band.

the shift of the continuum edge dominates, and for the thinnest perovskite layers in the $I_{n=1}$ configuration the exciton ground state experiences the strongest blue-shift, which gradually weakens as the layer thickness increases. The limiting case for $n \rightarrow \infty$ is the 3D perovskite MAPbI_3 with a fundamental exciton peak at $\lambda = 753$ nm, or 1.65 eV [154].

Figure 5.3(c) depicts a linear transmittance measurement of a 2D RPP flake with $I_{n=2}$. The key feature is the sharp dip at $\lambda = 570$ nm, which is due to the fundamental exciton resonance. Given that the binding energy for $I_{n=2}$ is 0.25 eV [151], the electronic bandgap, above which the absorption into the continuum of the conduction band starts, can be identified as $\lambda \sim 512$ nm. For energies below the exciton ground state, there is almost no linear absorption, but as we shall see later, there may be significant two-photon absorption. Given the much smaller real and imaginary parts of ϵ at the excitonic resonance of the $I_{n=1}$ flake, choosing $I_{n=2}$ instead of $I_{n=1}$ is perhaps somewhat surprising. The reason for this choice is that the results of the nonlinear characterisation with the analysis described below yielded the highest optical nonlinearity for $I_{n=2}$. It was also found that there is no qualitative difference

[†]The author acknowledges Xiao Chi for conducting the ellipsometry measurements on which the curves in Figure 5.3(a),(b) are based and Dr Ibrahim Abdelwahab for performing the transmittance measurements in Figure 5.3(c).

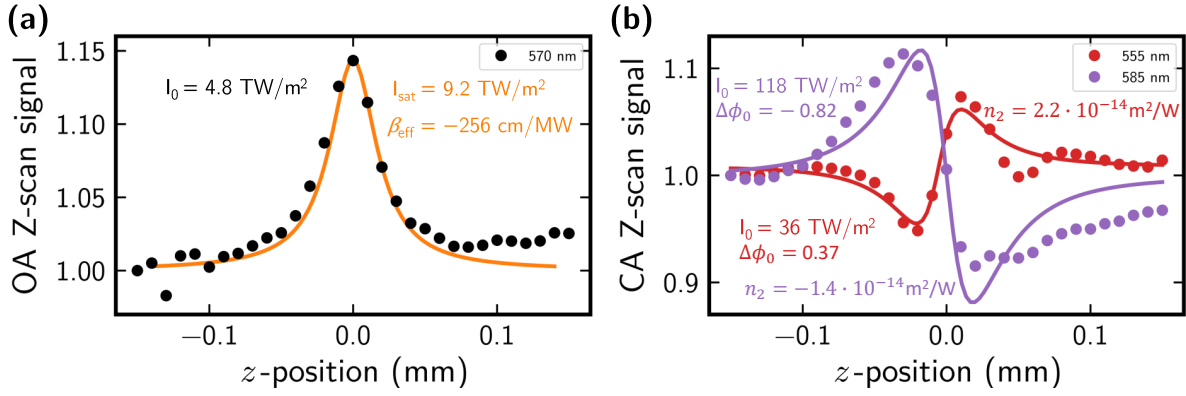


Figure 5.4 – Z-scan measurements on an $I_{n=2}$ flake around its excitonic resonance. (a) Open-aperture (OA) Z-scan on resonance demonstrating strong saturable absorption. The orange curve was obtained via a saturation model at $\beta_{\text{eff}} = -\alpha/I_{\text{sat}} = -256 \text{ cm/MW}$. (b) Closed-aperture (CA) Z-scan above (red) and below (purple) the excitonic resonance. The CA traces switch from a self-focussing trough-peak to a self-defocussing peak-trough signature, which demonstrates that there is a sign change of the nonlinear refractive index n_2 at resonance. Note that the CA data was normalised by the OA data for each scan to correct for nonlinear absorption. This figure was re-drawn from Ref. [135].

in the nonlinear behaviour of the flakes $I_{n=1/2/3/4}$ around their excitonic resonances, which can be seen in the Supplementary Information in the published version of this work [135]. Hence, we shall only discuss $I_{n=2}$ as a representative example in the remainder of this section.

At the outset of the nonlinear optical characterisation, let us discuss Z-scan measurements at and around the excitonic resonance of an $I_{n=2}$ flake, where the strongest nonlinearity is expected due to the resonant enhancement. A single laser beam with an approximately Gaussian profile, a pulse duration of 150 fs, and a repetition rate of 1 kHz was used to probe the sample. More details on the Z-scan setup can be found in the Appendix A. The datasets are shown in Figure 5.4. In order to find the maximum value of the effective nonlinear absorption coefficient β_{eff} , an open-aperture (OA) Z-scan was taken on resonance at $\lambda = 570 \text{ nm}$, which is shown in Figure 5.4(a). The increased transmittance at the high-intensity data points around the focus indicates saturable absorption, which was quantified using a numerical simulation with the simple two-level saturation model Eq.(4.25) discussed in the previous chapter. The simulated OA trace with the given sample thickness, peak intensity, and the complex refractive index as fixed inputs is shown as the orange curve, yielding a saturation parameter of $I_{\text{sat}} = 9.2 \text{ TW/m}^2$. In order to compare this result to the literature

in a straightforward way, the effective nonlinear absorption coefficient was calculated as $\beta_{\text{eff}} = -\alpha/I_{\text{sat}} = -256 \text{ cm/MW}$.

The analysis was conducted in this way for two reasons. Firstly, the saturation model is linked to a somewhat realistic, if crude, modelled behaviour of the light intensity throughout the sample, contrary to a fit with a negative two-photon absorption coefficient β_{TPA} , which would imply unphysical gain, as explained in Section 4.4 in the previous chapter. Secondly, even with a sensible interpretation of the negative β_{TPA} method, also explained in Section 4.4, not all OA data could be consistently fitted, particularly with our Rayleigh range of $z_R = 0.02 \text{ mm}$ as a fixed parameter.

Our on-resonance β_{eff} value of -256 cm/MW is very large compared to many other values reported in the literature, such as $-8 \text{ cm/MW}^\dagger$ for an ITO sample in its epsilon-near-zero region [125]. In terms of its absolute value, it is 5 orders of magnitude larger than the two-photon absorption coefficient of Si with about 2 cm/GW at $\lambda \sim 800 \text{ nm}$ [155]. However, larger values of β_{eff} have been published, for example $\beta_{\text{eff}} = -1620 \text{ cm/MW}$ in thin films of MAPbI_3 at 532 nm , albeit with a much longer pulse duration of 40 ns and a different analysis method and definition of β_{eff} [156]. Another example of a larger value was reported in a polycrystalline film of a 1D EAPbI_3 perovskite which was found to have a nonlinear absorption coefficient β of up to -310 cm/MW , however, again obtained with a different analysis method [157].

It cannot be emphasised enough that all these examples from the literature, and including our own work, use different procedures and sometimes slightly different definitions to arrive at a final estimate for a nonlinear absorption coefficient. This is indicative of the problematic situation faced by many researchers, where a complex saturation behaviour is to be quantified with a single number for easy comparability. We have discussed the most common approaches in detail in the previous chapter, highlighting the differences between them, as well as the issues with some of them. Here, the various values of β and β_{eff} were quoted despite these

[†]Due to the different definition of β_{eff} in that work, the value needs to be multiplied by ~ 3 before comparing it to our value, as explained in Section 4.4.

differences in approach, as they are all related to a change of the transmittance per unit sample thickness and incident intensity, and are thus at least still useful for a crude order-of-magnitude estimate of the relative strength of the nonlinear absorption or saturation in a particular sample.

Let us now consider the refractive nonlinearity in our $I_{n=2}$ flake around its excitonic resonance. The datasets shown in Figure 5.4(b) were obtained by taking closed-aperture (CA) Z-scans and dividing the data by the corresponding OA scans to remove the nonlinear absorption and isolate the nonlinear refraction, as described by Sheik-Bahae and colleagues [58]. In practical terms, this was achieved by integrating only over sufficiently few pixels in the centre of our CMOS camera to achieve the effect of an aperture with a transmittance of $S = 0.02$ for the CA dataset, while integrating over all pixels to obtain the OA dataset. Each dataset was normalised by a low-intensity data point, for which nonlinear effects are negligible, before dividing the CA data by the OA data. The CA data taken above resonance (red) exhibits a trough-peak signature, which is characteristic of self-focussing. The opposite is the case for the CA data taken below resonance (purple), where a pre-focal transmittance peak is followed by a post-focal trough of the transmittance, indicating that the behaviour has changed to self-defocussing. Note that in the immediate vicinity of the resonance at $\lambda = 570$ nm the CA data was dominated by nonlinear absorption and no n_2 value was extracted, as is explained below in the discussion of Figure 5.5. In terms of the absolute magnitude of the nonlinear refractive index n_2 , our values are again larger than many of those in the literature. Returning to the aforementioned study of an ITO layer in its epsilon-near-zero region [125] with a maximum reported value of $n_2 = 1.1 \cdot 10^{-14} \text{ m}^2/\text{W}$, it becomes apparent that our perovskite flake can compete with a high-nonlinearity state-of-the-art sample and even slightly surpasses its nonlinear refractive index. Note that the maximum and minimum values $n_2 \sim \pm 10^{-13} \text{ m}^2/\text{W}$ in Figure 5.5(b) are even larger. Comparing our n_2 values to those found in high- n_2 chalcogenide glasses, such as As_2Se_3 with $n_2 = 1.3 \cdot 10^{-17} \text{ m}^2/\text{W}$ at $\lambda = 1500$ nm, or just silica with $n_2 = 2.6 \cdot 10^{-20} \text{ m}^2/\text{W}$ [158] further demonstrates the strength of the nonlinear refraction.

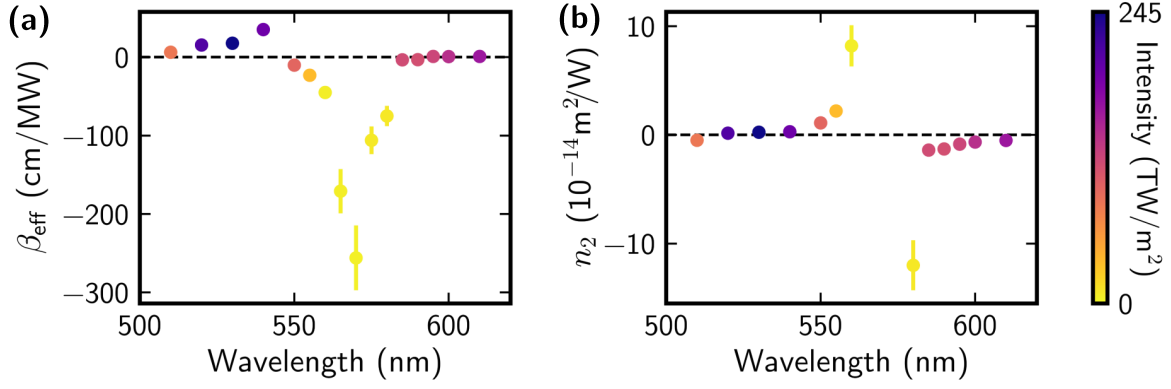


Figure 5.5 – Nonlinear parameters of the $I_{n=2}$ flake as a function of the wavelength. (a) The nonlinear absorption parameter β_{eff} is strongly enhanced at the excitonic resonance. Moreover, there is a sign change, indicating the transition from two-photon absorption to saturable absorption. (b) The nonlinear refractive index n_2 undergoes a sign change from self-focussing above the excitonic resonance to self-defocussing below it. In general, nonlinear refraction is strongly enhanced near the resonance. The measurements away from resonance were taken at relatively high intensities to ensure an acceptable signal-to-noise ratio. This figure was re-drawn from Ref. [135].

At the end of this section, let us take a look at the wavelength dependence of β_{eff} and n_2 in our $I_{n=2}$ flake to get an overview of its nonlinear absorption and refraction around the fundamental excitonic resonance. Figure 5.5(a) shows that the resonantly enhanced β_{eff} drops by an order of magnitude from its peak value at $\lambda = 570$ nm within only 10 nm on either side. Then, β_{eff} crosses zero at $\lambda \sim 550$ nm and ~ 590 nm, which corresponds to a transition from saturable absorption to two-photon absorption, which is expected for sub-bandgap energies. Note, that – as can be seen in Figure 5.5(b) – n_2 is non-zero at those zero crossings of β_{eff} , which is of interest for experiments and applications where one wants to achieve a nonlinear phase shift with as little nonlinear absorption as possible. One example demonstrating this need is self-phase modulation in Si waveguides, where two-photon absorption generates undesirable free carriers impacting the performance by free-carrier-absorption loss [159, 160]. The overall behaviour of n_2 as a function of the wavelength is characterised by a strong enhancement near the excitonic resonance at $\lambda = 570$ nm, as was the case for β_{eff} , however, with the difference that n_2 exhibits a sign change at resonance, where therefore in good approximation the third-order nonlinearity is purely imaginary. Naturally, the CA Z-scan data in that region is dominated by nonlinear absorption, which makes it impossible to extract n_2 with the division method explained above, because it requires $\Im(\chi^{(3)}) < \Re(\chi^{(3)})$ for the error introduced by

that method to remain below 10% [58]. In practical terms, as this condition needs to be checked using the results of the very analysis whose validity it determines, only the CA/OA datasets which were reasonably close to the shape of the ideal purely refractive CA curves, such as the fits in Figure 5.4(b), were analysed for n_2 , and no n_2 values were extracted from the scans close to the resonance where this was not the case.

Finally, comparing the shapes of β_{eff} and n_2 as a function of the wavelength in Figure 5.5 with the real and imaginary part of the dielectric function ε in Figure 5.3, it becomes apparent that β_{eff} is a mirror image of $\Im(\varepsilon)$ and n_2 is a mirror image of $\Re(\varepsilon)$, which has also been observed in semiconductor quantum wells [161]. This can be intuitively rationalised by considering that the ε curve would be flat in absence of the excitonic resonance. As the exciton state becomes more and more populated at high intensities, the linear absorption and refraction must gradually approach the approximately constant value that they would assume without the exciton. In other words, the peaks and dips in the ε spectra are ironed out with increasing laser intensity, and the nonlinear coefficients associated with this process must therefore form essentially the mirror image of the ε curves.

In summary, studying the example of $I_{n=2}$, we have seen that the 2D RPP flakes exhibit an impressively strong third-order nonlinearity at their fundamental excitonic resonances, allowing for intensity-dependent transmittance changes of several per cent and phase-shifts of $\sim \pi/4$ within only several tens of nanometres of nonlinear interaction volume and on a timescale of ~ 100 fs. The strong enhancement of the nonlinear absorption and refraction is narrowly confined to the spectral width of the excitonic resonance, but even off resonance, decent values of β_{eff} and n_2 were observed. Given these results, the question arises how the nonlinear performance of the 2D RPP flakes can be improved even further. We chose to place thin flakes on an array of plasmonic antennae to benefit from the local field enhancement and achieve even higher nonlinear coefficients. The results of this study are presented in the following section.

5.3 Perovskite Flakes Coupled to Al Nanoantennae

Combining a high-nonlinearity thin film with plasmonic antennae to improve and tailor the nonlinear refraction and absorption has recently been demonstrated by Alam and co-workers for the example of an ITO film in its epsilon-near-zero region [41]. Building on their previous results with the highly nonlinear ITO film referenced above [125], they achieved a ~ 2000 -fold enhancement of the nonlinear refraction compared to a bare ITO film by exploiting two mechanisms. Firstly, they used the fact that the antenna resonance shifts when the refractive index of the ITO layer changes at high intensities, which in their case led to a unity-level difference between the low-intensity and high-intensity refractive index of the system close to the initial antenna resonance. Secondly, at a given input laser intensity, the presence of plasmonic structures with their near-field enhancement boosts nonlinear effects, which scale super-linearly with the electric field. Therefore, the composite system exhibits a higher effective nonlinearity compared to the situation without the plasmonic structures [41].

In the case of our sample, the main motivation for the design was the latter mechanism, with the idea of first demonstrating a plasmonically-driven enhancement of the nonlinear response of the 2D RPP flake before attempting to specifically tailor the response. Figure 5.6 gives an overview of our hybrid perovskite-antenna sample and its linear optical transmission properties. The schematic in Figure 5.6(a) shows that the $I_{n=1}$ 2D RPP flakes with a thickness of ~ 30 nm were placed on arrays of Al antennae on the quartz substrate. Due to the operation in the visible part of the spectrum, Al was chosen as an antenna material instead of Au, which suffers from losses due to interband transitions at those wavelengths [89], while Al does not and is in fact well described by a simple Drude model from the UV to the mid-IR [94]. The 80×80 μm antenna arrays are composed of cross-shaped antennae with dimensions $l = 149$ nm, $w = 50$ nm, $t = 30$ nm, and an antenna spacing of $d = 250$ nm. As most flakes were slightly larger than the antenna arrays, the flakes are slightly bent at the edge of the array, which induces some strain, potentially affecting the Z-scan measurements that follow. This effect could not be reliably tested and we sought to mitigate its impact by

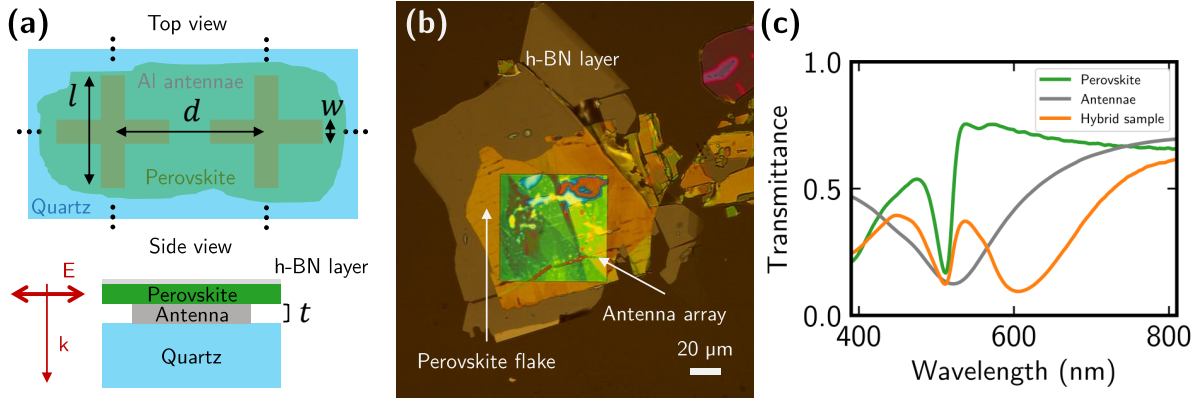


Figure 5.6 – Hybrid perovskite-antenna sample.[†] (a) Schematic of the hybrid perovskite-antenna sample. The perovskite flakes with a thickness of ~ 30 nm were placed on square arrays of cross-shaped Al antennae on the quartz substrate, rather than directly on the substrate as in the sample shown in Figure 5.2. The antennae arrays are $80 \times 80 \mu\text{m}$ wide and the antenna dimensions are $l = 149$ nm, $w = 50$ nm, $t = 30$ nm, $d = 250$ nm. As in the sample without antennae, the perovskite flake was covered under a layer of h-BN for improved long-term stability. The system is irradiated at normal incidence. (b) Optical microscopy image of a hybrid perovskite-antenna sample with an $I_{n=1}$ 2D RPP flake. The image is very similar to that in Figure 5.2(c), with the difference that large parts of the flake are placed on a square antenna array. (c) Linear optical characterisation data in the visible waveband. The transmittance curve of the $I_{n=1}$ 2D RPP flake (green) exhibits a sharp dip at $\lambda = 515$ nm, which is caused by the excitonic resonance, and strong absorption above 400 nm due to the continuum of states above the bandgap. The Al antennae are characterised by a broad plasmonic resonance centred around $\lambda = 515$ nm (grey). A transmittance measurement on the hybrid antenna-perovskite sample (orange) shows an interesting succession of local minima and maxima hinting at a hybridisation of the plasmonic and excitonic modes.

positioning the focussed laser spot near the centre of the array. As in the sample presented above in Figure 5.2, the flake was covered in a thin layer of h-BN for better long-term stability.

Figure 5.6(b) shows a top-view optical microscopy image of the composite system. It demonstrates that the morphology of the flake on the antenna array is not homogeneous, but characterised by several grooves, craters, and bumps, which lead to artefacts in the Z-scan data. In the experiments, it was attempted to avoid sudden jumps of the signal by using a large laser spot size to average over a big array region, but without exceeding the array size to avoid artefacts from the array edge. Moreover, a low-intensity scan containing

[†]The author acknowledges Dr Ibrahim Abdelwahab for fabricating the sample, for taking the optical microscopy image in Figure 5.6(b), for taking the linear characterisation data in Figure 5.6(c), and Dr Zhaogang Dong for performing the e-beam lithography to define the Al antennae.

the noise due to the morphology was subtracted from the Z-scan data, which lead to good results for the OA measurements; however, the quality of the CA data was not sufficient for a meaningful analysis. Therefore, the nonlinear characterisation results presented later in this section are restricted to the nonlinear absorption of the hybrid perovskite-antenna system and no n_2 values are reported.

Linear optical characterisation results are shown in Figure 5.6(c). The transmittance spectra of a separate $I_{n=1}$ flake without antennae (green) and of a separate antenna array (grey) possess the expected features. In the case of the flake, there is a sharp dip at $\lambda = 515$ nm due to the fundamental excitonic resonance, interband absorption above the continuum edge for $\lambda \lesssim 400$ nm, and high transparency for $\lambda > 515$ nm. For the antennae, there is one broad dip in the transmittance spectrum around $\lambda = 515$ nm, which is attributed to the fundamental plasmonic resonance of each antenna, potentially with some inhomogeneous broadening due to fabrication imperfections. The transmittance spectrum of the composite perovskite-antenna system (orange) exhibits a double-dip feature indicating coupling and the formation of hybrid modes.

This formation of hybridised modes is further underpinned by FDTD simulations with the software Lumerical, shown in Figure 5.7(a). In the top panel, the simulated transmittance spectra of the antenna array are visualised for varying antenna lengths l . The dark green low-transmittance feature corresponding to the fundamental plasmonic resonance shifts to longer wavelengths as the antenna length is increased. The situation in the bottom panel, depicting the simulated transmittance spectra for the hybrid perovskite-antenna system, is somewhat more complicated. The absorption of the perovskite flake at the excitonic resonance and the interband absorption above the bandgap appear as vertical dark green bands independent of the antenna length, which follows the behaviour of an uncoupled perovskite flake, which was modelled as a layer with interpolated refractive index data from ellipsometry measurements. Also, for long antenna lengths, where the plasmonic resonance is at much longer wavelengths than the excitonic resonance of the flake, the plasmonic mode is approximately in the same location as for the simulations without the flake in the top panel. However, as the antenna size

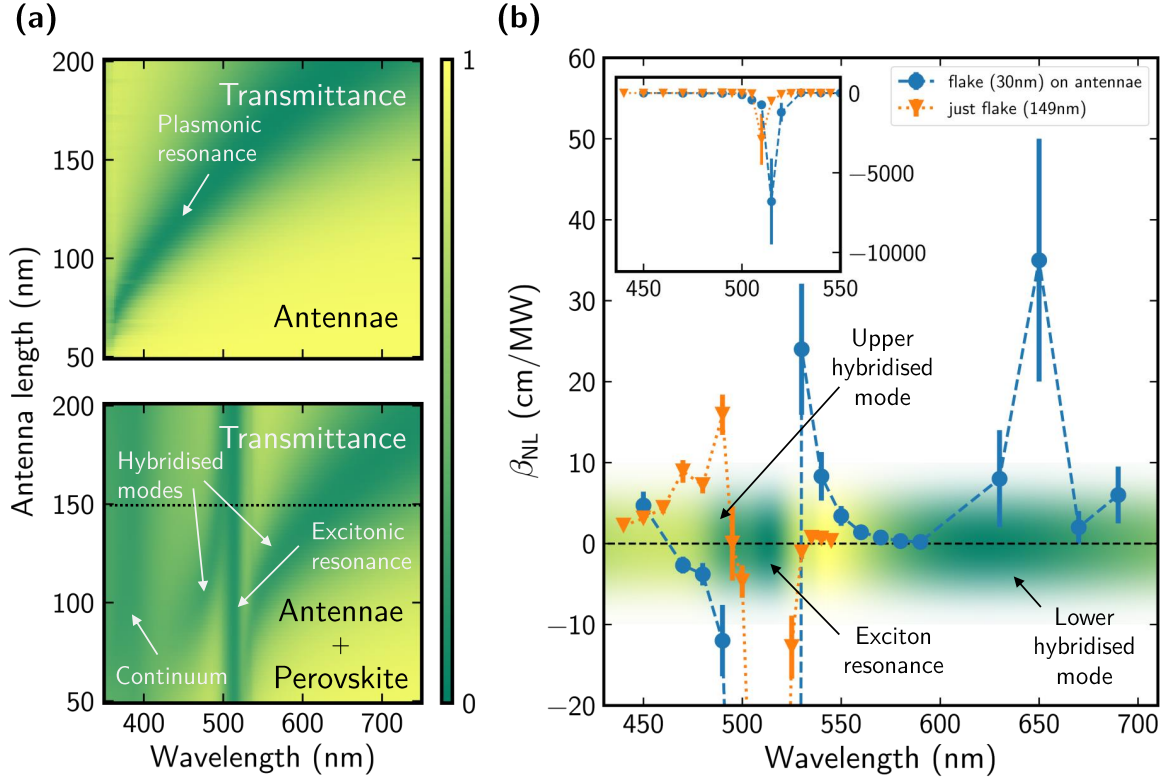


Figure 5.7 – Hybrid plasmonic-excitonic modes and nonlinear absorption.[†] (a) Simulated transmittance maps of just the antenna array (upper panel) and the antenna array covered by an $l_{n=1}$ perovskite flake (lower panel). The top panel shows that the transmittance dip associated with the plasmonic resonance shifts to longer wavelengths with increasing antenna length, as expected. In the lower panel, the above-bandgap absorption of the continuum of states in the perovskite flake and the absorption at its excitonic resonance appear as low-transmittance features independent of the antenna size. At the point of intersection of the plasmonic and excitonic modes, the transmittance map exhibits an avoided crossing, which is interpreted as two hybridised modes. (b) Nonlinear absorption coefficient β_{NL} as a function of the wavelength for a hybrid flake-antenna system (blue circles) and just the flake (orange triangles). In both cases, there is a strongly negative β_{NL} at the excitonic resonance, indicating saturable absorption, which persists in the hybrid system for energies above the resonance. This is attributed to the upper hybridised mode. Another resonant enhancement of the nonlinear absorption coefficient occurs around 650 nm and is attributed to the lower hybridised mode. The coloured ribbon is a cross-section of the lower transmittance map in (a) at the dashed line corresponding to the antenna length on the sample. There is a reasonable agreement between the predicted position of the hybridised mode and the features of β_{NL} .

is chosen such that the uncoupled plasmonic resonance approaches the excitonic resonance of the flake, which is the case for $l \sim 149$ nm (dashed line), the two modes do not intersect, but there is an avoided crossing, which indicates coupling of the modes. Note, however, that the

[†]The author acknowledges Dr Shang Sun for conducting the simulations in Figure 5.7(a).

simulations only predict a partial hybridisation since the fundamental exciton resonance is far from being completely absorbed into the two hybridised modes. A possible explanation for this partial hybridisation is that only the regions of the flake touching the antennae allow for a strong enough coupling to allow for mode hybridisation, while the excitons in the remaining volume of the flake are not sufficiently strongly coupled for hybridisation due to the rapidly decaying plasmonic near-field of the antennae. Despite this caveat, the results suggest that the perovskite flake will not only benefit from the local plasmonic field enhancement, but that two hybridised modes should be observable around the fundamental excitonic resonance.

These two propositions were tested with a series of OA Z-scan measurements at various wavelengths to obtain an overview of the nonlinear absorption behaviour in the visible part of the spectrum. As mentioned earlier, the CA Z-scan data quality was not sufficient to perform an analogous analysis for nonlinear refraction. Figure 5.7(b) depicts the nonlinear absorption coefficient β_{NL} as a function of the wavelength for the hybrid sample (blue circles), and for another sample with an $\text{I}_{n=1}$ flake without the antenna array (orange triangles) for comparison. Note that the nonlinear absorption coefficient is called β_{NL} and not β_{eff} to emphasise that the coefficient was determined with an analysis procedure different to that used earlier. This was necessary to allow for a meaningful and consistent analysis across the wavelength range, given the complicated saturation behaviour at the excitonic resonance. A detailed explanation of this analysis method including a justification, supplementary measurements, and ideas for alternative analysis procedures is therefore given separately in the next section.

Returning to the behaviour of β_{NL} , the dominant feature for both the hybrid perovskite-antenna system and the perovskite flake without antennae is the strongly negative peak at the excitonic resonance at $\lambda \sim 515$ nm, which is shown in the inset. The fact that the maximum absolute value of both curves does not appear at the same wavelength may possibly be due to a slight shift of the exciton mode in the hybrid sample, but it is also possible that the true minimum of β_{NL} is located between $\lambda = 510$ nm and $\lambda = 515$ nm and the apparent detuning is in fact due to an asymmetric peak shape and a too large wavelength step size.

Given the small size of the apparent detuning, this is unimportant for our discussion as long as we can make a plausible argument that the true minimum was not missed by a significant amount – otherwise it is possible that the flake on its own was actually a stronger saturable absorber than the hybrid perovskite-antenna system and we just happened to sample the nonlinear absorption with a too coarse resolution in wavelength space. Therefore, let us calculate the spectral width of the pulses from their duration of ~ 150 fs to establish a characteristic value for the wavelength resolution. Using $\Delta t \cdot \Delta f \sim 0.32$ [162], which marks the narrowest possible spectrum that could reasonably be expected in our laser system, namely that of an unchirped sech^2 pulse, we find $\Delta f \approx 2.1$ THz. Given that $\lambda = 515$ nm corresponds to a frequency of ~ 582 THz, by adding or subtracting 1 THz on either side, we find that the input pulses with a central wavelength of 515 nm range from 514 nm to 516 nm. Therefore, we can see that about one pulse spectrum fits between the data points with the pump beam at $\lambda = 510$ nm and 515 nm. Thus, we cannot rule out that we missed a region where β_{NL} of the sample with just the flake was larger than that of the hybrid sample. However, this would only be possible if the β_{NL} curve of the bare flake was very strongly peaked, in fact, more strongly so than the hybrid flake-antenna sample which does not seem likely from inspecting the data in the inset of Figure 5.7(b). Consequently, it indeed appears reasonable to claim that the hybrid flake-antenna sample is the stronger saturable absorber near the excitonic resonance. Note that this statement relies on accepting β_{NL} as a good figure of merit for saturable absorption strength and on a sufficiently weak dependence of the nonlinear response on the flake thickness, given that the thicknesses of the two flakes differ by a factor of 5. We have not yet conducted a systematic thickness study. In absolute terms, the values of β_{NL} are on the order of cm/kW, which is an order of magnitude larger than the already enormous values from the $\text{I}_{n=2}$ flake discussed earlier; however, further statements on the differences between the analysis methods are needed to qualify this comparison. These will be provided in the following section.

The clearest discrepancy between the two samples occurs at wavelengths around 485 nm, just above the excitonic resonance, where the flake without antennae exhibits two-photon

absorption with a corresponding positive β_{NL} , while the hybrid perovskite-antenna sample is a strong saturable absorber at those wavelengths. We attribute this to the upper hybridised mode of the coupled system which, as we hypothesise, is of mainly excitonic nature because of its proximity to the exciton mode and causes saturable absorption. The coloured band in Figure 5.7(b), which reproduces the slice of the simulated transmittance map in Figure 5.7(a) corresponding to the antenna size on the sample (dashed line), allows for a direct comparison of the spectral features of the nonlinear absorption and the spectral position of the hybridised modes. To the left of the darkest green strip corresponding to the exciton resonance, there is a dark green region indicating the upper hybridised mode. Comparing its spectral extent to the region where β_{NL} of the two samples are of opposite sign, we find that there is a significant overlap but that the upper hybridised mode as indicated by the simulated transmittance spectrum does not extend as far to the short wavelengths. Similarly, for the lower hybridised mode, the peak of β_{NL} in the hybrid perovskite-antenna system is not centred on the dark green feature in the simulated transmittance spectrum, however, the overlap is again acceptable. A possible reason for the spectrally shifted nonlinear response, compared to the modes of the simulated transmittance, is that the simulations assume linear refractive index values. Due to a nonlinear refractive index shift at high intensities, the antenna resonance and hence the hybrid modes may be at slightly different wavelengths in the nonlinear characterisation as compared to the low-intensity simulations [41]. As for the sign of β_{NL} , it can be seen that at the lower hybridised mode the nonlinear absorption does not switch to saturable absorption, but that the two-photon absorption is enhanced. We suspect that the lower hybridised mode is mainly of plasmonic nature, given its spectral distance to the exciton resonance and its proximity to the uncoupled plasmonic resonance, and hence does not possess a significant saturable component.

To summarise, we have seen that compared to a 2D RPP $\text{I}_{n=1}$ flake without antennae, the third-order nonlinear response of the hybrid perovskite-antenna system is enhanced at the excitonic resonance, subject to the caveats mentioned in the discussion above. Moreover, the nonlinear optical characterisation supports simulation results, which predicted the existence of

two hybridised modes on either side of the excitonic resonance. It was found that at the upper hybridised mode, the nonlinear optical response switches to saturable absorption instead of two-photon absorption, which the flake without antennae exhibited at those wavelengths. The lower hybridised mode, however, was not found to cause a sign change of β_{NL} ; instead, the two-photon absorption was enhanced. This is consistent with the partial hybridisation interpretation mentioned above, where the lower hybridised mode is dominated by the nearby plasmon mode while the upper hybridised mode inherits most of its character from the nearby exciton mode.

5.4 Analysing the Exciton Saturation Behaviour

So far, the technical detail of the analysis method for β_{NL} has been avoided to allow for a streamlined presentation of the results. In this section, we shall not only see how the nonlinear absorption coefficient β_{NL} was obtained from the Z-scan data, but also a detailed justification why this analysis method was chosen for the hybrid perovskite-antenna sample. Figure 5.8 shows OA Z-scan data of the hybrid perovskite-antenna system at the excitonic resonance at $\lambda = 515$ nm, which gives a direct insight into the difficulty of a meaningful analysis for this case of a complicated saturation behaviour. The OA Z-scan data plotted as a function of the sample position in Figure 5.8(a) is unusually broad compared to the Rayleigh range of ~ 60 μm , even for low input powers where the nonlinear transmission change just emerges from the noise floor. Note that a lens with a longer focal length was used for the Z-scan measurements of the hybrid perovskite-antenna system, leading to a longer Rayleigh range z_R of 60 μm compared to 20 μm for the experiments with the bare flakes in the first half of this chapter.

Let us consider the red dataset as an example for highlighting the extremely large width of the OA data. The data was fitted (dashed red) with the theory curve from Eq.(4.19)

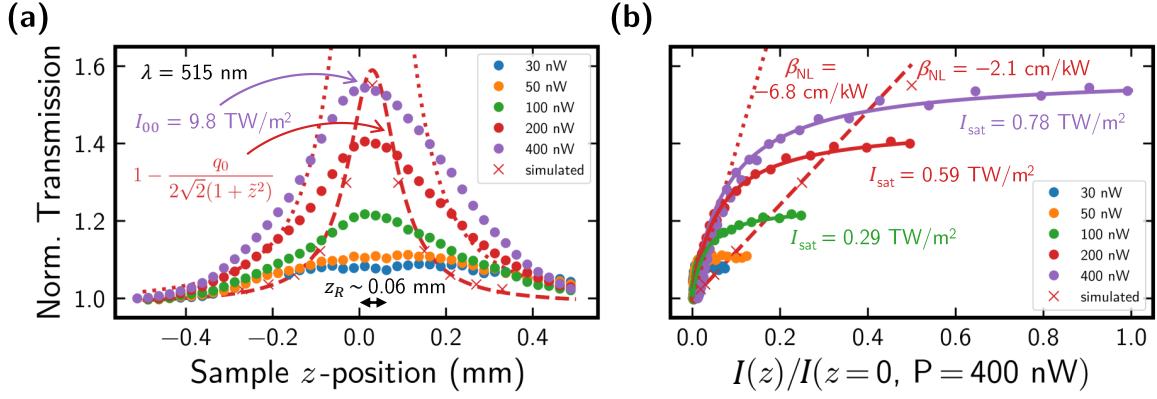


Figure 5.8 – Saturation of the nonlinear absorption on resonance. (a) OA Z-scan traces on the $I_{n=1}$ perovskite flake at $\lambda = 515 \text{ nm}$ for various input powers and an otherwise unchanged experiment. The red dashed line indicates the result of a first-order fit to the red data. The red crosses denote simulation results with a saturation model taking the flake thickness, incident peak intensity, and the refractive index of the flake as inputs. Both curves are much narrower than the measured data. The red dotted line indicates a first-order OA theory curve with a β_{NL} such that the agreement with the data far away from the focus ($|z| \gg 0$) is optimised, disregarding the data around the focus. (b) Transformation of the data in (a) such that the z -position is mapped onto the peak intensity at that position. The saturation behaviour of the measured data and the disagreement with the first-order curves, which are straight lines as a function of the intensity, is obvious. Interestingly, the simulation with the two-level saturation model (red crosses) barely deviates from the red dashed line. A simple phenomenological saturation model (solid lines) does agree with the data, however, the required saturation parameter is different for each dataset. The red dotted line represents the β_{NL} value corresponding to the initial linear trend in the data, which makes its meaning clearer than in (a).

presented in the previous chapter,

$$T_{OA} = 1 - \frac{q_0}{2\sqrt{2}(1 + \tilde{z}^2)} \quad (5.1)$$

where $q_0 = \beta_{NL} I_0 L_{eff}$ and $\tilde{z} = z/z_R$. Even though the shape of the fit function resembles the data, the fact that z_R is fixed at its correct value leads to a poor agreement between the fit and the data. The nonlinear transmission change is overestimated near the focus, but underestimated for the low-intensity data far away from the focus. Note that a seemingly good fit could have been achieved by allowing z_R to vary freely during the fitting instead of fixing it to the correct value. However, as we have shown in the previous chapter in Section 4.4.2 such an approach would lead to wrong or meaningless parameter values. This dataset is a very good example to showcase how problematic such a careless analysis is, because it would

lead to easily obtained parameter values without any obvious issues, completely obscuring the complexity of the saturation behaviour which we will be analysing in the remainder of this section.

Since the data is strongly broadened and a simple fit with Eq.(4.19) does not work, it seems reasonable to use the two-level saturation model Eq.(4.25) from the previous chapter, which supports gradually broadening fit curves as the peak intensity increases beyond the characteristic saturation intensity. Interestingly, this approach does not lead to a good agreement with the data either, despite the fact that it could be successfully applied to the data in Section 5.2. The red crosses in Figure 5.8(a) are data points simulated with this model, where all input parameters, such as the complex refractive index or the layer thickness were set to the appropriate values and the saturation intensity was set such that the corresponding β_{eff} would lead to the same q_0 value as for the dashed red line. Note that comparing the red crosses to the red dashed line shows that essentially no broadening is achieved by choosing the saturation model instead of the simple fit with Eq.(4.19). The reason behind this extremely similar behaviour is that the sample exhibits a strong linear absorption on resonance, which means that the saturation intensity I_{sat} in the saturation model must be set to a very large value compared to the input peak intensity in order to achieve the transmittance change given by the data. In other words, the broadening in the two-level saturation model only occurs as one approaches the maximum possible transmittance increase, which is very large for this sample at $\lambda = 515$ nm, since it is almost completely opaque at low intensities, leaving ample room for a nonlinear transmittance increase. However, the data already broadens very early, at low powers and transmittance changes, which cannot be captured by the two-level saturation model. Therefore, in contrast to the pure flake data in the previous section, which could be well explained with the saturation model, there must be an additional mechanism to cause this extreme broadening at levels far below where it would be expected in that model. Exploring this complicated saturation behaviour and various approaches for a meaningful analysis is at the heart of this section.

In order to make some progress with the analysis, we need to move beyond the standard methods and either improve our understanding of the data by gaining some partial information in a simplified analysis, or by attempting to explain the data with a more complex model. One simple approach following the former strategy is to use Eq.(4.19) as a linear relationship between intensity and nonlinear transmittance change, but such that it agrees well with the low-intensity data far away from the focus. The approach is indicated by the dotted red line in Figure 5.8(a), which vastly overestimates the nonlinear transmittance change near the focus and seems to offer no benefit at all. However, while the benefits of this method are far from obvious in Figure 5.8(a), transforming the Z-scan data by mapping the z -position onto the corresponding intensity

$$I(z) = \frac{I(z=0)}{1 + \tilde{z}^2} \quad (5.2)$$

where $\tilde{z}$ is the z -position of the sample in units of the Rayleigh range, offers further insights that vastly improve understanding, as shown in Figure 5.8(b). Now that the data is plotted as a function of the intensity at each scan point, it becomes apparent that all datasets coincide in the low-intensity limit, but branch out at higher intensities and saturate at different values of the nonlinear transmittance change. The dotted red line does in fact correspond to the initial linear slope common to all datasets, which is probably the best single figure of merit that can be extracted from the data, as we shall explain below. In each dataset, the rightmost data points correspond to the nonlinear transmission changes at the focus ($z = 0$), where the intensity is highest for each scan. Note that the branching-out of the data implies that there is not simply an intensity-dependent saturation, but a separate power and area dependence. To see this, consider for example the data points around $I(z) \sim 0.2 I_{\max}$ in Figure 5.8(b). The green dataset corresponds to a scan where the pump power was at a level that requires to move close to the focus to achieve the given intensity. The same intensity is achieved in the red and purple datasets for progressively larger spot sizes, meaning that they were pumped at higher average powers. However, despite the same intensity, the relative transmittance change is different and increasing with the laser spot area. In other words, doubling the input power and the illumination area is a more efficient way of achieving saturable absorption and

hence an increase of the relative transmittance, even though this should not be the case according to any intensity-dependent model that does not consider input power and beam area separately. As we shall see later in the discussion of Figure 5.9, the saturation behaviour is even more complicated because the nonlinear transmittance change on resonance is also dependent on the time interval between the laser pulses.

The complexity of the saturation behaviour is clearly far beyond what can be investigated and understood with Z-scan measurements only. Therefore, instead of attempting to fit the data with more complex models, which could not be based on well-understood mechanisms at this stage of the investigation, it is reasonable to step back as much as possible and find figures of merit to characterise the saturation behaviour phenomenologically in terms of nonlinear transmittance changes observed at certain input parameter values. Bearing this in mind, the meaning of the dotted red line in Figure 5.8 can be understood as characterising the only common feature of the data, which is its initial linear slope implying a nonlinear absorption coefficient of $\beta_{\text{NL}} = -6.8 \text{ cm/kW}$. Note that this is the peak value in Figure 5.7(b), which is over twice as large as the one for the bare $\text{I}_{\text{n}=1}$ flake, with $\beta_{\text{NL}} = -2.9 \text{ cm/kW}$. Both values are enormous and to the knowledge of the author no larger value has been reported in the literature to date. However, as a word of caution, this is a consequence of using the initial linear slope and we should not be claiming to have found the strongest known saturable absorber. Note that the bare $\text{I}_{\text{n}=1}$ flake only has a $\beta_{\text{eff}} = -\alpha/I_{\text{sat}} = -116 \text{ cm/MW}$ [135] when analysed with the saturation model, compared to the much larger value of $\beta_{\text{NL}} = -2.9 \text{ cm/kW}$ from the initial slope of the data as a function of the intensity. Of course, as we have just shown, the two-level saturation model fails for the hybrid perovskite-antenna sample, so the value of $\beta_{\text{eff}} = -\alpha/I_{\text{sat}}$ is not defined. However, it can be stated that we have observed an extremely strong saturable absorption phenomenon, where a nonlinear transmittance increase of over 20% was achieved with a peak intensity below 1 TW/cm^2 .

Let us now return to our discussion of Figure 5.8(b). It is clear that characterising the data with a linear slope is only a good approximation at low intensities, and a characteristic intensity where the saturable absorption trend begins to saturate should be stated to avoid

implying unrealistically large nonlinear transmission changes at higher intensities.[†] In this context, it also becomes clear that the β_{NL} extracted from the first-order fit indicated by the dashed red line is effectively averaging over a saturating curve and hence completely misrepresents how the nonlinear transmittance changes with intensity. Moreover, it implies higher transmittance changes than what was measured at high intensities. The simulated data points from the saturation model (red crosses) plotted against intensity confirm that the saturation trend of the data is much stronger than predicted by the simple saturation model for the given parameters; the red crosses are barely below the dashed red line, meaning that the I_{sat} parameter of the model to achieve the observed transmittance change given the linear absorption data is much higher than the Z-scan data would suggest. This supplements our argument above, where we commented on the reason why the two-level saturation model could not explain the broadening of the data in Figure 5.8(a).

