Nonlinear spectroscopy of plasmonic nanoparticles

Julian Obermeier, Thorsten Schumacher & Markus Lippitz

To cite this article: Julian Obermeier, Thorsten Schumacher & Markus Lippitz (2018) Nonlinear spectroscopy of plasmonic nanoparticles, Advances in Physics: X, 3:1, 1454341, DOI: 10.1080/23746149.2018.1454341

To link to this article: https://doi.org/10.1080/23746149.2018.1454341
Nonlinear spectroscopy of plasmonic nanoparticles

Julian Obermeier, Thorsten Schumacher and Markus Lippitz
Experimental Physics III, Universität Bayreuth, Bayreuth, Germany

ABSTRACT
The plasmon resonance of a metal nanoparticle increases the optical field amplitude in and around the particle with respect to the incoming wave. In consequence, optical effects that are nonlinear in their field amplitude profit from this increased field. In general, a plasmonic structure can react nonlinearly by itself and it can also enhance the effect of the nonlinearity in its environment, which we consider as plasmonic nanoantenna. In this paper, we review third-order nonlinear effects such as third-harmonic generation, pump-probe spectroscopy, coherent anti-Stokes Raman scattering and four-wave mixing of and near plasmonic nanostructures. All these processes are described by very similar equations for the nonlinear polarization, but the underlying physics differs.

1. Introduction
The collective oscillation of the conduction electrons in a noble metal nanoparticle comes under certain conditions to a resonance, the plasmon resonance. Its spectral position and shape depends on size, geometry and material of the nanoparticle, as well as on the dielectric properties of the surrounding. The linear optical properties such as absorption, dispersion and scattering have
been investigated in the last 20 years in great detail and are well understood. With increasing understanding, more and more complex structures have been designed, consisting of intriguing shapes of single particles or arrangements of multiple particles. Recent reviews cover diverse aspects of this still growing field [1–9].

Nonlinear effects are enhanced inside and near plasmonic nanostructures due to their nonlinear amplitude dependence. This causes the generation of new frequencies and ultrafast, transient variations of the fundamental wave’s intensity. Technologically, this is relevant for ultrafast switching and signal generation in optical telecommunication. As spectroscopic tools nonlinear processes give information beyond linear absorption spectroscopy, e.g. on ultrafast dynamics and the interaction between states. In plasmonics this has led to, e.g. the understanding of the electronic dynamics after laser excitation and to vibrational spectroscopy of single molecules near plasmonic antennas. In this review we will concentrate on one class of processes, namely third-order nonlinearities, inside and near single nanostructures or homogeneous arrays of structures, i.e. we do not cover nonlinear meta-surfaces and nonlinear waveguides. Other perspectives are given in recent reviews [10–24].

We start with a brief overview of nonlinear optical processes and how they are described by a nonlinear susceptibility. Then the special situation of nonlinear optics with plasmonic nanostructures is introduced. The three main sections of this review are concerned with pump-probe spectroscopy, third-harmonic generation (THG) and finally four-wave mixing (FWM) in general. We use the section on pump-probe spectroscopy to introduce the ultrafast dynamics in metal nanoparticles which was mainly investigated by this technique. For all nonlinearities, we discuss spectroscopic experiments on the metal nanoparticles themselves, but also on their use as antennas for adjacent nonlinear processes.

2. Third-order nonlinear optical processes

The polarization $\mathbf{P}$ describes the response of matter to an applied external optical field $\mathbf{E}$

$$\mathbf{P} = \varepsilon_0 \chi \mathbf{E}$$

$$\mathbf{P} = \varepsilon_0 \left( \chi^{(1)} \mathbf{E} + \chi^{(2)} : \mathbf{E} \mathbf{E} + \chi^{(3)} : \mathbf{E} \mathbf{E} \mathbf{E} + \chi^{(4)} : \mathbf{E} \mathbf{E} \mathbf{E} \mathbf{E} + \cdots \right),$$  \hspace{1cm} (1)

where the susceptibility $\chi$ is the proportionality constant and $\chi^{(n)}$ are its Taylor coefficients. In general, both the electric field $\mathbf{E}$ and the polarization $\mathbf{P}$ are vectorial quantities which do not need to point in the same direction, i.e. the susceptibility $\chi^{(n)}$ is a tensor of rank $n + 1$. To simplify the notation, we limit us here to a scalar discussion. The susceptibility $\chi^{(1)}$ describes the linear response of the material which dominates at low amplitudes of the field. The corresponding linear polarization
Figure 1. Sketches of $\chi^{(3)}$ processes: third-harmonic generation (THG), pump-probe spectroscopy (PP), coherent anti-Stokes Raman scattering (CARS) and non-degenerate four-wave mixing (FWM). All processes start and end at a real state that can be populated (solid horizontal line). Intermediate states can be virtual states (dashed horizontal line). The interaction between light and matter occurs in time from left to right within each diagram. The last downward arrow symbolizes the emission process. Other downward arrows symbolize incoming complex-conjugate fields that count negative when calculating the emission frequency.

\[
P^{(\text{linear})} = \varepsilon_0 \chi^{(1)} E
\]

is source of an emitted wave that interferes with the incoming wave. This interference leads to a reduced field amplitude of the transmitted beam. The extinction spectrum $\alpha(\omega)$ is thus directly related to the linear polarization $P^{(\text{linear})}$ and the linear susceptibility $\chi^{(1)}$:

\[
\alpha(\omega) = \frac{\omega}{c} \text{Im} \left( \chi^{(1)} \right).
\]

This homodyne interference is possible as the generated polarization and the incoming field oscillate at the same frequency.

