Single photon multiclock lock-in detection by picosecond time stamping

Lukas A. Jakob,[1] William M. Deacon,[1] Oliver Hicks,[2] Ilya Manyakin,[1] Oluwafemi S. Ojambati,[1] Michael Traxler,[3] Jeremy J. Baumberg[1,*]

[1] *Nanophotonics Centre, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK*
[2] *Cambridge Consultants, Cambridge CB4 0DW, UK*
[3] *GSI Helmholtzzentrum für Schwerionenforschung GmbH, Planckstraße 1, 64291 Darmstadt, Germany*
*jjb12@cam.ac.uk

Abstract: Extracting signals at low single-photon count rates from large backgrounds is a challenge in many optical experiments and technologies. Here we demonstrate a single-photon lock-in detection scheme based on continuous photon time stamping to improve the signal-to-noise ratio by more than two orders of magnitude. Through time-resolving the signal modulation induced by periodic perturbations, 98% of dark counts are filtered out and the <1 count/s contributions from several different nonlinear processes identified. As a proof-of-concept, coherent anti-Stokes Raman measurements are used to determine the vibrational lifetime of few molecules in a plasmonic nanocavity. This detection scheme can be applied to all single-photon counting experiments with any number of simultaneous modulation frequencies, greatly increasing signal-to-noise and resolving physical processes with picosecond time resolution while keeping photon dosage small. Open instrumentation provided here enables low-cost implementation.

1.   Introduction

Transient or time-resolved optical experiments often require elaborate experimental setups designed to measure very low signal intensities over ultrashort time scales, and hence they often suffer from poor signal-to-noise ratios [1]. Typically, the linear response dominates over any perturbation, giving small induced changes in the signal. Increasing the strength of a repetitive perturbation (for instance the optical pulse intensity) to enhance the nonlinear signal is often problematic since this can damage the samples, preventing stroboscopic measurement. For probing single nanostructures or individual quantum systems, the strong perturbation needed to obtain clear signals unfortunately often induces irrevocable structural changes such as bond cleavage [2], atomic displacements [3], reshaping [4] or ablation [5].

A technique commonly applied to extract such weak signals from a noisy background is lock-in detection. By introducing a modulation to the sample, the amplitude and phase of the emerging signal can be determined using phase-sensitive heterodyne detection while noise at other frequencies is rejected [6]. For instance, in all-optical experiments such as four-wave-mixing in semiconductor optical amplifiers [7] or stimulated emission from single nanocrystals [8], the pump pulse train is amplitude-modulated at high frequency f_{mod}. In scanning near-field optical microscopies, this modulation is provided by the vibration of a tip above the sample [9]. Optical lock-in detection is also used for stimulated-emission-depletion microscopy to enhance contrast in fluorescence microscopy [10], and in many other scenarios.
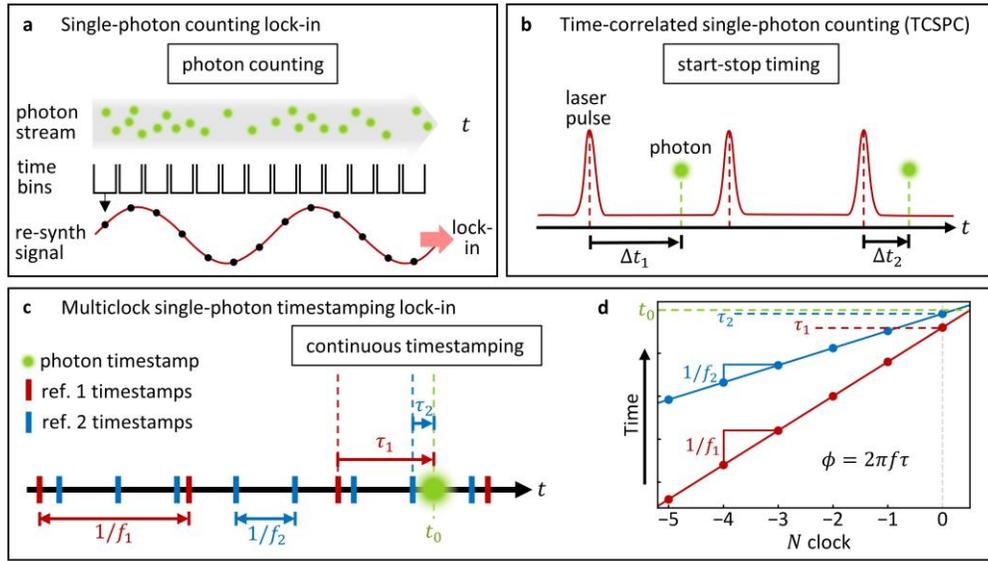
Fig. 1. Comparison of single-photon detection techniques. (a) Previous single-photon lock-in techniques: photons are counted in time bins to recover an analogue signal for analysis with a standard phase-sensitive lock-in algorithm. (b) Time-correlated single-photon counting: the time interval Δt between a trigger pulse and photon detection events is recorded to resolve the time decay of an optical signal, typically fluorescence. (c,d) Multiclock single-photon lock-in technique developed here: Continuous timestamping of photon arrival times (t_0) and multiple reference signals with frequency f_i allows identification of the phase τ_i of every individual photon compared to all the reference frequencies.