Despite the complexity of the saturable absorption behaviour at $\lambda = 515$ nm, which appears to be extremely challenging to model, the individual datasets in Figure 5.8(b) do follow a clear trend which can be fitted phenomenologically with the function

$$T_{\text{OA}}(I) = 1 + \frac{\Delta T_0}{1 + \frac{I_{\text{sat}}}{I}} \quad (5.3)$$

which is indicated by the solid lines. This formula was derived with a quick calculation from first principles involving $I_{\text{transmitted}} = I_{\text{in}} e^{-\alpha L} \approx I_{\text{in}}(1 - \alpha L)$ with the sample thickness L and $\alpha = \alpha_0 \cdot (1 + I/I_{\text{sat}})^{-1}$ from Eq.(4.25). Eq.(5.3) is obtained by calculating T/T_0 with the intermediate result $T = 1 - \alpha L$. When fitting the data, we found that different parameter values for ΔT_0 and I_{sat} are needed for each dataset, which means that it is unclear which parameters should be reported for this wavelength. Therefore, we conclude that the high complexity of the saturation behaviour on resonance makes the nonlinear absorption

[†]The increase in transmission is caused by saturable absorption, while the data plotted against intensity saturates in the sense that it has an initial slope, but approaches a terminal value at high intensities. In order to avoid confusion between these two contexts in which the words saturation or saturating are used, we shall speak of saturable absorption when referring to the physical mechanism, and of a saturating trend or curve when referring to the Z-scan data as a function of the intensity.

coefficient of $\beta_{\text{NL}} = -6.8 \text{ cm/kW}$, which is indicated by the dotted red line in Figure 5.8(b), the best parameter to quantify the nonlinear absorption of the hybrid perovskite-antenna sample given the circumstances. However, it must be stated that this value is only meaningful at low intensities – in this case below $\sim 0.5 \text{ TW/m}^2$ – and cannot be extrapolated to predict the nonlinear transmittance change at higher intensities.

The complexity of the saturable absorption at $\lambda = 515 \text{ nm}$ is further underpinned by a comparison of Z-scan data taken at various repetition rates. Figure 5.9(a) shows OA Z-scan traces on resonance at $\lambda = 515 \text{ nm}$ for such a repetition rate study, where the peak power was kept constant. Note that the Z-scans were taken in a mixed order, shown in the legend, to rule out that this increase originated from accumulated optical damage, gradual flake degeneration, or a linear drift of the sample. The first scan is slightly shifted to the left, which could in fact be due to a drift of the stage on which the sample was mounted, but this does not alter the fact that the saturable absorption strength is increasing as a function of the repetition rate. While the similarity of the data at repetition rates of 2 kHz and 5 kHz might indicate that the cumulative effect has reached its maximum extent, it cannot be ruled out that this is just due to a power fluctuation. Again, this does not affect the main result, which is that the nonlinear transmittance change increases with repetition rate, and the conclusion that a cumulative effect on a time scale of $\sim 1 \text{ ms}$ plays a key role in the saturable absorption behaviour on resonance.

For comparison, an off-resonance repetition rate study at $\lambda = 480 \text{ nm}$ is shown in Figure 5.9(b). In contrast to the on-resonance data in Figure 5.9(a), all OA Z-scan traces approximately coincide within the typical noise fluctuations. In particular, no systematic increase of the nonlinear transmittance change as a function of the repetition rate can be observed, which suggests that cumulative effects between pulses can be neglected away from the resonance.

It is interesting to juxtapose the repetition rate studies and data from the main dataset that was used to determine the β_{NL} values in Figure 5.7(b). The latter are shown in Figure 5.9(c),(d). As can be seen in Figure 5.9(c), the data at $\lambda = 515 \text{ nm}$ exhibits a strong saturation trend as a function of the intensity and only the initial low-intensity region can be

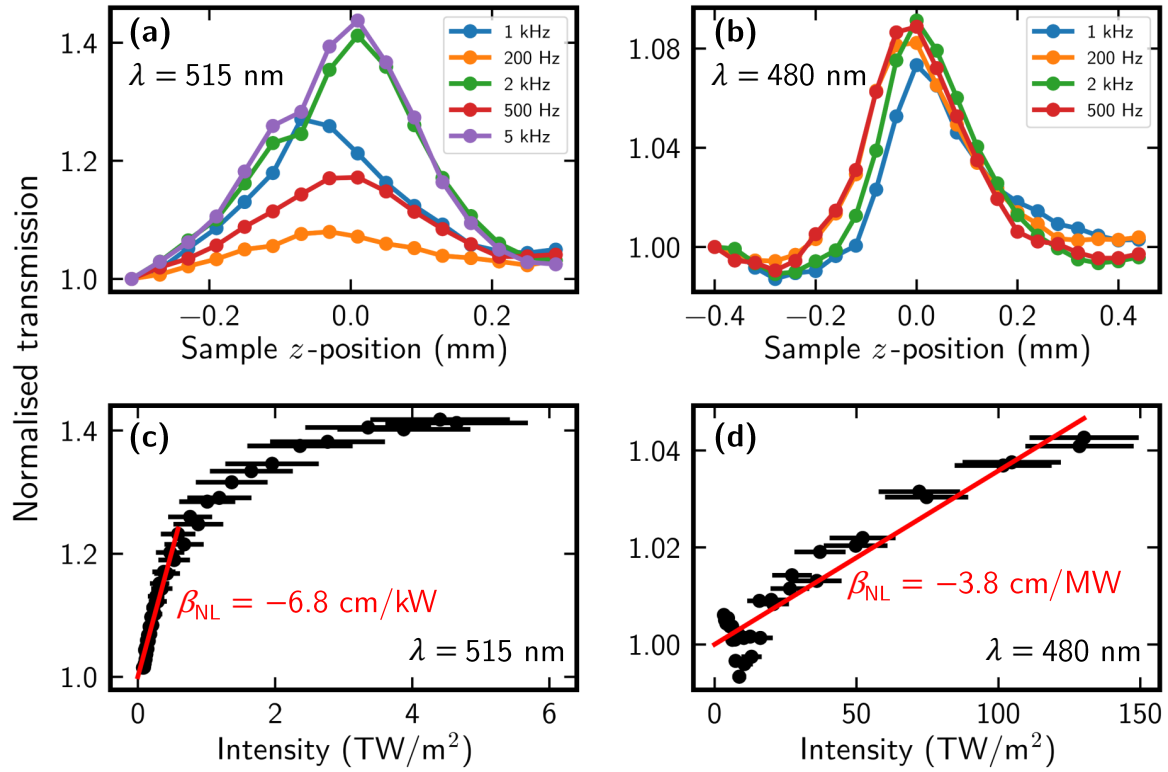


Figure 5.9 – Repetition rate study and saturation behaviour on and off resonance. (a) OA Z-scan data on resonance ($\lambda = 515$ nm) for varying repetition rates at constant peak power. The nonlinear transmission change clearly increases as the time between pulses is reduced. The measurements were taken in direct succession and in the order of the legend to identify potential degradation of the sample between measurements. (b) OA Z-scan data away from resonance ($\lambda = 480$ nm) for varying repetition rates at constant peak power. There is no significant impact of the repetition rate on the saturable absorption, implying that the system relaxes to its unperturbed state between the laser pulses. (c) OA Z-scan data on resonance as a function of the peak intensity at a repetition rate of 10 kHz. The nonlinear absorption value is extracted via a linear fit to the data in the low-intensity regime. (d) The OA Z-scan data at $\lambda = 480$ nm is much better approximated by a linear function due to the much higher saturation intensity off resonance compared to the resonant case.

used for a linear fit, as explained in the discussion of Figure 5.8. However, for the off-resonant data in Figure 5.9(d) the nonlinear transmittance change is almost linear as a function of the input intensity even at much higher intensities than those in Figure 5.9(c).

The juxtaposition of the data in Figure 5.9 allows for the preliminary conclusion that the extreme saturation of the saturable absorption on resonance is accompanied by slow cumulative effects, whereas the saturation of the nonlinear response is almost absent when slow cumulative effects can be neglected. Of course, in the latter case the nonlinear increase in

transmittance must saturate at a certain intensity level as well, but this is only expected at much higher intensities than those in the experiment, unless optical damage makes it impossible to observe the full saturation curve. We speculate that the saturation of the saturable absorption off-resonance is linked to the gradual increase in average population of the relevant excited states and might be modelled well with the simple two-level saturation model used for the sample without the antenna array discussed in the first half of this chapter. The fact that this model failed on resonance, as shown earlier in Figure 5.8 might be explained by combining all the results in this section, proposing that a cumulative effect depletes the exciton state sufficiently to put an early limit to the achievable increase in transmittance. For this interpretation to be plausible, and consistent with the ms time scale of the cumulative effect obtained from the repetition rate study, exciton lifetimes should be at least on the order of 1 ms. In the literature, however, time-resolved measurements on flakes with the same chemical composition demonstrated that the excitons, even those in trap states at the grain boundaries, decay much faster, with lifetimes on the order of 1 ns [163]. Moreover, a recent study of singlet and long-lived triplet excitons in the same 2D RPPs sandwiched between two glass slabs reported triplet emission time scales of no more than 50 ns and a coupled rate equation analysis with the longest time scale on the order of 10 μ s, corresponding to the non-radiative decay time of the triplets [164]. It can only be concluded that the key mechanisms behind the observed saturation behaviour remain elusive and further studies are necessary to solve this puzzle.

Despite the open question concerning the physical mechanisms involved in the on-resonance saturation behaviour, it is possible to phenomenologically study the wavelength dependence of the saturable absorption, which is the last major part of the analysis that can be conducted with the data at hand. The idea behind this is that even without a detailed knowledge of the underlying physical processes, it may be possible to find consistent figures of merit, allowing one to still gain quantitative information about the change in saturation behaviour as the exciton resonance is approached. Figure 5.10 presents the results of two phenomenological approaches to extract saturation parameters from the complete dataset at

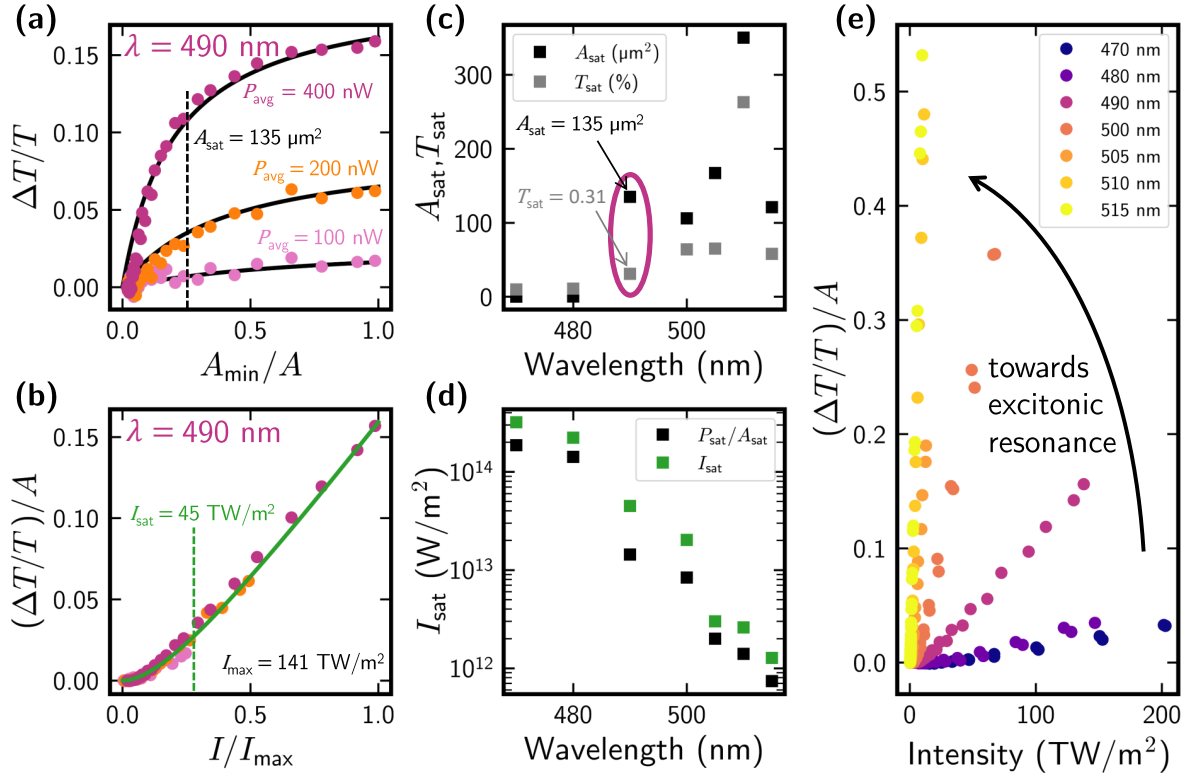


Figure 5.10 – Alternative saturation analysis. (a) OA Z-scan data as a function of the inverse illumination area. Here, the transmission data was plotted as the relative normalised change, rather than the normalised transmission, which amounts to a y -axis shift of 1. The black curves stem from a saturation model with saturation parameters P_{sat} , A_{sat} , and T_{sat} , which follow all datasets at different average powers well for only one set of global saturation parameters. (b) The same OA Z-scan data divided by the illumination area and plotted against the peak intensity at each data point. All datasets approximately follow an initially quadratic trend, which transitions into a linear trend around a characteristic intensity value I_{sat} . (c) Summary of the saturation parameters A_{sat} and T_{sat} from the model in (a) as a function of the wavelength. Both parameters strongly increase towards the excitonic resonance at $\lambda = 515$ nm. (d) Summary of the saturation intensities extracted from the models in (a) and (b) as a function of the wavelength. There is an approximately exponentially falling trend towards the resonance at $\lambda = 515$ nm. (e) Visualisation of all OA datasets to highlight that away from resonance, only a weak increase in transmission is achieved, whereas the effect is much larger on resonance and occurs at much lower input intensities.

each wavelength, containing Z-scans with various input powers but otherwise constant experimental parameters. The first approach applied to the dataset at $\lambda = 490$ nm, which was chosen as a representative example of the intermediate case between the on- and off-resonant experiments, is shown in Figure 5.10(a), where the model

$$\frac{\Delta T}{T} = T_{\text{sat}} \frac{1}{1 + \left(\frac{P_{\text{sat}}}{P}\right)^2} \frac{1}{1 + \frac{P_{\text{sat}}}{P} \frac{A}{A_{\text{sat}}}} \quad (5.4)$$

with three saturation parameters T_{sat} , P_{sat} , and A_{sat} was manually adjusted to fit the data. Note that the black curves from the model are not separate fits, but belong to the same set of parameters $T_{\text{sat}} = 0.31$, $P_{\text{sat}} = 310 \text{ nW}$, and $A_{\text{sat}} = 135 \text{ } \mu\text{m}^2$ but with the appropriate average power values P inserted for each subset of the data. Unlike in the earlier Figures 5.8 and 5.9 the data is plotted as a function of the inverse illumination area to decouple area and power dependence rather than only consider them combined as the intensity. Another difference is that the y -axis was shifted by 1 to explicitly represent the data as the relative *change* of the transmittance, which slightly simplifies the expression (5.4).

Figure 5.10(b) visualises the second phenomenological approach to quantify the saturation behaviour, applied to the same dataset. This time, the data was plotted as a function of the intensity and the y -coordinate was divided by the area in units of the minimum beam area for each data point. As a result, the data points at the focus have the same y -coordinate as in 5.10(a), but those corresponding to larger beam areas are suppressed. The reason for transforming the data in that way is that it was found to make all datasets at the same wavelength coincide on one curve (solid green), which starts with a quadratic trend that transitions into a linear trend at a characteristic intensity. Therefore the expression of the green curve is

$$\frac{\Delta T}{T \cdot A} = \text{const.} \cdot \frac{I}{1 + \frac{I_{\text{sat}}}{I}} \quad (5.5)$$

with the characteristic saturation intensity I_{sat} . Even though the physical interpretation of this property of the data, namely to coincide on the curve (5.5) for the given data transformations, is not obvious, it constitutes a possible way of extracting a characteristic saturation intensity from the data at each wavelength.

Now we are in a position to summarise the saturation parameters of both approaches in an attempt to characterise the wavelength dependence of the saturation behaviour. Figure 5.10(c) shows the saturation parameters A_{sat} and T_{sat} from the first approach as black and grey squares, respectively. Both the predicted maximum achievable transmittance change T_{sat} and the characteristic beam area A_{sat} below which the separate area dependence starts becom-

ing important, increase by at least an order of magnitude over the investigated wavelength range as the excitonic resonance is approached. There are large fluctuations in both trends; however, these are not large enough to affect this conclusion. The ratio of the saturation parameters P_{sat} and A_{sat} , which can be understood as a saturation intensity of the first approach, is plotted as black squares in Figure 5.10(d) and juxtaposed with the saturation intensity I_{sat} from the second approach, visualised as the green squares. Both are characterised by an exponential decrease towards the excitonic resonance, which is perhaps the most dependable outcome of this study. The fact that I_{sat} is larger than $P_{\text{sat}}/A_{\text{sat}}$ by about a factor of 2 at all wavelengths has no obvious origin and is likely due to the different ways in which the saturation parameters are extracted.

Finally, Figure 5.10(e) summarises the complete saturable absorption data in one plot to serve as an equivalent representation of the complete saturation behaviour, supplementing the understanding of the wavelength dependence of the saturation parameters. The key result is that far away from resonance, a large intensity is required to achieve an increase in transmittance and the maximum achievable transmittance change is low, while on resonance much smaller intensities suffice to produce an enormous transmittance change of $\sim 50\%$. It is this behaviour that we have attempted to quantify with the saturation parameters from the phenomenological models.

5.5 Conclusion

Let us briefly review the key points of this chapter. First, the perovskite crystal structure, as well as the chemical composition of the layered 2D perovskite flakes on our samples was discussed. This was followed by information on the sample fabrication process and the physical background of the optical properties. The alternating layers of inorganic 2D perovskites and organic spacer molecules form a multiple quantum well structure, which supports strongly confined room-temperature-stable excitons with large binding energies. This is not only due to quantum confinement but more importantly dielectric confinement,

as the Coulomb screening is greatly reduced in the part of the exciton wavefunction extending into the organic layers. The spectral position of the exciton resonance, in the vicinity of which the nonlinear optical response was found to be greatly enhanced, can be tuned through large parts of the optical waveband by varying the perovskite layer thickness. The two main effects responsible for this are the decrease in exciton binding energy and the red-shift of the continuum edge with increasing perovskite layer thickness. The latter effect dominates and the highest energy of the fundamental exciton resonance is found for the stack with perovskite monolayers, abbreviated as $I_{n=1}$, with the energy decreasing as the perovskite layer thickness is increased.

Z-scan measurements were conducted on a sample with an $I_{n=2}$ flake, as it showed the highest optical nonlinearity and was representative of the qualitative behaviour of the third-order nonlinear optical coefficients β_{eff} and n_2 around the fundamental excitonic resonance for all samples investigated. The nonlinear absorption coefficient β_{eff} became negative around the excitonic resonance, indicating saturable absorption instead of two-photon absorption elsewhere below the continuum edge. The nonlinear refractive index parameter n_2 exhibited a sign change on resonance, switching from self-focussing to self-defocussing, and was enhanced in magnitude around the excitonic resonance. Peak values for both $\beta_{\text{eff}} = -256 \text{ cm/MW}$ and $n_2 \approx \pm 10^{-13} \text{ m}^2/\text{W}$ were found to be higher than most values in the literature.

Next, a sample with a layered 2D perovskite flake on an antenna array with its plasmonic resonance tuned to the fundamental exciton resonance of the flake was studied with the aim to observe a stronger nonlinear response due to plasmonically enhanced local fields. After a presentation of this hybrid perovskite-antenna sample, linear transmission measurements and FDTD simulations were presented, indicating coupling between the two subsystems and the formation of two hybridised modes, even though it is more accurate to speak of a partial hybridisation since the exciton mode was still present in the hybrid perovskite-antenna sample. In the subsequent Z-scan study it was found that the saturable absorption at the fundamental exciton resonance is stronger than for a flake without antennae. Moreover, the upper hybridised mode was identified through a sign change of the nonlinear response

at wavelengths expected from simulations. Similarly, the lower hybridised mode was identified by its enhancement of the two-photon absorption around the wavelengths predicted by simulations. No analysis of the nonlinear refraction was performed on the hybrid perovskite-antenna sample due to insufficient data quality.

The nonlinear absorption coefficient for the hybrid perovskite-antenna sample had to be obtained differently to that for the perovskite flake without antennae, because the saturation behaviour of the hybrid system was too complex for the simple two-level saturation model used for the bare flake. In order to distinguish higher values of the nonlinear absorption coefficient due to the analysis method from those due to the actual behaviour of the sample and to qualify and understand the meaning of the nonlinear absorption coefficient used in the analysis of the hybrid system, the saturation behaviour was studied in more detail. It was found that on resonance, the Z-scan traces are broadened and saturate faster than predicted by a simple two-level absorption model. When plotted as a function of intensity, it could be seen that on resonance the strength of the saturable absorption depends on excitation area and power separately. Moreover, it was shown with repetition rate dependent measurements that cumulative effects on time scales of 1 ms are relevant on resonance, adding to the complexity of the saturation behaviour. Therefore, the investigation concluded with the result that simple phenomenological linear fits to the Z-scan data plotted against intensity were the best way to extract the nonlinear absorption coefficient. As a final part of the analysis, further phenomenological models were applied to the data to quantify the wavelength dependence of the saturation behaviour, where an exponential reduction of the characteristic saturation intensity as a function of the wavelength by over two orders of magnitude was found, as the exciton resonance was approached.

At the end of this chapter it may be concluded that the nonlinear Z-scan measurements combined with parameter scans have revealed a surprising wealth of information about the excitons in multilayer 2D perovskite flakes. Of course, many open questions, particularly concerning the saturation mechanisms in the hybrid perovskite-antenna sample remain, and the limitations of the insights achievable with simple Z-scan measurements have become

clear. Despite this, systematic thickness-dependent studies and particularly a two-dimensional parameter scan of the repetition rate and the wavelength around the excitonic resonance would be helpful to better understand the saturation behaviour of the hybrid perovskite-antenna sample. Moreover, dual-beam experiments with a pump beam exciting charge carriers above the bandgap and a probe beam on the exciton resonance could be conducted to observe a potentially strong modulation of transmittance as a function of the pump power. This idea was also discussed in a recent proposal by Xie and colleagues [44] for an all-optical and electrical switch, which was published after the measurements for this chapter had been completed. The work, which builds on the experimental results by Alam and co-workers [41] has certain similarities to our work, such as the cross-shaped antennae and the occurrence of mode hybridisation. On the other hand, the structure proposed by Xie and colleagues uses ITO instead of a 2D perovskite flake, the wavelengths of operation are in the IR instead of in the visible, and the article is a theoretical proposal without experimental results [44]. Finally, it would be interesting to consider a hybrid perovskite-antenna sample with bar-shaped antennae similar to the sample in Ref. [41] instead of cross-shaped antennae, such that the plasmonic enhancement can be turned on and off by means of the polarisation of the input laser and a direct comparison of the nonlinear parameters in both cases is conveniently possible. All these possibilities demonstrate that the results so far are merely a first step towards a thorough understanding of the strong optical nonlinearities offered by the hybrid 2D perovskite-antenna system.

Chapter 6

Nonlinear Frequency Mixing in GaN

After the discussion of nonlinear absorption and refraction related to the optical Kerr effect, where only the nonlinear response of the material at the pump wavelength was of interest, let us now focus on nonlinear interactions where new optical frequencies are generated. In this chapter, a Gallium nitride (GaN) film on a sapphire substrate is studied as a model system which supports both second-order and third-order nonlinear frequency mixing processes due to its non-vanishing bulk $\chi^{(2)}$. Etching approximately wavelength-sized cones into the GaN film is found to both concentrate the local optical fields and improve the coupling to the relevant tensor components, leading to an increased production of second-order nonlinear signal. Beyond being a preliminary result in its own right, this study prepares the ground and serves as a reference for the frequency mixing results with Au nanoantennae laid out in the following chapter. The GaN project in this chapter was a collaboration with the group of Prof Dr Anthony Bennett from Cardiff University. All individual contributions are acknowledged in footnotes below the corresponding figures.

6.1 Motivation

In order to understand the relationship between dielectric nanoparticles and other nonlinear light sources, it is helpful to give a brief overview starting with macroscopic nonlinear crystals. Achieving efficient frequency conversion has been at the heart of nonlinear optics research since its earliest days following the invention of the laser [38, 165], which for the first time made high-intensity coherent optical radiation available to researchers. This made it possible to observe and systematically study frequency mixing processes [38], such as Second Harmonic Generation (SHG), which was first demonstrated [166] in a non-phase-matched quartz crystal only one year later. It soon became apparent that phase-matching, explained in detail in Chapter 2, dramatically improved the nonlinear conversion efficiency [167], extending the coherent interaction volume throughout the entire nonlinear crystal rather than the typically only μm -sized coherence length defined by the dispersive properties of the crystal. This sparked research interest in negative uniaxial crystals with convenient phase-matching and large nonlinear susceptibilities, such as potassium dihydrogen phosphate (KDP) [168] and lithium niobate [63, 169]. Soon afterwards, optical fibres were explored as alternative nonlinear media, with the key benefits of large interaction volumes and modal confinement of the nonlinear output [170]. After early demonstrations [171] of self-phase-matched processes in fibres, such as Stimulated Raman and Brillouin Scattering, which is a photon-phonon interaction, efficient Four-Wave-Mixing with phase-matching via fibre birefringence [172] and via modal dispersion [173] was achieved. The common principle in the nonlinear platforms mentioned so far – macroscopic crystals and fibres – is to accumulate the nonlinear output over a large effective interaction volume, where phase-matching ensures that the fields add up in phase. Optical cavities containing a nonlinear crystal allow for a further multiplication of the effective interaction volume [174], as the photons pass through the crystal many thousand times before leaving the cavity, depending on the reflectivity of the mirrors.

However, with the advent of nanofabrication and plasmonics, which was discussed in Chapter 1, and the availability of femtosecond pulsed lasers with extremely high peak powers, a new

approach emerged. The basic idea was to achieve efficient nonlinear mixing in nanoscopic volumes in order to pave the way for low-power and small-footprint nonlinear devices [37], compensating for the reduced interaction volume by exploiting the superlinear scaling of frequency mixing and the strongly enhanced local fields of plasmonic nanostructures. However, despite the many advantages of nanoplasmonic systems, such as the aforementioned giant field enhancement and fast response time on the order of tens of femtoseconds [37], ohmic losses are severely impacting their performance and suitability for practical applications [46]. Therefore, workers in the field have recently been investigating dielectric nanoparticles to circumvent plasmonic losses while retaining a moderate field enhancement due to electric and magnetic Mie-resonances [46]. Essentially, in this strategy, which attempts to combine the best of both worlds, a large part of the field enhancement that plasmonics would offer is traded in for an increased interaction volume extending throughout the nanoparticle, and strongly reduced absorption loss which implies a higher damage threshold [29, 46]. More details will follow in Section 6.6; at this stage it suffices to say that both the plasmonic and the dielectric approach are active areas of research and have their own advantages and disadvantages. As far as this thesis is concerned, the remainder of the current chapter is dedicated to work on nonlinear optics in dielectric nanostructures, while results on nonlinear mixing in plasmonic nanoparticles will be presented in the following chapter.

When choosing a dielectric material for a nanoparticle supporting a strong concentration of optical fields, it seems reasonable to first look for materials with a high refractive index [29]. However, depending on the application it is not necessarily optimal to choose the dielectric with the highest possible refractive index at a particular frequency, as a high index is linked to strong absorption via the Kramers-Kronig relations [29]. In semiconductors, which are the most popular dielectrics in the visible waveband, the high refractive index is a consequence of a continuum of interband transitions, suggesting a relationship between the bandgap and the refractive index [29]. Indeed, the so-called Moss relation establishes an inverse relationship between the bandgap E_g and the fourth power of the refractive index, following a scaling argument whereby all energy levels in a dielectric have to be multiplied by a factor of $1/\epsilon^2$,

which agrees well with data from most dielectrics from the NIR to the UV [175, 176]. This means that in order to avoid compromising the frequency mixing performance in the visible part of the spectrum, it is preferable to choose a material with a lower refractive index but larger bandgap and hence a transparency window extending towards shorter wavelengths, in particular for direct-bandgap semiconductors which have a sharp absorption edge at E_g [29]. GaN is very well suited for nonlinear applications with its large second-order nonlinearity on the order of Lithium niobate [177] and a large bandgap of ~ 3.4 eV [178] leading to a transparency window ranging from 365 nm in the UV to 13.6 μm in the mid-IR [179]. In comparison, GaAs, another III-V direct-bandgap semiconductor, has a transparency window from only 873 nm to 13 μm [179], but a larger intrinsic second-order nonlinearity of 740 pm/V in its $\chi_{321}^{(2)}$ element [38], compared to 33 pm/V in the $\chi_{333}^{(2)}$ element of GaN [180]. On the other hand, GaN has the advantage of a high thermal conductivity and a small thermal expansion coefficient [181], which promises good stability in optical experiments with intense laser beams. As we are interested in bridging large regions of the optical spectrum and mostly operate at high pump powers to compensate for the moderate field enhancement in dielectric structures compared to plasmonic ones, those properties make GaN an ideal candidate for our frequency mixing experiments. In order to present the results in a systematic way, we shall first consider a plain, unstructured GaN film, and only introduce the wavelength-sized cones in Section 6.6.

6.2 Sample Structure

The sample under study is based on a commercial GaN template from Kyma Technologies, Inc. with a $3\text{ }\mu\text{m} \pm 10\%$ layer of GaN grown on a 430 μm c-plane sapphire substrate via hydride vapour phase epitaxy (HVPE). The sample structure is schematically visualised in Figure 6.1(a). Note that cone-shaped structures were etched into the GaN layer, which will be described in detail in Section 6.6. Until then we will only consider the flat region next to the cone, where the thickness of the GaN layer was reduced to $\sim 1.5\text{ }\mu\text{m}$ because of the

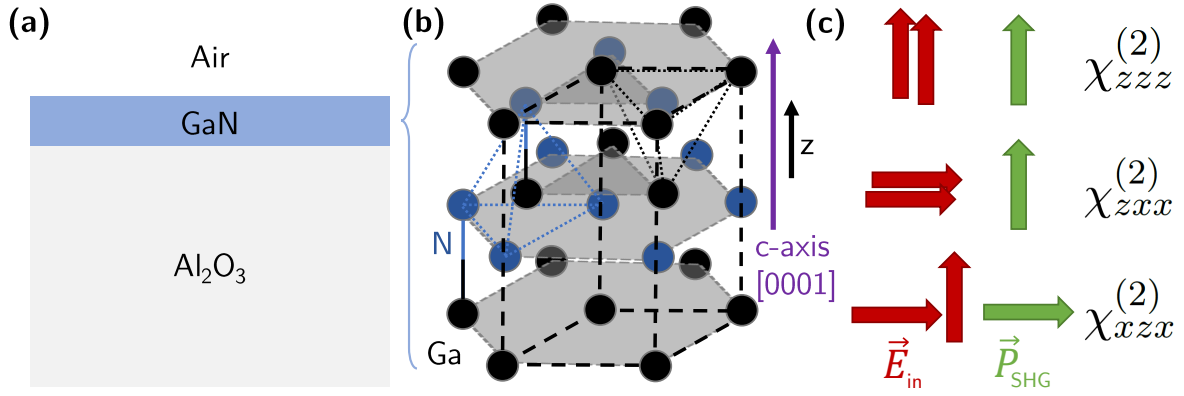


Figure 6.1 – Sample structure and $\chi^{(2)}$ tensor components. (a) A $3\ \mu\text{m} \pm 10\%$ GaN layer is grown along the c -axis on a $430\ \mu\text{m}$ sapphire substrate with an hcp crystal structure. (b) The GaN crystallises in a wurtzite structure consisting of an hcp lattice for both the Ga and N sites, where each atom of one species is surrounded by a tetrahedron of atoms from the other species (dotted lines) [182]. The unit cell is indicated by the black dashed lines. (c) In a wurtzite crystal the non-zero elements of the $\chi^{(2)}$ tensor are $\chi_{zzz}^{(2)}$, $\chi_{zxx}^{(2)}$, $\chi_{xzx}^{(2)}$, and $\chi_{zyy}^{(2)}$, $\chi_{yzy}^{(2)}$ [183]. The arrows visualise the resulting second harmonic polarisation in the GaN layer and the participating polarisation components of the fundamental electric field for an x -polarised (analogous for y polarisation) laser beam at normal incidence. Only the xzx (and yzy) contributions lead to radiation propagating along the z -axis; the others represent dipoles *polarised* along the z -axis, which radiate into the GaN slab.

etching process. The idea is to explore the nonlinear optical properties of the unstructured GaN before increasing the complexity of the system by adding the cones.

In order to understand the second-order nonlinear properties of the GaN layer, it is important to remember that it possesses a wurtzite crystal structure [183], which is drawn in Figure 6.1(b). Each atomic species crystallises in its own hcp sublattice and the coordination between them is such that any given atom is surrounded by a tetrahedron of atoms of the other species [182]. When placing the wurtzite lattice in a cartesian coordinate system where the z -axis is parallel to the c -axis of the hcp lattices, the non-zero $\chi^{(2)}$ tensor elements of the wurtzite lattice are $\chi_{zzz}^{(2)}$, $\chi_{zxx}^{(2)}$, $\chi_{xzx}^{(2)}$, and $\chi_{zyy}^{(2)}$, $\chi_{yzy}^{(2)}$ [183], which is visualised in Figure 6.1(c) for the xz -plane, but of course the yz -plane could have been chosen equally well. The key insight at this stage is that only the $\chi_{xzx}^{(2)}$ and $\chi_{yzy}^{(2)}$ tensor elements lead to a nonlinear polarisation oscillating in the c -plane of the crystal, which corresponds to the xy -plane of the GaN slab, as it is grown along the c -axis. All other contributions to the second-order nonlinear response will oscillate along the z -axis and hence radiate into the slab, which is

only useful if that radiation is to be passed on to another element on the chip. However, in this work, the sample is illuminated from the top at normal incidence, and the nonlinear signal is detected along the z -axis in both directions, so only the in-plane oscillations are of interest. Let us now take a closer look at the field distributions and the coupling to the relevant tensor elements inside the GaN layer.

6.3 Coupling Input Fields and Active Tensor Elements

In the experiments presented in this chapter, the sample is excited with one or two Gaussian beams focussed onto the sample at normal incidence. The second-order nonlinear response of the GaN layer, as measured in this work, can be understood by discussing in more detail the case of second-harmonic generation (SHG), where only one excitation beam is needed; the sum-frequency generation (SFG) and difference-frequency generation (DFG) signals are generated in a very similar way with two excitation beams of different wavelengths. Figure 6.2(a) depicts the $|E_z|^2$ distribution of an x -polarised NIR excitation beam, propagating along z and focussed with a 0.4 NA objective, which was used in most of the experiments. The contours indicate the $|E_x|^2$ distribution, which is essentially the same as the intensity contours of the beam, given that $|E_x|^2 \gg |E_z|^2$. The longitudinal field component E_z in a Gaussian beam is typically negligible, but may become important near a focus [2]. Using $\nabla \cdot \vec{E} = 0$ it can be estimated as

$$E_z(x, y, z = 0) = -i \frac{2x}{kw_0^2} E_x(x, y, z = 0) \quad (6.1)$$

in the focal plane with

$$E_x(x, y, z) = \frac{E_0}{\sqrt{1 + \frac{z^2}{z_R^2}}} \exp\left(-\frac{x^2 + y^2}{w_0^2 \left(1 + \frac{z^2}{z_R^2}\right)}\right) \cdot e^{ikz - i \arctan\left(\frac{z}{z_R}\right) + ik \frac{x^2 + y^2}{2R(z)}} \quad (6.2)$$

where $R(z) = z(1 + z_R^2/z^2)$ is the radius of curvature of the wavefronts, w_0 is the beam radius at the focus, $z_R = kw_0^2/2$ is the Rayleigh range, and the z -dependence of E_x can be

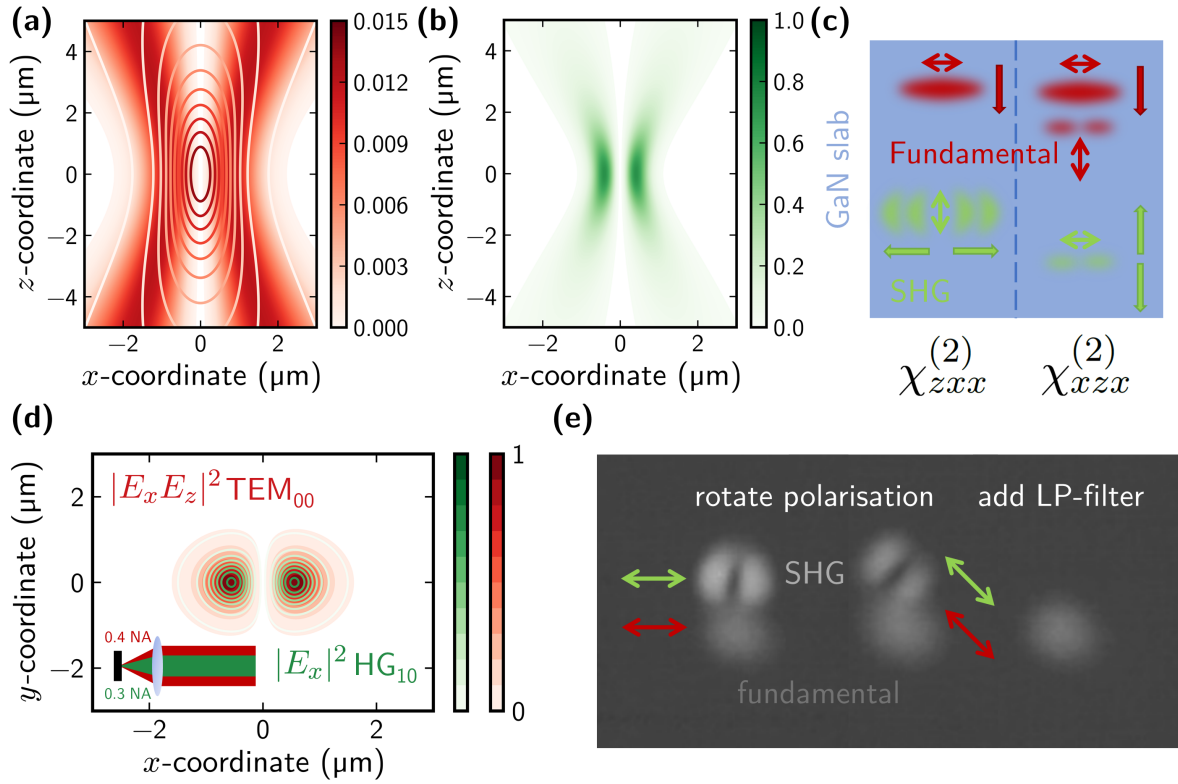


Figure 6.2 – Field components in Gaussian beams and nonlinear mixing. (a) Visualisation of $|E_z|^2$ in an x -polarised 0.4 NA fundamental Gaussian beam at $\lambda = 1500$ nm and unit field amplitude $|E|$. The contours indicate the corresponding distribution of $|E_x|^2$. (b) $|E_x|^2$ distribution of a 0.4 NA HG_{10} beam at 750 nm. Note the similarity to the $|E_z|^2$ distribution at 1500 nm. (c) The two main contributions to SHG in the GaN slab involve the $\chi_{zxx}^{(2)}$ and $\chi_{xzx}^{(2)}$ tensor elements. The former leads to a polarisation density at the second harmonic that is aligned with the z -axis and hence radiates into the slab, while the latter leads to an x -polarised SHG source term where the fundamental E_x and E_z overlap, supporting radiation along the z -axis in a HG_{10} mode. (d) The strong overlap of the normalised SHG source term (red) and the HG_{10} mode at the second harmonic (green) is the reason for most of the x -polarised SHG coupling to that mode. Note that the divergence angle of the second harmonic beam was lowered to the equivalent of about 0.3 NA for optimum overlap, meaning that the collimated second harmonic beam is narrower than the fundamental beam. (e) Reflected fundamental and SHG intensity distributions imaged with a Si detector and with a slight angle between both beams to achieve a separation of the spots on the camera. Rotating the polarisation causes the characteristic double-lobe pattern of the SHG to rotate as well, providing evidence for the HG_{10} argument presented above. A reference image with a long-pass filter was taken to demonstrate that the circular spot is indeed the fundamental beam and not a TEM_{00} contribution at the second harmonic.

used to extend the E_z distribution up to about a Rayleigh range away from the focus within good approximation [2]. The characteristic X-like shape of the intensity distribution of the longitudinal electric field strongly resembles that of the E_x component of a HG₁₀ mode [2], which is drawn for a beam at the second harmonic in Figure 6.2(b). Note that for both plots only the distributions in vacuum are drawn and the GaN-air interface is neglected because it has no bearing on the arguments developed here, given the relatively low 0.4 NA. For large numerical apertures, the interface cannot in good approximation be neglected due to the complicated distribution of focal fields [2]. The interplay between the field components inside the GaN slab is represented schematically in Figure 6.2(c). On the left, we can see the interaction related to the $\chi_{zxx}^{(2)}$ tensor element, which is only driven by the E_x component of the fundamental field, represented by a single ellipse to indicate its TEM₀₀ mode profile. The oscillations at the second harmonic are parallel to the z -axis, leading to radiation into the xy -plane, and thus confined to the slab as mentioned earlier in the discussion of Figure 6.1. There is also a weak contribution due to $\chi_{zzz}^{(2)}$ with the same orientation of the oscillating dipoles at the second harmonic, which is omitted in this figure due to its insignificance. On the right of Figure 6.2(c) we can see the much more interesting contribution due to $\chi_{xzx}^{(2)}$, which involves both the TEM₀₀-distributed E_x component and the approximately HG₁₀-distributed E_z component of the fundamental beam. These fields couple to an x -polarised field at the second harmonic, which is propagating along z and originates from a source term distributed according to $|E_x E_z|^2$. Figure 6.2(d) demonstrates the near-perfect overlap of the nonlinear source term $|E_x E_z|^2$, where E_x and E_z are taken from the fundamental beam, and a HG₁₀ mode at the second harmonic, albeit with a slightly reduced NA[†] to compensate for the smaller mode profile at the focus. Therefore, we expect the second harmonic intensity detected in the far-field along the z -axis to propagate exclusively in a HG₁₀ mode. An experimental demonstration that this is indeed the case is shown in Figure 6.2(e), which is a combination of three images of the light reflected back from the sample,

[†]The source term is maximally overlapped with the HG₁₀ mode at the second harmonic when the latter is emitted at NA $\approx$ 0.3 and hence does not fill the complete aperture. This can be seen when focussing a HG₁₀ mode at the second harmonic and checking the field distribution in focus for varying NA.

recorded with a Si detector. In order to be able to see both the fundamental and the SHG, the fundamental wavelength was set to 1300 nm, which leads to a weak image due to two-photon absorption in the silicon. On the left, the fundamental polarisation was horizontal, such that the longitudinal field component, and hence also the nonlinear source term linked to $\chi_{xxx}^{(2)}$ and the far-field SHG pattern exhibit a horizontally aligned double-lobe structure. Upon rotation of the input polarisation by 45 degrees, the far-field SHG pattern is found to rotate by the same amount, as shown in the middle. Finally, in order to prove that the two spots belong to the fundamental and second harmonic, a long-pass filter was introduced to absorb the SHG signal, with only the fundamental spot remaining on the camera, as can be seen on the right.

These results confirm the crystal orientation stated by the manufacturer and shed light on the inner workings of the second harmonic interaction inside the GaN layer. It has become clear that the z -component of the excitation fields is crucial for the second-order response detected in the far-field. Equipped with this knowledge we are in a better position to interpret and understand the experimental results that follow.

6.4 Frequency Mixing in the GaN Slab

Given the large transparency window and the nonzero second-order and third-order nonlinear susceptibilities of GaN, it is possible to observe a wide range of nonlinear processes, such as second-harmonic generation (SHG), sum-frequency generation (SFG), difference-frequency generation (DFG), and four-wave-mixing (FWM) covering large parts of the optical spectrum. In the light of setup limitations and the long-term aim to work towards photon sources producing pairs of visible and NIR photons, we will mainly be investigating mixing processes with a pump beam at the edge of the visible spectrum, around $\lambda = 850$ nm, and a second input beam to seed the process either in the NIR, or – for the reverse process – in the visible waveband. Details of the experimental setups can be found in Appendix A. Efficiency calculations of those nonlinear processes within the stimulated emission tomography (SET)

framework [56] will be presented in the separate analysis Section 6.5, where absolute values for the frequency conversion efficiencies and photon pair generation efficiencies in the quantum limit will be estimated. The SHG signal from the NIR seed beam is a very useful side product, which will be analysed for additional insights into the GaN system.

6.4.1 Dual-Beam Study with Tilted Sample

As a first study, the sample is pumped at $\lambda = 840$ nm and 1500 nm, while the reflected second- and third-order nonlinear signals are detected under variation of the sample tilt angle. The motivation behind this experiment is to vary the E_z component of the fundamental field, compare the second-order signal to the third-order signals, and build on the arguments established in Section 6.3. Figure 6.3(a) shows the spectra of SHG of the NIR seed beam, degenerate four-wave-mixing (DFWM) where two pump photons at $\lambda = 840$ nm are annihilated and one photon in the NIR seed field at $\lambda = 1500$ nm and one photon at $\lambda = 583$ nm are created in each elementary interaction, and SFG, where for each photon generated at $\lambda = 543$ nm one pump photon and one seed photon are annihilated.

