Shot noise limited heterodyne detection of CARS signals

M. Jurna$^a$, E. Böttner$^c$, J.P. Korterik$^a$, C. Otto$^b$, I. Rimke$^c$, H.L. Offerhaus$^a$

$^a$ Optical Sciences group, MESA+ Institute for Nanotechnology, Faculty of Science and Technology (TNW), University of Twente, the Netherlands;

$^b$ BioPhysical Engineering group, MESA+ Institute for Nanotechnology, Faculty of Science and Technology (TNW), University of Twente, the Netherlands;

$^c$ APE GmbH, Berlin, Germany;

ABSTRACT

We demonstrate heterodyne detection of CARS signals using a cascaded phase-preserving chain to generate the CARS input wavelengths and a coherent local oscillator. The heterodyne amplification by the local oscillator reveals a window for shot noise limited detection before the signal-to-noise is limited by amplitude fluctuations. We demonstrate an improvement in sensitivity by more than 3 orders of magnitude for detection using a photodiode. This will enable CARS microscopy to reveal concentrations below the current mMolar range.

Keywords: Nonlinear optics, parametric processes, coherent anti-Stokes Raman scattering, CARS, spectroscopy, four-wave mixing, heterodyne.

1. INTRODUCTION

Imaging and visualization of biological processes in complex environments, such as living cells, is a subject of wide interest.$^1,2$ Chemical selectivity, high spatial resolution and high sensitivity are the main parameters of interest. The combination of vibrational spectroscopy and nonlinear microscopy provides a direct and noninvasive technique to localize and identify structures of different chemical composition.$^3$ Coherent Anti-stokes Raman Scattering (CARS) is a well known technique for vibrational spectroscopy.$^4–6$ CARS is a four-photon process, where a pump photon of frequency $\omega_p$, a Stokes photon of frequency $\omega_s$ and probe photon of frequency $\omega'_p$ (often taken the same as the pump frequency$^7$) interact with the sample and generate an anti-Stokes photon of frequency $\omega_{as} = \omega_p - \omega_s + \omega'_p$. The CARS signal is resonantly enhanced when the difference frequency $\omega_p - \omega_s$ coincides with a molecular vibrational level transition. By tuning the difference frequency, the unique vibrational spectrum of a molecule can be obtained and used for imaging. The probing of the internal vibrational structure of molecules, enable the CARS process of label-free identification and chemically selective microscopy. The resonant enhancement of the CARS process makes it faster than Raman and, compared to IR spectroscopy, the short wavelengths involved ensure a higher spatial resolution. Another advantage of CARS over the commonly used spontaneous Raman scattering is the one-photon-fluorescence-free monochromatic signal that can be detected with a high collection efficiency. Furthermore, the CARS signal intensity depends quadratically on the pump field intensity and linearly on the Stokes field intensity so that pulsed excitation can result in signal rates much higher than in spontaneous Raman scattering. CARS spectroscopy and microscopy have therefore attracted considerable attention.$^4–6,8$ Implementation has been limited due to the complexity of the required systems and the low yield of the highly nonlinear interaction. The low CARS yield limits it to detection of substances with concentrations in the mMolar regime. Fluorescence spectroscopy of labeled species offers higher signal levels for situations where the autofluorescence is low and introduction of (large, disruptive and often toxic) labels can be tolerated. In biological and food-related systems, autofluorescence is generally high and labels are preferably avoided but mMolar concentrations are unrealistically high for many of the interesting substances.$^9,10$

In this paper we demonstrate that CARS spectroscopy is possible at the shot noise limit using a compact setup that exploits the phase coherence of the CARS process for heterodyne detection.$^{11}$ The setup includes

---

Further author information: (Send correspondence to H.L. Offerhaus)