Several versions of phase-sensitive detection have been implemented for single-photon counting detectors [11–16]. However, these rely on binning photon counts in successive time intervals to create a continuous signal for traditional lock-in analysis instead of analysing each photon individually (Fig. 1a). The detection challenge is to identify ~ 1 cts/s (counts per second) caused by the linear or nonlinear optical processes of interest under a large background. Here, we present a new approach to single-photon lock-in detection utilising continuous picosecond photon time stamping of each individual photon to temporally resolve the evolution of an optical signal undergoing modulation. By recording the arrival time of each signal photon at the detector alongside time stamps from synchronised reference clocks (Fig. 1c), the time dynamics of the signal is extracted. For each photon, the phase (τ_i) and frequency (f_i) of each reference clock at the photon arrival time (t_0) is accurately determined from a linear fit to the timestamp data of the previous N clock detections (Fig. 1d). In the post-experimental data analysis, experimental noise and background signals can be easily removed through temporal gating, implementing lock-in amplification individually for every photon and each reference clock suitable for increasing the signal-to-noise ratio of single-photon counting experiments by many orders of magnitude (here with factor >100).

The signal-processing technology to convert the arrival time of an electronic pulse to a digital time stamp is well established and widely used in high-energy physics [17,18]. Field programmable gate array (FPGA) boards are capable of performing logical operations with digital electronic signals in real time with bandwidths exceeding 100 MHz. Time-to-digital converters (TDC) implemented with FPGAs can now achieve timing precision < 10 ps [19] while it is possible to parallelise devices using 264 channels or more on one board [20].

Combining picosecond time stamping with pulsed optics holds enormous potential to improve existing and enable new applications. Although the proposed experimental design can increase signal-to-noise in all single-photon experiments, it is particularly advantageous in the areas of quantum correlation, time-of-flight spectroscopy, and scanning near field microscopy.

A particular new capability provided by continuous photon time-stamping is the ability to compare photon signals to multiple reference clocks at the same time. For instance here this allows single-photon lock-in synchronisation simultaneously to: (1) the optical pulse repetition rate, (2) the laser power modulation (here two different on-off periodic modulations), and (3) extra triggers (here pulse delay scan). This retrieves the maximum possible information content of each detected photon. In comparison, start-stop photon detection schemes for time-correlated signal photon counting (TCSPC) synchronise only to optical pulses, greatly limiting their use (Fig. 1b) [21]. We also note some similarities to LIDAR technology where multiple laser repetition rates are used to avoid distance ambiguity. However, these measurements are performed in quick succession, while here we resolve a signal modulating with several frequencies at once in our setup.