The first observation is that the DFWM signal, which is a third-order nonlinear process, decreases, while the two second-order processes SFG and SHG increase with the sample tilt angle. These trends can be seen more clearly in Figure 6.3(b), where the counts of each peak were integrated up and plotted against the tilt angle. While the decrease of the DFWM signal can be attributed to the single factor of less emission within the acceptance angle of the objective lens, when the sample is at a larger angle, there are two possible reasons for the strong increase in the second-order signal. Firstly, the E_z component coupling to $\chi_{xx}^{(2)}$ is growing with the sample tilt angle, leading to a stronger emission of second-order signal through the top of the GaN layer at small angles. Secondly, the nonlinear polarisation due to $\chi_{zzz}^{(2)}$ and $\chi_{zxx}^{(2)}$ is now oscillating at a finite angle with respect to the beam axis, which, according to the radiation characteristics of dipoles should lead to some emission along the beam axis, although this effect is expected to be weakened by total internal reflection at the GaN-air interface.

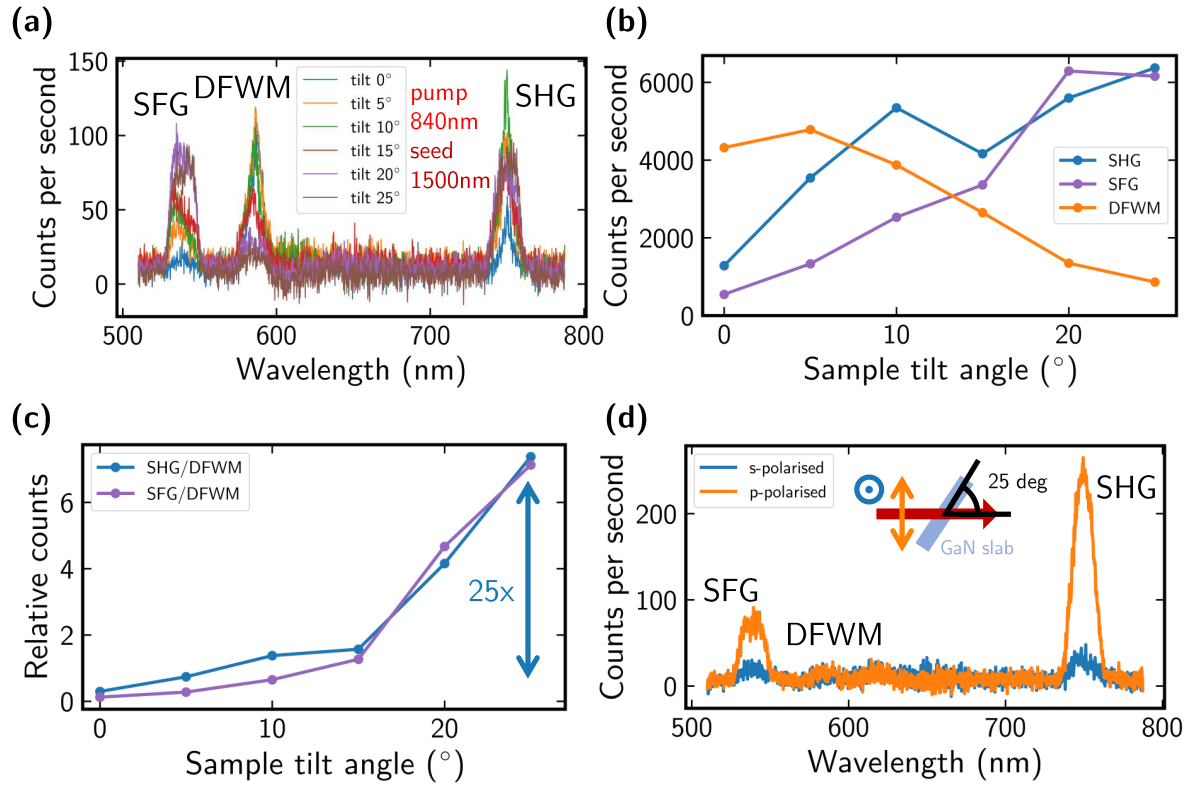


Figure 6.3 – Impact of sample tilt angle on nonlinear signal. (a) Spectra of SHG, DFWM, and SHG signals at varying sample tilt angles. The third-order signal decreases due to a less efficient collection when the sample is at an angle with respect to the objective lens. However, the second order signals increase due to a stronger field component along the z -axis of the GaN layer coupling to the $\chi_{xxx}^{(2)}$ tensor element, as well as a gradually better emission in the backward direction of the nonlinear polarisation resulting from $\chi_{xxx}^{(2)}$ and $\chi_{zzz}^{(2)}$. (b) Integrals over the SHG, DFWM, and SHG peaks in (a) as a function of the sample tilt angle to show more clearly the respective trends. (c) Taking the DFWM signal as a proxy for the collection efficiency at a given tilt angle, the intensity build-up of SHG and SFG inside the GaN slab at a tilt angle of 25° is estimated to be about 25 times the intensity at normal incidence. This strong increase can be attributed to the reasons laid out in (a). (d) Comparing s- and p-polarised excitation demonstrates that the contributions from $\chi_{zzz}^{(2)}$ and hence also from $\chi_{zzx}^{(2)}$ have only a weak impact on the increased radiation at the second harmonic propagating along the beam axis, establishing the stronger E_z component coupling to the $\chi_{xxx}^{(2)}$ tensor element as the main factor.

If we accept that the DFWM signal tracks the collection efficiency of the objective lens as a function of the tilt angle, a rough estimate of the SHG and SFG enhancement inside the GaN slab can be determined by considering the ratio of second-order to third-order signal, plotted in Figure 6.3(c). The growth of the enhancement accelerates when increasing the tilt angle and at a tilt angle of 25 degrees reaches about 25 times the value at normal incidence. Finally, in order to decide which of the two driving factors behind the increase in second-order signal dominates in this experiment, two separate spectra at a sample tilt angle of 25 degrees were taken, one with a seed beam perpendicular to the plane of incidence (s-polarised) and one parallel to it (p-polarised), while the pump beam remained p-polarised. While the drop in SFG signal for the s-polarised seed beam is expected as $\chi_{zxy}^{(2)} = \chi_{zyx}^{(2)} = 0$ in a wurtzite crystal [183], the SHG signal shows that the E_z component coupling to $\chi_{xxz}^{(2)}$ must be the main driving factor behind the observed increase of the second-order signal. This is because both for s- and p-polarisation there are oscillations along the c-axis of the GaN layer at the second harmonic, but only for p-polarisation the E_z component is increased and only in that case the collected SHG signal is enhanced. The fact that the DFWM peak has almost completely disappeared is due to the different alignment for this measurement. The objective lens suffers from chromatic aberration and hence, given the different scaling of SFG and DFWM with the pump intensity, the z -position of the sample can be adjusted to maximise either signal, in this case SFG, while the other one may be weakened.

6.4.2 SHG Wavelength Scan

Next, a single beam study was conducted with only the NIR beam pumping the GaN film at wavelengths varying between 1200 nm and 1600 nm. The goal of this experiment was to obtain reference data for a similar wavelength scan on the GaN cones in Figure 6.10 to identify potential resonant features. The reason for only using a single beam instead of conducting SFG measurements with a fixed pump beam was to simplify the experimental apparatus and to avoid signal fluctuations related to the relative alignment of the input beams when performing the wavelength scan.

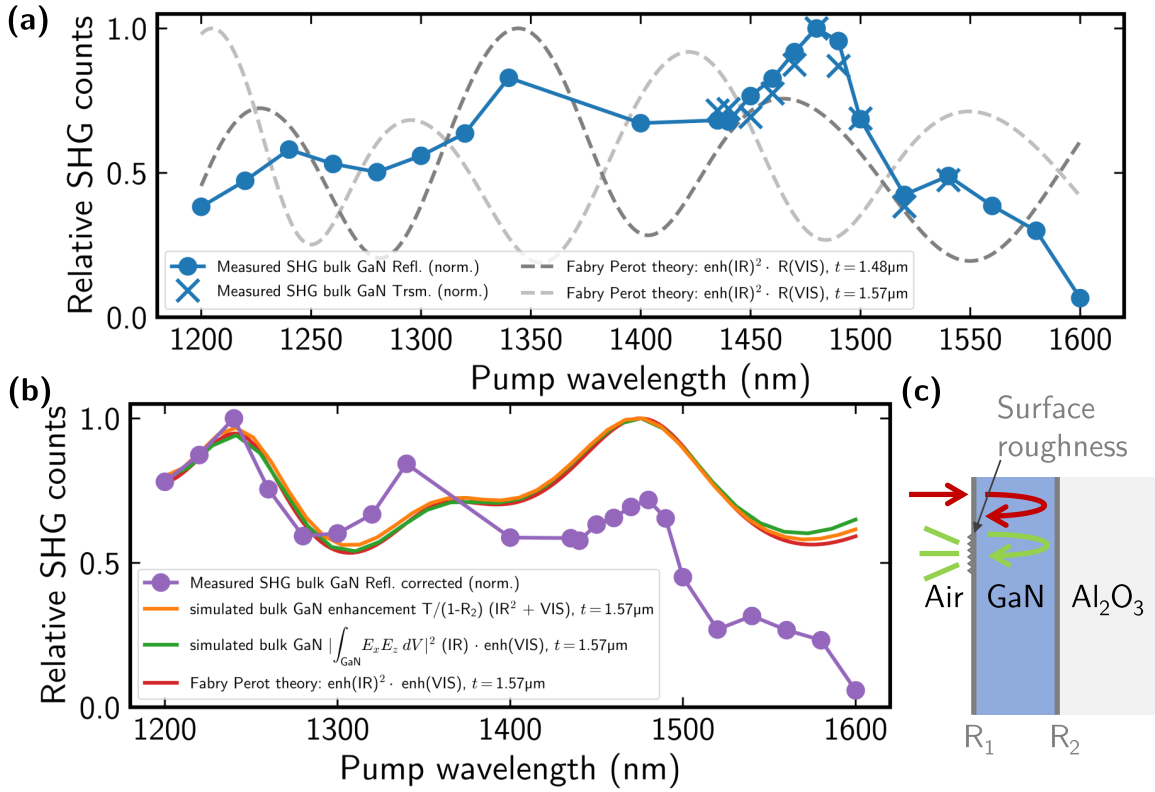


Figure 6.4 – Wavelength scan of SHG intensity. (a) The measured SHG intensity (blue dots) reflected from the sample exhibits several peaks as a function of the wavelength, which point towards a Fabry-Perot (FP) model of the GaN layer. A first attempt with a thickness of $1.48\ \mu\text{m}$ (dashed dark grey) leads to maxima with correct spacing and alignment, however, the modulation depth is unrealistically large and within FP theory a local maximum in the reflectance corresponds to a local minimum in transmittance, contradicting the measured transmitted SHG (blue crosses), which shares its local maximum with the reflected SHG data rather than exhibiting a local minimum. (b) A better agreement is achieved when correcting the reflected SHG data for setup dispersion (purple) and comparing it to the joint intensity enhancement at the fundamental and SHG wavelengths. FDTD simulations agree extremely well with the analytical FP result (red), both when taking the transmittance at the fundamental and second harmonic wavelengths as a proxy for the intensity enhancement inside the slab (orange) and when integrating the nonlinear source term with field data at the fundamental wavelength (green). Note that the modelled thickness had to be slightly adjusted to $1.57\ \mu\text{m}$ and that the FP reflectance at that thickness has its local minima aligned with the maxima in the data (light grey in (a)). (c) Schematic representation of the GaN layer as a FP cavity and scattering due to surface roughness at the GaN-Air interface. The initially surprising result that the measured reflectance is best modelled by the FP joint transmittance can be explained by scattering at the first interface, which, if dominating over the FP reflectance at that interface, acts as a probe of the joint intensity enhancement inside the slab.

Both the transmitted and the reflected SHG signals were detected as a function of the excitation wavelength. The results are presented in Figure 6.4. When examining the reflected SHG data (blue dots) in Figure 6.4(a), we find regularly spaced local maxima which are typical

for layered systems where multiple reflections give rise to this succession of constructive and destructive interference according to Fabry-Perot (FP) theory [116, 184]. As a starting point for the thickness of the GaN layer, we chose $1.5\ \mu\text{m}$ because of the specified thickness of about $3\ \mu\text{m}$ and because approximately $1.5\ \mu\text{m}$ were etched away, as mentioned above. Note again that the GaN slab measurements were taken on the unstructured, uniformly etched region next to the cones – the fabrication process will be explained later in section 6.6 – rather than on a separate sample. Using refractive index data of GaN and sapphire, the reflectivity of the front and back surface was calculated with Fresnel's equations at all wavelengths of interest, which determines the reflectance R , transmittance T , and internal intensity enhancement of the GaN layer at all wavelengths for a given thickness. A summary of the FP model including the interpolated refractive index data is given in Appendix B.

A reasonable attempt to explain the reflected SHG data in Figure 6.4(a) is to multiply the squared FP intensity enhancement inside the GaN slab at the fundamental wavelengths by the FP reflectance at the corresponding SHG wavelengths. Adjusting the modelled thickness to $t = 1.48\ \mu\text{m}$ to match the local maxima of the measured SHG signal as closely as possible leads to the dashed dark grey curve. The short modulations are due to the reflectance at the second harmonic, while the underlying larger modulation which lifts up the central maximum and lowers the outer local maxima results from the enhancement at the pump wavelengths.

Despite appearing reasonable at first, there are two main issues with this approach. The first problem is the large modulation depth of the FP reflectance curve, which is at least two times larger than that of the data. The other problem is that according to FP theory, maxima of the reflectance not only correspond to minima in the transmittance, but more importantly, to minima of the internal intensity enhancement due to destructive interference [116]. In other words, the measurement of the reflected SHG would have pointed us to minima of the SHG intensity build up, rather than the maxima. This theoretical argument is supported by transmitted SHG data (blue crosses) which forms a local maximum at the reflected SHG maximum around $1480\ \text{nm}$ rather than a local minimum. Therefore, when adjusting the GaN thickness of the model such that maxima of the field enhancement at the second harmonic

match up with the SHG data, which occurs at $t = 1.57 \mu\text{m}$, the FP reflectance curve is exactly offset by half a modulation period (light grey), but the joint intensity enhancement curves are in much better agreement with the data, as shown in Figure 6.4(b). Thus, despite trying to explain the wavelength dependence of the *reflected* SHG signal, the intensity enhancement inside the slab, which by FP theory is proportional to the *transmittance* [116], seems to explain the data much better.

Note that the reflected SHG data in Figure 6.4(b) is the same as in Figure 6.4(a) but was corrected for setup dispersion in order to see the FP interference pattern more clearly (purple), however, there is still an almost linear underlying trend which might be due to an alignment drift of the setup. Shown in red is the joint intensity enhancement predicted by FP theory, where the intensity enhancement at the fundamental was squared, as in the grey curves in Figure 6.4(a), because the SHG intensity quadratically depends on the pump power. This analytical result is supplemented with two datasets from a Finite-Difference-Time-Domain (FDTD) simulation with the commercial software Lumerical. The orange curve was obtained by multiplying the squared transmission data at the fundamental with the transmission at the second harmonic, divided by $(1 - R_2)$ at each wavelength. According to FP theory, the internal intensity enhancement can be determined indirectly via this approach due to the aforementioned proportionality of transmitted and circulating intensities [116] which can also be seen in Appendix B. The difference to the red curve is only that the transmittance was not calculated from the interpolated refractive index data and Fresnel's equations, but obtained from the simulation. For the green curve, the calculated intensity enhancement at the fundamental was replaced by a direct integration of the nonlinear source term inside the GaN layer according to the simulated electric field data.

The good agreement of the three curves is expected, as the layered system is exactly described by FP theory. A possible explanation for the good agreement of the measured SHG reflection data with the joint intensity enhancement – proportional to the FP transmittance, not the reflectance – is schematically drawn in Figure 6.4(c). Roughness at the GaN-air interface leads to scattering, which, according to the data, seems to be stronger than the FP

reflectance; hence why the SHG signal in reflection, just like in transmission is proportional to the SHG build-up in the GaN layer. Even if the linear reflectance at the second harmonic wavelength were stronger than the effective scattering at the GaN-air interface, the destructive interference of the SHG in the GaN slab would be key, because we are not dealing with a linear scattering problem, but with a nonlinear one, where the light input at the fundamental is to be maximised and the outcoupling of the second harmonic from within the GaN layer to be optimised for a local maximum of the measured SHG. This explanation which relies on some scattering at the first interface but also some FP oscillations in the GaN slab is certainly quite crude, however, there does not seem to be a simple alternative explanation avoiding more detailed simulations of the nonlinear interaction, as mentioned in Appendix B. Remaining with the scattering-based explanation, it is possible that the underlying linear trend in the SHG data (purple) in Figure 6.4(b) is caused by a decreasing scattering efficiency of the interface at longer wavelengths, although there is no further evidence to support this speculation for now.

6.4.3 Dual-Beam Frequency Mixing: Up- and Down-Conversion

Let us now return to frequency mixing with two input beams and present several parameter sweeps to prove that the nonlinear mixing signals discussed in this chapter behave as expected and are not spurious signals that happen to be at the correct wavelength. This is less of a concern for the up-conversion signals, which are easy to obtain experimentally, have a high signal-to-noise ratio even without time-consuming alignment optimisation, and appear in the spectra only if both input beams are launched simultaneously; however, the NIR signals of DFWM and DFG are hard to obtain due to our smaller and much less efficient IR detector and a lower grating efficiency in our spectrometer at those wavelengths. Often, even after lengthy optimisation of the setup and applying considerable input powers, the resulting signals are only weakly above the noise, with signal-to-noise ratios below 5.

Therefore, any weak signal appearing around the expected wavelength during the alignment procedure could easily be mistaken for the desired nonlinear signal, if rigorous tests of the

expected properties were not undertaken to confirm the nonlinear process by which it was generated. As far as the expected wavelength is concerned, a further problem is that any fluctuations in the pump fields in the visible have a big impact on the nonlinear signal in the NIR, as a fluctuation of 1 nm around $\lambda = 500$ nm corresponds to a 3 nm change around $\lambda = 1500$ nm due to energy conservation, making sole reliance on the wavelength information of a single spectrum to confirm the nature of a signal even more questionable. Consequently, the following discussion will be focussing on the properties of the NIR signals and only briefly cover the signals in the visible waveband.

Second-Order Processes

Figure 6.5 shows the second-order mixing processes SFG and DFG in the GaN slab, where for SFG the pump beam at the transition from visible to IR at $\lambda = 840$ nm interacts with a seed beam at $\lambda = 1500$ nm to generate photons at $\lambda = 543$ nm. The DFG process is the corresponding down-conversion process pumped at $\lambda = 543$ nm, where the input beam at the transition from visible to IR now plays the role of the seed rather than that of the pump. Due to frequency instability of the visible pump beam, the input beam at $\lambda = 840$ nm was slightly shifted to $\lambda = 850$ nm to achieve stable operation of the laser system while mixing almost the same wavelengths. The effect of scanning the wavelength of the seed beam on the SFG signal while keeping the pump and seed power constant is shown in Figure 6.5(a). As expected from energy conservation, the SFG signal is shifting to longer wavelengths when the seed wavelength is increased. Moreover, the signal strength does not exhibit resonant features and is largely constant, except for some small fluctuations that are linked to variations of the alignment. Each column of the plot is a spectrum from a separate measurement where a slit was shifted between two gratings to select the seed wavelength, with the alignment optimised and the power measured for normalisation of the spectrum at each measurement. This is the reason for the decreasing signal-to-noise ratio towards the edges of the plot.

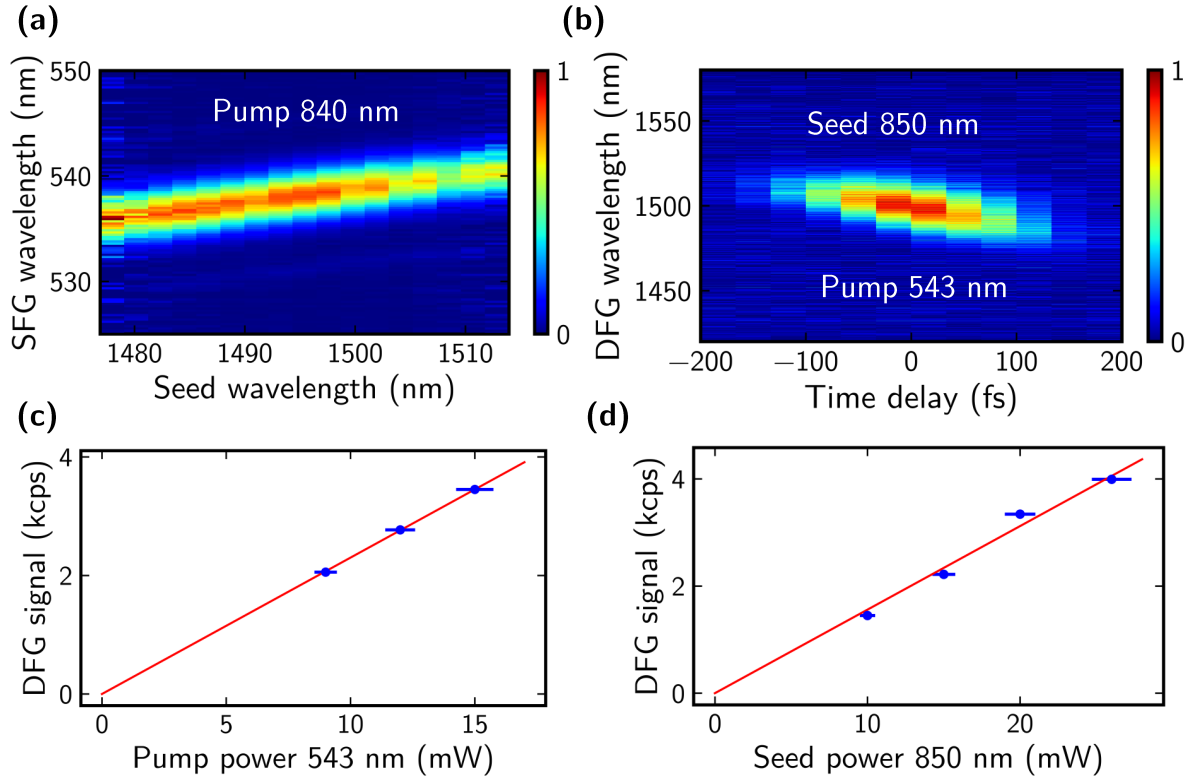


Figure 6.5 – Second-order frequency mixing in the GaN slab. (a) Normalised SFG spectra with varying seed wavelength and constant pump beam. The shift of the SFG peak is due to energy conservation. No resonant features appear in the relative intensities of the spectra. (b) DFG spectra, where the roles of seed and signal are swapped compared to SFG. The DFG signal only appears when the pump and seed pulses coincide in time, with a characteristic width of about 200 femtoseconds, implying an approximate pulse duration of 150 femtoseconds. The spectral shift of the signal indicates a chirp of the pulses. (c) Linear power dependence of the DFG signal when varying the pump power at constant seed power. (d) Linear power dependence of the DFG signal when varying the seed power at constant pump power. The DFG process is thus demonstrated to be linear in both input fields.

For the DFG signal it was not possible to conduct a frequency scan in this fashion because too much seed power was lost at the slit between the two gratings and an acceptable signal yield could only be achieved when removing it. Instead, a time delay scan and power scan of the input fields were measured. Figure 6.5(b) pictures the DFG spectra under variation of the time delay between the seed and pump pulses centred around $\lambda = 850$ nm and $\lambda = 543$ nm, respectively. The time signature of the signal is consistent with a pulse duration of ~ 150 fs and proves that the frequency mixing process involves both input fields. There is a frequency drift as a function of the delay time which is due to chirp of at least one of the input fields; in this case it is likely the pump, as its beam path contains two gratings, but other dispersive

elements in the beam path, such as lenses, are contributing to this phenomenon as well. Figure 6.5(c) shows the DFG signal as a function of the pump power, while the seed power is kept constant. The trend is linear, which is also the case for the power scan in Figure 6.5(d), where the seed power is varied at a constant pump power. These results confirm that the signal is linear in both input fields, which is expected for DFG, where one pump photon is annihilated and one seed photon is stimulating the process for each signal photon generated.

Third-Order Processes

An analogous study of third-order frequency mixing led to the results presented in Figure 6.6. The process observed here is DFWM, with two pump photons at $\lambda = 840$ nm annihilated and a pair of photons at $\lambda = 583$ nm and 1500 nm created in each elementary event; either of the latter two fields was stimulated with a second input – the seed beam. Scanning the seed wavelength at a fixed pump wavelength and power is shown in Figure 6.6(a). The spectra were taken together with those in Figure 6.5(a); both are in fact sections of the same spectra, as the pump and IR seed input configuration simultaneously creates a second-order and third-order signal in the GaN layer. Again, the nonlinear signal, in this case DFWM, shifts with the seed beam in accordance with energy conservation. Measuring the reverse DFWM process with the seed beam in the visible waveband leads to the data in Figure 6.6(b)–(d). The DFWM signal appears only when both pulses interact with the sample at the same time, which gives rise to the time delay dependence shown in 6.6(b). The chirp is stronger than in 6.5(b), supporting the hypothesis that mainly the pump beam is responsible for the shape of this trace. It is likely that alignment imperfections are the reason for its slight asymmetry. The power series in Figure 6.6(c) and (d) are analogous to Figure 6.5(c) and (d), with the expected difference being the quadratic pump power trend in Figure 6.6(c). This is confirming that two pump photons are participating in each elementary interaction.

Having characterised and confirmed the genuineness of the various second-order and third-order processes, we are now in a position to answer questions about their absolute and

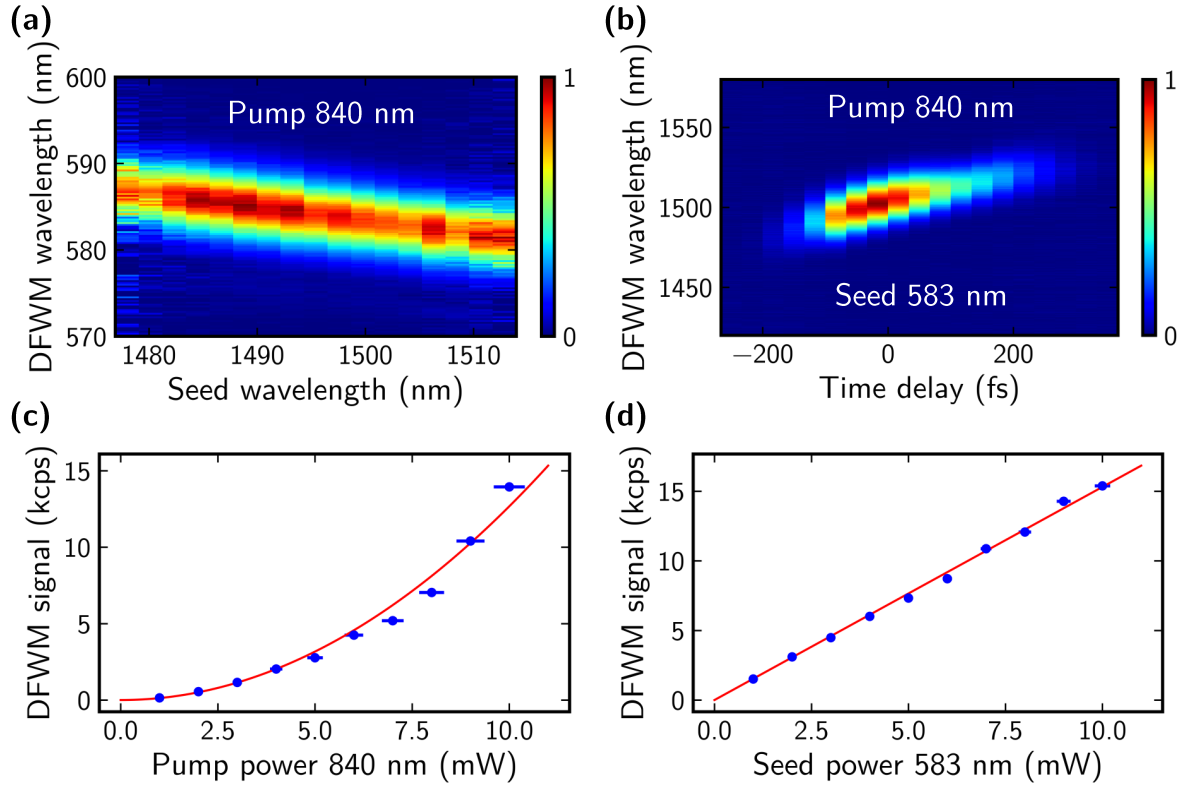


Figure 6.6 – Third-order frequency mixing in the GaN slab. (a) Normalised Degenerate Four-Wave-Mixing (DFWM) spectra with varying seed wavelength and constant pump beam. The shift of the DFWM peak is due to energy conservation. No resonant features appear in the relative intensities of the spectra. Data taken together with the spectra in Figure 6.5(a). (b) DFWM spectra in down-conversion configuration, where the roles of seed and signal are swapped compared to (a). Similarly to Figure 6.5(b), the mixing signal only appears when the pump and seed pulses coincide in time. The spectral shift is stronger than in Figure 6.5(b) and a slight asymmetry of the plot hints at an alignment drift with the delay stage position. (c) The quadratic power dependence of the DFWM signal when varying the pump power at constant seed power confirms that two pump photons are annihilated in each elementary DFWM event. (d) Linear power dependence of the DFWM signal when varying the seed power at constant pump power.

relative efficiencies, as well as estimate the photon pair production rates in the corresponding spontaneous processes via the stimulated emission tomography (SET) technique [56].

6.5 Conversion Efficiency and SET Analysis

Since the conversion efficiency analysis and the estimation of the spontaneous photon pair production rates expected in SPDC or SFWM measurements with the SET technique [56] are based on the same experimental results but constitute distinct aspects of the analysis,

this section will be structured as follows. First, the assumptions in the SET paper [56] are compared to our experimental parameters to justify the applicability of the SET analysis. Next, the relevant aspects of the setup are explained to clarify how the photon numbers for determining the frequency conversion efficiencies and the SET calculations were obtained. Also, all values immediately relevant to the analysis are summarised in Table 6.1 in that subsection. After that, the SET calculations are discussed in detail for the example of DFG and the expected SPDC photon pair creation rate. Next, the frequency conversion efficiency is discussed for the example of SFG and compared to recent results in the literature. Finally, we shall compare the obtained frequency conversion efficiencies to each other to demonstrate the reciprocity of up- and down-conversion processes, which will also be justified theoretically.

6.5.1 SET Assumptions

In a SET analysis, spontaneous photon pair generation rates are estimated from the corresponding stimulated experiment with an additional input beam seeding the process, as explained in Chapter 2. The centrepiece of this analysis is Eq.(2.30) which states that the number of stimulating photons is essentially the constant of proportionality between the stimulated and the spontaneous photon counts [56]. It is of course necessary to check if the experimental conditions violate the assumptions under which this result was derived. The input beams consist of ~ 150 fs pulses with a repetition rate of 80 MHz, which means that at mW average powers there are on the order of 10^8 photons in each pulse with negligible relative number fluctuations. Therefore, the assumption of a negligible contribution from the unseeded processes is easily satisfied. Also, it can be safely assumed that the interaction of each pulse is identical. Next, the assumption of a sharply peaked $\vec{k}$ -distribution of the seed pulses needs to be examined. Within the effective interaction volume, extending one Rayleigh range around the focus, there is only a moderate directional spread in $\vec{k}$ despite the relatively high 0.9 NA of the focussing objective.

The more critical aspect is the spread of $|\vec{k}|$, or the pulse spectrum. In the paper, a continuous-wave (CW) experiment with an approximate coherence time of 10 μ s is ana-

lysed as a representative example which easily satisfies the assumptions. For our experiment with approximately Fourier-limited pulses the coherence time is eight orders of magnitude shorter and it is not obvious that this situation is similar enough to a CW experiment to introduce an error less than an order of magnitude. In the time domain, the CW and pulsed experiments differ by the rate of change of the pulse envelope within one optical cycle.[†] The largest wavelength appearing in our mixing experiments is 1500 nm, which means that each pulse contains 30 optical cycles. For a Gaussian envelope of e^{-t^2/τ^2} , where τ is the pulse duration, the maximum slope is $1/\tau$ and thus the maximum relative change of the envelope function within one cycle is $1/\tau \cdot T = \lambda/c\tau \approx 3\%$. In the CW case with a coherence time of 10 μs this maximum relative change would have been 10^{-9} . It is reasonable to say that our pulses are long enough not to add any problems to the analysis.

Finally, the assumption of a constant energy density in the interaction volume needs to be discussed. For Gaussian beams striking a nonlinear medium this is not true, unless the in-plane extent of the medium is at least an order of magnitude smaller than the spot size. However, it is possible to find an effective interaction region with constant energy density inside, which produces an equivalent amount of interaction and for which the $1/e^2$ spot size is a natural choice. Moreover, this aspect is also relevant for the example in the SET paper [56], where it was not mentioned as problematic to the analysis.

As for the energy density being constant in the longitudinal direction, a similar argument to the coherence time discussion above applies, namely that for a GaN slab thickness of 1.57 μm there is a maximum relative change of the optical fields of $\sim 3\%$, which is completely unproblematic for conducting an order of magnitude estimation. Therefore it can be concluded that there are no obvious obstacles preventing the application of the SET framework to our measurements.

[†]The author would like to thank Dr Paul Kinsler for suggesting this argument.

6.5.2 Analysis Framework and Overview of Results

Before providing explicit values for the photon pair generation rates expected in spontaneous processes, it is necessary to comment on their meaning. In each SET calculation, we will be comparing two experiments with the same pump configuration but with the seed beam on and off, respectively, as explained in Chapter 2. The spectral width of the signal in the seeded experiment is determined by the spectral width of the seed beam due to energy conservation. The coherence length of GaN, which is at least[†] $2.5 - 3 \mu\text{m}$ in our wavelength range [183] is larger than the GaN slab thickness. Therefore, in absence of any significant phase-matching, the spectrum of the spontaneous signals is expected to be much broader than the stimulated ones, since vacuum fluctuations throughout the spectrum can interact with the pump photons. Thus, the SET prediction of the spontaneous photon pair creation rate refers to pairs created within the seeded bandwidth, which corresponds to an unseeded experiment where suitable bandpass filters are placed in the signal and idler arms before the coincidences are counted.[‡]

Let us now discuss the relevant experimental details and the way in which the photon numbers for the analysis were obtained. A schematic overview of the key aspects is depicted in Figure 6.7(a). The pump and seed beams are focussed onto the GaN sample with the same objective, and the nonlinear signal is collected with a second objective behind the sample for this set of measurements. The complete setup is shown in detail in Appendix A in Figure A.4. Not all photons generated by the nonlinear process in question – here, SFG is shown as an example – are collected and sent to the spectrometer as a free-space beam. The estimation of the fraction that is collected allows for converting the expected *collected* photon counts in a spontaneous experiment to the number of photons *generated* and will be treated in the next subsection, when the photon pair generation rates inside the sample are discussed. In order to relate the photon counts from the spectra to the number of photons sent towards

[†]Follows from the GaN coherence lengths given in Ref. [183] for SHG with a 1064 nm fundamental beam and the fact that the dispersion for our wavelength combinations is flatter – see Figure B.1.

[‡]The author would like to thank Dr Alex Clark for pointing this out.

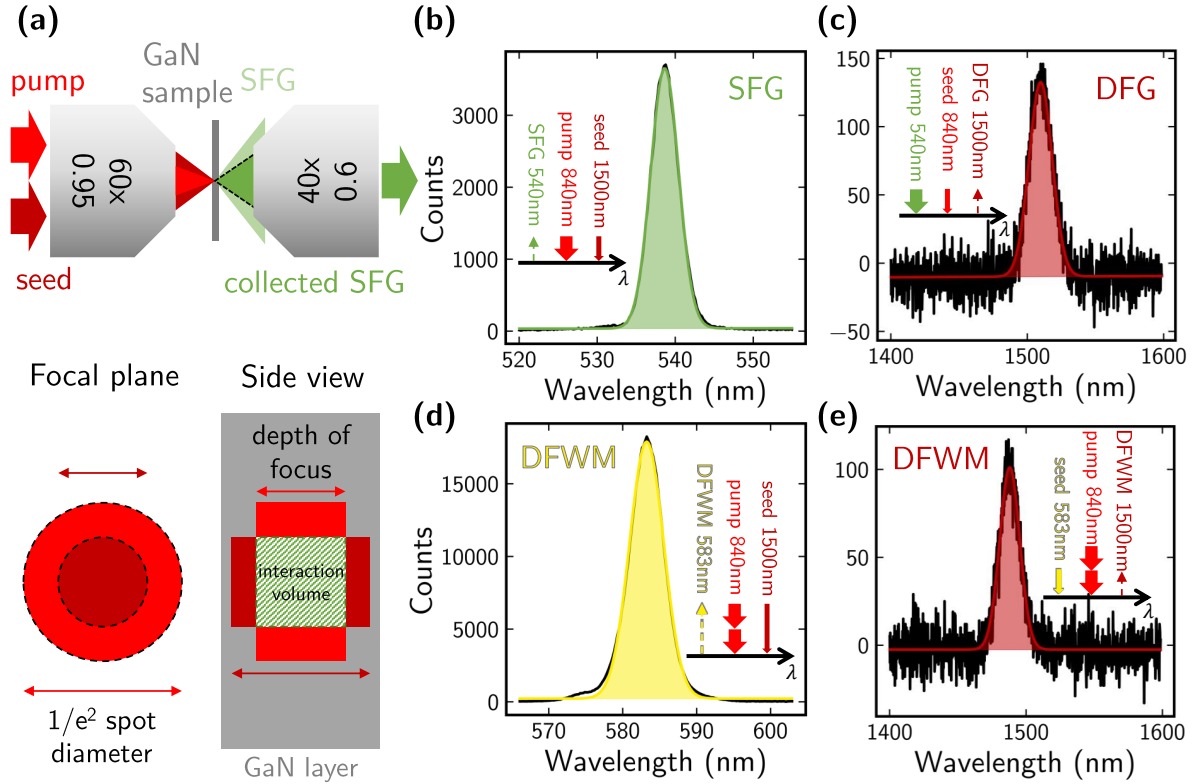


Figure 6.7 – Efficiency analysis. (a) Schematic of the experimental situation for the case of SFG measured in transmission. Pump and seed beams are focussed onto the GaN sample and the nonlinear signal is partially collected with a second objective behind the sample and sent to the spectrometer. The nonlinear interaction volume is taken as the overlap of the cylinders defined by the depth of focus and the $1/e^2$ diameter of the pump and seed beams. (b)-(e) Spectra of the nonlinear up-conversion and down-conversion signals used for the efficiency and SET analysis. The spectral positions of the involved photons are schematically drawn in the insets. (b) SFG, (c) DFG, (d) DFWM up-conversion, (e) DFWM down-conversion.

the spectrometer, calibration measurements with a known photon number were conducted, which are presented in Appendix C.

Due to the various experimental uncertainties, only an order of magnitude estimate will be given here, which allows for simplifying approximations and assumptions to be made. In the sample there is an imperfect overlap of the two input beams, which is visualised in the lower part of Figure 6.7(a). Here, the effective nonlinear interaction volume will be approximated by the shared focal volume of both beams, where the focal volume is taken as a cylinder whose length is the depth of focus, and whose base area is given by the $1/e^2$ spot size of each beam. This is an essential aspect of the analysis, as not all pump and seed photons are participating in the interaction, which, if neglected, would lead to serious errors. Additional

information about the beam size calculations can be found in Appendix C. For each process, a spectrum was taken, which is shown in Figure 6.7(b)-(d) where an inset indicates the wavelengths of the input beams and output photons for each spectrum.

We are now in a position to summarise our analysis framework. Given the average input powers of the pump and seed beams and the spectrometer counts, we need to find the number of pump and seed photons interacting, as well as the number of photons generated in the nonlinear process in order to perform the SET calculations and to state the frequency conversion efficiency. On the input side, this link is established via the transmittance of the focussing objective and the calculation of the intersection of the focal volumes approximated as cylinders. On the collection side, the fraction of photons generated inside the sample collected by the second objective, the transmittance of the collection objective, and a calibration measurement of the detection beam path including the spectrometer are needed to link the measured counts to the number of photons generated in the sample.

Let us now state the input and output photon numbers for the four nonlinear processes shown in Figure 6.7 as well as the SET estimate of the number of spontaneous photon pairs, obtained by simple division according to Eq.(2.30), and the interaction volumes. Note that the SET estimate was multiplied with the repetition rate of 80 MHz since the nonlinear interaction with the given ratio of stimulating and generated photons happens in each pulse. The information is summarised in Table 6.1 below; a more detailed version with intermediate steps is included in Appendix C. In the next three subsections we will further discuss the expected spontaneous photon pair generation rates, the frequency conversion efficiency, and the reciprocity between up- and down-conversion.

6.5.3 Spontaneous Photon Pair Generation Rates

In order to avoid redundancy, we shall only discuss the SET estimate of SPDC photon pairs from the DFG results as an example. The corresponding numbers in Table 6.1 mean that about one photon pair every ten seconds would have entered the collection objective if we

had switched off the seed beam in the DFG experiment and kept the pump beam at 2 mW average power. Note also, that due to the bandwidth argument above, those numbers refer to photon pairs within the spectral width of the DFG signal and the seed photons, which are roughly ranging from $\lambda = 1495\text{-}1525$ nm and from $\lambda = 845\text{-}855$ nm, respectively. In an unfiltered experiment, the photon pairs would be distributed over a broader spectrum, and assuming perfect detection efficiency, coincidence rates above 0.1 Hz would be observable; however, a meaningful a priori estimate is difficult. Let us therefore estimate the photon pair generation rate inside the sample within the given bandwidth.

From the discussion in Section 6.3 it is clear that the main DFG contribution, which is due to $\chi_{zxx}^{(2)}$, is radiated into the slab, as shown in Figure 6.2(c). For a rough estimate of the ratio of this contribution compared to the one related to $\chi_{xzx}^{(2)}$ which propagates along the beam axis it is helpful to consider Figure 6.2(a) showing that $E_z/E_x \sim 1\%$, meaning that two orders of magnitude are lost due to the unfavourable emission pattern. Moreover,

	kcps Fig. 6.7	ph/s collected	ph/s stimulating	ph/s spontaneous	eff. pump power (mW)	interaction length (μm)	interaction area (μm^2)
SFG	77	$7 \cdot 10^6$	$11.6 \cdot 10^{15}$	0.05	5	0.85	0.53
DFG	1.6	$32 \cdot 10^6$	$2.7 \cdot 10^{16}$	0.1	2	1.25	0.53
DFWM up	484	$24 \cdot 10^6$	$11.6 \cdot 10^{15}$	0.17	5	0.85	0.53
DFWM down	10	$200 \cdot 10^6$	$13.2 \cdot 10^{15}$	1.21	5	0.87	0.25

Table 6.1 – Efficiency analysis. The measured counts for all processes are taken from the spectra in Figure 6.7 and normalised to an integration time of 1 second. Dividing by the overall spectrometer detection efficiency and the transmittance of the second objective at the respective wavelengths leads to the number of collected photons. The number of stimulating photons is taken as the input photon number of the seed beam times the transmittance of the first objective and the fraction of the seed beam intersecting with the pump beam. Dividing collected by stimulating photons and multiplying with the repetition rate of 80 MHz leads to the number of photons in the spontaneous process. In order to compare these values, the pump power in the interaction volume and the dimensions of the latter need to be taken into account.

the overlap[†] of the TEM_{00} and HG_{10} modes for the E_x and E_z components needs to be considered, which can be roughly estimated by inspection to contribute a further order of magnitude. Assuming a generous factor of 2 lost before collection due to the higher 0.95 NA of the focussing objective compared to the 0.6 NA of the collection objective, we arrive at an estimated intrinsic SPDC pair production rate of ~ 200 Hz.

It is useful to compare this value to a recent SPDC study with a 300 nm LiNb film, pumped with a 685 nm CW laser at 9 mW, where 2.8 coincidences per second were detected in the interval $\lambda \sim 1350$ nm - 1400 nm [185], which demonstrates that detecting Hz-level coincidence rates at powers similar to our experiment would be a competitive result. Considering that the GaN layer thickness is about 4-5 times thicker, our pump power was about 5 times smaller than in Ref. [185], and assuming about one order of magnitude is lost in the detection of both signal and idler photons, Hz-level rates seem to be well achievable. Note that just losing one order of magnitude from photon generation to detection is only realistic with new m-plane samples where the photon output is emitted along the beam axis of the pump beam, which we shall return to in more detail at the end of Section 6.7.