H.L.Offerhaus: E-mail: H.L.Offerhaus@TNW.Utwente.nl
Figure 1. Schematic of the OPO. The resonator consists of a 30mm long Brewster-angled LBO crystal inside a temperature controlled oven, two concave mirrors, M1 (100mm ROC) and M2 (250mm ROC), two plane mirrors M3, M4 and a 6% output coupler (OC). A 200mm lens focuses the 532nm pump beam in the LBO crystal.

a green pumped OPO in which a signal and idler wavelength are created. The phase of the signal and idler individually are undetermined. The idler is used for the generation of the CARS whereas the signal is used as the local oscillator. Although these wavelengths bear no direct phase relation to each other, we demonstrate that the CARS output wavelength is phase-locked to the signal wavelength creating an appropriate scheme for heterodyne detection. Analysis of the noise properties of the local oscillator reveal a window that allows for shot noise limited detection of CARS signals. Attenuation measurements demonstrate the detection of CARS signal levels well below the detector dark (noise) level and the current detection limits. The setup is inherently jitter-free and makes use of every single wavelength that is generated.

2. SETUP

Green-pumped Optical Parametric Oscillators (OPOs) have been demonstrated previously\textsuperscript{12,13} and a picosecond synchronously green-pumped OPO\textsuperscript{14} (now commercially available) is an ideal source for the tunable wavelength for CARS spectroscopy and microscopy: it provides wavelengths in the optimal region (near infrared), synchronized pulses and reasonable bandwidth.\textsuperscript{15} Lithium Triborate (LBO) LiB\textsubscript{3}O\textsubscript{5} is a good choice for the nonlinear crystal. This crystal has a high damage threshold, low linear absorption, low cost, and due to the zero beam walk-off, LBO crystals can be made long to compensate the moderate nonlinear coefficient of 0.8pm/V. The non-critical-phase-matching means that all parts remain fixed during scanning (no rotating parts). Since the beam pointing is defined by the flat output coupler it is independent of the signal wavelength and no adjustments are required in the CARS setup. Only the OPO cavity length has to be adjusted for the variation in round trip time.

A layout of the setup of the home-built OPO can be seen in Figure 1. The pump source is a modelocked Nd:YVO\textsubscript{4} based laser (Vanguard by Spectra-Physics or a Paladin by Coherent) at 1064nm, which is partially doubled to 532nm. The repetition rate is close to 80 Mhz. The beam is expanded to a waist radius of 3.4mm at the focusing lens. The OPO cavity is singly resonant for the signal wave and is formed by two concave mirrors, two flat mirrors and a 6% plane output coupler. The two extra plane mirrors were introduced to reduce the resonator size to fit a breadboard of 90 x 30cm. The mirrors are highly reflective (R>99.9%) over the range 740-930 nm and the transmission at 532nm is higher than 98%. The output coupler was placed on a precision translation stage with a motor-controlled micrometer. The LBO crystal is 30mm long and Brewster angled (59°) for the signal. The crystal is cut for a type I (e→o+o) non-critical-phase-matching in the xy plane (θ=90° and φ=0°). The three beams propagate along the x direction, where the incoming pump beam is polarized along the y-axis and the out coming signal and idler beam are polarized along the z-axis. The temperature of the crystal can be varied between ambient and 200°C with a stability of 0.1°C. The concave mirrors have a radius of curvature (ROC) of 100 (M1) and 250mm (M2), respectively. The angle of incidence on the second concave mirror (M2) is 18° to compensate for astigmatism introduced by the Brewster-angled LBO crystal.\textsuperscript{16} The signal
bandwidth is reduced by a Lyot filter.