The lowest order nonlinear susceptibility $\chi^{(2)}$ is responsible for second-harmonic generation and sum- and difference-frequency generation. Second-order nonlinearities require breaking the centrosymmetry, which can be achieved in different ways. In nonlinear optical crystals, for example, the symmetry is broken microscopically by the crystal lattice. In contrast, plasmonic materials as silver and gold have centrosymmetric crystal structures, but the nanostructure can break the symmetry. This can be the intended shape, for example, similar to a V or L. It could also be manufacturing artefacts such as surface roughness and tilted side walls or the substrate beneath the sample that break the symmetry. As the latter are difficult to quantify and to model, the experimenter needs to make sure that they do not dominate the intended asymmetry. A detailed discussion of plasmonic second-order nonlinearities can be found in Refs. [14,15].

In this review, we will concentrate on the next higher order nonlinear polarization $P^{(3)} = \varepsilon_0 \chi^{(3)} EEE$ and its effects in and near plasmonic nanostructures: THG, CARS and pump-probe spectroscopy. These effects are summarized as FWM effects and depicted in Figure 1. The generated nonlinear polarization oscillates at the sum of the three input frequencies. Complex conjugate input
fields are counted as negative frequencies and depicted as downward arrows in Figure 1.

In THG, one optical wave at frequency $\omega$ generates a new optical field at frequency $3\omega$:

$$P^{(THG)}(3\omega) = \epsilon_0 \chi^{(3)} E(\omega)^3. \quad (4)$$

This is a coherent process, i.e. the phase of the new field is related to the phase of the fundamental wave, but homodyne interference is in contrast to linear absorption not possible. The detected signal intensity $I^{(THG)}$ is thus proportional to the third power of the fundamental intensity $I(\omega)$

$$I^{(THG)} \propto |P^{(THG)}(3\omega)|^2 = |\epsilon_0 \chi^{(3)} E(\omega)^3|^2 \propto I(\omega)^3. \quad (5)$$

In pump-probe spectroscopy, a pump-beam is absorbed and its influence on the absorption of a second probe beam is monitored. Here it is the intensity of the pump beam, not the field amplitude that enters. The probe beam appears only linear in field amplitude as in linear absorption spectroscopy. All together pump-probe spectroscopy can be written as

$$P^{(pp)}(\omega_{\text{probe}}) = \epsilon_0 \chi^{(3)} E(\omega_{\text{pump}}) E^*(\omega_{\text{pump}}) E(\omega_{\text{probe}})$$

$$= \epsilon_0 \chi^{(3)} |E(\omega_{\text{pump}})|^2 E(\omega_{\text{probe}}). \quad (6)$$

From this notation, it is obvious that pump-probe spectroscopy does not depend on the phase relation between pump- and probe-beam. The generated nonlinear polarization oscillates at the frequency of the incoming probe beam which again allows homodyne detection. The observed signal is thus proportional to the amplitude of the interference between nonlinear polarization and probe field and not to the square of the nonlinear polarization itself as in THG. The signal scales linearly both in pump and in probe power.

In CARS, the pump field also enters twice, as in pump-probe spectroscopy, but here it enters twice with the same phase. Both pump arrows point up in Figure 1. The nonlinear polarization is given by

$$P^{(CARS)}(\omega_{\text{CARS}}) = \epsilon_0 \chi^{(3)} E(\omega_{\text{pump}}) E(\omega_{\text{pump}}) E^*(\omega_{\text{Stokes}}) \quad (7)$$

and oscillates at a new frequency $\omega_{\text{CARS}} = 2\omega_{\text{pump}} - \omega_{\text{Stokes}}$. Either one detects its square (similar to THG) or one supplies an additional field acting as local oscillator to produce a signal that is linear in the generated polarization. The intermediate state of lowest energy in Figure 1 is a vibrational state of a molecule. It enhances the efficiency of this process which can also take place without a real state at this energy. In this review, we only call processes that are resonantly enhanced CARS.

In the most general case of FWM, all three input waves differ in frequency. The examples above are cases of degenerate four-wave mixing (DFWM) where some
frequencies coincide. In this review, we use the term ‘four-wave mixing’ only for the non-degenerate case and stick to the more specialized terms otherwise. An experimentally accessible variant of FWM is depicted in Figure 1 where three similar but not identical waves produce a fourth wave which again is spectrally close to the input waves. This process scales linearly in all three input powers and also the phase relation between all waves is relevant.