In the next section, the enhancement of the nonlinear conversion efficiency due to cones etched into the GaN layer will be studied, however, from the preliminary results no benefit beyond one order of magnitude could be found. Therefore, a GaN cone on our sample is expected to support an SPDC pair generation rate of up to 2 kHz under the circumstances outlined above, which would be a competitive performance, as we have seen in the previous paragraph. A summary of the contributing factors is given in Table 6.2. Of course, even higher generation rates could be achieved with an increased pump power; however, in order to find the maximum achievable pair generation rate, the damage threshold of the GaN sample needs to be evaluated in a separate study.

[†] TEM_{00} and HG_{10} are orthogonal modes and hence have zero overlap. This argument is about the second-order polarisation not being created everywhere in the TEM_{00} profile, in particular not in the high-intensity region in the centre.

0.1 Hz	spontaneous, collected
x100	E_x/E_z
x10	overlap
x2	collection loss
x10	cone
2 kHz	spontaneous, generated

Table 6.2 – Count rate estimate. Determining photons actually generated in SPDC from photons expected to be collected in SPDC.

6.5.4 Frequency Mixing Efficiency

Let us now consider the frequency conversion efficiency and compare it to the literature in order to put the numbers obtained above into perspective and to confirm that they are realistic. For the purposes of this discussion we will work with the simple ratio

$$\eta = \frac{\text{\#output photons}}{\text{\#stimulating photons}} \quad (6.3)$$

at a given pump configuration. In all processes, MHz generation rates were achieved at mW input powers, which corresponds to a lower bound on nonlinear conversion efficiencies in the range of $\eta \sim 10^{-10} - 10^{-8}$ as the photon energies are on the order of 1 eV. For the sake of easy comparability to literature values and given that third-order processes are less interesting than second-order processes in a $\chi^{(2)}$ -material, we shall consider the SFG performance in more detail as an example.

With 7 Mph/s collected at 540nm, the output power is at least 2.5 pW at a pump and seed power of 5 mW each, leading to an efficiency of $\eta > 5 \cdot 10^{-10}$, which is five orders of magnitude lower than recent reports of SFG and SHG in dielectric nanopillars [186,187]. Taking into account all correction factors in Table 6.2, in particular those related to the unfavourable alignment of the $\chi^{(2)}$ tensor elements, we arrive at an efficiency of $\eta \sim 10^{-6}$. This is still one order of magnitude below the efficiencies referenced above, which were reported on disks of GaAs for SHG at similar wavelengths and pulse durations. Taking into account the ~ 20 times higher nonlinear susceptibility of GaAs mentioned earlier in section

6.1, the efficiencies are approximately equal; however, the effective interaction regions and the peak intensities need to be accounted for as well.

In the article by Liu and colleagues [186], the highest efficiency of $\eta \sim 2 \cdot 10^{-5}$ was achieved at a peak intensity of $\sim 3 \text{ GW/cm}^2$, a spot area of $\sim 30 \text{ }\mu\text{m}^2$ and an effective interaction length of $\sim 0.3 \text{ }\mu\text{m}$, given by the height of the nanopillars where the nonlinear interaction takes place in that work. This is in contrast with the $\sim 100 \text{ GW/cm}^2$, spot area of $\sim 0.5 \text{ }\mu\text{m}^2$ and effective interaction length of $\sim 0.9 \text{ }\mu\text{m}$ in this work. With a factor of 30^2 for the intensity, where the square is necessary due to the quadratic power dependency of SHG, $1/60$ for the area, and 3^2 for the interaction length – the scaling with the interaction area and length will be justified in the following subsection – there remains a discrepancy of a factor of ~ 135 , which can be attributed to the field enhancement in the nanopillar. An on-resonance SHG enhancement of up to four orders of magnitude inside the nanopillars compared to unpatterned regions is reported in the article [186], but it is difficult to say how thick the effectively contributing region of the GaAs substrate is, given that the measurements were taken in reflection due to the absorption of SHG by the GaAs. This leaves an unexplained discrepancy of two orders of magnitude between our frequency conversion efficiency, and what Liu and colleagues would have likely obtained with their sample under our experimental conditions, ignoring any potential damage to the sample. In other words, our intrinsic SFG efficiency estimate may be up to two orders of magnitude too optimistic, probably because we were too generous with the discount factor due to the $\chi^{(2)}$ tensor elements. Therefore, this discussion can be concluded by noting that the value of $\eta \sim 10^{-8} - 10^{-6}$ in the GaN slab is likely a realistic estimate of the intrinsic SFG efficiency under our experimental conditions.

6.5.5 Reciprocity of Up- and Down-Conversion

To conclude this efficiency analysis section, let us compare up-conversion and down-conversion processes in our GaN sample to test if they are reciprocal, which is theoretically expected due to arguments developed in Ref. [38]. GaN is approximately lossless in its transparency window, where the experiments presented in this chapter are conducted. Therefore, $\chi^{(2)}$ and

$\chi^{(3)}$ possess full permutation symmetry, where the frequency arguments of the susceptibilities may be swapped as long as the corresponding cartesian indices are swapped as well [38], and we expect nonlinear mixing processes in this regime to be equally efficient for up- and down-conversion. In particular, for the second-order pair SFG and DFG we have

$$\chi_{xxz}^{(2)}(\omega_{\text{VIS}} = \omega_{\text{pump}} + \omega_{\text{IR}}) = \chi_{xxz}^{(2)}(\omega_{\text{IR}} = -\omega_{\text{pump}} + \omega_{\text{VIS}}) \quad (6.4)$$

and for the DFWM processes

$$\chi_{xxxx}^{(3)}(\omega_{\text{VIS}} = \omega_{\text{pump}} + \omega_{\text{pump}} - \omega_{\text{IR}}) = \chi_{xxxx}^{(3)}(\omega_{\text{IR}} = \omega_{\text{pump}} + \omega_{\text{pump}} - \omega_{\text{VIS}}). \quad (6.5)$$

These equations do not rely on Kleinman's symmetry, which establishes further identities where frequency arguments and cartesian indices may be swapped independently, whenever dispersion can be neglected [38]. As the refractive index of GaN does not vary strongly between $\lambda = 540$ nm and $\lambda = 1500$ nm, see Figure B.1, Kleinman's symmetry could actually be applied here, with the same implications for the reciprocity of the aforementioned processes, but it is helpful to establish that this does not hinge on the assumption of negligible dispersion.

In order to compare the results for the different processes in Table 6.1 we need to correct for the different pump powers and interaction volumes. We need to consider the effective pump power given by the input pump power times the fraction of the pump spot that is covered by the spot of the seed beam. As for the interaction volumes, the resulting photon number scales as $n_{\text{ph,NL}} \propto A^{1-\nu} l^2$, where A, l are the interaction area and length, and ν is the order of the nonlinear process. The area scaling is based on the argument that for a ν -th order process, doubling the intensities leads to 2^ν times as much signal. Therefore, halving the interaction area at constant pump power and interaction length results in $2^{\nu-1}$ times the signal because the intensity is doubled, but the interaction volume is halved. The scaling with the interaction length follows from the phase-matching term of the nonlinear output intensity, which is equal to l^2 for perfect phase-matching [38]. As the interaction lengths in this experiment are not

larger than half the coherence length in the GaN slab, as we have seen earlier in this section, this is a reasonable approximation. When comparing the second-order processes SFG and DFG, the correction factor for SFG is 0.4 for the pump and $(125/85)^2$ for the interaction length, meaning that the photon numbers for the corresponding spontaneous processes are only a factor of ~ 2 apart, which is an acceptable result for an order of magnitude estimate. An analogous comparison for the third-order processes leads to a discrepancy of only a factor of ~ 1.6 after corrections. Both results favour the IR signal, which could be due to a systematic error, such as an underestimation of the detection efficiency in the IR. However, within the achievable accuracy we have confirmed the reciprocity of up- and down-conversion for second- and third-order processes.

6.6 Frequency Mixing in GaN Cones

High-index dielectric nanoparticles have been studied extensively in recent years as alternatives to lossy plasmonic structures for nonlinear optics applications. While the latter provide an extremely strong local field enhancement of two orders of magnitude or higher [188,189], they have the disadvantage of high losses due to interband transitions in the visible part of the spectrum [89] and Drude losses [190] caused by intraband transitions of quasi-free electrons in the conduction band, which dominate in the infrared, growing with increasing wavelength and carrier density [191]. The key idea about dielectric nanostructures, which was already briefly mentioned earlier in Section 6.1, is that the whole particle volume rather than just a thin surface layer with a thickness of the skin depth can be used for nonlinear interactions, which compensates for the lower field enhancement compared to noble metal nanoparticles [46,186,192]. In dielectric nanoparticles, the field enhancement is the result of Mie resonances, which exhibit strong displacement currents originating from bound charges, rather than the coherent charge density oscillations in plasmonic systems [29]. The field enhancement inside a particle is particularly strong at the magnetic dipole (MD) resonance, where circular displacement currents induce an oscillating magnetic dipole field in the centre of the particle [193]. This magnetic dipole is not only useful for the design of low-loss

metamaterials with an artificial magnetic response [194], but the strong light-matter interaction also allows for efficient harmonic generation [186, 195, 196]. Of course, higher-order modes also provide a significant field enhancement which boosts the nonlinear conversion efficiency [197]. In this work we would like to create single free-standing nanostructures of a pre-determined size and design in the GaN sample described in Section 6.2, meaning that the various bottom-up growth techniques are not suitable due to the insufficient separation of the individual structures and their growth in particular sizes and morphologies linked to the respective growth process [198, 199]. Therefore, and due to the lattice mismatch near the sapphire interface it was decided to etch cones into the GaN layer, such that nanoscopic volumes were defined in the upper part of the layer where the lattice quality is highest [200]. In order to create a refractive index contrast between the structure and the rest of the GaN layer, the etching was done at an angle, leaving only a thin support region between the structure and the layer, which will be explained in the next subsection. This is a simple way of creating reasonably well separated nanoscopic volumes without the need for additional lift-off and wafer-bonding processes [201, 202], which makes it interesting to investigate whether a good frequency mixing performance is achievable with such a straightforward procedure. Let us now cover the fabrication process in more detail.

6.6.1 Cone Fabrication

The cones were defined in the GaN layer with a Faraday-cage angled-etching (FCAE) technique [203–206], where even suspended nanostructures can be carved out of a high-quality bulk crystal in a simple process, as opposed to alternatives where a thin-film multilayer structure – which is harder to grow at the same quality for some materials – is used to define the nanostructures. Let us now discuss the working principle of FCAE, as explained in Ref. [206], which is illustrated in Figure 6.8(a,b). An aluminium wire mesh is arranged in a tent-like shape and placed on an aluminium base-plate to form the Faraday cage. With the sample placed inside, the cage is inserted into an inductively coupled plasma reactive ion etching (ICP-RIE) chamber. Upon entering the approximately field-free region of the cage, the ions

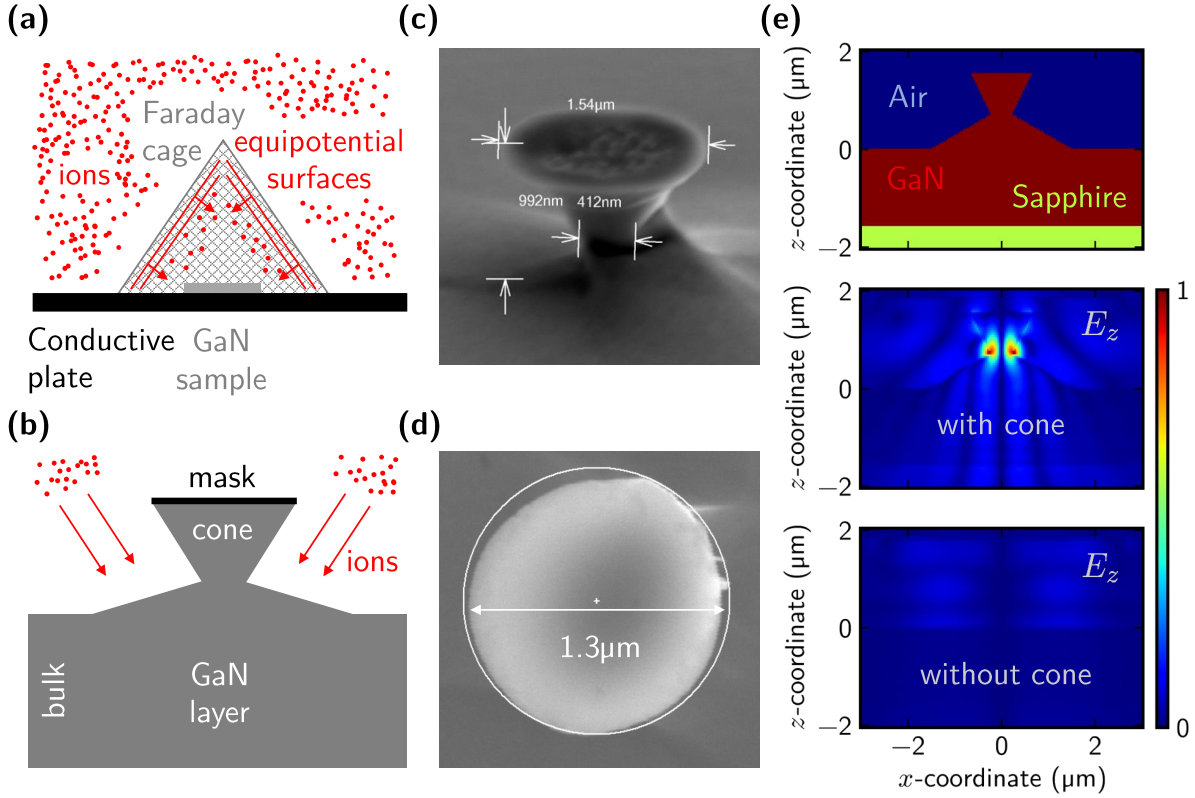


Figure 6.8 – Etching of GaN cones.[†] (a) Schematic of the FCAE technique. A wire mesh acting as the Faraday cage is placed on the baseplate of an inductively coupled plasma reactive ion etching (ICP-RIE) chamber. The equipotential surfaces inside the Faraday cage are parallel to its surface, meaning that the trajectories of the ions inside it are following the surface normals. This allows for angled etching of the sample at the centre of the Faraday cage, with the angle defined by the shape of the cage. (b) Thanks to the angled incidence of the ions it is possible to etch cones into the GaN at masked regions, rather than the pillars that would have been obtained at normal incidence. Varying the duration of the etching process allows for a control of the diameter of the neck of the cone. (c) Angled SEM image of a GaN cone on the sample. While the circular shape of the top of the cone is of good quality, the neck region is not rotationally symmetric and exhibits strong variations of the etch depth. (d) Top view of the smallest cone on the sample without major deviations from the ideal shape. This cone is modelled and investigated in the remainder of this chapter. (e) FDTD simulations of the cone in (d). The cone was replicated as well as possible following the information extracted from the SEM images (c,d). The top panel shows a cross-section of the simulation domain. Below are two plots of the absolute value of E_z at 1450 nm, with and without the cone. There is a strong field concentration around the neck.

move on trajectories perpendicular to the equipotential plane defined by the cage surface. This allows for direct control of the etching angle profile through the shape of the cage. In order to generate arrays of nano-cones it is necessary to first define the array with a hard mask and to carry out a cage-free vertical etch resulting in an array of nano-pillars. Then,

[†]The author acknowledges Dr JP Hadden for fabricating the sample, for taking the SEM image in Figure 6.8(c) and Dr Andrea Jacassi for taking the SEM image in Figure 6.8(d).

the sample is placed inside the Faraday cage for angled etching, after which the mask is removed. The plasma parameters of the ICP-RIE chamber can be identical for both etching steps, however, it is important to set the etch time such that a sufficiently wide supporting region remains underneath the cone. Note that angled etching is also possible with focused ion beam (FIB) milling, however, the Faraday cage technique is not only faster, but also avoids FIB milling problems such as Ga^+ implantation and re-deposited material [205].

A typical resulting cone is shown in Figure 6.8(c) in a scanning electron microscopy (SEM) image. While the overall shape of the cones above the supporting neck was etched with high quality, the neck itself exhibits an irregular shape with varying etch depths at different directions of the incoming ions. An additional issue was that the cone diameters were systematically smaller than the nominal cone widths, which is probably due to mask erosion [204]. The most extreme example of this is a cone of nominal width 2 μm , which was measured to be only 1.3 μm wide. This cone, shown in another SEM image in Figure 6.8(d) was chosen for the frequency mixing measurements due to its small size. Smaller nominal widths led to too thin supporting regions, which broke during either the etching or the mask removal process, such that the corresponding cones were lost. Therefore, it was not possible to observe the magnetic dipole resonance of a cone in our wavelength range of $\lambda \sim 500\text{--}1500$ nm and we decided to investigate the smallest available cone, which was expected to offer at least a moderate field enhancement due to other low-order resonances.

In order to better understand the distribution of optical fields inside the cone, an FDTD simulation was conducted with the software Lumerical. The top panel of Figure 6.8(e) shows a cross-section of the simulation domain with the cone in the centre modelled after Figure 6.8(d) and etching angle and neck size information taken from Figure 6.8(c). As the E_z component is of crucial importance for second-order processes in the GaN layer, which was established earlier in this chapter, the lower two panels depict a comparison of the $|E_z|$ distribution with and without the cone at a wavelength of 1450 nm. This wavelength corresponds to the highest measured SHG enhancement in the cone, as shown later in the experimental results in Figure 6.10(c). The key insight is that there is a strong concentration

of the fields in the neck region and that the field enhancement in the GaN slab due to the cone is only very moderate and less than one order of magnitude. The enhancement in the neck region is approximately one order of magnitude, which is much lower than in the plasmonic antennae that will be presented in the next chapter, but reasonable for a dielectric resonator.

This overview of the cones in the GaN layer can be concluded by stating that the cones significantly affect the field distribution, providing a moderate overall field enhancement and increasing the longitudinal electric field component in particular, which is expected to improve the second-order frequency mixing efficiency. However, there are two main shortcomings which are due to the early stage of this project and of the fabrication routines. The smallest surviving cone is too large to observe the magnetic dipole resonance in our wavelength range and the neck region is of irregular shape, which makes predictive simulations difficult due to the strong field concentration and hence the main contribution to the nonlinear source terms in precisely that region. Still, this study can establish analysis routines and generate insights to inform subsequent sample designs and experiments. Next, additional simulation results are discussed to build a better understanding of the cones, helping us interpret the subsequent experimental results.

6.6.2 Simulations

So far, we have seen that an increased longitudinal field component is expected in the neck region of the cone when pumping in the NIR near $\lambda = 1500$ nm. In order to put this result into context and to establish which wavelengths promise the highest nonlinear conversion efficiency, the scattering cross-section was simulated from $\lambda = 600$ nm to $\lambda = 3000$ nm, which is shown in Figure 6.9. Figure 6.9(a) compares the scattering cross-sections of the cone as it is on the sample (solid green), the cone with the connecting base but without the rest of the GaN slab (orange), and the bare cone in vacuum (blue), all normalised to the highest value of the three curves. The aim of this is to determine the position of the Mie resonances of the isolated cone in contrast with the more complicated resonances of

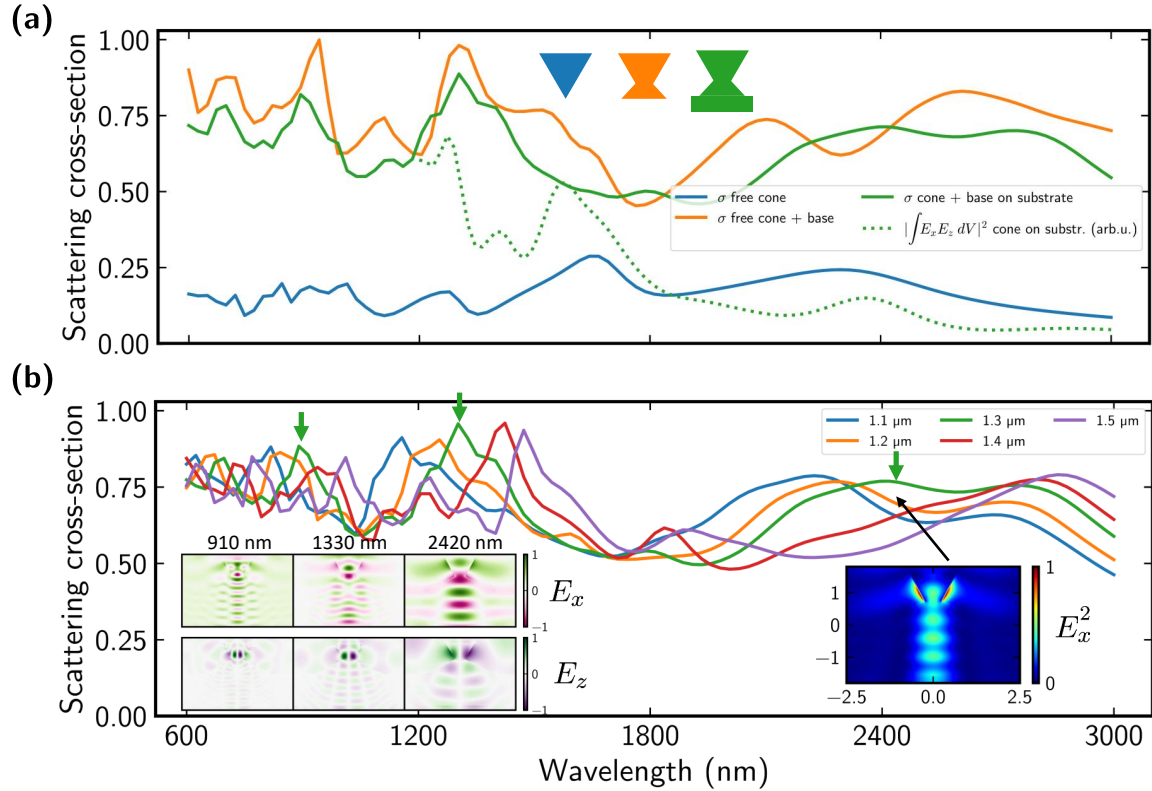


Figure 6.9 – Scattering cross-section and cone modes. (a) Simulated scattering cross-section as a function of the wavelength. The bare cone (blue) exhibits a spectrum dominated by Mie-modes with the magnetic dipole resonance at $\lambda = 2300$ nm. Adding the connecting base of the cone, but not the GaN layer and substrate leads to a more complicated spectrum, probably due to mixing of modes in subregions of the structure. Including the GaN layer and its interface with the sapphire substrate (solid green) adds an extra modulation due to FP interference in the GaN slab. The integrated second-order nonlinear source term inside the cone (dotted green) differs significantly from the corresponding scattering cross-section and is in fact most similar to the scattering cross-section of the bare cone. (b) Scattering cross-section of the cone and base on the GaN layer and substrate for varying cone radii, showing that the spectra are impacted by small fluctuations in the fabrication process, such as different degrees of mask erosion. When the cone size is reduced, the scattering maxima shift to lower wavelengths. The significant change of morphology of lowest-order mode for small variations of the cone size suggests subtle mode mixing processes in the cone and base. The inset shows the E_x and E_z distributions in the xz -plane at the local maxima of the scattering cross-section (green arrows). While the fields at $\lambda = 2420$ nm are following the pattern of a magnetic dipole and at $\lambda = 1330$ nm that of an electric dipole, the distribution at $\lambda = 910$ nm is more complicated and cannot be easily attributed to a Mie mode. The E_x^2 distribution at $\lambda = 2420$ nm – which is essentially the intensity – shows that the cone acts like an optical funnel, focussing the light into the GaN slab.

the composite systems in order to highlight the effect of the substrate and the importance of the connecting region. Of course, due to the deviation of the measured cone from its idealised modelled shape it is not possible to make exact predictions, but it is still worthwhile

to make a number of qualitative statements, as long as a detailed analysis of complicated features, which might be made obsolete by small deviations in the experiment, is avoided. The first observation is that the magnetic dipole resonance of the bare cone is shifted to longer wavelengths when adding the connecting region, which is unsurprising because the effective size of the structure is increased significantly. An interesting feature of the composite systems is the double peak structure around $\lambda = 2500$ nm which likely results from the lowest two resonances of the cone being pushed together or from a hybridisation-type mixing of the lowest order modes of the cone and base subregions. The higher order modes of all structures are very complicated and not easy to interpret. However, it is reasonable to say about the cone as it is on the sample (green) that the higher-order modes are expected to be concentrated in the main volume of the cone or in the neck region some distance away from the plane where the connecting region transitions into the GaN slab because the orange and green curves are very similar up to $\lambda \sim 1500$ nm, whereas there are big discrepancies at the lower-order modes at longer wavelengths, suggesting that those lower-order modes extend into the GaN slab. Finally, the integrated second-order nonlinear source term (dotted green) is scaled and compared to the scattering spectra in order to establish if the latter are a good indicator of an enhancement of the relevant fields inside the cone as it is on the sample. There is a reasonably good agreement with the spectrum of the bare cone down to $\lambda = 1200$ nm, however, only some peaks are aligned with those in the scattering spectra of the cone as it is on the sample. Therefore, the simulated scattering cross-section of the cone on the sample can give some indication of interesting modes, but is to be treated with caution when trying to predict wavelengths with maximum field enhancement in the cone.

Figure 6.9(b) shows the simulated scattering cross-sections of the cone on the sample under small variations of the cone diameter. Even small variations of less than 10% are demonstrated to strongly impact the position of the scattering peaks. It is interesting to consider the lowest-order mode, which drastically changes the morphology of its spectrum for these small variations of the cone size, which suggests that there could indeed be coupling of modes of the subregions.

To conclude the discussion of the simulations, let us consider the field distributions at the maxima of the scattering cross-section corresponding to the cone on the sample, indicated by the green arrows. At $\lambda = 2420$ nm the electric fields in the xz -plane are distributed as in a magnetic dipole and at $\lambda = 1330$ nm they follow the pattern of an electric dipole. These assignments refer to the patterns of the Mie modes in a spherical dielectric particle, with the analytical expressions and plots shown in Appendix D. The fields at $\lambda = 910$ nm correspond to higher-order modes which cannot easily be linked to the known spherical modes.

Considering E_x^2 , which is essentially the intensity, it can be seen that the cone focusses the light into the GaN slab, acting as an optical funnel. This concentration of the fields in the GaN slab is expected to increase the nonlinear conversion efficiency, potentially even more than the field enhancement in the cone itself, as the participating region in the slab has a larger volume and the field enhancements are only moderate in both cases. In summary, the simulations provide interesting information on the mode structure and field distribution inside the cone, confirming that the magnetic dipole resonance is far away from the wavelengths accessible in this experiment. The enhancement of the nonlinear conversion efficiency in the experiment is expected to be determined by the focussing action of the cone as well as by higher-order modes located in the cone and the neck region, however, no reliable predictions of optimal mixing wavelengths can be made due to the large impact of small size fluctuations.

6.6.3 Experimental Results

The experimental investigations of the GaN cones are divided into a study of the SHG wavelength dependency relative to the bare GaN slab in Subsection 6.4.2 in order to identify any resonant features, and a polarisation study to test if the cone modifies the polarisation distribution of the incident light significantly.

SHG Wavelength Scan

A simple test of the effect of the cone on the nonlinear conversion efficiency is a study of the reflected SHG for varying pump wavelengths, as presented earlier for the bare GaN slab in Figure 6.4. In comparison, the reflected SHG of the cone (blue dots) in Figure 6.10(a) exhibits two distinct peaks in the scanned spectral range, dominating the FP modulations. Pursuing a similar strategy to earlier and attempting to model the data with the joint intensity enhancement inside the GaN calculated from the transmittance (orange) is unsuccessful. This is likely due to the complicated shape of the structure, which strongly deviates from the flat interfaces assumed in the FP model linking transmittance and field enhancement. Replacing the field enhancement at the pump wavelengths with an integral of the nonlinear source term throughout the GaN volume (brown) does not resolve this issue, which suggests that the field enhancement term at the second harmonic needs to be replaced as well in order to better explain the data. Figure 6.10(b) shows an attempt to model the data with a combination of the integrated nonlinear source term and the simulated reflectance of the structure at the second harmonic (solid red) without involving the FP model. This time, the data shown in the plot (purple) was corrected for dispersion of the optical components in the detection beam line to only show the pure response of the sample as well as possible. The corrected data can be well understood as the corresponding GaN slab data in Figure 6.4(b) with an additional double peak structure due to the cone. While the solid red curve does still not explain the data well, the simulated reflectance at the second harmonic (dotted red) possesses a double peak structure very similar to the one in the data.

In order to separate off the pure enhancement effect of the cone, the SHG cone data was divided by the SHG data of the bulk GaN slab in Figure 6.4, resulting in the blue enhancement spectrum in Figure 6.10(c). This is possible because both datasets were taken in the same experiment; they are presented in separate figures for the sake of a more logical narrative, where the bare GaN film is studied before introducing the cone. Indeed the scaled reflectance at the second harmonic (dotted red) follows the enhancement spectrum surprisingly

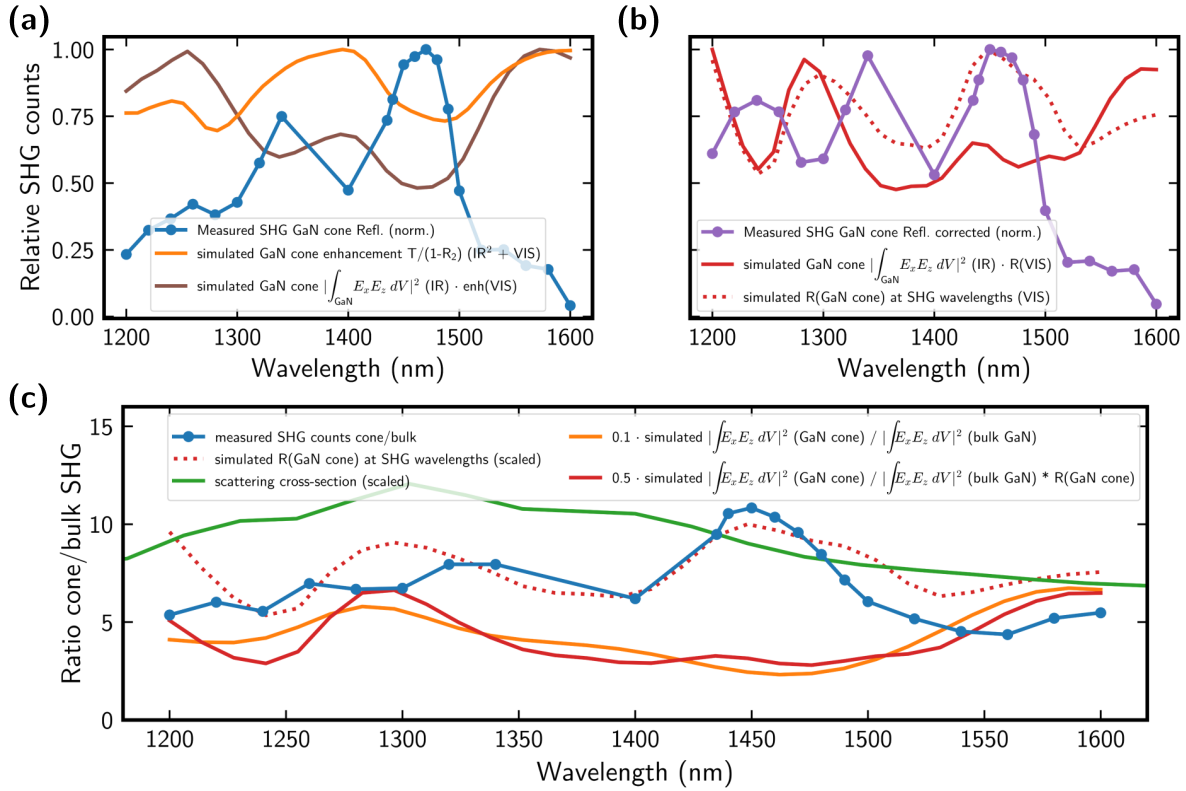


Figure 6.10 – SHG enhancement in GaN cones compared to the bulk region. (a) The reflected SHG intensity of the GaN cone (blue) is modulated due to FP interference, similarly to the reflected SHG spectrum of the bare GaN layer in Figure 6.4(a) but with the difference of two pronounced peaks in the middle of the range. Contrary to the case of the bare GaN slab, the joint field enhancement derived from the simulated transmission (orange) and the integrated nonlinear source term (brown) cannot explain the data well. (b) Analogously to Figure 6.4(b) the SHG data was corrected for setup dispersion (purple) to show the FP maxima more clearly. While an attempt to model the data with a combination of the simulated nonlinear source term and the reflectance at the second harmonic (solid red) fails to convince, the reflectance on its own (dotted red) shares the two pronounced peaks in the middle, albeit with the left peak at a slightly too short wavelength of ~ 1300 nm. (c) Dividing the SHG data from the cone by the bulk SHG data in 6.4(a) removes setup influences and shows the impact of the cone on the nonlinear conversion efficiency as a function of the wavelength (blue). Both datasets were taken in the same experiment and appear in different figures for a more logical presentation. The ratio of the integrated nonlinear source terms (orange) and that same ratio multiplied with the linear reflectance at the second harmonic (solid red) do not agree well with the data. The scaled scattering cross-section from Figure 6.9 cannot explain the data either, given that its peak is too broad and not at the right wavelength. However, the scaled linear reflectivity at the second harmonic (dotted red) does follow the measured SHG ratio remarkably well, supporting a similar interpretation to earlier, where the measured SHG is a proxy of the SHG field in the GaN layer times the reflectivity at the GaN-air interface.

well. Alternative attempts to model the enhancement spectrum with a ratio of the squared integrated nonlinear source terms in the cone and in the bulk (orange) cannot reproduce the key features, even when multiplied with the reflectance at the second harmonic (solid

red). Moreover, the simulated scattering cross-section (green) from Figure 6.9 does not match the measured enhancement spectrum either. The local maximum of the simulated scattering cross-section is located at a wavelength ~ 150 nm below the peak of the enhancement spectrum. Even allowing for a generous shift due to potential size fluctuations related to fabrication inaccuracies, it has to be noted that the peak of the scattering cross-section is several times too wide spectrally, confirming that the main peak in the measured enhancement curve is not due to a Mie resonance of the cone.

Summarising all observations made in this discussion, it appears most likely that an interpretation similar to the one adopted in the discussion of Figure 6.4 is also the best explanation for the SHG data with the GaN cone. Scattering at the GaN-air interface, characterised via the linear reflectance of the cone at the second harmonic, seems to determine the strength of the reflected SHG signal, even though the reflectance is obtained by simulating the scattering of a wave incident from the air side and not from within the GaN. However, the larger underlying trend of the measured enhancement curve follows the simulated scattering cross-section of the cone, which indicates that Mie resonances might affect the measured SHG, however, the enhancement is barely an order of magnitude and spectrally very broad. It is worth noting that, leaving aside any resonances, just the extra material that the cone provides is expected to increase the SHG signal by a factor of ~ 4 , given that the cone is about as tall as the GaN layer underneath it and the squared scaling of a phase-matched SHG process with the propagation length, which was established in the reciprocity analysis in Section 6.5.

Polarisation Study

After this analysis of the impact of the cone on the SHG spectrum, which did not reveal any promising wavelengths with a strong enhancement of the nonlinear conversion efficiency, let us focus on polarisation studies with two pump beams to see if the cone modifies the coupling to the different nonlinear tensor elements. For the first of these studies, the pump beam was set to $\lambda = 840$ nm, the seed beam was at $\lambda = 1500$ nm, and the transmitted up-conversion

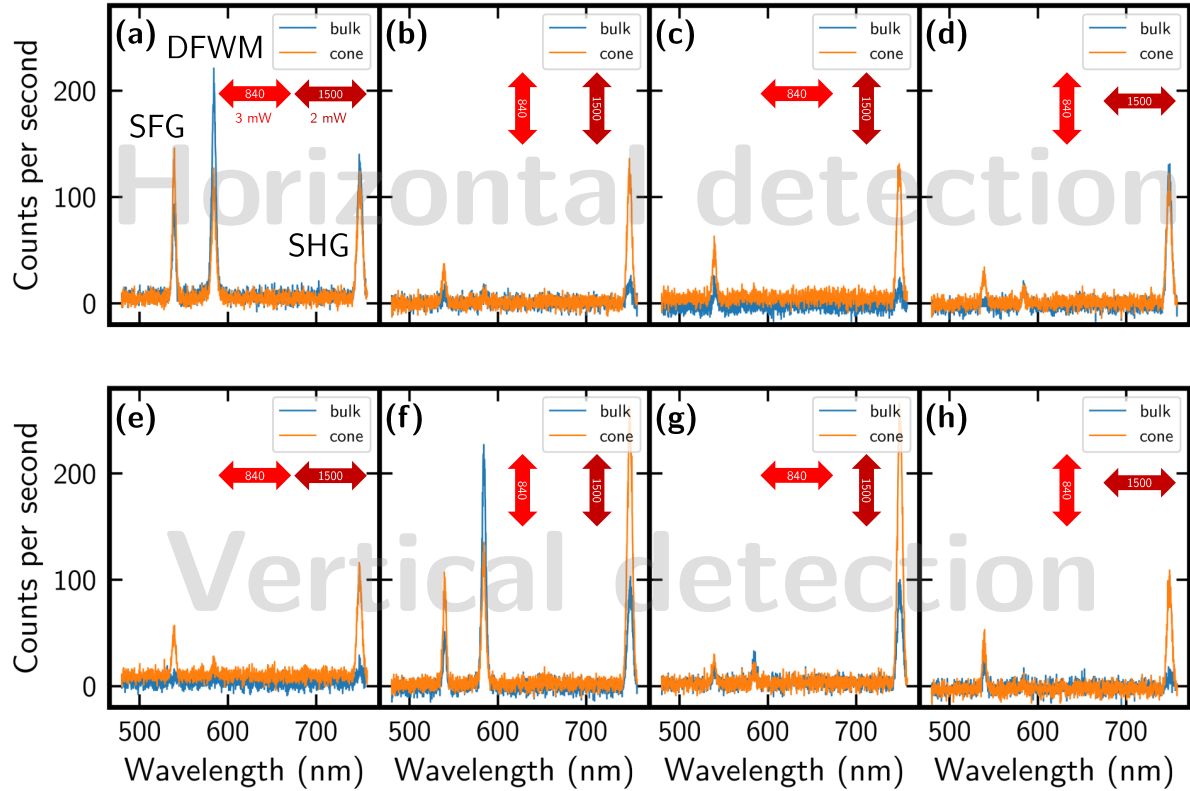


Figure 6.11 – Up-conversion signals for different polarisation configurations. (a)-(d) Transmitted SHG, SFG, and DFWM signals with inputs at 840 nm and 1500 nm for different combinations of the input polarisations and a horizontally oriented linear polariser before the spectrometer. The DFWM signal is strongly suppressed for all configurations except HHH, indicating that the interaction involves only the diagonal $\chi_{xxx}^{(3)}$ or $\chi_{yyy}^{(3)}$ elements. The cone slightly weakens the DFWM signal in (a) and strongly increases the SHG signal in (b) and (c), where the input field at 1500 nm is vertically polarised. The SFG signal is increased in all configurations. (e)-(h) The same signals for a vertically oriented linear polariser before the spectrometer. Again, the DFWM signal is suppressed in all configurations except VVV, where it is weakened by the cone. The SHG and SFG signals are boosted in (e) and (f) where the inputs have the same polarisation, but the cone only improves the SHG signal in (g) and (h) while the SFG signal is mostly unchanged. The discrepancies between the results with horizontally and vertically polarised detection suggest an asymmetric shape of the cone.

signals were collected for various input and detector polarisations. The resulting spectra where only the horizontally polarised photons were detected are shown in Figure 6.11(a)-(d), the analogous spectra for vertical detection can be seen in Figure 6.11(e)-(h). The DFWM signal is only strong for the HHH and VVV configurations shown in Figure 6.11(a) and (f), respectively, which demonstrates that the third-order response only involves the diagonal $\chi_{xxx}^{(3)}$ and $\chi_{yyy}^{(3)}$ tensor elements. The weak DFWM signals in all other spectra are likely due to imperfections in the preparation of the input polarisation states. Interestingly, the cone

weakens the DFWM signal in Figure 6.11(a) and (f) by about 50% which might be due to scattering losses. Let us now turn to the more complicated behaviour of the second-order signals SFG and SHG. While in the SHG spectra the output polarisation is identical to the input polarisation in the GaN bulk region, consistent with a coupling via the $\chi_{xxz}^{(2)}$ and $\chi_{yzy}^{(2)}$ tensor elements established earlier in Section 6.3, the cone allows for vertically polarised emission with a horizontally polarised pump beam and vice versa. The mechanism behind this cannot be polarisation conversion of the pump light, as this would have affected the DFWM signal as well. Therefore, the most likely explanation is improved scattering of z -polarised SHG which is generated via the $\chi_{zxx}^{(2)}$ and $\chi_{zyy}^{(2)}$ tensor elements. Moreover, the cone does not enhance the horizontal emission for a horizontal pump, but more than doubles the vertical emission when the pump beam is vertically polarised. This asymmetry hints at an irregular shape of the cone, which is consistent with the SEM images in Figure 6.8, where particularly the neck region was not rotationally symmetric. The SFG signal is expected to behave similarly to the SHG signal, which is experimentally confirmed except for some small discrepancies which can be explained by slight variations of the overlap of the input beams and, in the configurations where the signal barely emerges above the noise floor, by imperfections in the preparation of the input polarisation states.

The other polarisation study was conducted on DFG, with the seed beam at a wavelength of $\lambda = 840$ nm, but the pump beam at $\lambda = 543$ nm. The resulting NIR signal was collected in transmission geometry at a fixed horizontal pump polarisation while the polarisation angle of the seed polarisation was scanned. Figure 6.12(a) displays the DFG dataset from the bulk region of the sample. There is a sinusoidal variation of the signal as a function of the pump polarisation angle, which is expected due to the linear dependence of DFG on the pump beam. The signal is maximised for parallel input polarisations and strongly reduced for orthogonal input polarisations, with a 72% visibility of the sine fit. Aside from imperfections in the preparation of the input polarisation states, the signal of the local minimum is potentially due to coupling of the E_z component of the pump field with the E_x component of the seed field via the $\chi_{xxz}^{(2)}$ tensor element. This is only a weak contribution to the signal because,

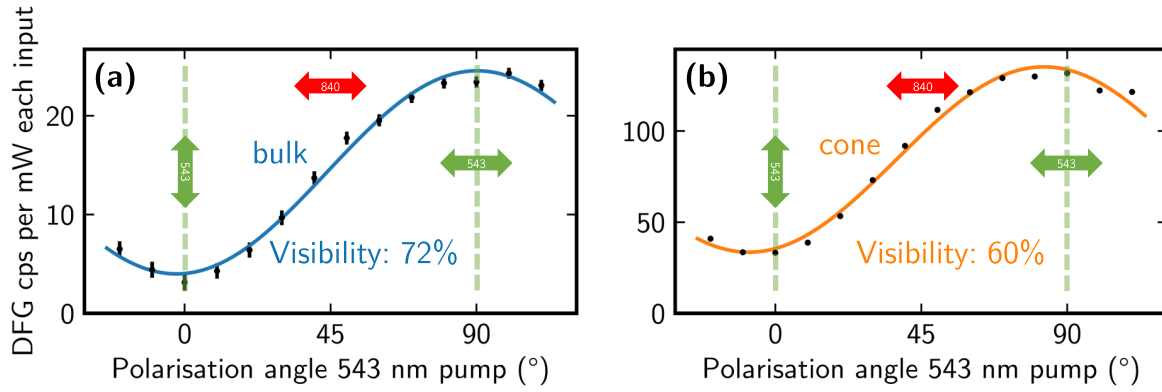


Figure 6.12 – DFG polarisation scan. (a) DFG signal of the unstructured bulk region of the GaN film as a function of the pump beam polarisation at a fixed horizontal polarisation of the seed beam. The signal is strongest for parallel polarisations and weakest for orthogonal polarisations of the inputs, with a 72% fringe visibility of the trace. (b) The same experiment on the GaN cone. The overall polarisation dependency is the same, but the fringe visibility is reduced to 60%. The relatively stronger signal for orthogonal input polarisations is potentially due to the increased longitudinal polarisation component and scattering of radiation by the cone which would otherwise be emitted into the slab.

due to the y -polarised pump field, the nonlinear polarisation is distributed similar to a HG_{01} mode instead of a HG_{10} mode, such that coupling to the fundamental TEM_{00} mode or a HG_{n0} mode with bad overlap is necessary. An analogous argument can be made for the E_y component of the pump field with the E_z component of the seed field.

The cone dataset shown in Figure 6.12(b) is similar to the bulk dataset except for a ~ 5 times higher count rate and a reduced visibility. As argued in the discussion of Figure 6.11, the extra signal in the local minimum is likely due to more scattering of the z -polarised DFG field and an enhancement of the longitudinal field components by the cone, which is consistent with the prediction of an enhanced E_z in Figure 6.8.