To demonstrate the capabilities as a source for CARS spectroscopy, the tunable signal beam (740-930 nm) is combined with the fundamental of the laser (1064 nm) for the detection of the CH vibrational stretch modes in organic molecules. These modes produce strong resonances and a large signal at low excitation power, so that a Forward-detected CARS setup can be used. The polarization of the pump and Stokes beams are aligned parallel and can be controlled by a λ/2-plate in both beams. The pump and Stokes beam are combined on a dichroic (HR: 1064 nm; HT: 700-900 nm) mirror and focused by a 0.65 NA objective lens in the sample. The collected and filtered CARS signal is detected on a Hamamatsu (R1463) photomultiplier tube. In Figure 2a the CARS vibrational spectrum of Toluene is shown, for a few tens of milliwatts in the pump and Stokes beam. The CARS signal is corrected for the variation in output power. Every second a measurement is taken of the pump wavelength and the CARS signal. The total vibrational scan is obtained over 800 wavenumbers. A high spectral resolution around 2 cm⁻¹ is obtained, due to the smooth and stable temperature scanning of the OPO. Comparison of the CARS vibrational spectrum of toluene with the spontaneous Raman spectrum, Figure 2b, shows that the vibrational frequencies can be assigned with a one-to-one correspondence.

![Figure 2. (a) CARS vibrational spectrum of toluene and (b) Raman spectrum](image)

### 3. PHASE PRESERVING NONLINEAR CHAIN

Figure 3 depicts the phase preserving nonlinear chain. We start with the passively mode-locked Nd:YAG laser generating ~15 ps pulses at a repetition rate close to 80 MHz. The output at 1064 nm is partially doubled to pulses of 12 ps at 532 nm. This doubling is phase coherent, where the phase of the 532 nm is given by the addition of the phases of the two photons that make up the 532 nm. The 532 nm synchronously pumps the (home built or commercial) Optical Parametric Oscillator (OPO) generating pulses of ~6.5 ps for the home built OPO pumped by the Vanguard laser and ~9 ps for the APE Levante Emerald OPO pumped by the Coherent Paladin laser at the signal and idler wavelength. The phase of the signal and idler separately are free but the addition of their phases is locked to the phase of the pump. The freedom for the phase of the signal ensures that the resonating signal has no phase restrictions and projects the carrier envelope phase slips and any other phase errors onto the non-resonating idler, ensuring smooth operation. The idler wavelength is combined with the 1064 nm to generate a CARS signal at ω_CARS = 2ω_pump − ω_Stokes, where the 1064 nm acts as the pump and the idler as the Stokes wavelength. The phase of the signal is given by (2 × phase_1064 − phase_idler). The CARS (or anti-Stokes) wavelength is equal to the signal wavelength from the OPO. Moreover, the phase of the resonant CARS signal is determined as (2 × phase_1064 − phase_idler + phase_χ(3)), where the last term is a constant for a fixed wavelength. The phase of the signal is therefore locked to the phase of the CARS signal except for path-length variations between the point of creation and the point where they interfere on the detector. The phase preservation means that the signal wavelength can interfere with the CARS signal in a predictable way and can thus be used for interferometric amplification.
Figure 3. Optical chain for the phase-preserved generation of wavelengths for the CARS process. The inset shows the energy diagram for both paths to the CARS wavelength.

4. HETERODYNE DETECTION

The interference of the Local Oscillator (LO, the OPO signal wavelength) with the CARS signal yields a total signal intensity on the detector of

\[ I_{\text{detector}} = \text{LO} + \text{CARS} + 2 \frac{\sqrt{\text{LO} \cdot \text{CARS}}}{H_{\text{CARS}}} \]  

(1)

where \( \text{LO} \) and \( \text{CARS} \) signify intensities. \( H_{\text{CARS}} \) refers to the Heterodyne CARS power. This interference term scales with the root of the LO power and the interferometric gain can be defined as \( H_{\text{CARS}}/\text{CARS} = 2\sqrt{\text{LO}/\text{CARS}} \). The noise in the optical signal is determined by the shot noise in the CARS signal itself and the shot noise introduced by the local oscillator. The introduced shot noise is the dominant term of the two (as \( \text{LO} \gg \text{CARS} \)) and scales with the root of \( \text{LO} \), just like the interference term and the interferometric gain. The interferometric amplification can thus be used to lift the signal above the detector noise without degrading the original signal-to-noise ratio as long as the local oscillator shot noise is the dominant term. To avoid \( 1/f \) noise, we bring the detection to a spectrally less noisy region, shifting the CARS wavelength 40 kHz with respect to the OPO signal wavelength.