Note that for a given material the value of $\chi^{(3)}$ depends on all three incoming frequencies $\omega_1, \omega_2, \omega_3$ and the outgoing frequency $\omega_4$, which is written as $\chi^{(3)}(\omega_4; \omega_1, \omega_2, \omega_3)$. To simplify the notation, we will omit the frequencies in the argument of $\chi^{(3)}$ in the following.

3. Plasmonic nanostructures

The plasmon resonance comes along with an increased field amplitude in the optical near-field around and inside the nanoparticle. As the nonlinear optical effects discussed above depend to the third power on the field amplitude, the plasmon resonance boosts such effects. This is obvious for the incoming fields, which are increased due to the plasmon resonance. But also the generated, outgoing field profits from the resonance. This can be seen as consequence of Lorentz reciprocity, i.e. that the direction of a beam can be reversed if certain precautions have been taken [25, 26]. From another perspective, one could also argue that the presence of the plasmon resonance changes the optical density of states which enters into the probability for emission processes [27].

In the preceding chapter on nonlinear optical processes, we did implicitly assume a homogeneous medium. When introducing plasmonic nanostructures this is not the case anymore. We need thus to distinguish between different situations:

- The nonlinearity stems from the medium around the plasmonic particle, which itself is assumed to respond linearly, thus to just modify the field amplitude and mode density. The particle acts as plasmonic nanoantenna to boost nonlinear processes in its environment. The nonlinear response is integrated over a certain volume, a hot spot, outside the particle.
- The plasmonic particle itself responds nonlinearly to the incoming field and the surrounding medium is assumed to respond linearly. With a skin depth of a few tens of nanometres, the optical field penetrates the nanoobject considerably. The nonlinear response is integrated over the volume of the particle, taking into account that the generated wave needs to be able to leave the particle.
- The optical nonlinearity can be caused by the surface of the particle, i.e. the interface between particle and environment. In this case, the nonlinear response is integrated over the particle surface only. This is mainly relevant for the case of second-harmonic generation [28].
Figure 2. Sketch of the band structure of gold. The energy scale is relative to the Fermi energy $E_F$, i.e. all states at negative energy are filled at zero temperature. The $sp$ band leads to the metallic behaviour of the free electrons. The electrons in the $d$ band need a certain minimum energy to be excited to the $sp$ band. The density of states (DOS) per energy interval is low for the free electrons and peaks about 2 eV below the Fermi energy due to the $d$ band. The right panel sketches the crystal structure and the Fermi surface in reciprocal space. Taken from [17] with permission of Springer.

4. Pump-probe spectroscopy of and by plasmonic nanoparticles

In a pump-probe experiment, a first ultrashort laser pulse, the pump pulse, modifies the optical properties of a sample which are then interrogated at a later time by a second pulse, the probe pulse. As the phase of the pump field does not enter in Equation (6), one also writes the process as pump-induced modification $\Delta \epsilon$ of the dielectric function $\epsilon(\omega)$. The complex-valued $\Delta \epsilon$ depends on pump- and probe-frequency and the temporal delay between the pulses.

4.1. Overview of time scales

The optical properties of solid materials are governed by their band structure. Gold has a filled $5d$ band and a partially filled $6sp$ band (Figure 2). The free electrons in the $sp$ band are described by a Drude model which allows excitations at arbitrary low frequency. The bound electrons in the $d$ band need an excitation energy of at least 1.9 eV (650 nm) to reach empty states in the $sp$ band at the Fermi energy.

The electronic dynamics in noble metal bulk and nanoparticles has been investigated in detail. For a review see Refs. [12,18,22,29,30]. In a gold nanoparticle, a near-infrared pump pulse excites a collective electron oscillation that dephases on a timescale of 10 fs [31]. The electron gas then still has an athermal energy distribution which thermalizes by further electron–electron scattering within a few 10 fs [32]. If the pump pulse has enough energy to excite also $d$ band electrons to the $sp$ band, Auger recombination will lead within a few 10 fs to a similar athermal energy distribution of the free electrons and again to a filled $d$ band. As soon as a temperature can be assigned to the electron gas, the following dynamics is described by the so-called two-temperature model [22,29,33]. As the heat capacity of the free electrons is low, the thermalized energy distribution
reaches typically a temperature of a few hundred to a few thousand Kelvin. On a timescale of about a picosecond the energy is transferred into the lattice by electron–phonon scattering and the lattice temperature increases. The lattice then cools down by heat conduction to the environment of the particle. These steps can be observed optically, see, e.g. Refs. [12,34]. The modified occupation $f(E)$ of the states in the $sp$ band changes the absorption for transitions within the $sp$ band and for those starting from the $d$ band. For the free electrons, the hot electron gas results in an increased damping rate as more states are accessible for electron-electron scattering [35]. Both together modify the extinction spectrum of a plasmonic nanoparticle as function of pump probe delay (Figure 3).