In this paper, coherent anti-Stokes Raman spectroscopy (CARS) is used as an example of a typical nonlinear experiment with low single-photon count rates that benefits from our scheme [22,23]. We thus briefly describe the measurements, while noting the general applicability of the technique. In CARS, molecular vibrations ν are excited by coherently pumping with two laser pulses (pump ω_p and Stokes ω_S, Fig. 2a,b), tuned so their frequency difference matches the vibration, $\omega_p - \omega_S = \nu$. Subsequently anti-Stokes scattering to $\omega_x + \nu$ of a time-delayed separate probe pulse at ω_x is used to identify the relaxing molecular vibrations and measure their vibrational lifetime. To access a domain capable of observing CARS from single molecules [22], nanoscale optical confinement is needed which is accessed here using plasmonic nanocavities [24]. These are based on a nanoparticle-on-mirror geometry (NPoM) in which an ordered molecular monolayer (here biphenyl-4-thiol, BPT) is sandwiched in a nm-thick gap between Au facets [25]. A background of non-resonant four-wave-mixing dominates the detected signal at $\omega_x + \nu$ [3,26] while additional noise is introduced by electronic dark counts and stray light. Our detection scheme here isolates < 1 cts/s rates of this nonlinear CARS signal using picosecond photon time stamping through a low-cost open-architecture FPGA board.
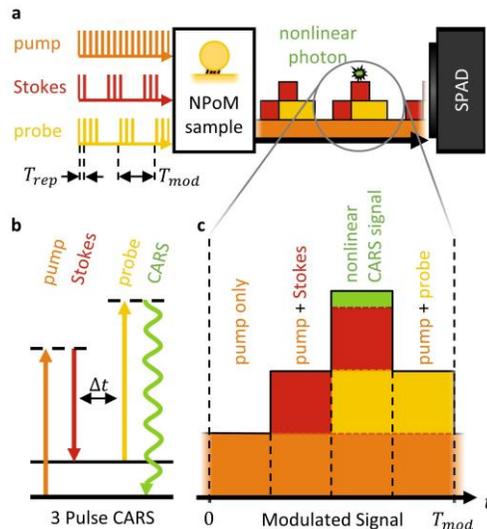

Fig. 2. (a) Schematic experimental setup of coherent anti-Stokes Raman spectroscopy (CARS). Three pulsed lasers with repetition period T_{rep} are focused on a nanoparticle-on-mirror (NPoM) nanocavity sample, with Stokes and probe modulated periodically with period T_{mod}. The emerging signal detected by a single-photon detector (SPAD) contains photons from the nonlinear CARS process, which need to be isolated from other contributions. (b) Excitation scheme of CARS, with pump and Stokes exciting a molecular vibration, probed by anti-Stokes scattering of the probe beam with time delay Δt. (c) Modulation of the detected SPAD signal. Only when all three beams are turned on can the nonlinear 3-pulse CARS signal be observed (green).

2. Experimental setup

The basis for detecting a nonlinear optical signal above other contributions is to identify and subtract the linear components and separate the multitude of other nonlinear signals. Here, this is achieved by modulating two laser beams exciting the sample (Fig. 2a). For CARS experiments, each of the three laser beams contributes to the total recorded count rate (Fig. 2c). Pump and probe pulses generate broadband contributions across the detected spectral window through electronic anti-Stokes Raman scattering. With pump and Stokes beams exciting the sample, a 2-pulse CARS process is also possible but cannot be time-resolved since the pump then both excites and probes. Only all three pulses together excite emission of the desired 3-pulse CARS signal, enabling investigation of the ultrafast vibrational dynamics. In systems with high signal and stability, these contributions can be separated by sequential acquisition of spectra without fast laser modulation. However with a fixed photon budget on samples such as nanoscopic structures and single molecules, a new approach is required to increase signal-to-noise and avoid damage.