The detector can be either a Photo-Multiplier Tube (PMT), Photo-Diode (PD) or Avalanche Photo Diode (APD). APD’s are generally used in photon counting mode, generating a pulse for each detected photon. The PMT and PD are connected to a transimpedance amplifier terminated at zero Ohm by an OPerational AMPlifier (OPAMP), as depicted in the inset in fig. 4. The PD can be considered as a current source with a current proportional to the detected intensity (typically 0.6 A/W). The dominant noise sources in the detector are the Johnson noise in the resistor \( \sqrt{4kT \cdot R \cdot BW} \) where \( BW \) represents the detection bandwidth in Hz, the amplifier current noise (dark current independent of \( R \)) and the shot noise of the number of electrons generated by the detected signal. The detected voltage scales linearly with the transimpedance resistor \( R \). By increasing the resistor to the point where it dominates the other noise sources, the signal-to-noise of the amplifier scales with \( \sqrt{R} \) and for sufficiently large output, the noise is dominated by the shot noise of the detected signal. The maximum \( R \) is limited by the required bandwidth of the amplifier which scales as \( 1/R \). A typical configuration yields a 1 MHz bandwidth for 1 MΩ transimpedance. We employed a Centronic BPX65 (1*1 mm sensitive surface) PD, Burr Brown OPA655P OPAMP and a Hamamatsu R1463 PMT. The PMT is similar to the PD except that the dark current is lower when compared to a resistor of similar gain. At a gain of 10-10\(^6\) a PMT typically has some
Figure 4. Noise as a function of local oscillator power. The noise is expressed in Noise Equivalent Power (NEP), Root Mean Square (rms) in 1.58 Hz bandwidth so that all lines can be compared directly. The dashed lines A and B have a slope of 1/2, 1 respectively. Continuous lines are a simulation, symbols represent measured data. Inset shows the detector electronics.

equivalent 100 photons · BW dark count noise (10·10³ sec⁻¹ dark counts). APDs have the lowest dark count rate (down to ≃10 photons · BW). For a more extensive treatment on low noise amplification see.²¹

When the amount of local oscillator power on the detector is increased, the noise (signal without any CARS signal on the detector) first shows the dark count noise level (flat) followed by a section showing the shot noise of the local oscillator. This shows up as a slope of 1/2 in a log-log plot of the noise versus the LO power (parallel to line A in fig 4). At some point the noise becomes dominated by the amplitude fluctuations of the local oscillator. The amplitude fluctuation are correlated to the local oscillator power and therefore scale with a slope of 1 (parallel to line B). Since this section of the noise scales faster than the heterodyne signal (scales parallel to line A), the signal-to-noise deteriorates. The section in which the noise follows line A is the window in which CARS signals can be detected limited only by the shot noise.

The gain dynamics in the OPO determine the amplitude fluctuations, in particular the saturation level. If the OPO is not strongly saturated the amplitude fluctuations can show up at (local oscillator) levels where the noise is not yet dominated by the shot noise, effectively closing the window for shot noise limited detection.

The level of the detected shot noise compared to the “real” optical shot noise scales inversely with the root of the Quantum Efficiency (QE), shifting that portion of the line upward. The position of the amplitude fluctuations is unaffected as it scales directly with the number of photons. This seems to improve the detection window by allowing higher levels of local oscillator but in reality deteriorates the detection sensitivity (see Fig. 7). PMTs tend to have a low QE (10% to 0.1%) for wavelengths towards the Near Infra Red (NIR) that are traditionally favored for biological CARS detection.³,²² PDs and APDs have a much better QE (typically up to 85%). For the detection of very low signals, therefore, APDs are best. For practical applications PDs are favored because they are large, cheap and robust. APDs are expensive and fragile and suffer from a reduced dynamic range. As long as a window exists for shot noise limited detection, the detector choice is irrelevant.