**4.2. Acoustical oscillations**

The impulsive heating of a plasmonic particle by a short laser pulse leads to a thermal expansion of the lattice and to an increased outward-bound pressure of the free electrons on the particle surface. Both effects together launch mechanical oscillations of the whole particle [38,39]. These oscillations periodically modulate the transient transmission signal as the particle size influences the extinction
cross section directly but also indirectly via the electron density and the plasma frequency. For a review see, e.g. Refs. [22,40].

In recent years, the relation between the shape of the particle and its acoustical mode spectrum has been investigated in detail [36,41–43]. Single particle spectroscopy allowed to investigate the influence of rare cases in particle geometry [44, 45]. The damping of the acoustical oscillations was found to depend on the acoustical mode and its interaction with the environment [36,37,46,47]. Two examples are depicted in Figure 4. In all cases, the intrinsic damping of the mechanical oscillation cannot be neglected. Even for very small particles with sizes of only a few nanometres, the elastic response can be measured and modelled [48,49].

An intriguing application of the acoustical oscillations is mapping of local elastic properties [50]. Although the coupling of a small spherical particle to the substrate is weak, a larger platelet emits a sizeable acoustical shock wave in the substrate. This wave travels with the velocity of sound away from the particle and is connected with a perturbation of the index of refraction. A part of a probe laser pulse is reflected at this perturbation and interferes with a more intense part that is reflected at the sample surface. In a pump-probe experiment one finds periodic variations in the reflected intensity that are connected to the distance between sample surface and sound wave. In this way, the local velocity of sound underneath the plasmonic platelet can be determined. This opens the door to nanoscale ultrasound imaging of natural and synthetic soft matter [51], starting from an optical nonlinearity in plasmonic particles.

**Figure 4.** Two examples of how the environment influences the acoustical oscillation of nanoparticles. (a,b) Lithographically defined gold nanodiscs of the same size but with a titanium adhesion layer of varying thickness show different eigenfrequencies of their acoustic mode. This can be modelled as transition from a free to a fixed boundary condition. Reprinted by permission from Macmillan Publishers Ltd. [36]. (c) A single gold nanorod immobilized on a glass surface shows two different oscillation modes: a breathing mode around 95 GHz and an extensional mode around 15 GHz. Changing one halfspace from air to water changes drastically the damping of the extensional mode, but leaves the breathing mode almost unaffected. This is attributed to hydrodynamic lubrication forces between rod and surface. Reprinted with permission from [37]. Copyright 2013 American Chemical Society.
Figure 5. A plasmonic nanoantenna increases the signal from acoustical breathing oscillations in another plasmonic particle. (a) Two gold nanodiscs of different size couple plasmonically and form a common hybridized mode. (b) After impulsive laser excitation, both discs oscillate mechanically. The small disc’s breathing mode modulates the common hybrid mode. Its amplitude (red) depends on the probe wavelength, but is about a factor of 10 larger than the corresponding signal of the small disc without antenna (black). Modified with permission from [53].

4.3. Plasmonic antennas for pump-probe spectroscopy

Pump-probe spectroscopy of plasmonic nanoparticles does not only allow to obtain information about the particle itself, but also about the environment of the particle. The increased amplitude of the optical field around a plasmonic particle should allow to use this particle as optical antenna to increase signals from pump-probe spectroscopy on objects in the environment. One can distinguish between two limiting cases [52]: the nanoobject under investigation scatters only very weakly so that it does not act back on the antenna, or the scattering is so strong that the nanoobject detunes the optical antenna and a new coupled system is investigated. The latter is the case if one uses a bigger gold nanodisc as antenna to investigate the breathing oscillations of a smaller gold disc. In this case, we could show [53] that a good antenna increases the nonlinear pump-probe signal of the small disc by a factor of 10, but the spectral response changes due to hybridization of the two plasmon modes (Figure 5). The spectrum of the acoustical modes in the gigahertz range is modulated on an optical carrier frequency that is not the plasmon resonance of the original particle but the hybridized mode of antenna and object under investigation.

In the other limit, the weakly scattering particle, the antenna resonance and the optical transition of the object under investigation have to coincide to obtain an antenna effect [52], similar to fluorescence emission enhancement. The same restrictions on the optimal distance hold: a too close distance between plasmonic particle and emitter leads to quenching, i.e. non-radiative processes. If the distance is too large, no enhancement is observable. Additional to these restrictions from linear fluorescence spectroscopy, in pump-probe spectroscopy, the acoustical breathing oscillations discussed above now become a background contribution which needs to be avoided or removed. No successful examples of antenna-enhanced pump-probe spectroscopy have been published to our
knowledge in this weakly scattering limit. Other nonlinear processes like CARS (see below) seem to be more accessible.