6.7 Future Measurements on New Cone Samples

The results of this chapter have informed the fabrication of a new sample with smaller cones, which was not yet available for measurements at the time of writing this thesis. The central idea is to create a magnetic dipole resonance in the wavelength range of interest to maximise the field enhancement available for boosting the nonlinear frequency mixing efficiency. Figure

6.13(a) shows an SEM image of a smaller cone on the new sample. Other than the smaller diameter of the top, the etching angle was changed to allow for a smaller supporting region, in order to better separate the cone from the GaN layer. Moreover, electron beam lithography rather than photolithography was used for patterning the mask, resulting in an improved control over the shape of the top of the cone.

Thanks to the smaller size of the cone, the magnetic dipole resonance is now within the wavelength range of interest, as can be seen in the FDTD simulations of the cone, depicted in Figure 6.13(b). The scattering cross-section of the cone in vacuum (dotted green) exhibits several peaks due to Mie resonances, which are shifted to longer wavelengths when the cone is connected to the GaN slab (dashed green), similarly to the larger cone in Figure 6.9(a). In the dashed green curve, the GaN slab was assumed to be semi-infinite in order to avoid FP fringes and to highlight the pure effect of connecting the cone with the slab. When adding the GaN-sapphire boundary, the FP modulations appear in the scattering cross-section spectrum (solid green), however, they do not significantly shift or distort the Mie modes, indicating a certain separability of the phenomena. In particular, contrary to the larger cone in Figure 6.9, the connecting neck region is small enough to not contribute its own low-order modes complicating the scattering cross-section spectra. Thus, the cone on the sample is in good approximation just a slightly larger version of the cone in vacuum, with some additional small ripples due to FP interference in the slab. It is interesting to note that even for the bare cone, the peaks of the integrated E^2 (dotted red) inside the cone are not always aligned with the peaks of the scattering cross-section for higher-order modes, meaning that the scattering cross-section is not necessarily a good predictor of the field enhancement.

Let us now take a closer look at the interplay between the magnetic dipole resonance and the FP interference by considering the shaded part of the spectrum, which is plotted in Figure 6.13(c). The scattering spectrum of the cone (green) is mostly flat in that region, which means that the transmittance (orange) and reflectance (blue) of the sample are almost perfectly aligned with the scattering cross-section, confirming the FP interference in the slab as the cause of the fast modulations. Moreover, the peaks of the integrated E^2 (red) are

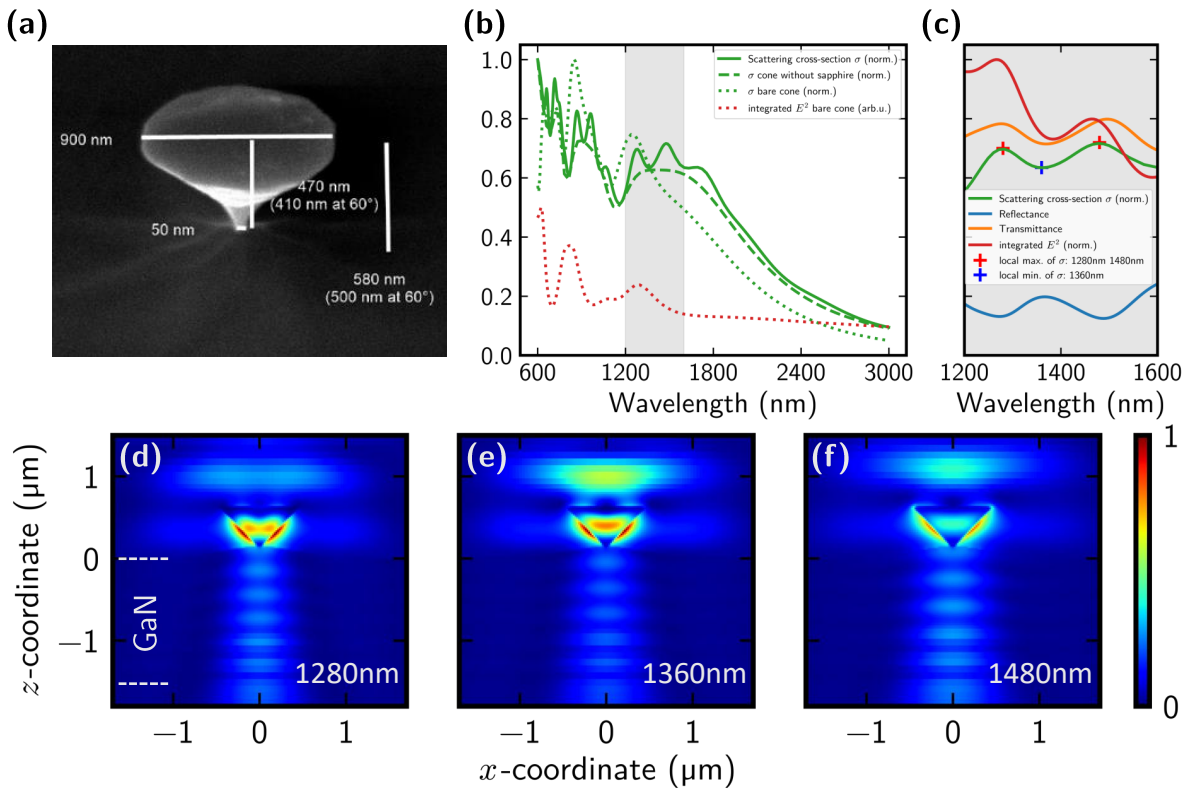


Figure 6.13 – GaN cones with smaller dimensions.[†] (a) SEM image of a cone on a new sample with an improved fabrication procedure. The neck size was reduced to ~ 50 nm for a better separation of the cone from the GaN layer and the top of the cone was defined via electron beam lithography, rather than photolithography. Smaller cone sizes are achievable at the new etch angles. (b) FDTD simulations of the scattering cross-section and the integrated intensity of the cone in (a). The scattering cross-section of the cone as it is on the sample (solid green) is composed of Mie modes with additional modulations due to FP interference in the GaN slab, which is highlighted by the same spectrum when simulated without the sapphire interface (dashed green). The corresponding spectrum of the bare cone (dotted green) exhibits the same Mie modes, but at shorter wavelength, which means that the substrate has the effect of shifting the resonances to longer wavelengths. The integrated intensity in the bare cone (dotted red) follows the scattering cross-section at the lowest-order resonance, but not at the higher-order modes. (c) Detail of the scattering cross-section (green) in the shaded area in (b). The two peaks coincide well with the peaks of the integrated intensity of the cone on the sample and also with the linear reflectance and transmittance. (d)–(f) Intensity distributions in the modelled cone at the local maxima and minima of the scattering cross-section marked with crosses in (c) to show the effect of the FP interference on the cone resonance. The distribution inside the cone is of a similar shape at all three wavelengths, while the intensity at the top and bottom of the simulation domain follows the FP resonances of the slab. The two phenomena appear to be largely decoupled, which is consistent with the reflectance in (c) following the scattering cross-section very closely and the similarity of the solid green and dashed green curves in (b). Moreover, the cone acts as a focussing element, similarly to the inset in Figure 6.9(b).

[†]The author acknowledges Dr JP Hadden for fabricating the sample and for taking the SEM image in Figure 6.13(a).

well aligned with the transmittance peaks, which is another characteristic property of FP cavities. Finally, let us consider the intensity distributions at the local maxima and minima marked by the red and blue crosses, which are shown in Figure 6.13(d)-(f). The intensity distributions inside the cone, as well as the focussing into the GaN slab are very similar at all three wavelengths. As expected from the discussion of Figure 6.13(c), the transmittance is highest at the local maxima, corresponding to the plots in Figure 6.13(d) and (f), while the reflectance is highest at the local minimum corresponding to the plot in Figure 6.13(e). The maximum intensity enhancement is approximately one order of magnitude, which is expected to improve the measured nonlinear signals by about the same amount.

Other than a better field enhancement via an optimised spectral position of the magnetic dipole resonance, the other key improvement consists in an optimised coupling between the input fields and the $\chi^{(2)}$ tensor elements, such that as much of the nonlinear output as possible is emitted into the far-field along the beam axis. This would be possible when coupling to the $\chi_{zzz}^{(2)}$ component in an m-plane GaN-on-sapphire sample [207], where the c-axis of the GaN layer is rotated by 90° compared to the c-plane sample investigated in this chapter.

6.8 Conclusion

In this chapter we have studied second-order and third-order nonlinear frequency mixing in a GaN film grown on a c-plane sapphire substrate. Due to the crystal structure and the experimental conditions, the longitudinal electric field component determines the measured nonlinear signal strength, which was confirmed in several independent measurements. SFG, SHG, DFWM, and DFG were observed and a SET analysis was carried out to estimate the photon pair generation rates in the corresponding spontaneous processes. The expected intrinsic SPDC pair generation rate in the GaN layer was estimated to be about 200 Hz, although this value might be up to two orders of magnitude too high, due to a possibly generous discount factor regarding the alignment of input fields and non-zero $\chi^{(2)}$ tensor

elements. Moreover, SFG was taken as an example for a detailed efficiency analysis, resulting in an estimated power conversion efficiency in the range of $\eta \sim 10^{-8} - 10^{-6}$, which was compared to similar systems in the literature.

After the discussion of the plain GaN film, approximately λ -sized cones etched into the film were introduced and their effect on the nonlinear signals was studied. An increase of the reflected SHG of $\sim 6-10$ was measured, which was attributed to scattering of the cone, the extra material provided by the cone and, to a small degree, the field enhancement inside the cone and its focussing of the incoming light into the slab. The cone was also found to increase the E_z component of the input fields, which affected the polarisation state of the second-order nonlinear signals. It was found that future experiments should focus on achieving an optimised coupling between the input fields and the $\chi^{(2)}$ tensor elements and the fabrication of smaller cones, such that as much of the nonlinear output as possible is emitted into the far-field along the beam axis and the magnetic dipole resonance becomes accessible in the wavelength range of interest.

Such experiments could lead to second-order nonlinear frequency mixing efficiencies on the order of $\eta \sim 10^{-5}$, or even $\eta \sim 10^{-4}$ for stronger pumping than in the experiments in this chapter. At those efficiencies, photon pair emission rates via SPDC in the cones would be brought to a competitive level for λ -sized interaction volumes [52]. Improvements of the simulations could involve two-stage procedures, where the nonlinear source terms are calculated explicitly via the tensor elements and are then inserted as source terms of a second simulation at the nonlinear output wavelength, which is more difficult but potentially more insightful than the comparison of the linear simulations at the various wavelengths, which was the approach in this chapter. Beyond studying single cones on m-plane samples, dense arrays and metasurfaces [208] of cones could be studied to improve the outcoupling of the nonlinear signals into the far-field. When working with second-order processes in c-plane samples, in-plane excitation schemes, either plasmonic [209] or dielectric [210], could be devised in order to optimise the coupling of the input fields to the active $\chi^{(2)}$ tensor

elements. Those prospects together with a large transparency window [179] make GaN nanostructures a promising system for high-efficiency low-loss nonlinear photon pair sources.

Chapter 7

Frequency Mixing in Gold Nanoantennae

In the last chapter we have seen how nanostructuring GaN, an essentially lossless dielectric in the visible and NIR, increased the nonlinear frequency mixing efficiency due to moderate field enhancement. The alternative presented in this chapter is to use metal nanoantennae which suffer from ohmic losses, but support extreme local field enhancement. After a brief overview of nonlinear optics with plasmonic nanostructures and the origin of the second-order nonlinear response in metals, the antenna sample is introduced together with simulations that predict the resonances of interest and the near-field distributions. The main part of the chapter is dedicated to the results and analysis of frequency mixing experiments. This includes an efficiency and SET analysis similar to Section 6.5 in the previous chapter, which provides an estimate of the strength of spontaneous processes. A discussion of the results and an outlook towards future studies with new triply resonant plasmonic antennae concludes the chapter.

7.1 Nonlinear Plasmonics

Harnessing the plasmonic field enhancement of metal nanostructures, the origin of which was explained in Chapter 3, has been a popular strategy for boosting the efficiency of nonlinear frequency mixing [37, 211–213]. This is because of the simultaneous benefits of extremely small interaction volumes in the sub-wavelength range and the large extent of the field enhancement, often several orders of magnitude, which is a promising combination for high-efficiency low-footprint nonlinear mixing devices [37]. The key design goals for nonlinear plasmonic nanostructures are the optimised energy transfer from the far-field to the small nanoscopic volume where the nonlinear interaction takes place [9], a strong enhancement and modal overlap of the interacting fields [48], and the orientation of the nonlinear medium, such that the largest tensor components of the nonlinear susceptibility are available for the interaction with the enhanced local fields [214]. While there have been many experiments with metal nanoantennae in a facilitating role, providing coupling between far- and near-fields and a strong local field enhancement for a nonlinear medium next to the antennae [215–219], it is also possible to achieve efficient frequency mixing with the optical nonlinearity of the antennae themselves [220, 221]. This is the approach pursued in the experiments of this chapter, with the idea of quantifying the expected photon pair generation rates of the bare antennae before introducing additional nonlinear materials in future studies.

As for the choice of material for the plasmonic antennae, gold and silver are the most popular for applications in the visible-NIR frequency range due to their large conductivity [222]. Gold has the disadvantage of higher optical loss caused by interband transitions in the green and blue part of the spectrum; however, for NIR applications it is often preferred to silver due to its high chemical stability [222]. As we will be studying the nonlinear mixing of visible and NIR frequencies and would like to push the limit of the longest participating wavelength towards the MIR in the future, gold is thus preferable to silver as a material for our antennae. In terms of its nonlinear optical properties, gold has been shown to exhibit a strong third-order nonlinearity [223] arising from the saturation of interband transitions and the redistribution

of conduction band electrons with their low specific heat capacity, heated up as a by-product of the excitation of 5d electrons [224]. Thus, while a decent efficiency is expected for third-order processes, gold seems to be a poor choice for nanoantennae supporting second-order nonlinear processes, given the face-centred cubic (fcc) crystal structure which possesses inversion symmetry; yet, second-order frequency mixing in gold has been observed in many experiments [225]. Given the importance of the second-order nonlinear response in the work presented below and the complicated situation in the literature, this initially surprising effect is explained in more detail in the following section.

7.2 Origin of the Second-Order Nonlinearity of Gold

There are several contributions and mechanisms responsible for the second-order nonlinear response of gold. Rudnick and Stern [226] presented a comprehensive analysis within the framework of a free electron gas model, which is summarised below. The second-order nonlinear currents of the conduction electrons can be separated into a longitudinal bulk component, which extends into the metal by the skin depth of several tens of nanometres, and parallel and normal surface currents confined to the gold-air interface by about one Thomas-Fermi screening length, which amounts to only a few nanometres. All three contributions are visualised in Figure 7.1(a). The bulk contribution is due to the Lorentz force of the linear magnetic field acting on currents driven by the linear electric field, which leads to a second-order current along the surface normal, as shown in Figure 7.1(b). Within about one screening length away from the surface, two further mechanisms appear. One of them is a nonlocal effect that only contributes significantly in regions where the electric field undergoes rapid spatial variation, which is the case in the immediate vicinity of the interface. Figure 7.1(c) visualises this rapid change of the electric field, where the decrease from the strong fields at the left to the middle is larger than that from the fields in the middle to the weak fields on the right. In this situation the local approximation breaks down and the current induced in the middle section is slightly stronger than the linear current that one would obtain by only considering the average field. This extra contribution, indicated by the green

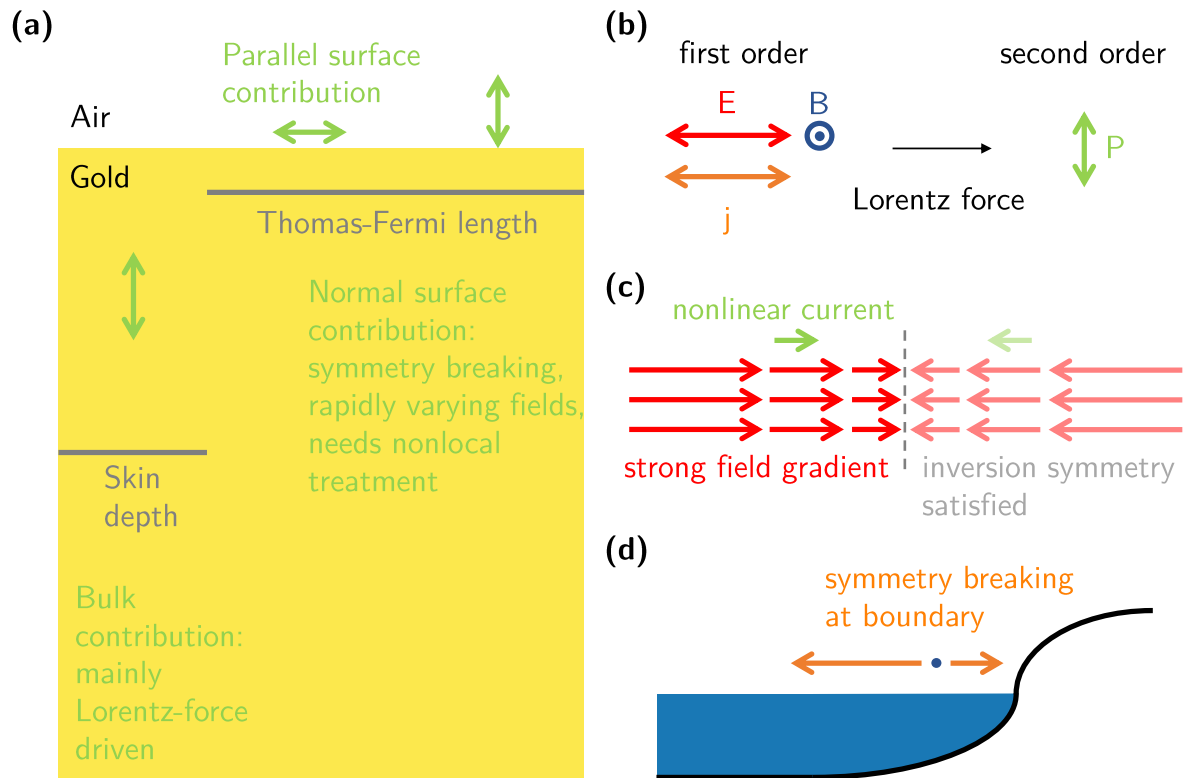


Figure 7.1 – Bulk and surface contributions to the second-order currents in gold. (a)

The second-order currents can be separated into bulk currents that occur within the skin depth, and parallel and perpendicular surface currents which decay about one screening length away from the surface. (b) First-order currents and the linear magnetic field interact via the Lorentz force, resulting in a second-order current. (c) Even in the absence of a boundary, second-order currents satisfying inversion symmetry are possible if there are strong field gradients leading to a nonlocal contribution. These rapid field variations occur within about one screening length from the interface. (d) An entirely different mechanism, which is directly rather than indirectly caused by the interface, is symmetry breaking. It is easier for electrons to flow into the medium than towards the boundary, causing an anharmonicity in oscillations normal to the surface within about one screening length.

arrow, is a nonlinear current which does not require symmetry breaking. Therefore, there is a second-order current which satisfies inversion symmetry and is hence allowed even in the bulk region of the gold. However, as the field gradients are weak in the bulk, this contribution is only significant near the interface. It is important to note that it is not the interface itself, which causes this effect, but the rapidly varying fields in its vicinity. In contrast, the second mechanism at the interface, shown in Figure 7.1(d), is the symmetry breaking which is due to the boundary itself. Electrons oscillating perpendicularly to the interface move in an anharmonic potential because they are less free to flow towards the boundary than

away from it. All three mechanisms give rise to the perpendicular surface component of the second-order current, but the rapid variation of the fields and the symmetry breaking at the interface dominate. The parallel surface current is not caused by the symmetry breaking, but by the strong field gradient at the surface.

In terms of concrete mathematical expressions for the nonlinear currents in the free electron gas model, the bulk contribution is proportional to $\nabla(\vec{E} \cdot \vec{E})$, whereas the surface components are proportional to $E_{\perp}E_{\perp}$ for the perpendicular component and $E_{\perp}E_{\parallel}$ for the parallel component. In the latter two terms, dimensionless phenomenological parameters a and b were introduced by Rudnick and Stern to account for the strength of the anharmonicity due to symmetry breaking and the impact of scattering loss. In a later work, Sipe and colleagues [227] reproduced these results with a simple hydrodynamic model and pointed out that the a parameter, which appears in the perpendicular surface term, is not constant as a function of the frequency, but may be resonantly enhanced by plasmonic modes. This finding supports the notion that the perpendicular surface term is of particular importance in our antennae at a plasmonic resonance, although there is room for error, given the crudeness of the model, which was pointed out by the authors [227].

It is important to note that the results so far were obtained with fairly simple electron gas models and for a flat, semi-infinite gold domain and that it is difficult to theoretically predict the relative strength of the second-order source terms in a concrete experiment. Therefore, it is interesting to consider experimental works which attempt to quantify and rank the various contributions and conveniently express them in terms of $\chi^{(2)}$ tensor elements linking the incident electric fields and the second-order polarisation response. One such work [228] used dual beam SHG measurements on a gold surface to separate the surface and bulk contributions of the second-order response. It was found that there is a weak but non-vanishing bulk contribution proportional to $(\vec{E} \cdot \nabla)\vec{E}$, which had been predicted in an earlier work by Bloembergen and co-workers [229], but which should vanish according to the hydrodynamic model used by Sipe and colleagues [227]. The other bulk term proportional to $\nabla(\vec{E} \cdot \vec{E})$ mentioned earlier was combined with the surface tensor elements [228], using

another result [230] showing that that component is experimentally indistinguishable from the surface $\chi_{\perp\parallel\parallel\parallel}^{(2)}$ and $\chi_{\perp\perp\perp\perp}^{(2)}$ elements. Therefore, the authors were able to express the whole second-order response in terms of the weak separable bulk component proportional to $(\vec{E} \cdot \nabla) \vec{E}$ and the effective surface nonlinear tensor elements $\chi_{\perp\parallel\parallel\parallel}^{(2),\text{eff}}$, $\chi_{\perp\perp\perp\perp}^{(2),\text{eff}}$, and $\chi_{\parallel\parallel\parallel\perp}^{(2),\text{eff}}$ [228]. The $\chi_{\perp\perp\perp\perp}^{(2),\text{eff}}$ element was found to be at least an order of magnitude stronger than the other second-order tensor elements when using the convention that the input fields are to be evaluated just inside the gold, and the separable bulk component was weakest regardless of the field convention in the definition of $\chi^{(2)}$ [228]. The authors concluded that surface contributions appear to be more important than bulk components, but that all contributions should be used if possible and that the relative weight of the contributions is unclear for nanoparticles instead of a flat surface [228]. Another article [231] published only one year later contributed a study of the various tensor elements for a 150 nm gold sphere. It was found that all aforementioned tensor components were needed to accurately describe the second-order response of the gold nanospheres, but that for certain polarisation configurations the bulk contribution was largest [231]. The importance of bulk contributions was also highlighted in another work [232] on the second-order response of gold nanoparticles of various shapes, where it was shown that the observed SHG dependence on the particle shape could be well explained with bulk contributions in a classical model, suggesting that there is a dipolar bulk contribution to the second-order response if the nanoparticle is sufficiently asymmetrically shaped.

In summary, we have seen that there are several distinct mechanisms causing the second-order response of gold. The particular case of nanoparticles is complicated and theoretical predictions should in general not be focussing on only one of the contributions, such as that of the $\chi_{\perp\perp\perp\perp}^{(2),\text{eff}}$ element only because it is dominating in certain measurements of unstructured films. In our case, the antenna design [49] is therefore based on the simple ideas of plasmonic resonances at the wavelengths of interest, with the accompanying field enhancement boosting the nonlinear conversion efficiency, simple constituent nanoparticle shapes with good coupling to the far-field, and a non-centrosymmetric overall arrangement of the nanoparticles forming

the antenna, to benefit from dipolar bulk contributions discussed in the work by Zeng and co-workers [232]. A detailed account of the concrete antenna design is given in the following section.

7.3 Sample and Simulations

The sample studied in the frequency mixing experiments in this chapter consists of arrays of gold nanoparticles on a $170\ \mu\text{m}$ glass coverslip. The antennae in each array are arranged in a square lattice and with a fixed spacing, a_0 . Figure 7.2(a) shows an SEM image of such an antenna array with a spacing of $a_0 = 750\ \text{nm}$ between the antennae. Note that the ensemble of one bar-shaped particle and two disk particles is referred to as an antenna. A drawing of the antenna design is depicted in Figure 7.2(b). The dimensions of the bar are $340 \times 80 \times 40\ \text{nm}$, while the disks are of the same thickness and have a diameter of $160\ \text{nm}$. All particles were grown on a $1.5\ \text{nm}$ Cr adhesion layer, which is shown in the cross-section on the top right of Figure 7.2(b). The gaps between the disks and the bar are $30\ \text{nm}$, however, there are fluctuations on the order of 10% , as can be seen in Figure 7.2(a). In order to

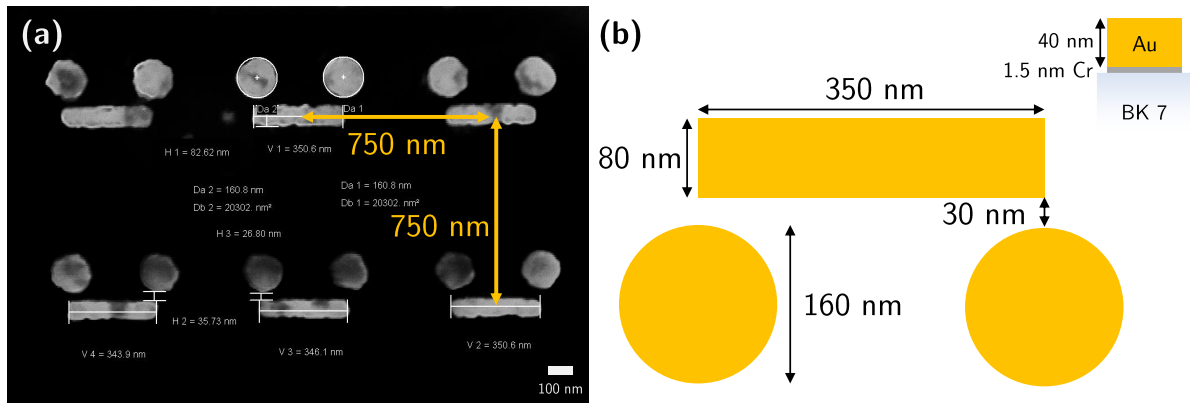


Figure 7.2 – Antenna sample and design.[†] (a) SEM image of an antenna array on the sample. The antennae are arranged in a square lattice with a spacing of $750\ \text{nm}$. Each antenna consists of a bar and two disk particles. (b) Top view of an idealised antenna with the dimensions of the particles and gaps. The cross-section in the top right corner shows the $1.5\ \text{nm}$ Cr layer between each nanoparticle and the glass substrate to improve the adhesion of the gold. Note that the lengths in the actual design are slightly longer to compensate for erosion effects during fabrication.

[†]The author acknowledges Dr Yi Li for fabricating the sample and for taking the SEM image in Figure 7.2(a).

take into account erosion during the fabrication, which leads to slightly smaller particles on the sample, the actual design lengths were set a few nanometres larger. Note that the little irregularities in the particle shapes in Figure 7.2(a) may be a positive rather than a negative factor for second-order frequency mixing, viewed from the perspective of symmetry-breaking.

The antenna design is the result of a study by Gennaro and colleagues [49] on the arrangement of gold nanoparticles for optimum emission of SHG from the antenna into the far-field, which revealed the following key points. The bar particle exhibits a dipolar resonance at $\lambda = 1500$ nm and facilitates efficient coupling of the fundamental fields to the near-field. Near-field coupling between the bar and the disks then allows for an efficient re-radiation of the SHG at $\lambda = 750$ nm via the disks which exhibit a dipolar resonance at that wavelength. The disks are placed on the same side of the bar to ensure that the overall arrangement is non-centrosymmetric and that the two disks radiate in phase.

Let us now explore the antenna design further by considering the simulated field distributions and scattering and absorption spectra. As in the previous chapter, the simulations were run with the software Lumerical. Figure 7.3(a) shows the distribution of E_z and $|E|^2$ just above the antenna. The dipole mode of the bar is efficiently excited at its fundamental resonance wavelength $\lambda = 1500$ nm when the pump beam is polarised along the bar (panels 1 and 5), but not when it is polarised perpendicular to the bar (panels 3 and 7). Note that all values are normalised to the case of x -polarised pumping at $\lambda = 1500$ nm, but the limits of the colourbars are set to the respective minimum and maximum value in each panel to improve data visibility. Panels 5 and 7 show that the local energy of the near-fields is almost two orders of magnitude higher for a polarisation along the bar rather than perpendicular to it. Pumping at $\lambda = 750$ nm (panels 2, 4, 6, and 8) leads to a high local intensity around the disks regardless of polarisation (panels 6 and 8), while only a small fraction of the field energy resides in the bar. However, some of the energy is coupled to the bar when the pump field is y -polarised, as can be seen in panel 8. The bar mode at $\lambda = 750$ nm is a quadrupole, which can be observed in panel 4, coupling to the dipole mode of the disks. This results in a high energy concentration in the gap between the particles, shown in panel 8.

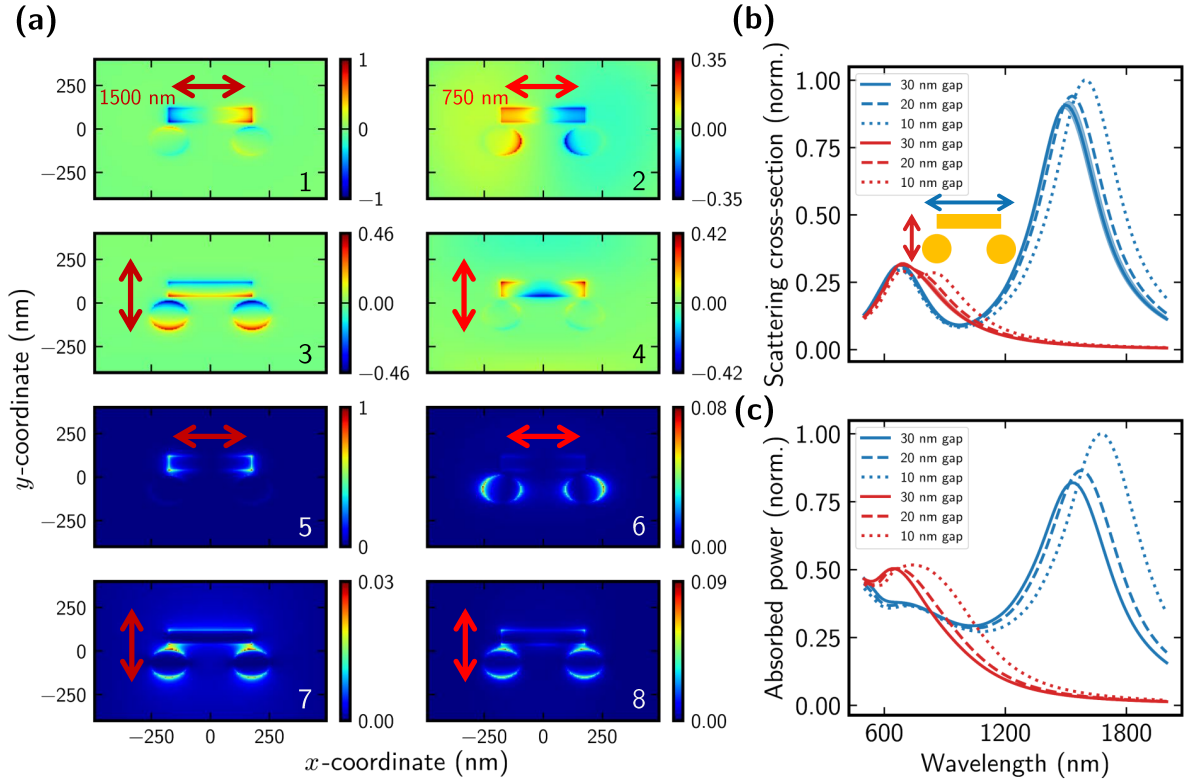


Figure 7.3 – Antenna field and intensity distributions and scattering cross-section. (a) E_z (panels 1-4) and $|E|^2$ (panels 5-8) distribution of the antenna at the design input wavelengths $\lambda = 1500$ nm (left column) and $\lambda = 750$ nm (right column) for horizontal and vertical input polarisation (arrows). The dipolar mode of the bar is well excited at its resonance $\lambda = 1500$ nm (panels 1 and 5), while at $\lambda = 750$ nm, a quadrupolar mode in the bar is supported (panel 4). The intensity is concentrated at the ends of the bar for x -polarised excitation at $\lambda = 1500$ nm (panel 5), while an excitation at $\lambda = 750$ nm leads to a high intensity at the disks due to their resonance, regardless of polarisation (panels 6 and 8). For y -polarisation, some intensity is passed on to the bar via near-field coupling of the dipolar modes of the disk and the quadrupole mode of the bar (panel 8). (b) Scattering cross-section for x -polarisation along the bar (blue) and y -polarisation (red). While x -polarisation leads to good coupling to the fundamental bar and disk resonances, only the disk is efficiently pumped for a y -polarised input. Narrowing the gap between the bar and the disks leads to a longer effective resonance wavelength of the bar and the disk. Additionally, a dip occurs at the disk resonance for a narrowing gap, indicating stronger coupling between the quadrupole mode of the bar and the dipole mode of the disk. (c) The absorption curves are in good agreement with the scattering cross-section in (b). An interesting extra feature is the underlying increase in absorption towards shorter wavelengths, which is due to the interband transitions in the gold.

The simulated scattering cross-section spectra in Figure 7.3(b) confirm these findings. While for an x -polarised input (blue) both the disk and bar dipole resonances are excited at $\lambda = 750$ nm and $\lambda = 1500$ nm, y -polarised pumping (red) only excites the disks at $\lambda = 750$ nm. The crucial feature in this graph is the dip in the y -polarised scattering spectrum, which emerges when the gap size is reduced to about 10 nm. The dip is caused by the spectrally narrow

quadrupole mode of the bar, which of course cannot be directly excited from the far field and is hence absent in the spectra for a larger gap size. So it is indeed the aforementioned near-field coupling from the disk dipole modes that excites the quadrupole mode in the bar, which only occurs for a sufficiently small gap size.

Another interesting observation is that the resonances slightly shift towards longer wavelengths, when the gap size is reduced. This could be due to a larger effective resonator length for the plasmonic oscillations when the particles are sufficiently close to each other. Note that the spectra for a 30 nm gap (solid lines) have a shaded area around them, corresponding to gap size fluctuations of ± 5 nm to visualise the impact of realistically expectable fabrication imperfections. Simulated spectra of the relative absorbed power are shown in Figure 7.3(c). The behaviour is very similar to that of the scattering cross-section spectra, with the difference that for wavelengths below the disk resonance the absorption strongly increases due to interband transitions in the gold. The local maxima are slightly shifted with respect to the cross-section spectra, which is likely due to the coarser mesh that was used in the simulation of the absorption spectra. This was necessary because of the high memory requirements of the 3D monitor that is needed for the latter simulations.

In summary, we have confirmed the key ideas behind the antenna design, in particular the near-field coupling between the quadrupolar bar mode and the dipolar disk modes. This makes the antenna particularly suitable for SHG pumped at $\lambda = 1500$ nm, given the optimised coupling to the dipolar bar resonance at that wavelength, the oscillations at the second harmonic which feed the quadrupole mode, the coupling to the dipoles of the disks and the efficient energy transfer from the disks back to the far-field. In the experiments below, the frequency mixing is generalised to dual-beam schemes at the same wavelengths as in the preceding chapter, where only the bar resonance at $\lambda = 1500$ nm but not the disk resonances will be used. This is to get a first idea of the frequency conversion efficiencies in this well-understood tried and tested nanoantenna system, even though the antenna design is not fully optimised for the experiments that follow. In the last section of this chapter, an adapted

design of a triply resonant antenna, which was not yet available when the measurements were carried out, is presented as an outlook.

7.4 Experimental Results

Now that we have studied the origin of the nonlinear response in the gold particles, as well as the design considerations and the resonant properties of the antennae, we are in a good position to discuss the frequency mixing experiments. As in the last chapter, we are mainly interested in mixing processes involving visible and IR photons in preparation of photon pair sources that connect those spectral regions. Therefore, the focus of this work is on difference frequency generation (DFG), degenerate four-wave mixing (DFWM), and sum-frequency generation (SFG), which are linked to the spontaneous photon-pair creating processes spontaneous four-wave mixing (SFWM) and spontaneous parametric down-conversion (SPDC), as was pointed out in the preceding chapter.

7.4.1 Output Polarisation Study

In order to use the resonance of the bar particle, the IR wavelength was set to $\lambda = 1500$ nm and the pump wavelength was chosen to be $\lambda = 840$ nm, such that the DFWM and SFG signals would occur around $\lambda = 583$ nm and $\lambda = 540$ nm, respectively. This choice of pump wavelength is due to constraints in our laser setup given that we intended to observe both the up-conversion and down-conversion process for each wavelength combination. Details of the frequency mixing setups can be found in Appendix A. Our InGaAs detector array for detecting the IR signals is less efficient and smaller than the Si detector array for the visible signals, which makes it experimentally much more challenging to observe DFG and DFWM seeded in the visible, compared to SFG and DFWM seeded in the IR. Therefore, in order to find the IR signals, it was important to reduce the number of experimental parameters that still need to be optimised, before searching those signals amongst the noise. Hence, the first experiment discussed in this section is a study of the nonlinear signals in the visible

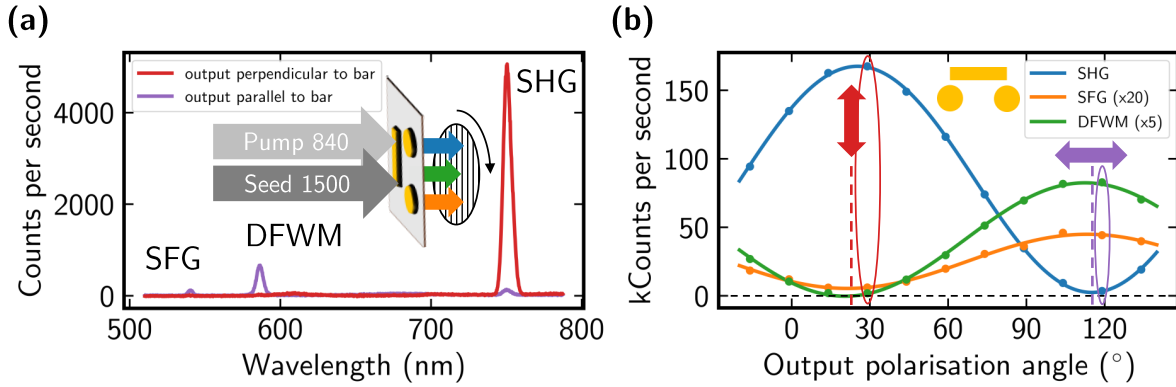


Figure 7.4 – Polarisation properties of the nonlinear antenna output in the visible. (a) Output spectra of the nonlinear signals from the antennae pumped at $\lambda = 1500$ nm and $\lambda = 840$ nm. When the detection polarisation is perpendicular to the bar (red spectrum), the SHG signal is maximised and DFWM and SFG suppressed. When only the light parallel to the bar is detected, the DFWM and SFG signals are maximised and SHG is strongly suppressed (purple spectrum). (b) A plot of the integrated counts of each signal as a function of the polariser angle reveals a sinusoidal dependence which is characteristic of a dipole antenna. The visibility of the SFG fringe is lower than for the other two signals, which may be due to the low signal-to-noise ratio as well as the modal properties of the antenna at $\lambda = 543$ nm.

waveband, emerging from the antenna when seeded at $\lambda = 1500$ nm and pumped at $\lambda = 840$ nm. The idea is to determine the output polarisation of these signals, which, by reciprocity, is equal to the optimum input polarisation for the reciprocal processes seeded in the visible.

A schematic of the experiment is shown in the inset of Figure 7.4(a), which also displays two output spectra, where a linear polariser oriented perpendicular and parallel to the bar particle was placed in front of the detector. The SHG signal is emitted with a polarisation perpendicular to the bar, as expected from the simulations presented in the last section. This is because the SHG populates the quadrupole mode of the bar, which couples well to the dipole mode of the disks that is perpendicular to the bar. On the other hand, the polarisation state of the DFWM and SFG signals is parallel to the bar, which could be due to a bright higher-order mode of the bar.

A systematic scan of the polariser angle, shown in Figure 7.4(b), reveals that each signal follows a sinusoidal trend, which is characteristic for emission from a dipole antenna and that indeed SHG, DFWM, and SFG have a preferential polarisation at right angles to each other. The higher visibility of the SHG fringe compared to the other signals is probably due

to the higher signal-to-noise ratio and the strong polarisation selectivity of the interparticle coupling process explained in Section 7.3. We conclude that in order to observe DFWM and DFG in the IR, these processes need to be seeded with a polarisation parallel to the bar.

7.4.2 DFG Measurements

This prediction was confirmed in DFG measurements, where the antennae were pumped at $\lambda = 543$ nm and seeded at $\lambda = 850$ nm. The slightly shifted pump and seed wavelengths were due to a better stability of the laser system for that combination. The wavelengths were chosen such that the seed is as close as possible to $\lambda = 840$ nm while ensuring a stable DFG output at $\lambda = 1500$ nm. Figure 7.5 shows the resulting DFG spectra for the various input polarisation configurations. Note that there was not only a DFG signal at $\lambda = 1500$ nm, but also a much broader and slightly fluctuating background, which is particularly well visible in Figure 7.7. This background originates from the gold and is likely due to a luminescence process [233,234], such as two-photon luminescence [235], or electronic inelastic light scattering [236], as was pointed out in an earlier work [237] on the same antennae used here. In order to isolate the DFG signal, a second spectrum with an intentional time delay between the input pulses was subtracted from all the spectra in Figure 7.5 for each polarisation configuration.

The strongest signal was found for cross-polarised inputs with the pump beam polarised parallel to the bar and the seed beam polarised perpendicular to it, as shown in Figure 7.5(a). This was expected following the discussion above, given the SFG emission with a preferential polarisation along the bar, and the simulation results suggesting that exciting the antennae perpendicularly to the bar at $\lambda = 750$ nm is best for indirectly exciting the bar particle. Here, we are seeding at $\lambda = 850$ nm, but this is reasonably close to the disk resonance, and the DFG signal suggests that there is still significant frequency mixing in the bar particle radiating off the DFG signal.

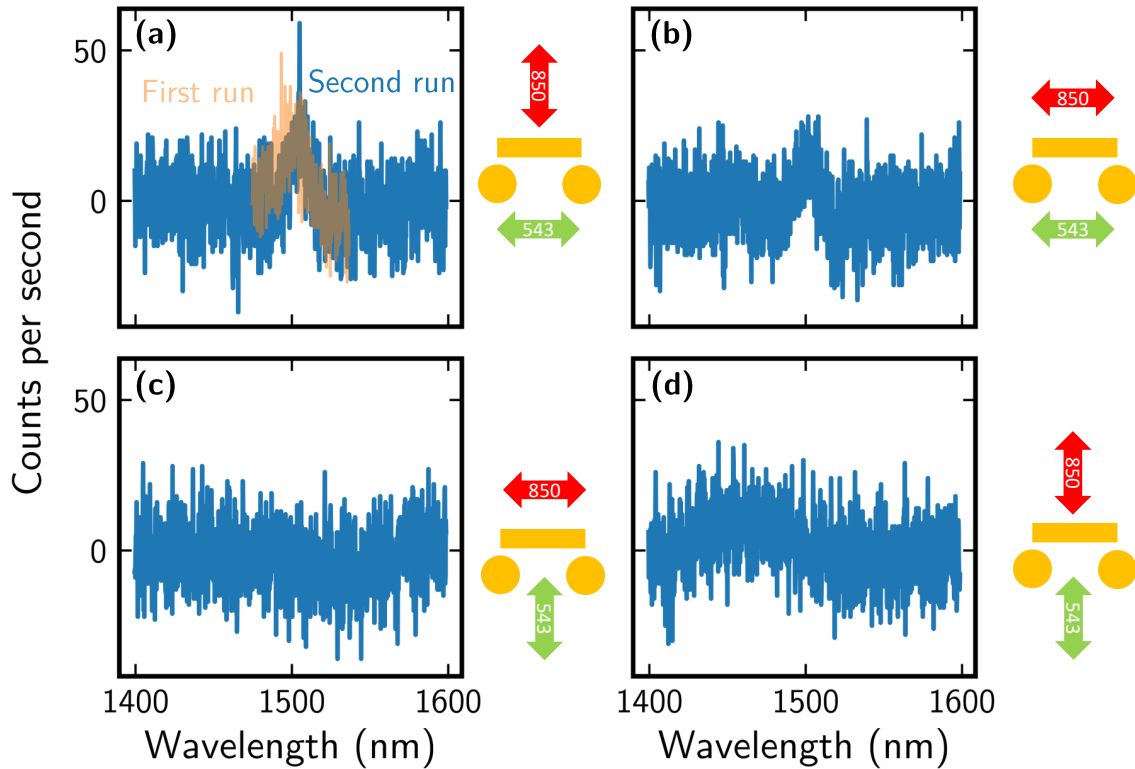


Figure 7.5 – DFG for different input polarisations. (a) When the seed and the pump beam are polarised perpendicular and parallel to the bar, respectively, the DFG signal is strongest, as expected from the up-conversion results in the previous subsection. The spectra in (b)-(d) were taken between the two measurements of this spectrum to demonstrate that the antennae were not damaged over the course of the measurements. (b) When the seed beam polarisation is rotated, such that both input polarisations are parallel to the bar, most of the signal remains. (c) When the pump polarisation is set perpendicular to the bar, the DFG signal vanishes. (d) The signal also disappears when both beams are polarised perpendicular to the bar. The negative counts and the shape of the remaining background are due to the subtraction of a reference spectrum where the input pulses were delayed by more than one pulse duration with respect to each other, to remove fluorescence as well as possible.