5. RESULTS AND DISCUSSION

We have tuned the OPO idler to 1578 nm so that the difference frequency between the 1064 nm and 1578 nm matches the C−H stretch vibration at 3060 cm⁻¹ in toluene. The OPO signal (and CARS signal) is at 802.7 nm. The toluene sample is prepared between two cover glasses and is approximately 15 µm thick. Tens of mWatts of 1064 nm and 1572 nm are focused using a 0.60 NA air objective. The CARS signal is collected with an 0.65 NA objective. Filters then remove the 1578 nm and 1064 nm.
To shift the CARS frequency we use an AOM that is placed in the 1064 nm, rather than the signal beam, because this wavelength does not change as the OPO is tuned for a particular vibrational frequency. The AOM is driven at 80.00 MHz, effectively shifting the 1064 nm pulses, at a repetition rate of 80.02 MHz (depending on the room temperature), by 20 kHz. This shift is translated to a shift of 40 kHz at the CARS wavelength due to the involvement of two photons at 1064 nm in the creation process. The CARS wavelength is combined on a detector with the OPO signal wavelength and the detected intensity is fed to a lock-in amplifier set to detect at 40 kHz. We integrate for 100 ms using a second order cut-off filter (BW = 1.58 Hz).

Figure 4 shows the detected noise levels for a photodiode with three different transimpedances and a PMT as a function of the local oscillator power. The noise levels are back calculated to noise equivalent input power (at 803 nm, 1.58 Hz bandwidth) for comparison. The shot noise line for the PMT lies substantially above that for the PDs due to the low quantum efficiency (0.3%). The PMT shows a large window for shot noise limited detection. The PD shows a window only when terminated by 1 MΩ or 10 MΩ. At lower termination the required local oscillator level introduces noise dominated by amplitude fluctuations. Above 10 MΩ the opamp is no longer able to follow the 40 KHz oscillations.
The dependence of the amplitude fluctuations level on the saturation of the OPO is demonstrated in Figure 5 where we plot both the OPO (signal) output power and the detected noise at the heterodyne frequency for a fixed level of local oscillator on the detector (2 µW). Pumping the OPO with the Paladin (up to 12 W of 532 nm) instead of our Vanguard (2.0 W of 532 nm) increases saturation. The decrease in the amplitude fluctuations enables shot noise limited detection for lower transimpedance (and corresponding higher frequency) until the input voltage noise dominates Johnson noise (below 50 kΩ). In Figure 6 the performance of the Levante Emerald OPO, pumped by the Coherent Paladin laser is shown. This OPO is designed to operate at much higher pump (and output) levels than the home built OPO used in Figure 4. Pumping with the Paladin laser we can saturate the Levante Emerald to the point that even lower transimpedance have a window for shot noise limited detection. The lower transimpedance implies the possibility of detection at higher modulation frequencies and therefore higher scanning speeds. At the 6.2W of pump power 1.9 W of signal and 1.0 W of idler power is available for the generation of CARS.

Figure 7 shows the measured heterodyne CARS signals for a fixed local oscillator level (50 nW for the photodiode and 2 nW for the PMT), where the CARS signal is varied by introducing attenuation filters in that part of the beam. For comparison we also show the direct detection of the CARS signal by removing the local oscillator and chopping the CARS beam with a mechanical chopper. The chopped signal shows the expected direct dependence on the CARS level with a slope of 1. The heterodyne signal depends only on the root of the CARS signal and thus decreases with a slope of 1/2. For increasing attenuation of the CARS level, the direct detection signal disappears in the detector noise whereas the heterodyne detection continues to lower levels before disappearing into the local oscillator noise. For detection by photodiode terminated with 10 MΩ, the minimum CARS signal that can be detected using the heterodyne detection is a factor 3500 lower than can be detected directly. Since the amount of local oscillator that we apply is such that the noise is limited by the shot noise (see Fig. 4) we detect the CARS signal with a precision limited only by the shot noise and the detection cannot be improved further. Detection of lower absolute signal levels is possible for longer integration times.