A widefield microscope guides these pulses onto the nanocavity sample (Fig. 3), which are generated at 820 nm (Stokes), 726 nm (pump) and 722 nm (probe) from an f_{rep} = 80 MHz pumped optical parametric oscillator (Spectra Physics OPO). The three lasers are spectrally tuned so that pump and Stokes pulses can resonantly drive the molecular vibration while the Stokes pulse is off-resonant to vibrations excited by the probe pulse (see supplementary Fig. S2a). With ultrafast pulses of 500 fs duration, the spectral resolution of the exemplar experiment here is 50 cm^{-1}. Stokes and probe beams are each modulated with electro-optic modulators (EOM) driven by function generators (HP 33120A) producing square wave output at variable frequency f_{mod}. The two function generators are phase locked and operate with a fixed phase difference of 90°. For CARS experiments, probe beam pulses are delayed by Δt with respect to the other two laser beams using a delay stage running in a continuous loop around $\Delta t = 0$ (f_{delay} = 1 Hz). The Stokes pulses are temporally aligned to the pump pulses (by optimising the 2-pulse CARS signal) initially, after which this zero relative delay is fixed. All three beams are spatially overlapped and co-focussed on the sample.
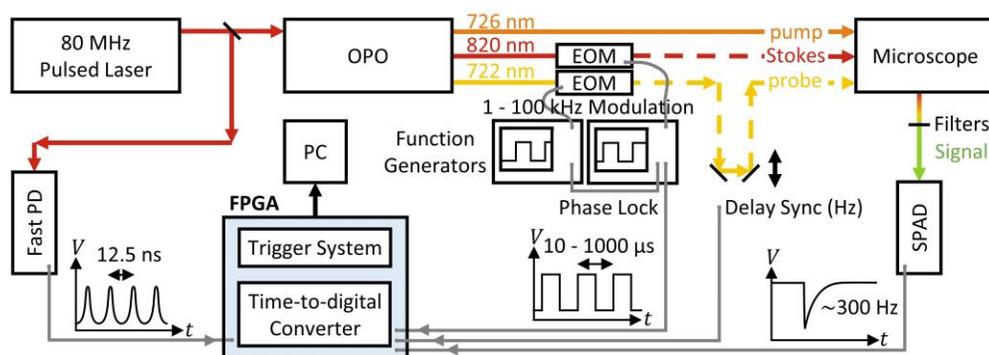
Fig. 3. Experimental setup for CARS experiments with photon time tagging. Three laser beams are produced by an 80 MHz pulsed pump laser and optical parametric oscillator (OPO). Two of these beams are 50 kHz modulated by electro-optic modulators (EOM) driven with square waves. The probe beam delay is continuously varied by moving a mirror, giving trigger synchronization pulses at the beginning of each loop. All beams are spatially overlapped and focused onto the sample in a microscope. A single-photon avalanche diode (SPAD) detects the spectrally filtered signal photons. A reference signal for the pulsed laser is provided by a fast photodiode (PD). The electronic signals (V-t diagrams) are passed to a central FPGA board, where the signal and clock timestamps are recorded digitally.

151 In the detection path, the laser light is blocked using spectral filters and a single-photon
152 avalanche diode (SPAD, Micro Photon Devices PDM $PD-100-CTD) detects the signal
153 photons. Each arriving photon leads to an electronic pulse with timing accuracy of 35 ps.
154 Typical count rates in our experiment range from 1 to 1000 cts/s obscured by at least 100 cts/s
155 of dark counts and stray light.
156   In addition to the single-photon counts, three reference signals are recorded. A fast
157 photodiode monitors the pulse repetition rate of the Ti:Sapphire pump laser (Spectra Physics
158 MaiTai, photodiode integrated into laser head) producing a digitally-conditioned f_{rep} =
159 80 MHz pulse train. Together with a synchronised TTL square-wave at f_{mod} and TTL trigger
160 pulses at the beginning of each delay stage loop (f_{delay}), all electronic signals are passed to the
161 FPGA board (Fig. 3). This board continuously converts the arrival times of all electronic
162 signals to digital timestamps that are streamed to a computer and saved. This time-to-digital
163 conversion is achieved by combining a fine time-to-digital converter (TDC) and a 32-bit coarse
164 counter (Fig. 4a). The fine TDC consists of a tapped delay line that provides 30 ps accuracy
165 with a range of 5 ns. The coarse counter simply counts increments of the 200 MHz internal
166 FPGA clock and hence covers the nanosecond to second regime.
167
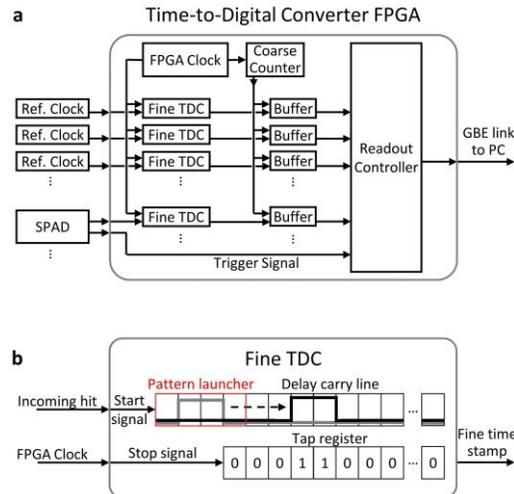
168
169 Fig. 4. Time-to-digital conversion with an FPGA. (a) Time stamps with ps precision are recorded by combining a fine
170 time-to-digital converter (TDC) and a coarse counter. Data is streamed to the computer via gigabit Ethernet (GbE)
171 when triggered by a signal photon. (b) In the fine TDC, an electronic signal is carried along a tapped delay line
172 connected to a tap register. When the signal is stopped by the FPGA clock, the register entry is recorded.