Interestingly, the DFG signal is almost as strong when both input polarisations are parallel to the bar, which can be seen in Figure 7.5(b). This means that the antenna is not as strictly polarisation-selective for frequency mixing at $\lambda = 850$ nm than it is for an input at $\lambda = 1500$ nm. However, for a pump polarisation perpendicular to the bar, the DFG signal is strongly suppressed regardless of the pump polarisation, as shown in Figure 7.5(c),(d). This highlights the importance of the pump polarisation for DFG and that studying the SFG output polarisation above in Figure 7.4 proved applicable to the corresponding down-conversion process.

The slight modulation of the background, particularly in the last spectrum of Figure 7.5, is caused by fluctuations of the luminescence background over time. One likely reason for this are small changes in the particle morphology, as it was necessary to operate near the damage threshold in order to collect the DFG signal with the setup at hand. However, we do not believe that the particles were significantly damaged during the measurement as the signal in Figure 7.5(a) was practically the same in the first and last measurement of the series; the blue spectrum was taken last and the orange spectrum first of all the measurements to prove that no significant optical damage occurred in this experiment.

7.4.3 Suppression of DFWM from the Substrate

Since DFG is a second-order process, it is clear that the signal must originate from the antennae and not from the glass substrate. However, more care needs to be taken for DFWM, which is a third-order process that can also occur in the substrate. Even the much smaller $\chi^{(3)}$ of glass compared to gold and the strongly enhanced local fields at the surface of the gold particles do not guarantee that the DFWM contribution from the antennae dominates, as the nonlinear interaction volume in the substrate is several orders of magnitude larger and the small regions in the glass close to the antennae also benefit from the plasmonic field enhancement. Therefore, it is useful to study the dependence of the DFWM signal on the input polarisation configuration, which is presented in Figure 7.6. The idea of this experiment is to exploit the polarisation selectivity of the antenna, particularly the bar particle at $\lambda = 1500$ nm, to separate the DFWM contributions from antenna and substrate.

Figure 7.6(a) shows the nonlinear spectra from the substrate (blue) and from the antennae on the substrate (orange) for both inputs polarised parallel to the bar. The substrate exhibits a strong DFWM signal, which weakens by more than half when moving the excitation beams onto the antenna array. This can be explained by the strong absorption of the bar particle at that wavelength and polarisation. At least part of the remaining signal is still generated in the substrate, but a significant contribution from the antennae cannot be ruled out, given that the near-field coupling at the pump wavelength is not strongly polarisation-selective,

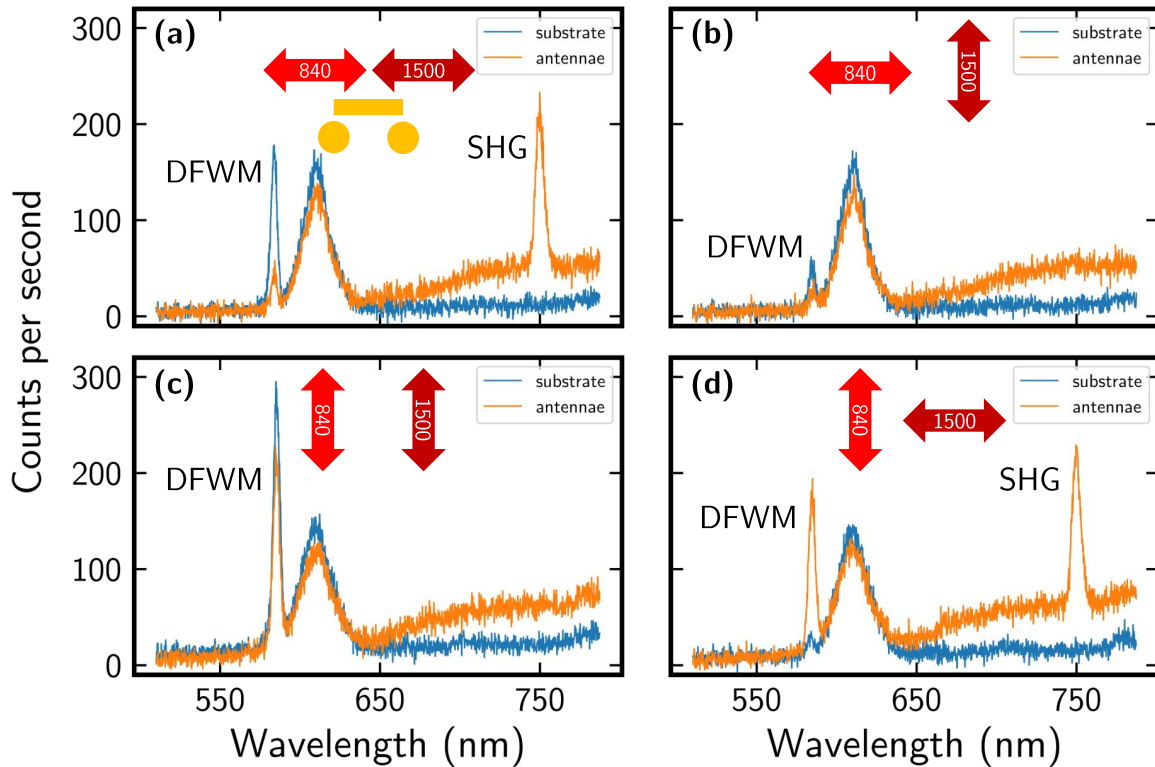


Figure 7.6 – Suppressing DFWM from the substrate. (a) For both input polarisations parallel to the bar, SHG from the antennae and DFWM from the substrate is observable. The DFWM peak is weakened by the antennae, as they absorb the seed beam at $\lambda = 1500$ nm very efficiently for this polarisation. (b) Rotating the seed polarisation such that it is perpendicular to the bar means essentially no coupling to the antennae, and hence a strong suppression of all frequency mixing processes. There is a weak residual DFWM signal from the substrate, probably due to imperfections in the preparation of the polarisation state. (c) When both input polarisations are perpendicular to the bar, there is again no frequency mixing from the antennae, but a strong DFWM signal from the substrate. It is barely weakened by the antennae, because of the weak coupling of the seed beam to the antenna. (d) For cross-polarised inputs with the seed beam polarisation aligned with the bar, the frequency mixing in the antennae is optimised, while DFWM in the substrate is suppressed. All spectra exhibit a noise peak, potentially due to scattered residual pump light, at $\lambda = 610$ nm, which, however, does not affect the conclusions drawn from this experiment.

as was found earlier in the discussion of Figure 7.5. Additional interesting features of the antenna spectrum in Figure 7.6(a) are the SHG peak, confirming the good coupling of the seed beam to the bar antenna, and a broad background signal extended towards the IR with an onset at around $\lambda = 650$ nm. Potentially, this background is resulting from the same processes as that in Figure 7.7.

Rotating the seed polarisation by 90 degrees results in the spectra shown in Figure 7.6(b). As expected, the SHG peak has vanished, since the bar particle is strongly polarisation-selective

at $\lambda = 1500$ nm. Of course, this dramatic reduction in coupling to the antenna near-field also weakens the DFWM signal from the antennae substantially. Moreover, the DFWM from the substrate is strongly suppressed for this cross-polarised input configuration, which indicates that the relevant non-zero tensor elements in the glass are $\chi_{xxx}^{(3)}$ and $\chi_{yyy}^{(3)}$ on the diagonal. This is consistent with remarks in Ref. [38] stating that in centrosymmetric isotropic materials, such as glasses, the restoring force of the nonlinear oscillations is always co-linear with the displacement vector and hence with the input polarisation.

Figure 7.6(c) shows that for both input polarisations perpendicular to the bar, there is only a strong DFWM signal from the substrate, which is only slightly reduced by the antennae. The difference between this situation and that in Figure 7.6(a) is that the bar particle is barely excited at $\lambda = 1500$ nm, and hence more of the incoming IR light reaches the substrate. Figure 7.6(d) shows the configuration necessary for suppressing the substrate contribution to DFWM and for coupling to the antenna as well as possible. The SHG peak is as strong as in Figure 7.6(a) and the DFWM signal does originate from the antennae and disappears when the input beams are moved onto the bare substrate. Therefore, this input configuration is used for all DFWM measurements to ascertain that the signal is really originating from the antennae.

To conclude the discussion of this set of measurements, it is interesting to note that the luminescence background is almost unchanged in all antenna spectra and hence does not seem to depend much on the coupling, which is unexpected. Moreover, a noise peak of unclear origin is present in all spectra in Figure 7.6 at about 610 nm. One possible explanation for this peak is scattered residual pump light. However, the behaviour of the DFWM and SHG signals is unaffected.

7.4.4 Spectrograms of Dual-Beam Signals

Let us now consider spectrograms of the dual-beam frequency mixing processes from the antennae, obtained by scanning the time delay between the input beams. This serves to

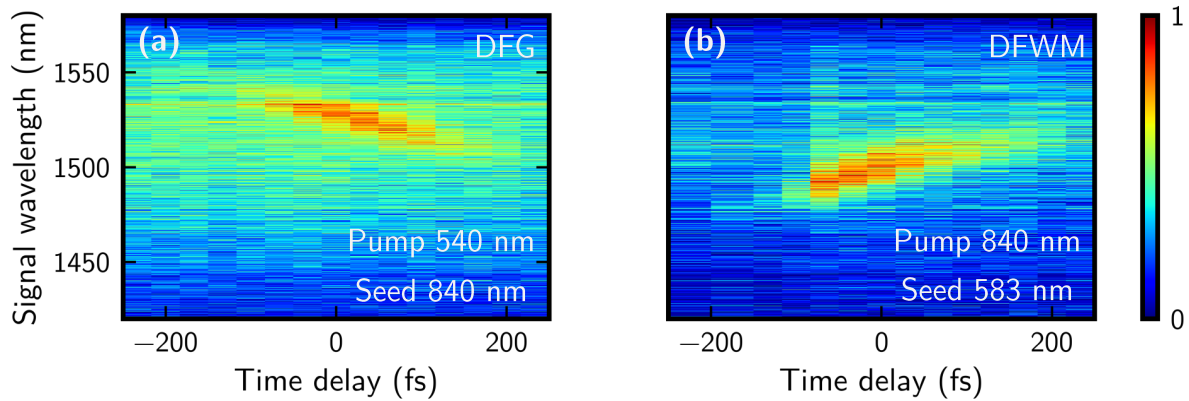


Figure 7.7 – Spectrograms of DFG and DFWM signals. (a) DFG signal from antennae pumped at $\lambda = 540$ nm and seeded at $\lambda = 840$ nm for various time delays between the pulses. While the DFG signal only appears when both pulses are overlapped in time, there is a strong luminescence background independent of the delay. (b) Changing the visible input wavelength to $\lambda = 583$ nm results in a DFWM signal at $\lambda = 1500$ nm. The inputs were cross-polarised to avoid a DFWM contribution from the substrate. Again, there is luminescence independent of the time delay.

distinguish the difficult-to-obtain signals in the IR from unwanted noise, similar to the scans in the previous chapter. The DFG spectrogram in Figure 7.7(a) demonstrates how the DFG signal only appears when both input pulses excite the antennae at the same time, unlike the underlying luminescence background, which persists for all delay times. The luminescence is spectrally much broader than it appears in this dataset and is suppressed by optical filters at the short wavelength end and by the rapidly decreasing efficiency of the InGaAs detector at the long wavelength end. As mentioned earlier, it is possible that this luminescence background extends up to the visible waveband, where it appears in spectra such as in Figure 7.6. Similar to the scans in the previous chapter, the DFG signal exhibits a chirp, which is likely due to dispersive elements in the input beam paths.

The spectrogram of the down-conversion DFWM in Figure 7.7(b), sitting on a broad and delay-independent luminescence background as well, also shows the presence of chirp. The frequency shift of the signal as a function of the delay time has the opposite orientation to the DFG trace, because in DFG the seed photons are annihilated, whereas in DFWM the pump photons are annihilated. Therefore, if at high delays the high-energy end of the seed pulse interacts with the low-energy end of the pump pulse, the signal photons in DFG must be of a higher energy than at zero delay, while the DFWM photons must lose energy compared to the

case of zero delay. The DFG and DFWM spectrograms of the GaN sample, shown in Figures 6.5(b) and 6.6(b), respectively, were obtained with analogous experimental configurations and demonstrate that the luminescence background does indeed originate from the gold.

7.4.5 Joint Spectral Density (JSD) Maps of Dual-Beam Signals

Next, frequency scans of the dual-beam signals were carried out, where an IR seed beam was spectrally filtered as described in the previous chapter and in Appendix A, and the slit used for spectral filtering was scanned throughout the pulse spectrum. Since the filter drastically reduced the signal strength, these measurements could only be taken on the strong up-conversion signals but not on the weak down-conversion signals. The main goal was to obtain joint spectral density (JSD) maps, which allow for the characterisation of the biphoton wavefunction within the stimulated emission tomography (SET) framework, as explained in Chapter 2. As a side result, the genuineness of the frequency mixing signals could be confirmed via the spectral drift as a function of the input wavelength.

Figure 7.8 shows JSD maps for the up-conversion processes DFWM and SFG. Since moving the slit used for spectral filtering changes the input power according to the spectral energy distribution of the pulse, each spectrum of the nonlinear signal was normalised by the measured input power, which is responsible for the increased noise at the edges of the spectrograms corresponding to the tails of the pulse. Some of the spectra stand out due to their higher overall counts, which is only partly due to power fluctuations of the laser and a drift of the sample with respect to the input beams. An additional reason is that during the scan the signal had to be optimised for each spectrum to correct for a slight drift of the seed beam with the position of the slit used for spectral filtering. At low signal-to-noise levels this optimisation procedure was difficult to carry out and subject to significant fluctuations.

The spectral trace of the DFWM scan in Figure 7.8(a) follows energy conservation, similarly to its counterpart in Figure 6.6; however, it seems to exhibit a broad resonance centred around $\lambda = 1505\text{--}1510$ nm. This can be attributed to the dipole resonance of the bar

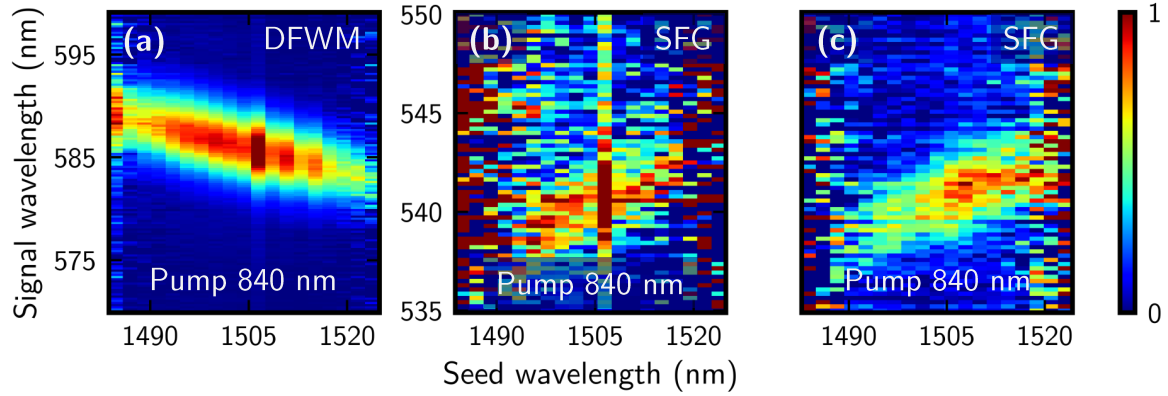


Figure 7.8 – Joint spectral density (JSD) maps. (a) DFWM signal from the antennae for a fixed pump and varying seed wavelength. The resulting JSD map indicates a potential antenna resonance centred around $\lambda = 1505$ nm. (b) The same spectra truncated in a region around $\lambda = 540$ nm show the corresponding SFG signal. The signal-to-noise ratio is low due to an optimisation of the DFWM signal strength at the expense of SFG. This trade-off between the two signals is due to chromatic aberration of the focussing objective. (c) Repetition of the SFG frequency scan with the input beams focussed such that the relative strength of SFG is maximised, leading to a spectrogram with a better signal-to-noise ratio.

particle, which is discussed in Figure 7.3. Note that the width of the resonance in the JSD map seems to be determined by the scanning range of the seed wavelength, which means that there could be a further uncompensated drift in the measurement and the actual bar particle resonance might be much broader. Given the simulation results in Figure 7.3(b) this is likely the case. Moreover, it is possible that some residual DFWM signal from the substrate is present in the JSD map despite the fact that the polarisation configuration was chosen as in Figure 7.6(d) to suppress DFWM from the substrate.

Another slice of the same spectra showing the SFG signal in this scan is depicted in Figure 7.8(b). The signal-to-noise ratio is very low because this experiment was optimised for the DFWM signal, which resulted in a weaker SFG signal due to chromatic aberration in the focussing objective. As DFWM is quadratic in the local pump intensity, the optimum focal position is different from the one for SFG, which is linear in both input fields. Another JSD map with an optimised SFG signal is shown in Figure 7.8(c). The signal-to-noise ratio is much improved and the trace seems to follow a resonance which can be attributed to the bar particle, as in Figure 7.8(a). Note, however, that the DFWM and SFG traces are slightly shifted with respect to each other which supports the hypothesis of imperfect substrate

DFWM suppression. On the other hand, we have demonstrated a way of obtaining JSD maps with our experimental setup, which allows for the characterisation of narrow resonances in the NIR on the order of $\lesssim \pm 10$ nm.

7.5 Efficiency and SET Analysis

So far, we have explored how the polarisation sensitivity of the antennae helps to find difficult-to-obtain nonlinear mixing signals in the IR, and how it can be used to separate the antenna DFWM signal from that of the substrate. Then, we presented scans to give an overview of the second- and third-order up- and down-conversion processes and to demonstrate that they are indeed the desired signals and not noise or a spurious signal at the wavelength expected from energy conservation. However, within the context of working towards nonlinear photon and photon pair sources, the key question is how efficient these processes are in the antennae, which will be discussed in this section.

Let us begin the analysis by noting that certain properties of the antenna array make it more difficult to determine meaningful efficiency values, compared to the GaN sample in the previous chapter. One key problem is that the pump spot diameter in our experiments is about 1 μm , which means that only a few antennae are excited, making the signal much more susceptible to fluctuations due to alignment. Moreover, the efficient coupling to each individual antenna is also much more sensitive to alignment fluctuations than the coupling to the GaN slab. In this analysis, we will follow a black-box approach, where we work with input and output photons without consideration for the details of the antenna array, which are only introduced in the discussion of each result in order to reduce complexity for the fundamental estimates. The rationale behind this is that we will still be able to make approximate estimates for nonlinear conversion efficiencies and spontaneous photon pair generation rates, particularly since we are only interested in an order-of-magnitude estimate at this stage. This will serve as a useful reference for later experiments with larger pump spots and arrays discussed in Section 7.7.

Analogously to the efficiency and SET analysis in Chapter 6, the experimental results of the second- and third-order frequency mixing processes are summarised in Table 7.1 before discussing the expected spontaneous photon pair generation rates and nonlinear conversion efficiencies. One general observation is that for the down-conversion processes, the number of photons picked up by the collection objective is 2 – 3 orders of magnitude higher than for the corresponding up-conversion processes. The reason for this is that the up-conversion spectra were taken with a 0.4 NA objective, rather than the 0.95 NA objective for the down-conversion spectra, which leads to a smaller spot size – see the illuminated area column – and a higher peak intensity. The larger spot size for the up-conversion experiments was necessary because of extreme fluctuations of the signal, possibly due to the increased alignment sensitivity near the $\lambda = 1500$ nm bar particle resonance, and operation close to the damage threshold of the antennae when using the 0.95 NA objective for the $\lambda = 1500$ nm seed beam. This is again caused by the good coupling to the bar particle at its dipole resonance at $\lambda = 1500$ nm, which makes it particularly easy to melt the particle and destroy the antenna resonance by closing the gaps, meaning that the damage threshold is very low at that wavelength.

	kcps	ph/s collected	ph/s stimulating	ph/s spontaneous	eff. pump power (mW)	illuminated area (μm^2)
SFG	0.8	$60 \cdot 10^3$	$6.4 \cdot 10^{15}$	$8 \cdot 10^{-4}$	2.0	3.6
DFG	3.4	$68 \cdot 10^6$	$9.2 \cdot 10^{15}$	0.6	6.0	0.22
DFWM up	4.0	$150 \cdot 10^3$	$3.2 \cdot 10^{15}$	$4 \cdot 10^{-3}$	3.0	3.6
DFWM down	2.6	$52 \cdot 10^6$	$5.3 \cdot 10^{15}$	0.78	2.4	0.25

Table 7.1 – Efficiency analysis. Signal counts for all processes were taken from appropriate spectra and the photon numbers calculated in the same way as in the previous chapter. The effective pump power and interaction area allow for a fair comparison of the different processes taking into account the experimental circumstances.

7.5.1 Spontaneous Photon Pair Generation Rates

Let us as a representative example discuss the expected SPDC photon pair generation rates from the DFG signal using the stimulated emission tomography (SET) framework [56] explained in Chapter 2 and applied analogously in Chapter 6. A key reason for choosing DFG is that its spontaneous analogue, SPDC, produces photon pairs at 840 nm and 1500 nm, given our experimental parameters. As explained in Chapter 1, the ability to generate spectrally separated photon pairs is crucial for potential future use in technologies such as quantum imaging [55]. At the mW-level average pump power in the experiment, about 0.6 pairs per second are expected to be collected if the seed beam is blocked. Note, that this is a conservative estimate, as the process may be seeded more efficiently by vacuum fluctuations than by the photons from the seed beam. The DFG signal is emitted via the bar particle at its dipole resonance, which can be approximated as a dipole antenna at a glass-air interface [238]. As a generous estimate, it is assumed that $\sim 10\%$ of the radiation is picked up by the collection objective. Therefore, we expect that about 6 SPDC photon pairs per second are generated in the antenna array.

Of course, it is desirable to increase this low photon pair generation rate. However, it is difficult to achieve this by increasing the input intensities, as the particles were pumped close to the damage threshold in our experiments, in contrast to the GaN sample. Working with a larger excitation spot size, such that more antennae are pumped close to their damage threshold seems to be the most promising approach to boost the spontaneous photon pair generation rate. Moreover, this approach offers the benefit of a more directional emission [239, 240] closer to the optical axis, increasing the fraction of pairs collected by the second objective. As a rough estimate, assuming that the radius of the interaction area can be increased by an order of magnitude by using a lens with a sufficiently long focal length, the interaction area would increase by two orders of magnitude, boosting the count rate estimate by a factor of ~ 100 .

Another way to increase the intrinsic efficiency is by optimising the design of the antennae themselves to support multiple resonances at the desired wavelengths to benefit from the plasmonic field enhancement. Such a design with a triple resonance at $\lambda = 550$ nm, 850 nm, and 1550 nm is presented in Section 7.6. Assuming an efficiency enhancement factor of ~ 10 for each resonance, there would be an enhancement factor of 100 compared to the antennae in this experiment, where only the bar particle resonance was used. Thus, we have found that with the current sample, about 600 SPDC pairs per second are a reasonable approximate limit of the sample investigated here, whereas 60 kHz and potentially higher pair generation rates are possible with the improved design that will be explained in Section 7.6. The SPDC pair generation rates and the improvements discussed above are summarised in Table 7.2. Comparing these SPDC count rates to the results from Santiago-Cruz and colleagues [185], as we have done in the previous chapter on GaN, again demonstrates that generating up to 600 Hz on our antenna sample is a competitive result, if the total detection loss of both signal and idler photons can be limited to about one order of magnitude in a coincidence measurement. This shows that while phase-matched large bulk crystals may be able to greatly outperform our antennae at specific wavelengths in terms of the photon pair generation rate [51], this is not necessarily the case if one wants to relax phase-matching constraints by working with thin crystals with a length on the order of ~ 100 nm. In particular, any improvement in the antenna design, as presented later in Section 7.6, may constitute substantial progress for this class of ‘phase-matching-free’ photon pair sources.

0.6 Hz x10	spontaneous, collected collection loss
6 Hz x100	spontaneous, generated large array
600 Hz x100	likely maximum for this sample new antenna design
60 kHz	likely maximum for new sample

Table 7.2 – Count rate estimate. Summary of SPDC pair generation rates likely achievable in the sample used in the frequency mixing experiments and in a new sample with an improved design, introduced later in Section 7.6.

7.5.2 Frequency Mixing Efficiency

Let us now discuss the nonlinear conversion efficiency η of the DFG process in our antenna array, using again the simple power ratio Eq.(6.3). The collected DFG signal amounts to 68 Mph/s, which establishes a lower bound of the generated DFG output power of 9 pW. At the mW input powers used in the experiment, this means that the minimum DFG efficiency is $\eta \sim 10^{-9}$. As explained in the previous subsection, this needs to be corrected for collection losses of about an order of magnitude, leading to an intrinsic DFG efficiency of the antenna array on the order of $\eta \sim 10^{-8}$.

This value of the DFG conversion efficiency can be improved with the strategies described in the previous subsection, but certain subtleties need to be taken into account. Increasing the excitation spot size would bring the collected DFG signal close to the generated DFG signal due to an improved emission pattern [239,240], as explained above. However, aside from a potential collective resonance at particular array spacings [241,242], the intrinsic efficiency would not be improved, as any increase in area by a given factor would not only increase the number of pumped antennae by the same factor, but also the required pump power to achieve the same local intensities. Moreover, the fraction of the array covered by antennae, which is $\sim 20\%$, reduces the efficiency by a factor of up to 5 compared to single-antenna excitation and needs to be compensated for by the more favourable emission pattern to result in an overall benefit. In contrast, improving the antenna design as shown later in Section 7.6 would be an effective way of increasing the intrinsic efficiency. As above, an enhancement of two orders of magnitude can be assumed as a rough estimate, leading to a DFG efficiency on the order of $\eta \sim 10^{-6}$. Comparing this to the GaN results in the previous chapter and the literature results on frequency mixing in dielectric nanostructures [52,186,187,243], the latter seem to be more suitable for applications, particularly considering their tolerance for higher pump powers, already achieved mixing efficiencies of $\eta > 10^{-5}$ and a higher potential upside for the nonlinear conversion efficiency. However, as explained in the previous subsection, the absolute achievable SPDC photon pair generation rates are likely competitive.

7.5.3 Reciprocity of Up- and Down-Conversion

Let us add to the efficiency analysis by checking the reciprocity of up- and down-conversion with the crude assumption that the antenna array behaves like an effective medium and the output signal scales with $A^{1-\nu}$, where A is the interaction area and ν is the order of the process, as established in the previous chapter. The numbers for the second-order processes in Table 7.1 can be put on an equal footing by multiplying the SFG counts by 3 to account for the lower pump power and by $3.6/0.22$ to compensate for the larger interaction area. Combining those effects leads to a compensation factor of $(3.6/0.22) \cdot 3 \approx 49$, while the expected spontaneous photon counts differ by a factor of ~ 463 , which means that for the second-order processes there is still an unexplained discrepancy of a factor of ~ 10 . For the third-order processes, the compensation factor is calculated in a similar fashion, just that the compensation factor for the interaction area needs to be squared due to the scaling law above, which was explained in Subsection 6.5.5 in the previous chapter. One obtains $(3.6/0.25)^2 \cdot 1.25^2 = 324$, while the expected spontaneous photon counts differ by a factor of ~ 195 , which is in reasonable agreement, within less than a factor of 2.

Given that the spacing between the antennae in the array is 750 nm and the spot sizes are so small that only a few antennae are excited, the effective medium approximation and the scaling law above are certainly oversimplifications. Consider, for example the case of a single antenna in the centre of the pump spot: reducing the illumination area at constant power then only increases the local intensity, but does not change the interaction area as long as the antenna is still fully covered by the excitation beams, leading to a scaling of $A^{-\nu}$ instead of $A^{1-\nu}$. This would bring the spontaneous counts deduced from the DFG and SFG experiments much closer together, but worsen the agreement between the DFWM processes. On the other hand, when the pump spot just covers some antennae in its peripheral region, any small reduction in size leads to a disproportionate loss of interaction volume, which may be more than what is compensated for by the increased intensity in the centre, such that the effective exponent in the scaling law may even be positive in some situations. This shows

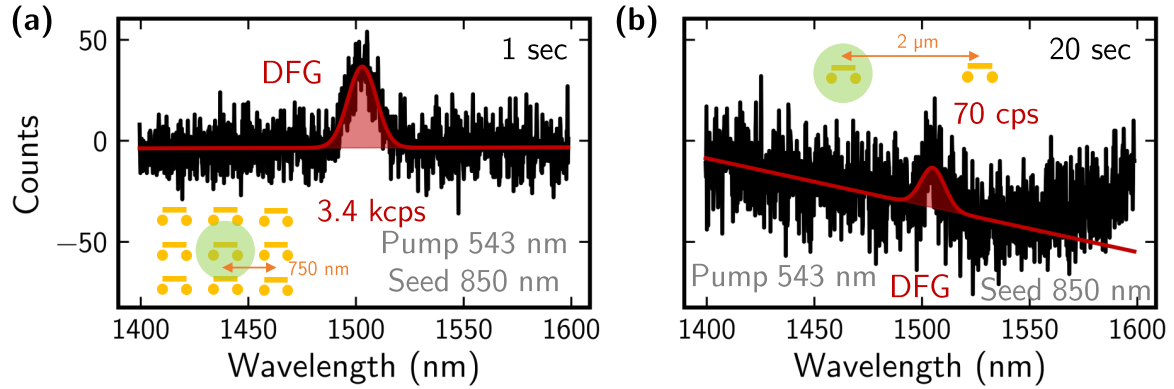


Figure 7.9 – DFG signal from an antenna array and a single antenna. (a) DFG spectrum on dense antenna array with 750 nm spacing. In the inset, a schematic of the array and the area of overlap of the input beams is shown as a green circle. (b) Repetition of the DFG experiment on an antenna array with 2 μm spacing, such that only one antenna is pumped. The luminescence background was slightly fluctuating between taking the subtracted reference spectrum and the DFG spectrum, which explains the curved background of the spectrum.

that it is difficult to make a reliable statement about the reciprocity of frequency conversion processes in the antennae; however, it seems likely that the reciprocity is not gravely violated by several orders of magnitude. It is important to note that unlike GaN, gold is not in good approximation a lossless medium in our frequency range, and thus the theoretically expected perfect reciprocity of up- and down-conversion processes, explained in Subsection 6.5.5 in the previous chapter, does not apply here. However, as stated above, despite the complexity of the system and the perhaps somewhat oversimplified analysis, a huge violation of 2 orders of magnitude or more seems very unlikely.

7.5.4 Frequency Mixing in a Single Antenna

Finally, let us compare DFG data from a dense and a sparse antenna array to provide more context to the above discussion about the number of antennae excited by the input beams. We shall also motivate that this measurement allows us to provide conservative estimates for conversion efficiencies and expected SPDC pair generation rates from a single antenna. Figure 7.9(a) shows the spectrum providing the DFG data in Table 7.1, which was taken on a dense antenna array with 750 nm spacing, on which all other data in this chapter was taken as well. In contrast, the spectrum in Figure 7.9(b) was taken on a sparse array with a lattice

constant of $2\ \mu\text{m}$. Both spectra were taken under the same experimental conditions in direct succession. The effective illumination area where both spots are overlapped was estimated to be $0.22\ \mu\text{m}^2$, corresponding to a beam radius of $r = 260\ \text{nm}$, which would only cover one antenna, regardless of whether the dense or the sparse array are illuminated. However, this estimate assumes a perfect beam quality with $M^2 = 1$, perfect positioning of the sample in the focal plane, and is derived from the spot size before the objective, with comes with an uncertainty on the order of 10%. Therefore, and given the ~ 50 times larger signal from the dense array it is likely that the effective radius of the interaction region is larger than 260 nm and that more than one antenna was excited on the dense array. If only one antenna was excited on both arrays, the discrepancy could be explained by a lower antenna quality or a too large inter-particle gap size on the sparse array. Alternative explanations are optical damage to the antenna during the measurement or a significant alignment drift, although the latter is very unlikely given the short acquisition time of the spectrum and the closed-loop piezoelectric stabilisation of the sample. In any case, given that the initially estimated radius of the interaction region $r = 260\ \text{nm}$ is almost an order of magnitude smaller than the antenna spacing on the sparse array, it can be said with high certainty that only one antenna was illuminated by the input beams on the sparse array.

Therefore, the sparse array spectrum in Figure 7.9(b) does allow for a lower bound estimate of the DFG efficiency from a single antenna, which is obtained by dividing the relevant efficiencies and generation rates in the discussion above by the factor of 50 by which the signal of the sparse array is smaller than that of the dense array. Including a $\sim 10\%$ collection efficiency, the intrinsic DFG efficiency of a single antenna is therefore at least $\eta \sim 2 \cdot 10^{-10}$. As for the SPDC photon pair generation rate from a single antenna, on the order of 0.1 Hz is expected for an antenna pumped at several mW with a pulse duration of 150 fs, a repetition rate of 80 MHz, and with an approximately diffraction-limited spot. With the new triple-resonance design introduced in the following section, the spontaneous pair generation rate from a single antenna may be increased to about 10 Hz.

7.6 New Triply-Resonant Antenna Design

In the experiments so far, only the resonance at $\lambda = 1500$ nm was used for the dual-beam frequency mixing experiments. The results suggest that despite the high efficiency per unit volume, the realistically achievable overall frequency mixing efficiencies are significantly worse than for the GaN sample in the last chapter. An additional problem for applications as a photon pair source is the high luminescence background, shown in Figure 7.7(a) and (b), which is absent in GaN. The luminescence background is expected to severely impact the counts-to-accidental ratio in coincidence measurements, and the heralding efficiency if the antenna is to be used as a heralded single photon source [50].

However, an improved antenna design with three resonances at all wavelengths of the nonlinear interaction might significantly improve the performance. Figure 7.10(a) shows an SEM image of such a newly designed antenna, which was not yet available at the time of the measurements for this thesis, where the disk particles of the antenna design in Figure 7.2 were replaced with rectangular particles. This has the effect of splitting the old disk resonance at $\lambda = 750$ nm into two new resonances around $\lambda = 550$ nm and $\lambda = 850$ nm, which can be seen in the simulated scattering cross-sections in Figure 7.10(b). The FDTD simulations were conducted with reduced particle sizes to take into account erosion effects, which are assumed to take away a margin of 5 nm around each particle. Moreover, the simulations were repeated for slightly different gap sizes to visualise the impact of small fluctuations on the antenna resonances. The predicted resonances look promising, however, the resonance of the small particles at $\lambda = 550$ nm is still relatively far away from the target wavelength. This may not be a problem due to the y -polarised resonance of the big particle, which would still provide resonant field enhancement at that wavelength. It will only be possible to draw more conclusions after taking FTIR extinction spectra of the sample to check if the predicted resonances do match those in the actual antennae.

The next step is to conduct dual-beam frequency mixing experiments to compare the efficiencies to those determined above. If an efficiency enhancement of several orders of magnitude

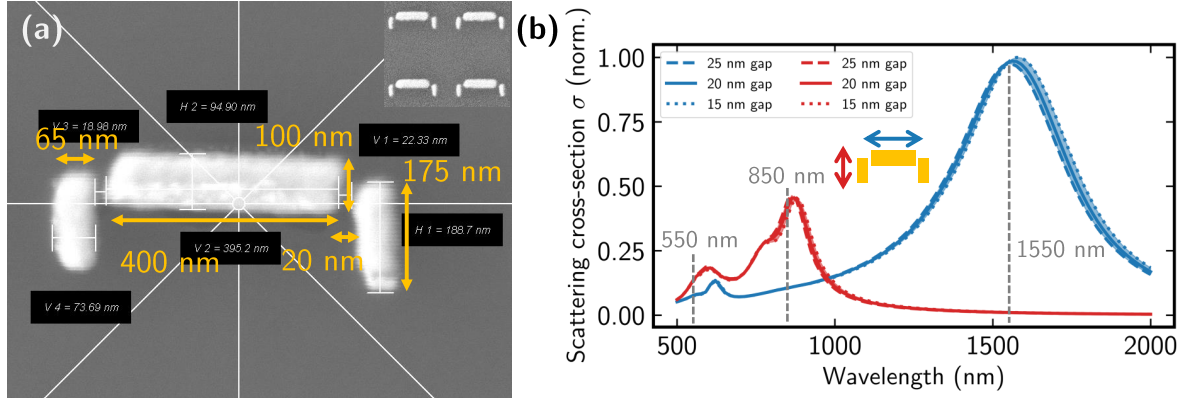


Figure 7.10 – New design for a triply resonant antenna.[†] (a) SEM image of a new antenna with two rectangular bars instead of disks, to support resonances at $\lambda = 850 \text{ nm}$ and $\lambda = 550 \text{ nm}$. The inset shows part of an array with the new antennae. (b) Simulated scattering cross-section spectra of the new antenna design taking into account an erosion of 5 nm around the particles. The shaded area indicates the impact of fluctuations of the inter-particle spacing due to fabrication imperfections, assuming a maximum deviation of 5 nm . There are three resonances at $\lambda = 550 \text{ nm}$, 850 nm , and 1550 nm , which promises to enhance SFG and DFG pumped at $\lambda = 850 \text{ nm}$ and seeded at either of the outer wavelengths.

[†]The author acknowledges Dr Xiaofei Xiao for fabricating the sample and for taking the SEM image in Figure 7.10(a).

is achieved, the luminescence problem mentioned above in the discussion of Figure 7.7 would be much less significant, as the same photon pair production rates could be achieved at lower pump powers with a proportionally lower luminescence signal.

7.7 Conclusion and Outlook

The purpose of this chapter was to introduce a system of gold nanoparticles for efficient nonlinear frequency mixing and to extend previous work [49] by systematic dual-beam studies. After a discussion of the second-order nonlinearity in gold and the considerations about the antenna design, dual-beam frequency mixing data was presented. First, preliminary experiments to showcase the polarisation properties of the antennae and to separate antenna and substrate contributions were discussed, followed by an overview of second- and third-order up- and down-conversion processes in the form of parameter scans to demonstrate the genuineness of the signals. After that, the nonlinear conversion efficiencies were discussed for the example of DFG, which, due to its link to SPDC, is of particular interest. An intrinsic

efficiency of $\eta \sim 10^{-8}$ and a spontaneous photon pair production rate of ~ 6 Hz was found for the antenna array on the sample. Reference measurements on a sparse array yielded a conservative estimate for frequency mixing in a single antenna about 50 times lower. Finally, in Section 7.6 a new triply resonant antenna design for improving the efficiency of the nonlinear mixing of the three participating wavelengths was introduced.

The obvious next step is to characterise the new triply resonant antennae and to test if the two additional resonances provide a performance boost. In the discussion of the DFG efficiency in Section 7.5 we have seen that unless the new antenna design provides at least an additional order-of-magnitude boost of the conversion efficiency per resonance, the antenna arrays are likely not competitive against dielectric particle systems [52, 186, 187, 243] in terms of the frequency mixing efficiency, highlighting the importance of these measurements. However, we have seen that in terms of the absolute SPDC pair generation rates, our antenna sample is likely competitive even without the new triple-resonance design, which might constitute a significant step forward for nanophotonic SPDC sources. Additionally, higher absolute photon pair generation rates could be achieved in experiments with large pump spot sizes to benefit from a collaborative resonance of all particles, as briefly mentioned in Section 7.5; however, it is then necessary to choose an appropriate spacing between the antennae [241, 242]. Spontaneous photon pair generation rates of ~ 60 kHz are expected after those improvements, as estimated in Section 7.5.

A promising alternative approach would be to coat the antennae with a high- $\chi^{(3)}$ polymer, such as MEH-PPV [244], using the plasmonic field enhancement to provide strong local fields, but benefiting from the nonlinear interactions throughout the volume of a nearby low-loss material. This strategy has been successfully used in the literature [245, 246] and can be readily applied to the new antenna design, with only small changes to the design parameters required to take into account the resonance shift [28] due to the higher refractive index of the polymer. Note that coupling the antennae to a high- $\chi^{(3)}$ material nicely links this work to the hybrid perovskite-antenna nonlinear metasurface presented in Chapter 5, which is highly

suitable for frequency mixing [134]. All those perspectives show that our antenna system is a promising platform for efficient low-footprint nonlinear frequency mixing.

Chapter 8

Conclusion

8.1 Summary

In this thesis we have seen how nonlinear optical processes can be strengthened using dielectric and plasmonic nanostructures. After an introductory overview of the research context in nanophotonics in Chapter 1, a selection of essential concepts in nonlinear optics and plasmonics was explained in Chapters 2 and 3 to underpin the experimental results. The first original results were presented in Chapter 4 covering Z-scan theory with a focus on data analysis in the presence of saturable absorption. It was found that the popular two-photon absorption Z-scan formulae cannot be directly used in that case, but that the data can be fitted as long as one is careful about the new meaning of the coefficient β , which in the original formulae denotes the two-photon absorption coefficient. A numerical framework for analysing more complicated saturation behaviour was introduced and applied to identify mistakes in alternative approaches in the literature.

In Chapter 5, the insights about Z-scan analysis were applied to data from 2D Ruddlesden-Popper-phase (2D RPP) perovskites, which were found to exhibit extremely strong saturable absorption with $\beta_{\text{eff}} = -256 \text{ cm/MW}$ at the fundamental exciton resonance of one of the flakes. Nonlinear refraction was also found to be strongly enhanced near the exciton

resonance, with a sign change from self-focussing to self-defocussing at resonance and peak values of $n_2 \approx \pm 10^{-13} \text{ m}^2/\text{W}$. Combining the 2D RPP flake with an array of Al nanoantennae to form a nonlinear metasurface was found to have two effects. Firstly, the exciton and plasmon resonances hybridised into two new modes, which was experimentally confirmed in linear and nonlinear characterisation measurements, however, only evidence for an imperfect hybridisation could be found. Secondly, the nonlinear absorption coefficient of the flake was found to be at least twofold enhanced in the presence of the antennae. However, as a caveat, this comparison was made with flakes of different thickness and is hence not yet a proof that the plasmonic field enhancement of the antennae caused a significant increase of the nonlinear absorption parameter.

Frequency mixing experiments on GaN cones and Au nanoantennae were shown in Chapters 6 and 7, respectively. In both cases, the genuineness of the mixing signals was confirmed with time delay scans and power scans that exhibited the expected power scaling. For the GaN experiment, the cone was found to enhance the SHG efficiency by about a factor of 10 compared to a bare GaN slab. However, complementary FDTD simulations showed that the enhancement of the nonlinear response is likely to originate from an optical focussing effect through the cone, and only to a small part from a Mie-resonance of the cone itself. Another finding was that the non-zero $\chi^{(2)}$ tensor elements in the sample require an electric field along the cone axis and thus the surface normal of the slab to yield a second-order response detectable with our free-space illumination scheme at normal incidence. The simulations also showed that this field component was enhanced near the neck region of the cone, which does explain part of the enhancement provided by the cone, but also suggests a strong sensitivity of the performance with respect to the fabrication accuracy. In the Au antennae experiment, the main achievement was the observation of a DFG signal in the telecommunications waveband, which to the knowledge of the author is the first time this mixing process has been shown in Au nanoantennae.

Both the GaN cone and the Au antenna sample were investigated with respect to their expected performance as an SPDC-based heralded single photon source using the SET technique.

In the case of the GaN sample, the SPDC pair creation rate was estimated to be around 200 Hz under experimental conditions analogous to the DFG measurements and taking into account collection losses as well as the fact that most photons are emitted into the slab. As the cone was only found to enhance SHG for an input around 1500 nm by a factor of 10, the pair creation rate from the cone could not be estimated higher than about 2 kHz given the experimental evidence. Using the SHG enhancement in an estimate based on DFG measurements is justified by the reciprocity of up- and down-conversion, which was demonstrated for $\chi^{(2)}$ and $\chi^{(3)}$ processes in the GaN sample by comparing the frequency conversion efficiencies. For the Au antennae, the measured SFG and DFG efficiencies did just about agree within a factor of 10 and hence reciprocity of up- and down-conversion could not be confirmed with high confidence, which is not unexpected given the losses in the antennae. An estimate of the intrinsic SPDC pair creation rate in the Au antenna array yielded a value of about 6 Hz for the sample and setup used in the experiments, however, it was estimated that about 60 kHz could be achieved, taking into account the additional enhancement with a new triply resonant antenna design and assuming illumination of hundreds of antennae in a large metasurface in future experiments. It was found that the performance hinges on the efficiency boost due to the triple resonance; if it significantly exceeds a factor of 10 per resonance, MHz on-chip photon pair generation rates could be within reach.