As special case of a plasmonic antenna for pump-probe spectroscopy, one can see photothermal microscopy [54]. A pump beam is absorbed in a nanoobject that is embedded in a transparent medium. The absorbed energy heats the environment of the nanoobject, which leads to a modified index of refraction. The probe beam is transmitted through this volume of different refractive index around the nanoobject. The signal is generated by scattering or by interference with a reference wave [55]. This scheme is special compared to all other processes that we discuss in this review as pump- and probe-field interact with different materials and as the effect on the probe beam is mainly dispersive. In the first experiments, the nano-object has been a plasmonic nanoparticle, but since then also semiconductor nanocrystals [56], molecules [57] and polymers [58] have been investigated. The surrounding medium can be water or a polymer, or, as shown recently, a liquefied gas under high pressure near its triple point, as in this situation the temperature-induced variation of the refractive index is largest [58]. The combination of an infrared pump beam and a visible probe beam leads to an optical resolution defined by the visible wavelength in combination with spectroscopic information of the infrared wavelength [59]. A recent review gives a more detailed overview on this technique [60].

5. Third-harmonic generation

5.1. Origin of the THG

In contrast to pump-probe spectroscopy, THG is a coherent process. The nonlinear polarization is coherent with the driving electrical field. Although THG is also a $\chi^{(3)}$-process, the underlying physics differs. The responsible processes have to conserve the phase relation, i.e. should be fast compared to dephasing times which typically are in the order of 10 fs. Moreover, a free electron gas does not have a restoring force, and therefore also no nonlinearity in the restoring force, as long as the particle surface is not taken into account [61–63]. Theoretical models predict an absolute value of $\chi^{(3)}$ for THG that is three to four orders of magnitude lower than for pump-probe spectroscopy [61]. Experimental observations (for an overview see [64]) find values in the order of $|\chi^{(3)}| \approx 0.2 \text{ nm}^2/\text{V}^2$ for THG and non-degenerate FWM in gold. For comparison, indium-tin-oxide (ITO) has a value of about 0.01 nm$^2$/V$^2$ [65], silicon about 2.3 nm$^2$/V$^2$ [64].

Experiments in the field started by investigating THG at a plain gold surface [66]. An ensemble of lithographically defined particles was used by Lamprecht et al. [67]. In 2005, we detected the third-harmonic emission from single gold nanospheres on a glass substrate when excited at 1500 nm wavelength [68]. The third-harmonic was thus close to the particle resonance. The size dependence of the signal did not allow to distinguish between a surface ($R^4$) and a volume ($R^6$) effect. The analysis was complicated by the finite skin depth which at least for
the larger particles was smaller than the radius, and by the size-dependence of the plasmon resonance.

Utikal et al. [69] used hybrid dielectric-plasmonic waveguides with different dielectric materials. A comparison of the extinction and the third-harmonic emission spectrum together with simulations of the field distribution allowed to identify the dominating source for THG. A similar method of comparing a structure in different environments was proposed to measure the nonlinear susceptibility of small nanoparticles, excluding contributions from the environment [70].

As THG is a quasi-instantaneous process, it allows to obtain information on the local electric field on the femtosecond timescale. Already in 1999 Lamprecht et al. [67] used a third-harmonic interferometric autocorrelation technique to determine the dephasing time of an ensemble of plasmonic nanodiscs to 6 fs. In a similar experiment, Hanke et al. [71] investigated the dephasing of a single bow-tie antenna. The dephasing time influences strongly the maximum third-harmonic amplitude of an antenna [72].

A plasmonic nanostructure can be used in different ways for THG. It can enhance the fundamental field [67,71,73], or it can enhance the emission of the generated third-harmonic [68], or both. As the fundamental field enters to the third power, this first use is especially beneficial. Bar-Elli et al. [74] compare these cases and also show an example of the third case, the double resonance. They use a semiconductor shell around a gold nanorod not only to shift the resonance of the rod, but also to use the excitons in the shell as resonant state to enhance THG.

5.2. The nonlinear Lorentz oscillator model and Miller’s rule

Knowing the nonlinear susceptibilities $\chi^{(3)}$ of the involved materials, it is sufficient to integrate Equation (4) over a plasmonic nanostructure to obtain its third-harmonic response. However, to be able to design structures with desired properties, it is helpful to have a more intuitive understanding of the signal generation. It turned out that a simple nonlinear Lorentz oscillator is sufficient to describe the nonlinear response of many plasmonic structures [75–77].

A Lorentz oscillator, i.e. a charged mass connected to a spring, is a simple model for the linear response of matter [78]. The nonlinear response is included when adding a nonlinear term $bx^3$ to the restoring force, leading to

$$\ddot{x} + \gamma_0 \dot{x} + \omega_0^2 x - bx^3 = -\frac{e}{m}E_0 e^{-i\omega t}. \quad (8)$$

Here, $\omega_0$ and $\gamma_0$ are eigenfrequency and damping of the oscillator of mass $m$. $\omega$ is the frequency of the driving field of amplitude $E_0$. The linear susceptibility $\chi^{(1)}$ is proportional to the extension $x$ of the spring if the medium has a number density $N$ of oscillators:
\[ \chi^{(1)}(\omega) = N e x = \frac{Ne^2}{\epsilon_0 m} \frac{1}{\omega_0^2 - \omega^2 - i\omega\gamma_0}. \] (9)