173

174 In a tapped delay line (as commonly applied in time-to-digital conversion [18]), each element
175 is connected to a register that records the state of the carried signal and hence the current
176 position of the signal within the carry line (Fig. 4b). To determine the photon arrival time, an
177 incoming electronic signal launches a pattern that travels along the delay line until it receives a
178 stop signal which causes the position of the signal to be read from the register. Since the stop
179 signal is provided by the FPGA clock, the arrival time of the signal is the time the signal spent
180 propagating along the delay line before the latest clock. The propagation distance is converted
181 to time in post-experimental data processing.
182   Since the laser pulse reference clock is at 80 MHz, recording every pulse arrival time would
183 require a data transfer rate >25 GB/s. Therefore, a trigger system was implemented to only
184 record reference time stamps when a signal photon is detected: every time a photon arrives at
185 the SPAD, a defined (and tuneable) number of most recent timestamps from all reference clocks

186 are sent to the host PC by the readout controller alongside the photon timestamp. This allows
187 the user to control the number of reference timestamps recorded for each signal event, and thus
188 optimise signal-to-noise by maximising the accuracy of the laser reference given the maximum
189 data transfer rate at each SPAD count rate. As we show below, even a stable repetitive laser
190 system experiences variations of <0.1% in cavity length, which if not tracked greatly reduce
191 the timing precision. Here, ten timestamps per reference clock sufficed to reach the optimum
192 time resolution and thus minimise the required data size.
193   The TDC was implemented on a Digilent Arty Z7 development board specifically
194 programmed for this application. We provide the FPGA software online [will be provided on
195 acceptance of the paper] under an open source licence to allow for low-cost implementation.
196 Here, we demonstrate the function of this system with three reference clocks and one signal
197 channel. However, the FPGA board supports up to eight input signals which can be assigned to
198 record either reference or signal channels, allowing for more complex experimental setups.
199
200

3. Results

202 With this setup, data files are acquired containing a list of arrival timestamps. For the first
203 demonstrations here, two reference clock signals are chosen, f_{rep} and f_{mod}. In general, any
204 periodic reference signal can be chosen as a clock enabling a plethora of different applications.
205 As an initial calibration experiment, a single laser pump beam modulated at $f_{mod} = 1$ kHz is
206 focused on the sample and spontaneous Stokes scattering from the nanocavity is recorded by
207 the SPAD (see Fig. S1 for spectrum detected here).
208
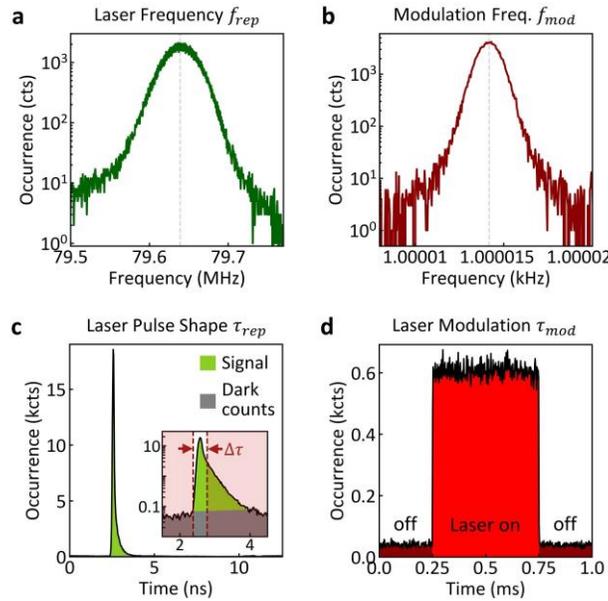
Fig. 5. Analysis of timestamp data from spontaneous Raman scattering. (a,b) Histograms of the clock frequency f for (a) the laser pulses and (b) the EOM modulation, over 100 s. (c) Reconstruction of the laser pulse shape discriminating signal (green) and dark counts (grey). Inset: 0.2 ns time window $\Delta\tau$ including 60% of signal counts (dashed). Bin size 25 ps. (d) Reconstruction of the laser modulation cycle, recording Raman scattering from a nanocavity with single laser pump turned off and on (laser power 2 µW). Bin size 2 µs.