8.2 Outlook

In general, the experimental research presented in this thesis is in its early to mid stages; however, many insights into the performance of the samples were gained and partial results obtained to inform the next steps for a smooth transition to mature stages and a successful completion of the projects. To conclude this thesis, let us outline the future course of the three main experimental pathways.

The Z-scan project with the 2D RPP flakes is the most advanced part of this thesis, with a substantial result in terms of the analysis of complicated OA Z-scan data as a standalone

piece of work, and published results on the optical nonlinearity in the bare 2D RPPs [135]. The main challenge ahead is to better understand the complicated on-resonance saturation behaviour of the hybrid perovskite-antenna metasurface. So far, the results are phenomenological in nature and more insight into physical mechanisms would significantly strengthen this body of work. As suggested in the conclusion of Chapter 5, a metasurface sample with bar antennae similar to the sample in Ref. [41] instead of cross antennae would be a useful tool to that end, allowing for a polarisation dependent coupling to the antenna resonance and hence a controlled study about the impact of the field enhancement. Dual-beam experiments [247] should be conducted to study the timescales of exciton formation and saturation of the exciton ground state; with one pulsed beam in the continuum above the electronic bandgap and the other beam at the excitonic resonance, the variation of the transmittance could be studied as a function of the time delay. Additionally, fluorescence lifetime measurements [163] should be conducted to check the timescale of the exciton decay. Fast modulation times of the strong observed transmittance change would be promising for optical switching applications [248]. It should be noted that nW average powers were sufficient to achieve modulations on the order of 1%, which renders this platform promising for low-power switching. Of course, the largest open question is the change of the saturation behaviour on ms timescales, which is difficult to investigate and for which to the best knowledge of the author no obvious mechanisms are known in the literature, as we have seen in Chapter 5. In contrast, an easier but nonetheless promising study is a frequency scan of the THG efficiency [134] near the double resonance of the metasurface, based on the idea that both THG and the optical Kerr effect are third-order processes. A metasurface with bar antennae would be particularly suitable for demonstrating the effect of the field enhancement in such an experiment, with a likely huge impact given the E^3 scaling of the THG output. Moreover, the obtained efficiencies could be compared to previous results on 2D RPP flakes without antennae [134].

The second stream of experiments is composed of frequency mixing studies with GaN nanostructures as a follow-up to the cone results in Chapter 6. The main issue with the sample

studied so far is the suboptimal compatibility of the experimental setup and the alignment of the non-zero $\chi^{(2)}$ tensor elements. Additionally, as we have seen in Section 6.6, the cones are too large for the magnetic dipole resonance to fall within the IR C-band, which marks the longest wavelengths available in our setup. The second point is addressed by the new sample presented in Section 6.7, which was fabricated but not yet available for measurements at the time of writing this thesis. The first problem can be remedied by either working with m-plane samples [207], or by in-plane beam launching and collection schemes. Only if both issues are addressed, a GaN cone etched directly into a slab can be used to its maximum effect, and a sensible assessment of its benefits for nonlinear mixing can be given. Despite its lower $\chi^{(2)}$ nonlinearity compared to GaAs in the visible and NIR wavebands [38, 180], GaN is interesting due to its large transparency window [179] and its higher damage threshold compared to the Au antenna metasurface; however, a systematic comparative damage threshold study is needed to confirm and quantify this preliminary experimental finding.

Next steps in the Au antenna project, which is the third family of experiments, are centred around the triply resonant antenna design introduced in Section 7.6. As explained in Chapter 7, optimising the field enhancement and mode overlap is the crucial ingredient for making the Au nanoantenna metasurface a technologically interesting photon pair source. The main limitation is the relatively low damage threshold of the antennae, which means that frequency conversion efficiencies cannot be boosted further by increasing the intensities significantly beyond the values used in the experiments presented in Chapter 7. This makes large-area metasurfaces [249] attractive, as more antennae can be excited at the maximum non-damaging intensity and the emission properties improved due to the in-phase emission of an array of dipoles compared to a single dipole [239]. However, even including these factors in the estimate of SPDC count rates as above and in Chapter 7, only kHz pair generation rates are expected. This highlights again the importance of the optimised design and fabrication quality of each antenna in order to create a useful photon source for quantum applications. Finally, the key advantage of the metasurface, namely the control over the output photon properties in a sub-wavelength nonlinear interaction volume [250], could be

exploited by designing a collective nanoparticle resonance with a narrow linewidth [241]. The idea is that phase-matching constraints are avoided due to the small interaction volume [48] and that the photons are created in frequency modes defined by the collective resonance. For a sufficiently narrow resonance, the resulting biphoton state would be factorisable and hence heralding would yield pure and indistinguishable single photons, similar to an earlier work on photonic crystal fibres [83]. Testing the joint spectral density and hence the expected purity of the heralded output would be easily achievable, since a SET measurement with a scanned spectrally filtered seed beam was demonstrated on the existing antenna metasurface in Chapter 7. Thus, despite their preliminary nature, the results in this thesis pave the way for designer nanoscopic photon sources in the future.

Appendices

Appendix A

Details of the Experimental Setups

Included here are schematic drawings and descriptions of the setups used for the experiments presented in this thesis.

A.1 Z-Scan

The Z-scan setup used for the measurements presented in Chapter 5 is shown in Figure A.1. Given the relative experimental simplicity of Z-scan, the experimental setup is very similar to the schematic in Figure 4.1(a). An imaging setup was added to allow for a precise positioning of the laser spot on the perovskite flake with several μm accuracy. Note that the imaging lens has a much longer focal length than the main lens focussing the laser beam, in order to achieve a sufficient magnification. In practical terms, for an ideal Z-scan measurement, the modulation of the transmittance as a function of the stage position should be several times stronger than the noise, but not too strong to avoid complex higher-order effects, which means that one typically aims for a signal on the order of several per cent. The signal-to-noise ratio can be improved by reducing background counts as much as possible, which means operating in the dark and with all non-essential sources of light switched off. Regulating the power and repetition rate allows one to reach the desirable strength of the

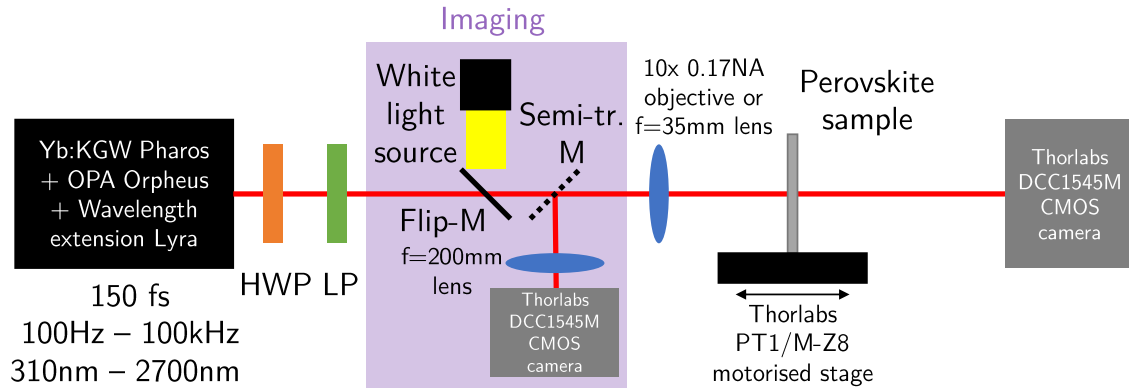


Figure A.1 – Z-scan experimental setup. The laser system can be tuned over a wide wavelength range and produces pulses of a duration of 150 fs at a repetition rate adjustable between 100 Hz and 100 kHz. The power is regulated by a combination of a half-wave plate (HWP) and a linear polariser (LP). Both closed- and open-aperture Z-scans can be taken with the camera by suitable integration of pixel values during post-processing of the images without requiring a physical aperture. The setup is similar to the Z-scan schematic in Figure 4.1(a) except for an additional imaging system consisting of a white-light source, a second lens, and a camera, for deterministically positioning the laser spot on the perovskite flake.

nonlinear effect, although high repetition rates should be avoided due to thermal effects, as mentioned in Chapter 4.

A.2 Frequency Mixing

The frequency mixing setup was used for both the GaN and Au nanoantenna metasurface samples, and hence for taking all the data in Chapters 6 and 7. However, depending on the target output wavelength and the desired nonlinear process, some modifications were necessary. Let us start with the down-conversion setup, shown in Figure A.2, which was used for measuring DFG in both GaN and the Au antennae. In practical terms, the alignment of the green 543 nm beam is greatly facilitated by its brightness, which makes it easy to align the transmission arm to the spectrometer and onto the closed slit. Note that the reflection arm is available but was not used for the DFG experiments so far. Moreover, no DFG wavelength scan was performed due to the weakness of the signal even when the slit in the wavelength selection part of the setup (orange) was removed. Note that 1-2 orders

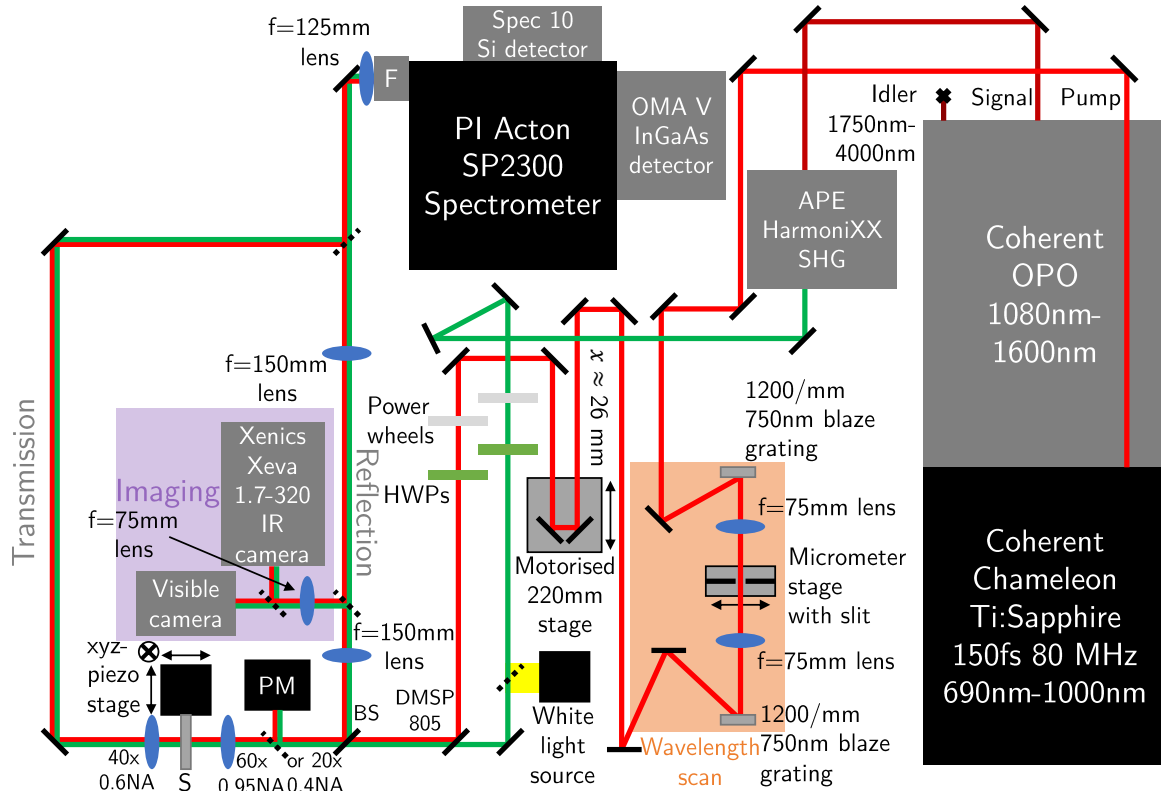


Figure A.2 – Down-conversion setup. Two input beams are generated in the laser and OPO, after which the IR beam from the OPO is frequency-doubled. In a typical DFG measurement, the pump wavelength is set to 850 nm and the signal wavelength from the OPO set to 1086nm, which is then converted to 543 nm. The pump beam is sent through a wavelength-selection setup (orange) consisting of gratings, lenses for collimating, and a slit. This is followed by a Thorlabs DDS220 stage with a retro-reflector to regulate the delay between the two beams. The overlap position in this specific configuration of the setup was at a stage position of around 26 mm. Power and polarisation can be set with power wheels and HWP's. Imaging the sample (S) is possible with a visible Logitech webcam and a scientific IR camera (purple), with light from a white light source or an adjacent IR light source (not shown). The power incident on the sample can be monitored with a powermeter (PM). The DFG signal can be picked up from the sample in transmission and reflection. Long-pass filters (F) before the spectrometer select the DFG signal, which is detected with a 1024x1 pixel liquid-nitrogen-cooled Princeton Instruments OMA V InGaAs detector.

of magnitude of signal are typically lost when inserting the slit for the SET-type wavelength scans.

The up-conversion setup is shown and described in Figure A.3. Note that the gratings in the wavelength-selection part of the setup need to be changed to minimise losses. The 75 mm lenses should be changed as well, since the anti-reflection coatings are different for this wavelength region around 1500 nm. When carrying out a wavelength scan by moving the slit on the micrometer stage in the orange-coloured part of the setup, the laser spot on

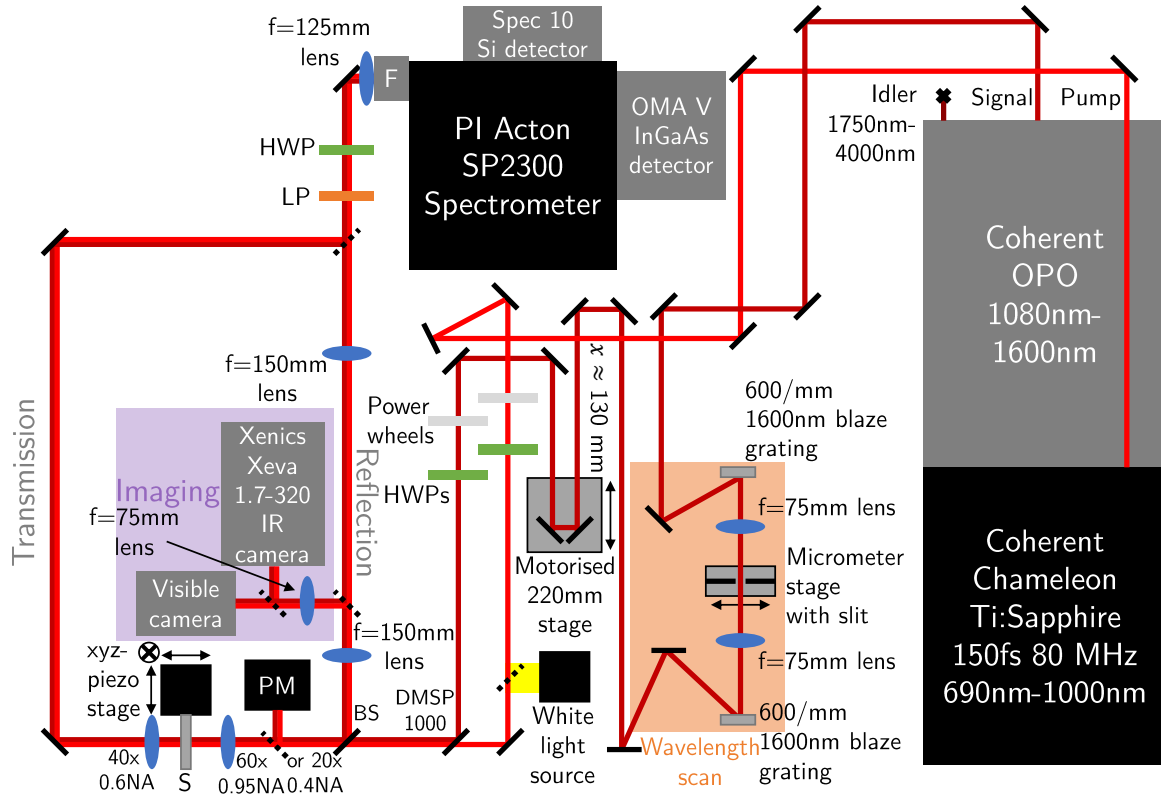


Figure A.3 – Up-conversion setup. Similar to the down-conversion setup in Figure A.2, except that the signal beam is sent to the wavelength-selection setup and is not frequency-doubled. The output polarisation of the up-conversion signals from the sample can be resolved with a linear polariser (LP), and a HWP allows for compensating for the polarisation-dependent efficiency of the spectrometer. The up-conversion signals were measured with the Princeton Instruments Spec 10 Si detector with 1340x100 pixels and the filters (F) were appropriate short-pass filters.

the sample (S) was found to slightly drift vertically, which needs to be compensated for by re-optimising the signal by adjusting the angle of the DMSP1000 at every change of the slit position. In the experiments, the transmission arm was used most of the time, however, for the tilt study and for the wavelength scan on the GaN sample, the reflection arm was used. In terms of the objective lenses, the 20x 0.4NA NIR apochromat by Mitutoyo has the advantage of a long working distance of about 2 cm and was hence used for the tilt study on GaN, but also for the up-conversion measurements on the Au antennae due to a better stability of the signal, which can be attributed to the larger laser spot size compared to the Nikon 60x 0.95NA objective.

Bypassing the wavelength-selection setup as shown in Figure A.4 requires a longer beam path of the pump beam to ensure that the overlap is within the range of the delay stage.

Appendix B

Fabry-Perot Model of the GaN Slab

Included here is a summary of the Fabry-Perot model which was used in Chapter 6 in an attempt to understand the wavelength dependence of the SHG signal shown in Figure 6.4, a study that was carried out to identify potential resonant features. The GaN slab is sandwiched between air and the sapphire substrate as shown in Figure 6.4(c) and the reflectance at each interface can be determined in the usual way with Fresnel's formulae

$$R_1 = \frac{(n_{\text{GaN}} - 1)^2}{(n_{\text{GaN}} + 1)^2} \quad R_2 = \frac{(n_{\text{GaN}} - n_{\text{Al}_2\text{O}_3})^2}{(n_{\text{GaN}} + n_{\text{Al}_2\text{O}_3})^2} \quad (\text{B.1})$$

and suitable refractive index data.[†] Since the wavelength scan was carried out in the range between $\lambda = 1200$ nm and 1600 nm, with the second harmonic ranging from $\lambda = 600$ nm to 800 nm it was sufficient to interpolate the data in those intervals, which is shown in Figure B.1 below. Knowledge of R_1 and R_2 at the relevant wavelengths allows for a calculation of the transmittance and reflectance of the GaN slab, as well as the field enhancement inside it. The corresponding formulae can be derived following the approach in Ref. [116], leading

[†]The data, used below in Figure B.1 was originally published in Refs. [251, 252] for GaN and sapphire, respectively, but for the purposes of plotting taken from <https://refractiveindex.info/?shelf=main&book=GaN&page=Barker-o> and <https://refractiveindex.info/?shelf=main&book=Al2O3&page=Malitson-o>, both accessed on 9 October 2020.

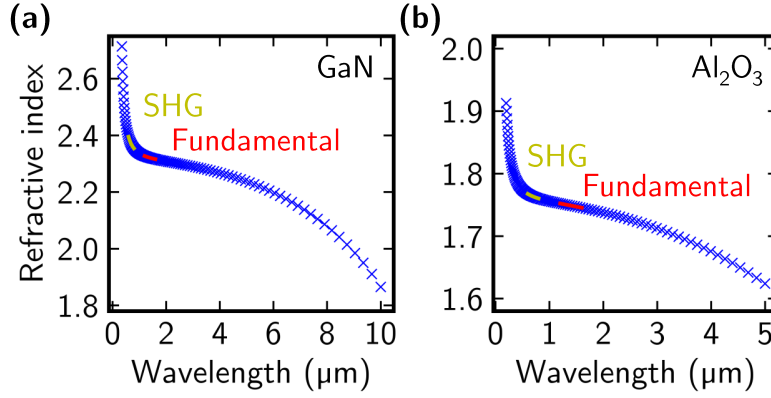


Figure B.1 – Refractive index data for the Fabry-Perot model. (a) GaN refractive index data [251] with the interpolated sections $\lambda = 1200$ nm to 1600 nm (red) and $\lambda = 600$ nm to 800 nm (yellow) relevant for the SHG wavelength scan. (b) The same for the sapphire refractive index data [252].

to the expressions

$$\frac{I_{\text{circ}}}{I_{\text{in}}} = \frac{1 - R_1}{(1 - \sqrt{R_1 R_2})^2 + 4\sqrt{R_1 R_2} \sin^2 \phi} \quad (\text{B.2})$$

for the field enhancement factor inside the slab with the single-pass phase $\phi = \frac{2\pi}{\lambda}nd$,

$$\frac{I_{\text{transm}}}{I_{\text{in}}} = (1 - R_2) \frac{I_{\text{circ}}}{I_{\text{in}}} \quad (\text{B.3})$$

for the transmitted intensity, and

$$\frac{I_{\text{refl}}}{I_{\text{in}}} = 1 - \frac{I_{\text{transm}}}{I_{\text{in}}} \quad (\text{B.4})$$

for the reflected intensity. The measured SHG signal is expected to be optimised whenever the maxima of the enhancement at the fundamental and second harmonic wavelength coincide. Hence, the SHG counts as a function of the wavelength were compared to

$$\left(\frac{I_{\text{circ,fundamental}}}{I_{\text{in,fundamental}}} \right)^2 \cdot \frac{I_{\text{circ,SHG}}}{I_{\text{in,SHG}}} \quad (\text{B.5})$$

where the enhancement factor of the pump is squared due to the relationship $I_{\text{SHG}} \propto I_{\text{fundamental}}^2$. This approach produced a reasonably good agreement with the measured SHG intensity variations as a function of the wavelength. The surprising finding was that the SHG

maxima in transmission and reflection coincided instead of being offset by half a period. In particular, attempting to explain the reflected SHG data with an expression proportional to $\frac{I_{\text{refl,SHG}}}{I_{\text{in,SHG}}}$ failed, possibly because the corresponding minimum of the intensity build-up of the second harmonic inside the slab could mean that the frequency mixing efficiency is actually reduced in this case. The problem is that SHG in the slab is more complicated than a simple monochromatic linear Fabry-Perot scenario, since the fields at the second harmonic are generated inside the slab during the build-up of the fundamental instead of being inserted. A more reliable prediction of the wavelength dependence of SHG emitted from the GaN slab would require a numerical simulation, which, however, goes beyond what is needed in this discussion.

Appendix C

Details of the SET Analysis

Included here are supplementary measurements and calculations necessary for the stimulated emission tomography (SET) analysis in Chapters 6 and 7.

C.1 Detection Efficiency Measurements

In order to estimate the number of photons generated on the sample from the measured spectra, it is necessary to take reference spectra with a known photon number to establish a link between the photon number and the detector counts. An alternative is to estimate the losses at each individual component in the collection beam path until detection and then to multiply the remaining photon number with the detector efficiency; however, some estimates, particularly for components inside the spectrometer, may be imprecise and hence the method with the reference spectra is likely to yield a far better estimate. For all SET measurements, the signal was collected via the transmission arm and hence the reference spectra were taken with an attenuated laser beam transmitted through both objectives. The attenuation was achieved with neutral density filters characterised separately at each wavelength. Thus, after dividing by this extra attenuation, which was necessary to avoid saturating and potentially

damaging the detectors, the overall detection efficiency via the transmission arm

$$\eta = \frac{\text{\#counts detected}}{\text{\#photons behind collection objective}} \quad (\text{C.1})$$

could be determined from the reference spectrum and a power measurement directly after the collection objective. A summary of η for the wavelengths of interest, for different polarisations, and for different focussing objectives is given in Table C.1 below.

λ		Without GaN		With GaN	
		Horizontal	Vertical	Horizontal	Vertical
1500 nm	20x	0.02%		0.005%	
	60x	0.02%		0.02%	
583 nm	20x	3.3%	4.4%	2.7%	3.9%
	60x	3.2%	4.5%	2.5%	3.5%
543 nm	20x	1.6%	2.2%	1.6%	2.3%
	60x	1.8%	2.2%	1.3%	1.7%

Table C.1 – Collection efficiencies η via the transmission arm. The second column refers to the magnification of the focussing objective. A Nikon CFI Plan Apo Lambda 40XC 0.6 NA objective was used for collection in all cases.

An extra set of reference measurements was taken for the case of the GaN sample inserted at the focus, since the slightly rough back surface caused a worse collimation of the beam after the collection objective, which needed to be taken into account when dealing with the GaN frequency mixing data. With a known η , it is only necessary to divide by the known transmittance of the collection objective at the wavelength in question and by an estimate of the fraction of photons collected by the objective in order to obtain the number of photons generated in the sample.

C.2 Beam Calculations

Other than the overall collection and detection efficiency, it was also necessary to estimate the spatial overlap and the number of photons interacting in the cylinder-shaped effective interaction region, defined by the waist size w_0 of the smaller beam and the depth of focus

$2z_R$ of the shorter depth-of-focus beam, as illustrated earlier in Figure 6.7(a). Since the aperture of the focussing objectives was not always completely filled, it was not sufficient to simply use the specified NA to determine the focussed spot size, but instead to evaluate

$$\text{NA}_{\text{eff}} = \sin \left(\arctan \left(\frac{w_{0,\text{in}}}{\text{efl}} \right) \right) \quad (\text{C.2})$$

which follows from simple trigonometry, as illustrated in Figure C.1. Next, the relation

$$w_{0,\text{focus}} = \frac{\lambda}{\pi \arcsin(\text{NA}_{\text{eff}})} = \frac{\lambda}{\pi \arctan \left(\frac{w_{0,\text{in}}}{\text{efl}} \right)} \quad (\text{C.3})$$

following from the divergence half-angle $\theta = \arcsin(\text{NA}_{\text{eff}}) = \lambda/(\pi w_0)$ of a Gaussian beam with minimum radius w_0 was used. Using $A = \pi w_0^2$ and $z_R = \pi w_0^2/\lambda$ the dimensions of the interaction regions could be determined.

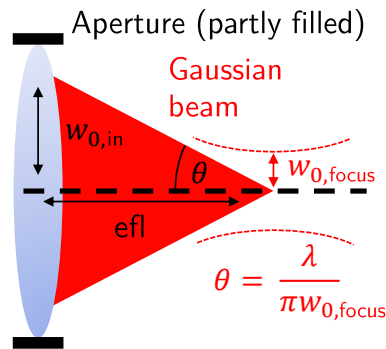


Figure C.1 – Gaussian beam in a partly filled aperture. Diagram for illustrating the quantities used in Eqs.(C.2) and (C.3). The effective numerical aperture NA_{eff} is calculated from the measured beam size $w_{0,\text{in}}$ before the objective and its effective focal length (efl).

The two focussing objectives used, a Mitutoyo 20x M Plan APO NIR (Thorlabs MY20X-824) and a Nikon CFI Plan Apo Lambda 60XC 0.95 NA have an effective focal length of 10 mm and 3.3 mm, respectively. The FWHM spot size just before the focussing objective was measured at all input wavelengths and the cylindrical effective interaction regions calculated.

C.3 SET Details for the GaN Sample

Let us now evaluate the expressions for the effective interaction volume for the four different frequency mixing measurements in GaN, shown in Figure 6.7(b)-(e). The results are presented in Table C.2 below and allow for an estimate of the fraction of incident photons participating in the nonlinear interaction.

Process		λ (nm)	Area (μm^2)	Length (μm)	Photons interacting
SFG	pump	840	0.53	1.25	100%
	seed	1500	0.64	0.85	83%
DFG	pump	543	1.51	5.55	33%
	seed	850	0.53	1.25	100%
DFWM up	pump	840	0.53	1.25	100%
	seed	1500	0.64	0.85	83%
DFWM down	pump	840	0.53	1.25	47%
	seed	583	0.25	0.87	100%

Table C.2 – Interaction volumes GaN. Summary of the interaction volumes in all 4 frequency mixing processes in GaN that were analysed with the SET method.

These values, together with the power of the input beams, the transmittance values of the objective lenses, and the overall detection efficiency η determined above allow for a calculation of the expected photon number in unseeded experiments with the SET technique. The results of the SET calculations are summarised in Table 6.1; a detailed version is shown below in Table C.3.

As stated in Chapter 6, the GaN data supports the notion of reciprocity between up-conversion and down-conversion processes when the correction factors for the different experimental conditions are taken into account. In terms of the expected count rates, the spontaneous counts estimated in Table C.3 refer to the expected collected photon number in an unseeded experiment; however, the actual number of photons generated in the sample in such an experiment is much higher. In Chapter 6, DFG is considered as an example for an order-of-magnitude estimate for the number of photons generated in the sample, since

it is the stimulated version of SPDC. The arguments in the main text lead to an expected SPDC pair generation rate of ~ 2 kHz.

Process	SFG	DFG	DFWM up	DFWM down
Spectrum kcps	77	1.6	484	9.9
η detection	1.3%	0.02%	2.5%	0.02%
T collection objective	84%	25%	82%	25%
Ph/s collected ($\times 10^6$)	7	32	24	200
Seed beam $\bar{P}$ (mW)	5.0	10.0	5.0	5.0
Fraction interacting	83%	100%	83%	100%
T focussing objective	37%	64%	37%	90%
Ph/s stimulating ($\times 10^{15}$)	11.6	27.0	11.6	13.2
Ph/s spontaneous	0.05	0.09	0.17	1.21
Pump beam $\bar{P}$ (mW)	5.0	6.5	5.0	10.0
Fraction interacting	100%	33%	100%	47%
Ph/s per mW pump	0.01	0.04	0.007	0.055
Interaction length (μm)	0.85	1.25	0.85	0.87
Interaction area (μm^2)	0.53	0.53	0.53	0.25
Correction factor	1	2.16	1	4.71

Table C.3 – SET calculations GaN. Details of the SET calculations summarised in Table 6.1. The first section of this table deals with determining the number of photons collected from the interaction volume using the measured spectra and calibration measurements. The second section determines the number of stimulating photons inside the interaction volume. The expected number of photons in the spontaneous process is found by applying the SET formula and dividing the photons from the stimulated process by the number of stimulating photons and multiplying by the repetition rate of 80 MHz, since the interaction with the determined ratio of stimulated to stimulating photons takes place in every pulse. The remaining two sections of the table deal with the comparability of the processes and show the correction factors to take into account the different pump powers and the differently sized interaction volumes. The correction factor is calculated from the relation $n_{\text{ph}} \propto A^{1-\nu} L^2$ explained and justified in Chapter 6, where A is the effective interaction area, L the interaction length, and ν the order of the process.

C.4 SET Details for the Au Nanoantenna Sample

The SET calculations for the Au nanoantenna sample are largely analogous to those for the GaN sample except for two main differences. The first one is that the interaction volume is defined by the nanoparticles and hence the effective interaction length is ~ 50 nm regardless of the depth of focus. The second difference is closely linked and concerns the treatment

of the metasurface as an effective medium. As shown in Chapter 7, the periodicity of the metasurface is 750 nm and the largest dimension of any nanoparticle is 350 nm, which in our wavelength range is approximately wavelength-sized and just at the boundary of what could still be treated as an effective medium with some justification. Of course, several unit cells per wavelength would be needed for this to be a good description, however, for the purposes of an order-of-magnitude SET estimate it seems to be acceptable to carry out the calculations with the effective-medium assumption as a rough first guidepost for future work. As in the case of the GaN sample, four different frequency mixing processes were investigated, with the relevant beam parameters in the experiment summarised below in Table C.4.

Process		λ (nm)	Area (μm^2)	Photons interacting
SFG	pump	840	3.6	100%
	seed	1500	4.2	86%
DFG	pump	543	0.22	100%
	seed	850	0.53	42%
DFWM up	pump	840	3.6	100%
	seed	1500	4.2	86%
DFWM down	pump	840	0.53	47%
	seed	583	0.25	100%

Table C.4 – Interaction areas Au nanoantenna array. Summary of the interaction areas in all 4 frequency mixing processes in the Au nanoantennae that were analysed with the SET method.

Note that the depth of focus is not included in the table due to the short interaction region, as explained above. The SET calculations shown below in Table C.5 were carried out analogously to those for the GaN sample.

As discussed in Chapter 7, the SET results support the reciprocity of up- and down-conversion for the third-order processes, but not the second-order processes. In terms of the expected SPDC pair generation rates, the expected SPDC detection rate from the DFG column in Table C.5 needs to be corrected for the fraction of photons not collected by the objective. As discussed in Chapter 7, a generation rate of ~ 6 Hz is expected. When including improvements in the antenna design, such as a triple resonance, and a larger illumination area,

SPDC count rates of 60 kHz are expected, which may be further increased if each antenna resonance boosts the mixing efficiency by more than an order of magnitude. For a single antenna, an expected SPDC pair generation rate of ~ 0.1 Hz was found; with the new triply resonant design shown in Section 7.6 this rate might be increased to ~ 10 Hz.

Process	SFG	DFG	DFWM up	DFWM down
Spectrum kcps	0.8	3.4	4.0	2.6
η detection	1.6%	0.02%	3.3%	0.02%
T collection objective	84%	25%	82%	25%
Ph/s collected ($\times 10^6$)	0.06	68	0.15	52
Seed beam $\bar{P}$ (mW)	2	8	1	2
Fraction interacting	86%	42%	86%	100%
T focussing objective	49%	64%	49%	90%
Ph/s stimulating ($\times 10^{15}$)	6.4	9.2	3.2	5.3
Ph/s spontaneous	0.0008	0.6	0.004	0.78
Pump beam $\bar{P}$ (mW)	2	6	3	5
Fraction interacting	100%	100%	100%	47%
Ph/s per mW pump	0.0004	0.1	0.001	0.14
Interaction area (μm^2)	3.6	0.22	3.6	0.25
Correction factor	1	16	1	207

Table C.5 – SET calculations Au nanoantennae. Details of the SET calculations summarised in Table 7.1. Analogous to Table C.3 above.

Appendix D

Mie-Modes in Dielectric Spheres

This is a comparison of several analytical field distributions from Mie-theory [253] for a dielectric sphere against the simulated field distributions in the GaN cone shown earlier in Figure 6.9. The direct juxtaposition of analytical theory and numerical simulations below in Figure D.1 is meant to illustrate and support the claim made in the discussion of Figure 6.9 that the cone modes at $\lambda = 1330$ nm and 2420 nm are predominantly of an electric dipole and magnetic dipole nature, respectively. The analytical expressions for the fields in question are easily evaluated from the general formulae in Ref. [101]. For an electric dipole, we have

$$\Re(E_x)|_{y=0} \propto 2 \sin^2 \theta \left(\frac{\sin \rho}{\rho^3} - \frac{\cos \rho}{\rho^2} \right) + \cos^2 \theta \left(\frac{\sin \rho}{\rho} \left(1 - \frac{1}{\rho^2} \right) + \frac{\cos \rho}{\rho^2} \right) \quad (\text{D.1})$$

and

$$\Re(E_z)|_{y=0} \propto 2 \sin \theta \cos \theta \left(\frac{\sin \rho}{\rho^3} - \frac{\cos \rho}{\rho^2} \right) - \sin \theta \cos \theta \left(\frac{\sin \rho}{\rho} \left(1 - \frac{1}{\rho^2} \right) + \frac{\cos \rho}{\rho^2} \right). \quad (\text{D.2})$$

For a magnetic dipole, we find

$$\Re(H_y)|_{y=0} \propto \frac{\sin \rho}{\rho} + \frac{\rho \cos \rho - \sin \rho}{\rho^3}. \quad (\text{D.3})$$

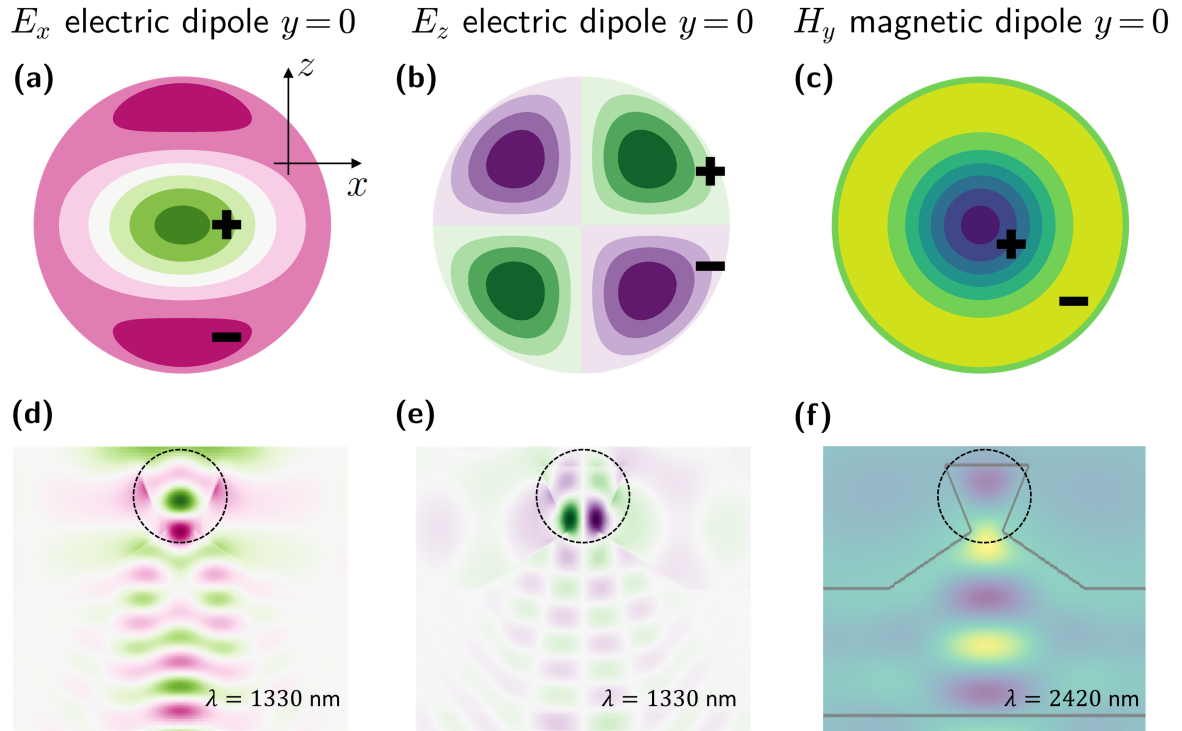


Figure D.1 – Analytical Mie theory and fields in cone. (a) E_x distribution in a dielectric sphere at the electric dipole mode. (b) E_z distribution in a dielectric sphere at the electric dipole mode. (c) H_y distribution in a dielectric sphere at the magnetic dipole mode. (d) E_x distribution in the GaN cone at $\lambda = 1330$ nm. Re-plot of the corresponding panel in the inset of Figure 6.9. (e) E_z distribution in the GaN cone at $\lambda = 1330$ nm. Re-plot of the corresponding panel in the inset of Figure 6.9. (f) H_y distribution in the GaN cone at $\lambda = 2420$ nm. GaN outline included for clarity, since the boundaries are not as obvious as in (d),(e). The dashed circle is a guide to the eye, indicating the area relevant for the comparison. The incident light is x -polarised in all cases.

Bibliography

- [1] Kirchain, R. & Kimerling, L. A roadmap for nanophotonics. *Nature Photonics* **1**, 303–305 (2007).
- [2] Novotny, L. & Hecht, B. *Principles of nano-optics* (Cambridge university press, 2012).
- [3] Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie* **9**, 413–468 (1873).
- [4] Gramotnev, D. K. & Bozhevolnyi, S. I. Plasmonics beyond the diffraction limit. *Nature photonics* **4**, 83–91 (2010).
- [5] Maxwell, J. C. A dynamical theory of the electromagnetic field. *Philosophical transactions of the Royal Society of London* 459–512 (1865).
- [6] Sturley, K. R. Radio technology. Encyclopedia Britannica, 20 Sep. 2019. URL <https://www.britannica.com/technology/radio-technology>. Accessed 26 May 2021.
- [7] Hertz, H. Ueber sehr schnelle electrische Schwingungen. *Annalen der Physik* **267**, 421–448 (1887).
- [8] Marconi, G. Wireless telegraphic communication. *Nobel Lecture* **11**, 198–222 (1909).
- [9] Novotny, L. & Van Hulst, N. Antennas for light. *Nature photonics* **5**, 83 (2011).
- [10] Engheta, N. Circuits with light at nanoscales: optical nanocircuits inspired by metamaterials. *Science* **317**, 1698–1702 (2007).

- [11] Bardeen, J. & Brattain, W. H. The transistor, a semi-conductor triode. *Physical Review* **74**, 230 (1948).
- [12] Shockley, W. The theory of p-n junctions in semiconductors and p-n junction transistors. *Bell System Technical Journal* **28**, 435–489 (1949).
- [13] Shockley, W. A unipolar "field-effect" transistor. *Proceedings of the IRE* **40**, 1365–1376 (1952).
- [14] Atalla, M. M. & Kahng, D. Silicon-silicon dioxide field induced surface devices. In *RE-AIEE Solid State Device Research Conference, Pittsburgh, PA. June 1960* (Carnegie Mellon University Press, 1960).
- [15] Moore, G. E. Cramming more components onto integrated circuits. *Electronics Magazine* **38** (1965).
- [16] Bossung, J. W. & Muraski, E. S. Optical advances in projection photolithography. In *Developments in Semiconductor Microlithography III*, vol. 135, 16–23 (International Society for Optics and Photonics, 1978).
- [17] Pease, R. Electron beam lithography. *Contemporary Physics* **22**, 265–290 (1981).
- [18] Reyntjens, S. & Puers, R. A review of focused ion beam applications in microsystem technology. *Journal of micromechanics and microengineering* **11**, 287 (2001).
- [19] Watt, F., Bettiol, A., Van Kan, J., Teo, E. & Breese, M. Ion beam lithography and nanofabrication: a review. *International Journal of Nanoscience* **4**, 269–286 (2005).
- [20] Wagner, C. & Harned, N. Lithography gets extreme. *Nature Photonics* **4**, 24–26 (2010).
- [21] Fan, D. & Ekinici, Y. Photolithography reaches 6 nm half-pitch using extreme ultraviolet light. *Journal of Micro/Nanolithography, MEMS, and MOEMS* **15**, 033505 (2016).

- [22] Saleh, B. E. & Teich, M. C. *Fundamentals of Photonics, Second Edition* (John Wiley & Sons, Inc., 2007).
- [23] Maier, S. A. & Atwater, H. A. Plasmonics: Localization and guiding of electromagnetic energy in metal/dielectric structures. *Journal of applied physics* **98**, 10 (2005).
- [24] Tkach, R. W. Scaling optical communications for the next decade and beyond. *Bell Labs Technical Journal* **14**, 3–9 (2010).
- [25] Soref, R. The past, present, and future of silicon photonics. *IEEE Journal of selected topics in quantum electronics* **12**, 1678–1687 (2006).
- [26] Maier, S. A. *et al.* Plasmonics – a route to nanoscale optical devices. *Advanced materials* **13**, 1501–1505 (2001).
- [27] Maier, S. A. Plasmonics: The promise of highly integrated optical devices. *IEEE Journal of selected topics in Quantum Electronics* **12**, 1671–1677 (2006).
- [28] Maier, S. A. *Plasmonics: fundamentals and applications* (Springer Science & Business Media, 2007).
- [29] Baranov, D. G. *et al.* All-dielectric nanophotonics: the quest for better materials and fabrication techniques. *Optica* **4**, 814–825 (2017).
- [30] Atwater, H. A. The promise of plasmonics. *Scientific American* **296**, 56–63 (2007).
- [31] Conway, J. A., Sahni, S. & Szkopek, T. Plasmonic interconnects versus conventional interconnects: a comparison of latency, crosstalk and energy costs. *Optics express* **15**, 4474–4484 (2007).
- [32] Huang, X., Jain, P. K., El-Sayed, I. H. & El-Sayed, M. A. Plasmonic photothermal therapy (PPTT) using gold nanoparticles. *Lasers in medical science* **23**, 217–228 (2008).

- [33] Heidarzadeh, H. Analysis and simulation of a plasmonic biosensor for hemoglobin concentration detection using noble metal nano-particles resonances. *Optics Communications* **459**, 124940 (2020).
- [34] Zhao, J., Zhang, X., Yonzon, C. R., Haes, A. J. & Van Duyne, R. P. Localized surface plasmon resonance biosensors. *Nanomedicine* **1**, 219–228 (2006).
- [35] Kinkhabwala, A. *et al.* Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nature Photonics* **3**, 654–657 (2009).
- [36] Dadosh, T. *et al.* Plasmonic control of the shape of the Raman spectrum of a single molecule in a silver nanoparticle dimer. *ACS nano* **3**, 1988–1994 (2009).
- [37] Kauranen, M. & Zayats, A. V. Nonlinear plasmonics. *Nature Photonics* **6**, 737 (2012).
- [38] Boyd, R. W. *Nonlinear optics* (Academic press, 2020).
- [39] Pors, A. & Bozhevolnyi, S. I. Plasmonic metasurfaces for efficient phase control in reflection. *Optics express* **21**, 27438–27451 (2013).
- [40] Dickson, W. *et al.* Dielectric-loaded plasmonic nanoantenna arrays: a metamaterial with tuneable optical properties. *Physical Review B* **76**, 115411 (2007).
- [41] Alam, M. Z., Schulz, S. A., Upham, J., De Leon, I. & Boyd, R. W. Large optical non-linearity of nanoantennas coupled to an epsilon-near-zero material. *Nature Photonics* **12**, 79 (2018).
- [42] Huang, Z., Baron, A., Larouche, S., Argyropoulos, C. & Smith, D. R. Optical bistability with film-coupled metasurfaces. *Optics letters* **40**, 5638–5641 (2015).
- [43] Zhang, Z. *et al.* All-optical switch and logic gates based on hybrid silicon-Ge₂Sb₂Te₅ metasurfaces. *Applied optics* **58**, 7392–7396 (2019).
- [44] Xie, Z. T., Wu, J., Fu, H. & Li, Q. Tunable electro-and all-optical switch based on epsilon-near-zero metasurface. *IEEE Photonics Journal* **12**, 1–10 (2020).