This linear susceptibility can be determined by measuring the linear extinction spectrum \(\alpha(\omega)\) (see Equation (3)). In this model, the nonlinear susceptibility \(\chi^{(3)}\) is fully determined by \(\chi^{(1)}(\omega)\)

\[ \chi^{(3)}(\omega_4; \omega_1, \omega_2, \omega_3) = \frac{b m}{3N^3\epsilon_0^4} \chi^{(1)}(\omega_1) \chi^{(1)}(\omega_2) \chi^{(1)}(\omega_3) \chi^{(1)}(\omega_4). \] (10)

This prediction of the nonlinear susceptibility \(\chi^{(n)}\) is known as Miller’s rule and discussed in many nonlinear optics textbooks for the case of transparent nonlinear crystals [78]. It also successfully describes the third harmonic emission of bow-tie antennas [75]. A bow-tie antenna is apparently a sufficiently simple plasmonic structure so that its linear response can be described by a simple harmonic oscillator, i.e. has a Lorentzian lineshape. This means that higher order quadrupolar modes are spectrally sufficiently separated or even damped out due to \(d\) band absorption of gold. The field distribution inside the antenna \(E(r)\) has a certain non-constant shape. The total linear polarization \(p^{(1)}\) is then obtained by integrating over the particle:

\[ p^{(1)} = \int_{\text{particle}} p^{(1)} \, dr = \epsilon_0 \int_{\text{particle}} E(r) \, dr. \] (11)

The total nonlinear polarization for THG is obtained by

\[ p^{(\text{THG})} = \int_{\text{particle}} p^{(\text{THG})} \, dr = \epsilon_0 \int_{\text{particle}} E(r)^3 \, dr. \] (12)

Miller’s rule states that \(\chi^{(1)}\) and \(\chi^{(3)}\) are related, but this only helps to predict \(p^{(\text{THG})}\) if also the integrals over \(E(r)\) and \(E(r)^3\) are related by a constant factor. This is only the case if the shape of the field distribution does not change with the fundamental frequency \(\omega\) or with any shape parameter of the structure that is tuned when comparing experiment and model. When tuning the laser frequency over the plasmon resonance or varying the gap of the bow-tie antenna, the field distribution does not change substantially. For very small gaps, the shape of the field distribution changes and the model fails (see Figure 6).

This limitation becomes obvious in the case of so-called dolmen structures: three plasmonic nanorods arranged in the shape of a U [76]. When shifting the middle rod out of its centred position, the nonlinear oscillator model correctly predicts the shape of the third-harmonic spectrum for all positions of the rod. This means that tuning the laser frequency over the resonance of the structure does not significantly change the field distribution. However, the model does not correctly predict the amplitude of these spectra as function of rod position.
Moving the rod changes the shape of the field distribution, which breaks the relation between the integral over first (Equation (11)) and third power of the field distribution (Equation (12)).

5.3. Imaging third-harmonic emission

When the plasmonic nanostructure is larger than the third-harmonic wavelength, it becomes possible to image the third-harmonic emission by a high-resolution optical microscope, either in real-space or in reciprocal space [72,79–81]. The latter gives information about the emission direction and also about coherence between different emitting points. As in a double-slit experiment, interference fringes in the angular intensity distribution signal coherence between the slits. Such experiments [79,80] demonstrate that the third-harmonic emission stems from the centre of short rod antennas. A surface or a hot-spot contribution would lead to distinctively different emission patterns and is thus much weaker, if present at all. A more detailed discussion is given in Ref. [79] and its supplementary material.

As we have seen above, the difference between linear absorption and THG is mainly the different exponent in the dependence on the electrical field $E$. Imaging of a plasmonic nanostructure allows to demonstrate this effect [79]. A plasmonic nanorod has not only a fundamental dipolar resonance, but also higher order resonances similar to standing waves in a resonator [82]. Even order modes with a node in the centre do not couple to plane waves by symmetry. If one excites a rod of suitable length around its third-order resonance, the third-harmonic emission shows a peculiar response: it switches from one emitting centre for fundamental
Figure 7. Switching of the third-harmonic emission. Top: Numerical simulations of the field distribution at a gold nanorod of 925 nm length. When the fundamental wavelength is varied around the third-order resonance at 1250 nm, the field distribution does not change visibly. However, due to the third power dependence, the field at the third harmonic changes drastically, from two hot spots below resonance to one hot spot above resonance. Bottom: When imaging the far-field emission pattern, one finds below resonance the interference stemming from two spatially separate hot spots, while above the resonance almost a pure dipolar emission pattern is observed. Modified with permission from [79] and its supplementary material.

wavelengths above the third-order resonance to two emitting centres for shorter wavelengths [79]. Numerical simulations show that the fundamental field almost does not change, but the steep amplitude dependence of THG makes the difference (Figure 7). This allows to change the third-harmonic near-field around the rod drastically by only slightly tuning the fundamental wavelength.