216 From the periodic clock timestamps, the frequency f of the clock at the time of arrival of each
217 photon at the detector is determined. For each of the detected photons, the clock frequency
218 slightly varies due to fluctuations in the period seen in the histograms of Fig. 5a,b. Over this
219 100 s data set, the average pulse repetition rate of the laser was f_{rep} = 79.64 MHz with a
220 standard deviation of 25 kHz. For the EOM modulation of the laser, a frequency of 1.00001
221 kHz is found $\pm$ 0.78 mHz as expected for this source. These vary little between successive data
222 sets, demonstrating the accuracy of our global clock.
223   The point in the clock cycle τ when each signal photon was detected is now calculated as
224 well. Extracting this value for all photons detected during a measurement allows us to
225 reconstruct the signal modulation through the clock cycle, of length set by the inverse of the
226 clock frequency. This projects all counts into a single clock cycle and hence repetitive but
227 extremely low count rates can be analysed by simply increasing the integration time.
228   We first apply this temporal reconstruction to the periodically pulsed laser. Using the laser
229 repetition rate to wrap the signal within the T_{rep} = 12.5 ns long clock period places most counts
230 near a specific time which depends on a system electronic delay from optical paths, cables and
231 latency (Fig. 5c). All photons emerging from the sample due to excitation by the periodic laser
232 pulses are detected within a $\Delta\tau \sim 200$ ps wide window, because Raman is a prompt process.
233 Even though the photons are emitted within the 500 fs optical pulse width, this peak is
234 broadened by the detection electronics. In contrast, electronic dark counts and stray light
235 photons are uncorrelated with the excitation pulses and hence give a constant background signal
236 spread over the whole period. Exploiting this property of dark counts allows us to distinguish
237 them from photons emitted by the sample and remove them from the data. This results in a
238 >98% reduction in dark counts and thus an increase of $T_{rep}/\Delta\tau > 60$ in signal-to-noise for low
239 count rate experiments (Fig. 5c). The approach to remove dark counts in the time domain is
240 similar to previous reports of time-gated single-photon counting which has been demonstrated
241 both with actively quenched detectors [27–29] and digitally by TCSPC [30].
242   The same concept of temporal reconstruction is now applied to determine which lasers are
243 exciting the sample when each signal photon was generated. This is crucial for separating the
244 linear and nonlinear components of the signal. To demonstrate this functionality in our
245 calibration case, the single laser pump is switched on and off at frequency 1 kHz and 50% duty
246 cycle while recording Raman scattering from the nanocavity sample (average laser power on
247 sample 2 µW). Consequently during the laser modulation cycle, two temporal regions with
248 constant count rates are observed, with the background for laser-off being simply identified
249 (Fig. 5d).
250   When the power of the laser is reduced by near hundred-fold (50 nW on sample), the
251 modulation depth of the signal is greatly reduced since the dark count rate is now larger than
252 the Raman signal (Fig. 6a). Dark counts outside the 200 ps window centred on the laser pulse
253 in Fig. 5c are now removed, reducing the count rate outside the laser window close to zero (Fig.
254 6b) with only a residual 2% unfiltered dark counts remaining. This dark count rejection
255 increases the modulation visibility by 400%, but in measurements with even lower laser power
256 or more stray light/dark counts, this enhancement can exceed 4000%. Analysing the
257 distribution of the noise in the modulation reveals two Poisson distributions for the laser on and
258 off states (Fig. 6c). These arise from the single-photon counting statistics in the experiment and
259 show that the signal-to-noise is now only limited by photon shot noise, which can thus be
260 improved by increasing the integration time to collect more photons.
261
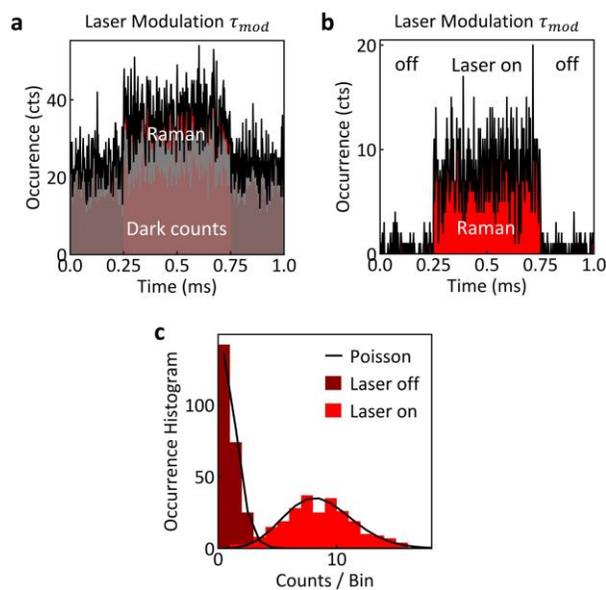
262