- [45] Chen, P.-Y., Argyropoulos, C. & Alù, A. Enhanced nonlinearities using plasmonic nanoantennas. *Nanophotonics* **1**, 221–233 (2012).
- [46] Kuznetsov, A. I., Miroshnichenko, A. E., Brongersma, M. L., Kivshar, Y. S. & Lukyan-chuk, B. Optically resonant dielectric nanostructures. *Science* **354**, aag2472 (2016).
- [47] Schmidt, C. *et al.* Multi-order investigation of the nonlinear susceptibility tensors of individual nanoparticles. *Scientific reports* **6**, 1–9 (2016).
- [48] Celebrano, M. *et al.* Mode matching in multiresonant plasmonic nanoantennas for enhanced second harmonic generation. *Nature nanotechnology* **10**, 412 (2015).
- [49] Gennaro, S. D. *et al.* The interplay of symmetry and scattering phase in second harmonic generation from gold nanoantennas. *Nano Lett* **16**, 5278–5285 (2016).
- [50] Hong, C. & Mandel, L. Experimental realization of a localized one-photon state. *Physical Review Letters* **56**, 58 (1986).
- [51] Kwiat, P. G. *et al.* New high-intensity source of polarization-entangled photon pairs. *Physical Review Letters* **75**, 4337 (1995).
- [52] Marino, G. *et al.* Spontaneous photon-pair generation from a dielectric nanoantenna. *Optica* **6**, 1416–1422 (2019).
- [53] Solntsev, A. S., Agarwal, G. S. & Kivshar, Y. Y. Metasurfaces for quantum photonics. *Nature Photonics* **15**, 327–336 (2021).
- [54] Ming, Y. *et al.* Photonic entanglement based on nonlinear metamaterials. *Laser & Photonics Reviews* **14**, 1900146 (2020).
- [55] Lemos, G. B. *et al.* Quantum imaging with undetected photons. *Nature* **512**, 409–412 (2014).
- [56] Liscidini, M. & Sipe, J. Stimulated emission tomography. *Physical review letters* **111**, 193602 (2013).

- [57] Nielsen, M. P. *Semiconductor nanoplasmonics*. Ph.D. thesis, Imperial College London (2017).
- [58] Sheik-Bahae, M., Said, A. A., Wei, T.-H., Hagan, D. J. & Van Stryland, E. W. Sensitive measurement of optical nonlinearities using a single beam. *IEEE journal of quantum electronics* **26**, 760–769 (1990).
- [59] Gu, B. *et al.* Characterization of saturable absorbers using an open-aperture Gaussian-beam Z-scan. *Physical Review A* **73**, 065803 (2006).
- [60] Miller, R. C. Optical second harmonic generation in piezoelectric crystals. *Applied Physics Letters* **5**, 17–19 (1964).
- [61] O'Brien, K. *et al.* Predicting nonlinear properties of metamaterials from the linear response. *Nature materials* **14**, 379–383 (2015).
- [62] Butet, J. & Martin, O. J. Evaluation of the nonlinear response of plasmonic metasurfaces: Miller's rule, nonlinear effective susceptibility method, and full-wave computation. *JOSA B* **33**, A8–A15 (2016).
- [63] Giordmaine, J. A. & Miller, R. C. Tunable coherent parametric oscillation in lithium niobate at optical frequencies. *Physical Review Letters* **14**, 973 (1965).
- [64] Kaneda, F. *et al.* Time-multiplexed heralded single-photon source. *Optica* **2**, 1010–1013 (2015).
- [65] Lounis, B. & Orrit, M. Single-photon sources. *Reports on Progress in Physics* **68**, 1129 (2005).
- [66] Diedrich, F. & Walther, H. Nonclassical radiation of a single stored ion. *Physical review letters* **58**, 203 (1987).
- [67] Sterk, J., Luo, L., Manning, T., Maunz, P. & Monroe, C. Photon collection from a trapped ion-cavity system. *Physical Review A* **85**, 062308 (2012).

- [68] Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance absorption by nuclear magnetic moments in a solid. *Physical review* **69**, 37 (1946).
- [69] Higginbottom, D. B. *et al.* Pure single photons from a trapped atom source. *New Journal of Physics* **18**, 093038 (2016).
- [70] Maunz, P. *et al.* Quantum interference of photon pairs from two remote trapped atomic ions. *Nature Physics* **3**, 538–541 (2007).
- [71] Michler, P. *et al.* A quantum dot single-photon turnstile device. *science* **290**, 2282–2285 (2000).
- [72] Santori, C., Pelton, M., Solomon, G., Dale, Y. & Yamamoto, Y. Triggered single photons from a quantum dot. *Physical Review Letters* **86**, 1502 (2001).
- [73] Santori, C., Fattal, D., Vučković, J., Solomon, G. S. & Yamamoto, Y. Indistinguishable photons from a single-photon device. *Nature* **419**, 594 (2002).
- [74] Gazzano, O. *et al.* Bright solid-state sources of indistinguishable single photons. *Nature communications* **4**, 1–6 (2013).
- [75] Burnham, D. C. & Weinberg, D. L. Observation of simultaneity in parametric production of optical photon pairs. *Physical Review Letters* **25**, 84 (1970).
- [76] Fasel, S. *et al.* High-quality asynchronous heralded single-photon source at telecom wavelength. *New Journal of Physics* **6**, 163 (2004).
- [77] Waks, E., Diamanti, E. & Yamamoto, Y. Generation of photon number states. *New Journal of Physics* **8**, 4 (2006).
- [78] Harris, S., Oshman, M. & Byer, R. Observation of tunable optical parametric fluorescence. *Physical Review Letters* **18**, 732 (1967).
- [79] Guo, X. *et al.* Parametric down-conversion photon-pair source on a nanophotonic chip. *Light: Science & Applications* **6**, e16249 (2017).

- [80] Gerry, C. & Knight, P. *Introductory quantum optics* (Cambridge university press, 2005).
- [81] Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- [82] Takesue, H. & Inoue, K. Generation of polarization-entangled photon pairs and violation of Bell's inequality using spontaneous four-wave mixing in a fiber loop. *Physical Review A* **70**, 031802 (2004).
- [83] Mosley, P. J. *et al.* Heralded generation of ultrafast single photons in pure quantum states. *Physical Review Letters* **100**, 133601 (2008).
- [84] Eckstein, A. *et al.* High-resolution spectral characterization of two photon states via classical measurements. *Laser & Photonics Reviews* **8** (2014).
- [85] Fang, B., Cohen, O., Liscidini, M., Sipe, J. E. & Lorenz, V. O. Fast and highly resolved capture of the joint spectral density of photon pairs. *Optica* **1**, 281–284 (2014).
- [86] Parker, S., Bose, S. & Plenio, M. Entanglement quantification and purification in continuous-variable systems. *Physical Review A* **61**, 032305 (2000).
- [87] Law, C., Walmsley, I. A. & Eberly, J. Continuous frequency entanglement: effective finite hilbert space and entropy control. *Physical Review Letters* **84**, 5304 (2000).
- [88] Grice, W. P., U'Ren, A. B. & Walmsley, I. A. Eliminating frequency and space-time correlations in multiphoton states. *Physical Review A* **64**, 063815 (2001).
- [89] Johnson, P. B. & Christy, R.-W. Optical constants of the noble metals. *Physical review B* **6**, 4370 (1972).
- [90] Weeber, J.-C. *et al.* Near-field observation of surface plasmon polariton propagation on thin metal stripes. *Physical review B* **64**, 045411 (2001).

- [91] Bozhevolnyi, S. I., Erland, J., Leosson, K., Skovgaard, P. M. & Hvam, J. M. Waveguiding in surface plasmon polariton band gap structures. *Physical review letters* **86**, 3008 (2001).
- [92] Ditlbacher, H. *et al.* Fluorescence imaging of surface plasmon fields. *Applied physics letters* **80**, 404–406 (2002).
- [93] Nezhad, M. P., Tetz, K. & Fainman, Y. Gain assisted propagation of surface plasmon polaritons on planar metallic waveguides. *Optics Express* **12**, 4072–4079 (2004).
- [94] Mendlowitz, H. Optical constants of aluminium. *Proceedings of the Physical Society (1958-1967)* **75**, 664 (1960).
- [95] Fang, N., Lee, H., Sun, C. & Zhang, X. Sub-diffraction-limited optical imaging with a silver superlens. *Science* **308**, 534–537 (2005).
- [96] Griffiths, D. J. *Introduction to Electrodynamics, fourth edition* (Cambridge University Press, 2017).
- [97] Kelly, K. L., Coronado, E., Zhao, L. L. & Schatz, G. C. The optical properties of metal nanoparticles: the influence of size, shape, and dielectric environment (2003).
- [98] Cole, J. R. & Halas, N. Optimized plasmonic nanoparticle distributions for solar spectrum harvesting. *Applied physics letters* **89**, 153120 (2006).
- [99] Khlebtsov, N. G. & Dykman, L. A. Optical properties and biomedical applications of plasmonic nanoparticles. *Journal of Quantitative Spectroscopy and Radiative Transfer* **111**, 1–35 (2010).
- [100] Mock, J. J., Smith, D. R. & Schultz, S. Local refractive index dependence of plasmon resonance spectra from individual nanoparticles. *Nano letters* **3**, 485–491 (2003).
- [101] Bohren, C. F. & Huffman, D. R. *Absorption and scattering of light by small particles* (John Wiley & Sons, 2008).

- [102] Luk'yanchuk, B. *et al.* The Fano resonance in plasmonic nanostructures and metamaterials. *Nature materials* **9**, 707–715 (2010).
- [103] Maier, S. A. The benefits of darkness. *Nature materials* **8**, 699–700 (2009).
- [104] Zhang, S., Genov, D. A., Wang, Y., Liu, M. & Zhang, X. Plasmon-induced transparency in metamaterials. *Physical review letters* **101**, 047401 (2008).
- [105] Spence, D. E., Kean, P. N. & Sibbett, W. 60-fsec pulse generation from a self-mode-locked Ti:sapphire laser. *Optics letters* **16**, 42–44 (1991).
- [106] Fisher, R. A., Kelley, P. & Gustafson, T. Subpicosecond pulse generation using the optical Kerr effect. *Applied Physics Letters* **14**, 140–143 (1969).
- [107] Wang, L., Hong, C. & Friberg, S. Generation of correlated photons via four-wave mixing in optical fibres. *Journal of optics B: Quantum and semiclassical optics* **3**, 346 (2001).
- [108] Xia, T., Hagan, D., Sheik-Bahae, M. & Van Stryland, E. Eclipsing Z-scan measurement of $\lambda/10^4$ wave-front distortion. *Optics letters* **19**, 317–319 (1994).
- [109] Ma, H., Gomes, A. & de Araújo, C. B. Measurements of nondegenerate optical nonlinearity using a two-color single beam method. *Applied physics letters* **59**, 2666–2668 (1991).
- [110] Sheik-Bahae, M., Wang, J., DeSalvo, R., Hagan, D. & Van Stryland, E. Measurement of nondegenerate nonlinearities using a two-color Z-scan. *Optics letters* **17**, 258–260 (1992).
- [111] Wang, J., Sheik-Bahae, M., Said, A., Hagan, D. J. & Van Stryland, E. W. Time-resolved Z-scan measurements of optical nonlinearities. *JOSA B* **11**, 1009–1017 (1994).
- [112] Oliveira, L. & Zilio, S. C. Single-beam time-resolved Z-scan measurements of slow absorbers. *Applied physics letters* **65**, 2121–2123 (1994).

- [113] Petrov, D., Gomes, A. & De Araújo, C. B. Reflection Z-scan technique for measurements of optical properties of surfaces. *Applied physics letters* **65**, 1067–1069 (1994).
- [114] Hermann, J. & McDuff, R. Analysis of spatial scanning with thick optically nonlinear media. *JOSA B* **10**, 2056–2064 (1993).
- [115] Chapple, P., Staromlynska, J. & McDuff, R. Z-scan studies in the thin-and the thick-sample limits. *JOSA B* **11**, 975–982 (1994).
- [116] Ismail, N., Kores, C. C., Geskus, D. & Pollnau, M. Fabry-Pérot resonator: spectral line shapes, generic and related Airy distributions, linewidths, finesse, and performance at low or frequency-dependent reflectivity. *Optics express* **24**, 16366–16389 (2016).
- [117] Weaire, D., Wherrett, B., Miller, D. & Smith, S. Effect of low-power nonlinear refraction on laser-beam propagation in InSb. *Optics Letters* **4**, 331–333 (1979).
- [118] Siegman, A. E. *Lasers* (University Science Books, 1986).
- [119] Samoc, M., Samoc, A., Luther-Davies, B., Humphrey, M. G. & Wong, M.-S. Third-order optical nonlinearities of oligomers, dendrimers and polymers derived from solution Z-scan studies. *Optical materials* **21**, 485–488 (2003).
- [120] Gu, B., Ji, W. & Huang, X.-Q. Analytical expression for femtosecond-pulsed Z-scans on instantaneous nonlinearity. *Applied optics* **47**, 1187–1192 (2008).
- [121] Van Stryland, E. W. & Sheik-Bahae, M. Z-scan. In *Characterization techniques and tabulations for organic nonlinear optical materials*, 671–708 (Routledge, 2018).
- [122] Wang, J., Gu, B., Wang, H.-T. & Ni, X.-W. Z-scan analytical theory for material with saturable absorption and two-photon absorption. *Optics communications* **283**, 3525–3528 (2010).

- [123] Gu, C. *et al.* Giant and multistage nonlinear optical response in porphyrin-based surface-supported metal–organic framework nanofilms. *Nano letters* **19**, 9095–9101 (2019).
- [124] Gurudas, U. *et al.* Saturable and reverse saturable absorption in silver nanodots at 532 nm using picosecond laser pulses. *Journal of applied physics* **104**, 073107 (2008).
- [125] Alam, M. Z., De Leon, I. & Boyd, R. W. Large optical nonlinearity of indium tin oxide in its epsilon-near-zero region. *Science* **352**, 795–797 (2016).
- [126] Hercher, M. An analysis of saturable absorbers. *Applied optics* **6**, 947–954 (1967).
- [127] Lee, Y. H., Yan, Y., Polavarapu, L. & Xu, Q.-H. Nonlinear optical switching behavior of au nanocubes and nano-octahedra investigated by femtosecond Z-scan measurements. *Applied Physics Letters* **95**, 023105 (2009).
- [128] Gu, B. *et al.* Z-scan theory of two-photon absorption saturation and experimental evidence. *Journal of Applied Physics* **102**, 083101 (2007).
- [129] Wang, K., Long, H., Fu, M., Yang, G. & Lu, P. Intensity-dependent reversal of nonlinearity sign in a gold nanoparticle array. *Optics letters* **35**, 1560–1562 (2010).
- [130] Kojima, A., Teshima, K., Shirai, Y. & Miyasaka, T. Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *Journal of the American Chemical Society* **131**, 6050–6051 (2009).
- [131] Liu, M., Johnston, M. B. & Snaith, H. J. Efficient planar heterojunction perovskite solar cells by vapour deposition. *Nature* **501**, 395–398 (2013).
- [132] Lin, Q., Armin, A., Nagiri, R. C. R., Burn, P. L. & Meredith, P. Electro-optics of perovskite solar cells. *Nature Photonics* **9**, 106 (2015).
- [133] Park, N.-G. Perovskite solar cells: an emerging photovoltaic technology. *Materials today* **18**, 65–72 (2015).

- [134] Abdelwahab, I. *et al.* Highly enhanced third-harmonic generation in 2D perovskites at excitonic resonances. *ACS nano* **12**, 644–650 (2018).
- [135] Abdelwahab, I. *et al.* Giant and tunable optical nonlinearity in single-crystalline 2d perovskites due to excitonic and plasma effects. *Advanced Materials* **31**, 1902685 (2019).
- [136] Li, W. *et al.* Chemically diverse and multifunctional hybrid organic–inorganic perovskites. *Nature Reviews Materials* **2**, 16099 (2017).
- [137] Smith, I. C., Hoke, E. T., Solis-Ibarra, D., McGehee, M. D. & Karunadasa, H. I. A layered hybrid perovskite solar-cell absorber with enhanced moisture stability. *Angewandte Chemie International Edition* **53**, 11232–11235 (2014).
- [138] Cao, D. H., Stoumpos, C. C., Farha, O. K., Hupp, J. T. & Kanatzidis, M. G. 2D homologous perovskites as light-absorbing materials for solar cell applications. *Journal of the American Chemical Society* **137**, 7843–7850 (2015).
- [139] Etgar, L. The merit of perovskite's dimensionality; can this replace the 3D halide perovskite? *Energy & Environmental Science* **11**, 234–242 (2018).
- [140] Soe, C. M. M. *et al.* New type of 2D perovskites with alternating cations in the inter-layer space, $(\text{C}(\text{NH}_2)_3)(\text{CH}_3\text{NH}_3)_n \text{Pb}_n \text{I}_{3n+1}$: Structure, properties, and photovoltaic performance. *Journal of the American Chemical Society* **139**, 16297–16309 (2017).
- [141] Chen, Y. *et al.* 2D ruddlesden–popper perovskites for optoelectronics. *Advanced Materials* **30**, 1703487 (2018).
- [142] Stoumpos, C. C., Mao, L., Malliakas, C. D. & Kanatzidis, M. G. Structure–band gap relationships in hexagonal polytypes and low-dimensional structures of hybrid tin iodide perovskites. *Inorganic chemistry* **56**, 56–73 (2017).

- [143] Kieslich, G., Sun, S. & Cheetham, A. K. Solid-state principles applied to organic–inorganic perovskites: new tricks for an old dog. *Chemical Science* **5**, 4712–4715 (2014).
- [144] Saidaminov, M. I., Mohammed, O. F. & Bakr, O. M. Low-dimensional-networked metal halide perovskites: the next big thing. *ACS Energy Letters* **2**, 889–896 (2017).
- [145] Stoumpos, C. C. *et al.* Ruddlesden–popper hybrid lead iodide perovskite 2D homologous semiconductors. *Chemistry of Materials* **28**, 2852–2867 (2016).
- [146] Ishihara, T., Takahashi, J. & Goto, T. Exciton state in two-dimensional perovskite semiconductor $(\text{C}_{10}\text{H}_{21}\text{NH}_3)_2\text{PbI}_4$. *Solid state communications* **69**, 933–936 (1989).
- [147] Ishihara, T., Takahashi, J. & Goto, T. Optical properties due to electronic transitions in two-dimensional semiconductors $(\text{C}_n\text{H}_{2n+1}\text{NH}_3)_2\text{PbI}_4$. *Physical review B* **42**, 11099 (1990).
- [148] Chemla, D., Miller, D., Smith, P., Gossard, A. & Wiegmann, W. Room temperature excitonic nonlinear absorption and refraction in GaAs/AlGaAs multiple quantum well structures. *IEEE Journal of Quantum Electronics* **20**, 265–275 (1984).
- [149] Keldysh, L. Coulomb interaction in thin semiconductor and semimetal films. *Soviet Journal of Experimental and Theoretical Physics Letters* **29**, 658 (1979).
- [150] Hanamura, E., Nagaosa, N., Kumagai, M. & Takagahara, T. Quantum wells with enhanced exciton effects and optical non-linearity. *Materials Science and Engineering: B* **1**, 255–258 (1988).
- [151] Blancon, J.-C. *et al.* Scaling law for excitons in 2D perovskite quantum wells. *Nature communications* **9**, 1–10 (2018).
- [152] Umebayashi, T., Asai, K., Kondo, T. & Nakao, A. Electronic structures of lead iodide based low-dimensional crystals. *Physical Review B* **67**, 155405 (2003).

- [153] Saparov, B. & Mitzi, D. B. Organic–inorganic perovskites: structural versatility for functional materials design. *Chemical reviews* **116**, 4558–4596 (2016).
- [154] Papavassiliou, G. C. & Koutselas, I. Structural, optical and related properties of some natural three- and lower-dimensional semiconductor systems. *Synthetic Metals* **71**, 1713–1714 (1995).
- [155] Bristow, A. D., Rotenberg, N. & Van Driel, H. M. Two-photon absorption and Kerr coefficients of silicon for 850–2200 nm. *Applied Physics Letters* **90**, 191104 (2007).
- [156] Kalanoor, B. S. *et al.* Third-order optical nonlinearities in organometallic methylammonium lead iodide perovskite thin films. *Acs Photonics* **3**, 361–370 (2016).
- [157] Liang, W.-Y. *et al.* High optical nonlinearity in low-dimensional halide perovskite polycrystalline films. *Optics Express* **28**, 24919–24927 (2020).
- [158] Lenz, G. *et al.* Large Kerr effect in bulk Se-based chalcogenide glasses. *Optics Letters* **25**, 254–256 (2000).
- [159] Liu, Y. & Tsang, H. Time dependent density of free carriers generated by two photon absorption in silicon waveguides. *Applied physics letters* **90**, 211105 (2007).
- [160] Liang, T.-K. & Tsang, H. K. Role of free carriers from two-photon absorption in Raman amplification in silicon-on-insulator waveguides. *Applied physics letters* **84**, 2745–2747 (2004).
- [161] Dakhlaoui, H. Linear and nonlinear optical absorption coefficients and refractive index changes in GaN/Al_xGa_(1-x)N double quantum wells operating at 1.55 μm . *Journal of Applied Physics* **117**, 135705 (2015).
- [162] Lazaridis, P., Debarge, G. & Gallion, P. Time–bandwidth product of chirped sech^2 pulses: application to phase–amplitude-coupling factor measurement. *Optics letters* **20**, 1160–1162 (1995).

- [163] Blancon, J.-C. *et al.* Extremely efficient internal exciton dissociation through edge states in layered 2D perovskites. *Science* **355**, 1288–1292 (2017).
- [164] Fujiwara, K. *et al.* Excitation dynamics in layered lead halide perovskite crystal slabs and microcavities. *ACS Photonics* **7**, 845–852 (2020).
- [165] Maiman, T. H. Stimulated optical radiation in ruby. *nature* **187**, 493–494 (1960).
- [166] Franken, P., Hill, A. E., Peters, C. e. & Weinreich, G. Generation of optical harmonics. *Physical Review Letters* **7**, 118 (1961).
- [167] Giordmaine, J. Mixing of light beams in crystals. *Physical Review Letters* **8**, 19 (1962).
- [168] Maker, P., Terhune, R., Nisenoff, M. & Savage, C. Effects of dispersion and focusing on the production of optical harmonics. *Physical review letters* **8**, 21 (1962).
- [169] Boyd, G., Miller, R. C., Nassau, K., Bond, W. & Savage, A. LiNbO₃: an efficient phase-matchable nonlinear optical material. *Applied physics letters* **5**, 234–236 (1964).
- [170] Stegeman, G. I. & Stolen, R. H. Waveguides and fibers for nonlinear optics. *JOSA B* **6**, 652–662 (1989).
- [171] Stolen, R. H. The early years of fiber nonlinear optics. *Journal of Lightwave Technology* **26**, 1021–1031 (2008).
- [172] Stolen, R., Bösch, M. & Lin, C. Phase matching in birefringent fibers. *Optics letters* **6**, 213–215 (1981).
- [173] Hill, K., Johnson, D. & Kawasaki, B. Efficient conversion of light over a wide spectral range by four-photon mixing in a multimode graded-index fiber. *Applied optics* **20**, 1075–1079 (1981).
- [174] Ashkin, A., Boyd, G. & Dziedzic, J. Resonant optical second harmonic generation and mixing. *IEEE Journal of quantum electronics* **2**, 109–124 (1966).

- [175] Moss, T. A relationship between the refractive index and the infra-red threshold of sensitivity for photoconductors. *Proceedings of the Physical Society. Section B* **63**, 167 (1950).
- [176] Moss, T. Relations between the refractive index and energy gap of semiconductors. *physica status solidi (b)* **131**, 415–427 (1985).
- [177] Xiong, C. *et al.* Integrated GaN photonic circuits on silicon (100) for second harmonic generation. *Optics express* **19**, 10462–10470 (2011).
- [178] Levinshtein, M. E., Rumyantsev, S. L. & Shur, M. S. *Properties of Advanced Semiconductor Materials: GaN, AlN, InN, BN, SiC, SiGe* (John Wiley & Sons, 2001).
- [179] Chowdhury, A., Ng, H. M., Bhardwaj, M. & Weimann, N. G. Second-harmonic generation in periodically poled gan. *Applied physics letters* **83**, 1077–1079 (2003).
- [180] Zhang, H. *et al.* Study of nonlinear optical effects in GaN:Mg epitaxial film. *Applied physics letters* **69**, 2953–2955 (1996).
- [181] Shibata, H. *et al.* High thermal conductivity of gallium nitride (GaN) crystals grown by HVPE process. *Materials Transactions* **48**, 2782–2786 (2007).
- [182] Galsin, J. S. Crystal structure of solids, Chapter 1. In *Solid State Physics*, 1–36 (Academic Press, 2019). URL <https://www.sciencedirect.com/science/article/pii/B9780128171035000013>.
- [183] Miragliotta, J., Wickenden, D., Kistenmacher, T. & Bryden, W. Linear-and nonlinear-optical properties of GaN thin films. *JOSA B* **10**, 1447–1456 (1993).
- [184] Perot, A. & Fabry, C. On the application of interference phenomena to the solution of various problems of spectroscopy and metrology. *The Astrophysical Journal* **9**, 87–115 (1899).

- [185] Santiago-Cruz, T., Sultanov, V., Zhang, H., Krivitsky, L. A. & Chekhova, M. V. Entangled photons from subwavelength nonlinear films. *Optics Letters* **46**, 653–656 (2021).
- [186] Liu, S. *et al.* Resonantly enhanced second-harmonic generation using III–V semiconductor all-dielectric metasurfaces. *Nano letters* **16**, 5426–5432 (2016).
- [187] Sautter, J. D. *et al.* Tailoring second-harmonic emission from (111)-GaAs nanoantennas. *Nano Letters* **19**, 3905–3911 (2019).
- [188] Kim, S. *et al.* High-harmonic generation by resonant plasmon field enhancement. *Nature* **453**, 757–760 (2008).
- [189] Schuller, J. A. *et al.* Plasmonics for extreme light concentration and manipulation. *Nature materials* **9**, 193–204 (2010).
- [190] Drude, P. Zur Elektronentheorie der Metalle. *Annalen der Physik* **306**, 566–613 (1900).
- [191] Naik, G. V., Shalaev, V. M. & Boltasseva, A. Alternative plasmonic materials: beyond gold and silver. *Advanced Materials* **25**, 3264–3294 (2013).
- [192] Smirnova, D. & Kivshar, Y. S. Multipolar nonlinear nanophotonics. *Optica* **3**, 1241–1255 (2016).
- [193] Kuznetsov, A. I., Miroshnichenko, A. E., Fu, Y. H., Zhang, J. & Luk’Yanchuk, B. Magnetic light. *Scientific reports* **2**, 492 (2012).
- [194] Ginn, J. C. *et al.* Realizing optical magnetism from dielectric metamaterials. *Physical review letters* **108**, 097402 (2012).
- [195] Shcherbakov, M. R. *et al.* Enhanced third-harmonic generation in silicon nanoparticles driven by magnetic response. *Nano letters* **14**, 6488–6492 (2014).

- [196] Gili, V. *et al.* Monolithic AlGaAs second-harmonic nanoantennas. *Optics Express* **24**, 15965–15971 (2016).
- [197] Grinblat, G., Li, Y., Nielsen, M. P., Oulton, R. F. & Maier, S. A. Efficient third harmonic generation and nonlinear subwavelength imaging at a higher-order anapole mode in a single germanium nanodisk. *ACS nano* **11**, 953–960 (2017).
- [198] Kang, M. S., Lee, C.-H., Park, J. B., Yoo, H. & Yi, G.-C. Gallium nitride nanostructures for light-emitting diode applications. *Nano energy* **1**, 391–400 (2012).
- [199] Kente, T. & Mhlanga, S. D. Gallium nitride nanostructures: Synthesis, characterization and applications. *Journal of Crystal Growth* **444**, 55–72 (2016).
- [200] Melton, W. A. & Pankove, J. I. GaN growth on sapphire. *Journal of crystal growth* **178**, 168–173 (1997).
- [201] Wong, W., Sands, T. & Cheung, N. Damage-free separation of GaN thin films from sapphire substrates. *Applied Physics Letters* **72**, 599–601 (1998).
- [202] Cho, H. K. *et al.* Laser liftoff GaN thin-film photonic crystal GaN-based light-emitting diodes. *IEEE Photonics Technology Letters* **20**, 2096–2098 (2008).
- [203] Boyd, G., Coldren, L. & Storz, F. Directional reactive ion etching at oblique angles. *Applied Physics Letters* **36**, 583–585 (1980).
- [204] Lee, J.-K. *et al.* Oblique-directional plasma etching of Si using a Faraday cage. *Journal of The Electrochemical Society* **156**, D222–D225 (2009).
- [205] Burek, M. J. *et al.* Free-standing mechanical and photonic nanostructures in single-crystal diamond. *Nano letters* **12**, 6084–6089 (2012).
- [206] Latawiec, P., Burek, M. J., Sohn, Y.-I. & Lončar, M. Faraday cage angled-etching of nanostructures in bulk dielectrics. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena* **34**, 041801 (2016).

- [207] Armitage, R. & Hirayama, H. M-plane GaN grown on m-sapphire by metalorganic vapor phase epitaxy. *Applied Physics Letters* **92**, 092121 (2008).
- [208] Marino, G. *et al.* Zero-order second harmonic generation from AlGaAs-on-insulator metasurfaces. *ACS Photonics* **6**, 1226–1231 (2019).
- [209] Bernal Arango, F., Kwadrin, A. & Koenderink, A. F. Plasmonic antennas hybridized with dielectric waveguides. *ACS nano* **6**, 10156–10167 (2012).
- [210] Wang, C. *et al.* Metasurface-assisted phase-matching-free second harmonic generation in lithium niobate waveguides. *Nature communications* **8**, 1–7 (2017).
- [211] Lippitz, M., van Dijk, M. A. & Orrit, M. Third-harmonic generation from single gold nanoparticles. *Nano letters* **5**, 799–802 (2005).
- [212] Danckwerts, M. & Novotny, L. Optical frequency mixing at coupled gold nanoparticles. *Physical Review Letters* **98**, 026104 (2007).
- [213] Hasan, S. B., Lederer, F. & Rockstuhl, C. Nonlinear plasmonic antennas. *Materials Today* **17**, 478–485 (2014).
- [214] Lee, J. *et al.* Giant nonlinear response from plasmonic metasurfaces coupled to inter-subband transitions. *Nature* **511**, 65–69 (2014).
- [215] Clark, H. A., Campagnola, P. J., Wuskell, J. P., Lewis, A. & Loew, L. M. Second harmonic generation properties of fluorescent polymer-encapsulated gold nanoparticles. *Journal of the American Chemical Society* **122**, 10234–10235 (2000).
- [216] Harutyunyan, H., Volpe, G., Quidant, R. & Novotny, L. Enhancing the nonlinear optical response using multifrequency gold-nanowire antennas. *Physical review letters* **108**, 217403 (2012).
- [217] Aouani, H., Rahmani, M., Navarro-Cía, M. & Maier, S. A. Third-harmonic-upconversion enhancement from a single semiconductor nanoparticle coupled to a plasmonic antenna. *Nature nanotechnology* **9**, 290 (2014).

- [218] Metzger, B. *et al.* Doubling the efficiency of third harmonic generation by positioning ito nanocrystals into the hot-spot of plasmonic gap-antennas. *Nano letters* **14**, 2867–2872 (2014).
- [219] Shibanuma, T., Grinblat, G., Albella, P. & Maier, S. A. Efficient third harmonic generation from metal–dielectric hybrid nanoantennas. *Nano letters* **17**, 2647–2651 (2017).
- [220] McMahon, M., Lopez, R., Haglund Jr, R., Ray, E. & Bunton, P. Second-harmonic generation from arrays of symmetric gold nanoparticles. *Physical Review B* **73**, 041401 (2006).
- [221] Kujala, S., Canfield, B. K., Kauranen, M., Svirko, Y. & Turunen, J. Multipole interference in the second-harmonic optical radiation from gold nanoparticles. *Physical review letters* **98**, 167403 (2007).
- [222] West, P. R. *et al.* Searching for better plasmonic materials. *Laser & Photonics Reviews* **4**, 795–808 (2010).
- [223] Boyd, R. W., Shi, Z. & De Leon, I. The third-order nonlinear optical susceptibility of gold. *Optics Communications* **326**, 74–79 (2014).
- [224] Hache, F., Ricard, D., Flytzanis, C. & Kreibig, U. The optical Kerr effect in small metal particles and metal colloids: the case of gold. *Applied Physics A* **47**, 347–357 (1988).
- [225] Butet, J., Brevet, P.-F. & Martin, O. J. Optical second harmonic generation in plasmonic nanostructures: from fundamental principles to advanced applications. *ACS nano* **9**, 10545–10562 (2015).
- [226] Rudnick, J. & Stern, E. Second-harmonic radiation from metal surfaces. *Physical review B* **4**, 4274 (1971).

- [227] Sipe, J., So, V., Fukui, M. & Stegeman, G. Analysis of second-harmonic generation at metal surfaces. *Physical Review B* **21**, 4389 (1980).
- [228] Wang, F. X. *et al.* Surface and bulk contributions to the second-order nonlinear optical response of a gold film. *Physical Review B* **80**, 233402 (2009).
- [229] Bloembergen, N., Chang, R. K., Jha, S. & Lee, C. Optical second-harmonic generation in reflection from media with inversion symmetry. *Physical Review* **174**, 813 (1968).
- [230] Sipe, J., Mizrahi, V. & Stegeman, G. I. Fundamental difficulty in the use of second-harmonic generation as a strictly surface probe. *Physical Review B* **35**, 9091 (1987).
- [231] Bachelier, G. *et al.* Origin of optical second-harmonic generation in spherical gold nanoparticles: Local surface and nonlocal bulk contributions. *Physical Review B* **82**, 235403 (2010).
- [232] Zeng, Y., Hoyer, W., Liu, J., Koch, S. W. & Moloney, J. V. Classical theory for second-harmonic generation from metallic nanoparticles. *Physical Review B* **79**, 235109 (2009).
- [233] Haug, T., Klemm, P., Bange, S. & Lupton, J. M. Hot-electron intraband luminescence from single hot spots in noble-metal nanoparticle films. *Physical review letters* **115**, 067403 (2015).
- [234] Roloff, L. *et al.* Light emission from gold nanoparticles under ultrafast near-infrared excitation: Thermal radiation, inelastic light scattering, or multiphoton luminescence? *Nano letters* **17**, 7914–7919 (2017).
- [235] Schuck, P., Fromm, D., Sundaramurthy, A., Kino, G. & Moerner, W. Improving the mismatch between light and nanoscale objects with gold bowtie nanoantennas. *Physical review letters* **94**, 017402 (2005).

- [236] Hugall, J. T. & Baumberg, J. J. Demonstrating photoluminescence from Au is electronic inelastic light scattering of a plasmonic metal: the origin of SERS backgrounds. *Nano letters* **15**, 2600–2604 (2015).
- [237] Gennaro, S. D., Li, Y., Maier, S. A. & Oulton, R. F. Double blind ultrafast pulse characterization by mixed frequency generation in a gold antenna. *ACS Photonics* **5**, 3166–3171 (2018).
- [238] Engheta, N., Papas, C. & Elachi, C. Radiation patterns of interfacial dipole antennas. *Radio Science* **17**, 1557–1566 (1982).
- [239] Lindfors, K. *et al.* Imaging and steering unidirectional emission from nanoantenna array metasurfaces. *ACS Photonics* **3**, 286–292 (2016).
- [240] Baselli, M. *et al.* Plasmon-enhanced second harmonic generation: from individual antennas to extended arrays. *Plasmonics* **12**, 1595–1600 (2017).
- [241] Kravets, V., Schedin, F. & Grigorenko, A. Extremely narrow plasmon resonances based on diffraction coupling of localized plasmons in arrays of metallic nanoparticles. *Physical review letters* **101**, 087403 (2008).
- [242] Rodriguez, S., Schaafsma, M., Berrier, A. & Rivas, J. G. Collective resonances in plasmonic crystals: Size matters. *Physica B: Condensed Matter* **407**, 4081–4085 (2012).
- [243] Liu, S. *et al.* An all-dielectric metasurface as a broadband optical frequency mixer. *Nature communications* **9**, 1–6 (2018).
- [244] Martin, S., Bradley, D., Lane, P., Mellor, H. & Burn, P. Linear and nonlinear optical properties of the conjugated polymers PPV and MEH-PPV. *Physical Review B* **59**, 15133 (1999).

- [245] Nielsen, M. P., Shi, X., Dichtl, P., Maier, S. A. & Oulton, R. F. Giant nonlinear response at a plasmonic nanofocus drives efficient four-wave mixing. *Science* **358**, 1179–1181 (2017).
- [246] Albrecht, G., Hentschel, M., Kaiser, S. & Giessen, H. Hybrid organic-plasmonic nanoantennas with enhanced third-harmonic generation. *ACS omega* **2**, 2577–2582 (2017).
- [247] Grinblat, G. *et al.* Ultrafast all-optical modulation in 2d hybrid perovskites. *ACS nano* **13**, 9504–9510 (2019).
- [248] Chai, Z. *et al.* Ultrafast all-optical switching. *Advanced Optical Materials* **5**, 1600665 (2017).
- [249] Verre, R. *et al.* Directional light extinction and emission in a metasurface of tilted plasmonic nanopillars. *Nano letters* **16**, 98–104 (2016).
- [250] Gennaro, S. D., Li, Y., Maier, S. A. & Oulton, R. F. Nonlinear pancharatnam- berry phase metasurfaces beyond the dipole approximation. *ACS Photonics* **6**, 2335–2341 (2019).
- [251] Barker Jr, A. & Illegems, M. Infrared lattice vibrations and free-electron dispersion in GaN. *Physical Review B* **7**, 743 (1973).
- [252] Malitson, I. H. & Dodge, M. J. Refractive-index and birefringence of synthetic sapphire. In *Journal Of The Optical Society Of America*, vol. 62, 1405–1405. Optical Society of America (American Institute of Physics Inc., 1972).
- [253] Mie, G. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Annalen der Physik* **330**, 377–445 (1908).

List of Publications and Presentations

Publications Discussed in this Thesis

Abdelwahab, I.*, **Dichtl, P.***, Grinblat, G., Leng, K., Chi, X., Park, I.H., Nielsen, M.P., Oulton, R.F., Loh, K.P. and Maier, S.A., 2019. Giant and Tunable Optical Nonlinearity in Single-Crystalline 2D Perovskites due to Excitonic and Plasma Effects. *Advanced Materials*, 31(29), p.1902685.

* co-first authors

Other Publications

Nielsen, M.P., Shi, X., **Dichtl, P.**, Maier, S.A. and Oulton, R.F., 2017. Giant nonlinear response at a plasmonic nanofocus drives efficient four-wave mixing. *Science*, 358(6367), pp.1179-1181.

Grinblat, G., Nielsen, M. P., **Dichtl, P.**, Li, Y., Oulton, R. F., and Maier, S. A., 2019. Ultrafast sub-30-fs all-optical switching based on gallium phosphide. *Science advances*, 5(6), eaaw3262.

Grinblat, G., Abdelwahab, I., Nielsen, M.P., **Dichtl, P.**, Leng, K., Oulton, R.F., Loh, K.P. and Maier, S.A., 2019. Ultrafast all-optical modulation in 2D hybrid perovskites. *ACS nano*, 13(8), pp.9504-9510.

Presentations

Dichtl, P., Abdelwahab, I., Grinblat, G., Leng, K., Chi, X., Park, I.H., Nielsen, M.P., Oulton, R.F., Loh, K.P. and Maier, S.A., 2019, September. Giant and Tunable Optical Nonlinearity in Single-Crystalline 2D Perovskites due to Excitonic and Plasma Effects. In *Frontiers in Optics* (pp. FTu5F-3). Optical Society of America.

Dichtl, P., Gennaro, S.D., Li, Y., Maier, S.A. and Oulton, R.F., 2019, September. Exploiting the Nonlinear Optical Response of Gold Nanoantennas for ultrafast pulse characterisation. In *Laser Science* (pp. JTu3A-47). Optical Society of America.

Dichtl, P., Abdelwahab, I., Grinblat, G., Leng, K., Chi, X., Park, I.H., Nielsen, M.P., Oulton, R.F., Loh, K.P. and Maier, S.A., 2020. Giant and tunable optical nonlinearity in single-crystalline 2D perovskites due to excitonic and plasma effects (Conference Presentation), *Proc. SPIE 11358, Nonlinear Optics and its Applications 2020*, 113580K (1 April 2020); <https://doi.org/10.1117/12.2555559>

Gennaro, S.D., **Dichtl, P.**, Li, Y., Maier, S.A. and Oulton, R.F., 2020. Double-blind ultrafast pulse characterisation by mixed frequency generation in a gold antenna (Conference Presentation), *Proc. SPIE 11358, Nonlinear Optics and its Applications 2020*, 113581L (1 April 2020); <https://doi.org/10.1117/12.2555945>

Acknowledgements

At the end of this thesis, I would like to thank the many people I had the pleasure of working with and who helped and supported me. Firstly, I would like to thank my supervisors, Prof Dr Stefan Maier and Prof Dr Rupert Oulton. Thank you Stefan for giving me the chance to work in your group, for your valuable feedback on my work, for your continuous support, for establishing collaboration opportunities, and for managing everything so well! Thank you Rupert for being such a supportive and approachable supervisor, for always making time even during busy times, for the many enjoyable discussions about physics, for all your valuable comments and suggestions for my work in general and for improving my thesis in particular, and last but not least for your legendary annual barbecues!

I would like to thank dstl for funding my studentship in the CDT, enabling me to pursue this research and Dr Chris Howle for accompanying my project. Furthermore, my thanks go to Dr Sylvain Gennaro and Dr Michael Nielsen for training me on the experiments and for all the tips and tricks they showed me. I am also indebted to Dr Nicholas Gsken for his lab advice, his helpful contributions to and discussions about the perovskite project, and for the great time we had in London. My thanks also go to Dr Xiaofei Xiao for his help and advice on my simulations. Moreover, I would like to thank Dr Ibrahim Abdelwahab for the successful collaboration on the perovskite project. For this project I also owe a debt of gratitude to Dr Gustavo Grinblat, Dr Piers Barnes, Dr Oskar Blaszczyk, and again Dr Michael Nielsen for general experimental advice and very helpful discussions on the Z-scan measurements. For the GaN project, I would like to thank John Yang for letting me join him on the project and for his help with some of the measurements and Prof Dr Anthony Bennett and Dr JP

Hadden from Cardiff University for the successful collaboration. In this context, I would also like to thank Dr Alex Clark for his insightful remarks, his very helpful advice, and for being such an engaging mentor of my cohort in the CDT for Controlled Quantum Dynamics. Furthermore, I would like to thank Prof Dr Ben Sauer for running the CDT – it has been a fantastic environment for pursuing my PhD and I believe I have had the opportunity to learn many things that I would not have learnt otherwise.

My thanks also go to Miranda Toora for handling the administrative side of the CDT so well and for her competent help whenever I had any administration-related problems. For the same reason, I would also like to thank Juraci Didone, Carolyn Dale, and Ievgeniia Kovryga from the cluster office.

Next, I would like to say thank you to all the students and postdocs in the group with whom I have not had the opportunity to work, but to interact in meetings and to share the office H722, for all the funny moments and the relaxed lunch breaks. Moreover, I would like to thank the students from my CDT cohort for the great time we had together, in particular my flatmates Alex Paige and Julia Bieniewska, and of course also my non-CDT flatmates Edd Sherlock and Istvan Kleijn, as well as my new flatmates in Hyndman House. My thanks also go to the members of the ICOS committee for all the fun times we had and the things we achieved together. In terms of extracurricular activities, music has always been closest to my heart, so I would like to thank everyone in Imperial College Choir for the fantastic time, particularly those with whom I was playing chamber music, and also everyone from the German Choir London and from the German Catholic Church St Boniface, particularly Father Andreas Blum, who let me accompany their Sunday services and allowed me to access the organ free of charge. Finally, I would like to express my gratitude to my friends and family in Germany, as well as to my uncle, aunt, and cousin here in London, and particularly to my parents, who always supported me throughout my journey.