5.4. Plasmonic antennas increase THG in dielectric particles

It is intriguing to use a plasmonic nanoantenna to increase THG in a neighbouring nonlinear material. Several efforts have been undertaken in this direction, both experimentally as well as theoretically, and the subject is still under debate [69,83–91]. One example is shown in Figure 8. The problem arises from the high nonlinearity of gold itself. Even if the field amplitude in the antenna is much weaker than in the nonlinear dielectric particle next to it, it is not obvious which sub-volume dominates. Moreover, a nonlinear dielectric also has a linear index of refraction which shifts the plasmon resonance spectrally. Ideally, one would switch $\chi^{(3)}$ of a material while keeping $\chi^{(1)}$ constant. A recent experiment tried to approach this scenario by photobleaching a polymer [89].

As Boyd et al. point out [64], the nonlinear susceptibility of gold is strongly influenced by the $d$-band absorption. For pump-probe spectroscopy, $\chi^{(3)}(\omega_1; \omega_2, -\omega_2, \omega_1)$ depends on pulse length and wavelength. Similar effects are to be expected also for $\chi^{(3)}(3\omega; \omega, \omega, \omega)$ for THG, although only few experimental values
Figure 8. A gold nanoring as antenna for THG in a silicon pillar. When excited at 1320 nm wavelength, the hybrid construct generates orders of magnitude more third-harmonic light than the silicon pillar or the gold ring alone. On resonance, the field is concentrated in the silicon part, which has a nonlinear susceptibility $\chi^{(3)}$ that is about an order of magnitude larger than that of gold [64]. Reprinted with permission from [83]. Copyright 2017 American Chemical Society.

exist [71,92]. This could explain some of the discrepancies between different experimental results on plasmonic enhancement of dielectric nonlinearities.

6. Other FWM techniques

6.1. Non-degenerate FWM

While pump-probe spectroscopy and THG are also FWM techniques, we focus now on more general approaches. Already in 2007 Danckwerts and Novotny [93] could demonstrate that in a nanoparticle dimer a FWM process leads to emission at the frequency $2\omega_1 - \omega_2$ when two laser pulses with frequency $\omega_i$ impinge on the sample. The signal amplitude depends strongly on the separation of the particles in the dimer, as a close distance brings the hybridized plasmon mode in overlap with the FWM signal. Moreover, the near-field distribution changes towards a more and more concentrated hot-spot.

As motivated by, e.g. Babinet’s principle [94], a hole in a metal film shows a plasmon resonance comparable to that of the inverse structure, a metal nanodisc. The spectral position of the resonance changes with the size of the hole. When shifting the resonance of a hole over the frequency $\omega_1$ of one of the laser pulses, a resonance in the FWM amplitude at $2\omega_1 - \omega_2$ was observed [95]. A small inclusion of vacuum can thus lead to a nonlinear optical response.

An even stronger effect can be observed when the plasmonic nanostructure supports a resonance at both laser frequencies, as demonstrated for a multi-particle arrangement [96]. A prerequisite is that both resonances lead to field enhancement in overlapping volumes.

No resonance at all was desired by Kravtsov et al. [97] in a sharp metal tip. A broad-band ultrashort laser pulse was used to launch surface plasmons travelling along the shaft towards the apex of the tip (Figure 9). With decreasing shaft
Figure 9. FWM at the apex of a gold nanotip. (a) An ultrafast laser pulse is spectrally and temporally shaped before it is launched along a gold nanotip. The field enhancement at the apex leads to FWM signals. (b) The FWM signal is detected as wing at the blue side of the spectrum. To facilitate detection, the laser spectrum was clipped beforehand. (c,d) Mapping the FWM signal as function of tip position shows enhanced signals at certain spots over gold islands. The enhancement is attributed to resonances that form between tip and island. Reprinted by permission from Macmillan Publishers Ltd. [97].

... diameter, the field amplitude increases which boosts nonlinear effects. FWM of three spectral components of the broad-band pulse leads to new frequencies on the blue side of the spectrum. This signal is spectrally separated and detected. As the tip shows no resonance, its response is instantaneous and only limited by the temporal length of the exciting pulse. When the tip is close to a gold nanostructure, a resonance forms and the response shows an intrinsic time constant of about 10 fs. This allows ultrafast spectroscopy on the nanoscale.