263 Fig. 6. Demonstration of dark count removal. (a) Reconstruction of laser-modulated spontaneous Raman scattering
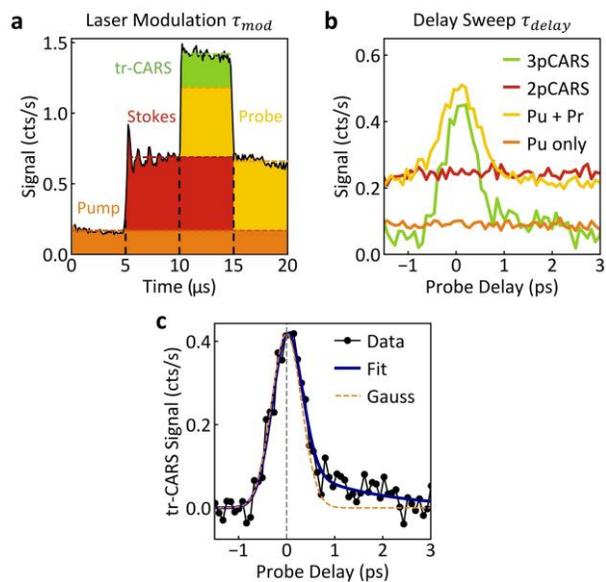264 from a nanocavity at low average laser power (50 nW) over 100 s, with Raman signal smaller than the dark count rate.
265 (b) Reconstruction of modulated signal after dark count removal. Bin size for (b,c) 2 µs. (c) Distribution of noise in
266 (b), with fitted Poisson distributions (lines) indicating the single-photon statistics of the experiment.

267

268

269

270 Fig. 7. Time-resolved CARS experiments. (a) CARS signal during 50 kHz laser modulation period (dark counts
271 removed). Contributions from the different lasers are marked in orange (pump), red (Stokes) and yellow (probe), with
272 the nonlinear signal CARS signal in green. Bin size 40 ns. (b) Extracted contributions from (a) locked-in to the delay
273 stage sweep as the probe (Pr) pulses are delayed with respect to pump (Pu) and Stokes pulses. (c) Time-resolved CARS
274 signal with exponential decay fit to extract a vibrational lifetime of 1595 fs of the BPT molecules in the nanocavity.
275 For negative delays, the signal follows a Gaussian matching the 500 fs pulse width. Average laser power on sample
276 per beam is 0.2 µW.