The PMT has a very low noise level and high gain so that the improvement by the heterodyne is not so impressive (only a factor of 7.5). However, the quantum efficiency of the PMT is so low that the minimum detectable signal is considerably higher than for the heterodyne detection using a photodiode. There are PMTs with better QE commercially available (QE up to 15% are quoted). For those PMTs the heterodyne detection minimum would approach the limit for the photodiode. Due to residual alignment errors this detected heterodyne signal is a factor 10 below the theoretical limit.
The noise-free heterodyne gain is achieved for a local oscillator level that is chosen such that the local oscillator noise is only slightly above the detector dark noise level. In practice this means that the levels of local oscillator required to lift the signal above the detector noise are always fairly small (nW) compared to the light levels driving the CARS (mW). The local oscillator can thus be sent collinear with the other beams through a sample without disturbing the CARS process, avoiding path length differences in scanning microscopy.

6. CONCLUSIONS

We have demonstrated greatly improved sensitivity through heterodyne detection. The levels of local oscillator that we employed are such that detection is possible at the shot noise level where the detector noise is no longer relevant. For detection with a photodiode this improved the detection sensitivity by orders of magnitude. For a PMT or APD the benefit from using heterodyne detection is lower since the detector noise levels are lower. The use of an APD in linear mode is inferior to the photodiode due to increased excess noise.\textsuperscript{23}

It is important to note at this point that the current detection limit for CARS signals is not caused by detector noise alone. For very small numbers of resonant molecules in the focal volume the non-resonant background can overwhelm the resonant signal, much like auto-fluorescence limits the detection of fluorescent probes.\textsuperscript{11} Potma et al show that since the signal at resonance is shifted in phase compared to the non-resonant signal, the interferometric signal can be used to separate the resonant part of the CARS signal from the non-resonant part. The non-resonant signal can be rejected if the phase of the local oscillator with respect to the CARS signal can be established. The (local) phase can e.g. be extracted from a microscopy image if it contains a non-resonant region and if the phase is sufficiently stable to record the image without drift. To reject the non-resonant part without knowledge of the phase and to avoid noise amplification of the non-resonant signal, the polarization directions of the incoming beams and the local oscillator can be adjusted so that only the resonant part of CARS signal is amplified by the optical interference. The overall loss of signal is compensated by the interferometric gain and the noise generated by the non-resonant part does not add to the noise in the resonant part. Such a combination of Heterodyne Interference and Polarization-CARS (HIPCARS) would be analogous to the HIPSOM setup demonstrated for linear scattering by metal nanoparticles in.\textsuperscript{24}

We believe that the ability to provide noise-free amplification makes this OPO preferable to previously published sources such as continuum from nonlinear fibers\textsuperscript{25,26} where the noise levels were not specified. Other approaches to heterodyne detection, such as the use of the non-resonant background as a local oscillator have also been reported.\textsuperscript{27,28} These approaches require broadband detection or spectral scanning (or temporal scanning for Fourier transform variations) to extract the oscillating part of the signal. One measurement yields a full spectrum which can be preferable in terms of identification but it generally requires many photons to build a spectrum that is sufficiently defined to extract lineshapes, which in turn requires high input powers or long measurement times. The jitter-free operation of this setup, the possibility to scan the wavelength without synchronization errors and the possibilities to distinguish resonant and non-resonant contributions will significantly enhance the applicability of CARS microscopy.

ACKNOWLEDGMENTS

We acknowledge E.O. Potma and C.L. Evans for useful discussions during the CARS workshop 2005 at the group of X.S. Xie at Harvard. This research is supported by NanoNed, a nanotechnology programme of the Dutch Ministry of Economic Affairs and partly financed by the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). We also acknowledge Coherent for the use of the Paladin laser.

REFERENCES