6.2. Extensions and variants of pump-probe spectroscopy

Pump-probe spectroscopy is a special case of FWM, thus a setup for general FWM spectroscopy is able to retrieve pump-probe signals. However, a closer control of the phase relations of the fields allows to obtain not only one but both components of the complex-valued pump-induced change $\Delta \epsilon$ in the dielectric function $\epsilon(\omega)$. The experimental problem in FWM spectroscopy is the separation
Figure 10. Plasmonically enhanced CARS of a single molecule in the centre gap of four gold nanodiscs (panel b). This arrangement of discs shows resonances at all involved wavelengths, thus maximally enhancing the FWM process. When covered with a diluted mixture of para-mercaptoaniline (p-MA) and adenine, Raman spectra of the single species can be detected (red and green curve in panel a). The rare observation of a mixed spectrum (blue) is consistent with an average 0.9 molecules per hot spot. Reprinted by permission from Macmillan Publishers Ltd. [101].

of the mixing signal. With bulk samples, the propagation direction can be used, which has to fulfill momentum conservation. This is not possible in a microscope objective. The experiments discussed above use a large spectral distance between the contributing fields. The experiment by Masia et al. [98,99] starts from a single rather narrow-band laser, but tags the fields by frequency-shifting each field independently using an acousto-optical modulator. The mixing field thus ends up at a certain frequency relative to the starting laser and can be demodulated there. As side effect, this allows to vary the phase of the local oscillator, so that both real and imaginary components of the nonlinear polarization can be detected.

In pump-probe spectroscopy one often measures a signal as function of pump-probe delay and then performs a Fourier-transform to obtain frequencies in the gigahertz range. This can also be turned around. Two narrow-band continuous-wave lasers are scanned relative to each other. When the frequency difference hits a resonance of the system, energy is transferred from one laser beam to the other. Such an experiment was demonstrated by Wu et al. [100] for the breathing mode oscillations of metal nanoparticles that were discussed in Section 4.2 in the time-domain.
6.3. Plasmonic antennas for FWM

Similar to THG and pump-probe spectroscopy, a plasmonic nanoparticle can also enhance other FWM processes in its environment. Especially interesting are variants of Raman scattering, as they yield chemical information on the nanoscale. Raman scattering itself is an incoherent inelastic scattering process [102] and does thus not fall in the scope of this review. Details on tip-enhanced Raman spectroscopy can be found in Ref. [103,104]. However, CARS and stimulated Raman scattering (SRS) are both coherent nonlinear $\chi^{(3)}$ processes that yield information on vibrational modes. CARS is the same process that in the Section 6.1 above has led to the mixing product at $2\omega_1 - \omega_2$, with the difference that a vibrational state exists at $E_{\text{vib}} = \omega_1 - \omega_2$. The first step in the process is thus resonantly enhanced. Scanning $\omega_1$ relative to $\omega_2$ allows to obtain a Raman spectrum. This process is also coherent in the number of involved molecules, i.e. the number of molecules enters squared in the signal amplitude. A multiparticle plasmonic structure has allowed to measure plasmon-enhanced CARS of a single molecule [101]. In this work, Zhang et al. used a similar structure as above for the double resonance enhancement of FWM. Investigating a mixture of two Raman-active molecules, they could show that the statistical occurrence of one Raman spectrum alone or a superposition of both is compatible with an average 0.9 molecules per plasmonic structure (Figure 10). Similar results have been obtained for a molecule near the touching point of two gold nanospheres held together by a silica coating [105,106].

Stimulated Raman scattering (SRS) is a pump-probe technique in which energy from a pump beam is transferred to a probe beam via a vibrational state [107,108]. It is very similar to CARS with the difference that in CARS the detected nonlinear polarization oscillates at a new frequency, while in SRS the nonlinear polarization oscillates at the optical probe frequency. The latter leads to homodyning with the incoming probe wave. Wickramasinghe et al. [109] used the field enhancement at the sharp gold tip of a scanning tunnelling microscope to locally increase the stimulated Raman scattering efficiency.

7. Conclusion and outlook

The examples discussed in the preceding chapters give an overview on third-order nonlinearities in and near plasmonic nanostructures. The field of nonlinear plasmonics is not as mature as linear plasmonics. To a large part this might be due to the additional experimental efforts connected with ultrashort laser pulses that are needed to provoke nonlinear effects. The idea of nonlinear plasmonics is simple: the high field amplitude in the hot spots will boost nonlinear processes. However, in some parts it is the clear demonstration of an otherwise straight forward idea that is lacking, in other parts it is our understanding of physics that needs to advance.
In the case of nonlinearities inside the particle, the ultrafast dynamics of the electrons is rather well understood, as demonstrated in the interpretation of pump-probe experiments. Similar physics should also describe the instantaneous nonlinearities responsible for THG. Here, several models have been proposed for the microscopic origin of the nonlinearity, but in our understanding a clear experiment is missing.

When it comes to plasmonic antennas for nonlinear spectroscopy, i.e. hybrid plasmonic structures connected to other nanoobjects such as molecules or dielectric particles, then the field is still in its infancy. A few promising examples have been demonstrated, but we are far away from routine applications of plasmonically enhanced third-order nonlinear spectroscopy. Perhaps more elaborate concepts of linear plasmonics using more evolved structures need to be matched to the right nonlinear process.

**Acknowledgment**

This publication was funded by the German Research Foundation (DFG) and the University of Bayreuth in the funding programme Open Access Publishing.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**ORCID**

Markus Lippitz [http://orcid.org/0000-0003-1218-6511](http://orcid.org/0000-0003-1218-6511)

**References**