277

278 In order to demonstrate the ability of the setup to detect a small nonlinear signal upon a large
279 background, CARS experiments are then carried out. The detected CARS spectra are shown in
280 supplementary Fig. S2b. Since two beams are now modulated, each at $f_{\mathrm{mod}} = 50$ kHz (to
281 improve signal-to-noise) but with one phase shifted by 90°, with addition of a third beam of
282 constant intensity, the signal modulation shows four distinct windows of different height (Fig.
283 7a) as expected from the experimental design (Fig. 2c). Dark counts in the CARS experiments
284 throughout the entire modulation period were removed as detailed above. The measured count
285 rate is highest when all three beams illuminate the sample and lowest when only the pump beam
286 is on. Analysing the different windows allows the contributions from each individual laser
287 combination to be determined (yellow, orange and red in Fig. 7a). Subtracting these values
288 from the count rate when all lasers illuminate the sample allows the nonlinear 3-pulse CARS
289 count rate to be extracted (green in Fig. 7a).
290   With this method to extract the nonlinear signal, time-resolved CARS measurements can
291 now be performed by delaying the probe pulse compared to pump and Stokes pulses. This is
292 achieved by mechanically scanning a delay line for the probe pulse. Conventionally slow scans
293 are performed, integrating until sufficient signal-to-noise is achieved at each time point,
294 however this produces strong artefacts in the delay scan due to transient changes in the emission
295 spectrum (caused by movement of Au atoms on the nanoparticle facet [31,32]) and damage to
296 the nanostructure. Hence, the delay stage is continuously scanned back and forth at $f_{\mathrm{delay}} =$
297 1 Hz and a third reference trigger is introduced into the FPGA from the scanning delay line.
298 This allows each photon detected to also be tagged with the time-delay at which it was
299 measured, thus building up the entire time-delay curve simultaneously, without any artefacts.
300   Time-resolved tracks for all contributions identified by the laser modulation are compared
301 in Fig. 7b. The 2-pulse CARS signal induced by pump and Stokes beams (red) stays constant
302 as well as the electronic Raman scattering from the pump (orange). While electronic Raman
303 scattering from the probe also leads to a constant signal, vibrational pumping by surface-
304 enhanced Stokes scattering of pump and probe photons adds a time-dependent signal to the
305 yellow contribution in Fig. 7a. This signal decays equally to both positive and negative delays
306 as pump or probe excite the sample identically. On the other hand, the 3-pulse CARS count rate
307 decreases for both positive and negative delay, but is not symmetric around 0 ps (Fig. 7c). When
308 the probe pulse arrives before the molecules are excited, the signal vanishes quickly with a rise
309 time corresponding to the pulse length of 500 fs (orange dashed, Fig. 7c). For probe pulses
310 arriving after pump and Stokes, the signal decreases exponentially. From an exponential fit, the
311 lifetime of the 1585 cm^{-1} vibration of BPT is estimated as 1595$\pm$420 fs. This signal is emitted
312 from only an estimated 100 molecules in the nanocavity gap, billions of times fewer than for
313 solution measurements. Previous attempts to measure this without single-photon detection
314 required average laser powers >4 µW per beam, at least ten-fold more than in this photon-
315 counting lock-in mode, and which is enough to perturb and destroy the nanocavity structures.
316 With safe powers of $I=0.2$ µW per beam employed here, the CARS signal is 1000-times smaller
317 (since it scales as I^3), below the limits of integrating detectors. As the signal strongly varies
318 from nanoparticle to nanoparticle, further ongoing experiments and theory are required for full
319 analysis, but beyond the scope of this article.

4. Discussion

To quantify the improvements made by eliminating dark counts, the window (within which counts are not removed) is centred on the pulse and increased in width (Fig. 5a inset). For each width, the percentage of dark counts that are removed is calculated as well as the photon events preserved (Fig. 8a). A window width of $\Delta\tau = 200$ ps retains ~60% of real photon events and only 1.6% of all dark counts. To remove a higher percentage of dark counts, the window width has to be reduced thus decreasing the signal counts. The ratio of signal counts to dark counts (Fig. 8b) shows that by filtering out dark counts, this ratio can be increased by almost two orders of magnitude from 1.3 in the unfiltered data to >100. Since a compromise between best signal-to-noise and maximum retained signal counts is demanded, we thus choose a window width of 200 ps which retains signal counts most efficiently (before green curve in Fig. 8a saturates) that increases the signal-to-noise by 60.
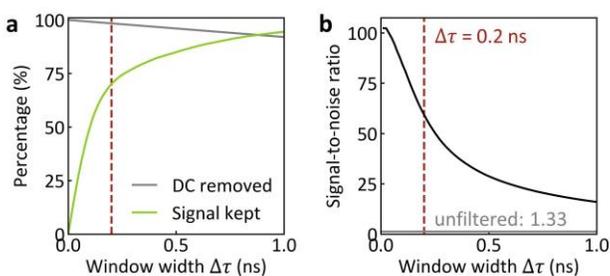
Fig. 8. Characterisation of dark count removal while varying filter window width $\Delta\tau$ (see Fig. 5c). (a) Percentage of signal counts preserved and dark counts removed for increasing window widths. (b) Ratio of signal counts to dark counts, compared to unfiltered data (horizontal line at 1.3).

Currently, limitations of the timing precision in this setup are observed in the distorted pulse shape. Even though the laser pulses are < 1 ps, the reconstructed pulse shape (Fig. 5c) has a width of ~200 ps. The SPAD has a nominal jitter of only 35 ps and the FPGA board has a timing precision of 30 ps (see Fig. S3), but further inaccuracies are introduced by the fast photodiode (internal in Spectra Physics Maitai pump laser) and amplification of the MHz clock signal by two amplifiers (Mini-Circuits ZFL-1000LN). Additionally, the detector electronic response adds a shoulder to the peak, thus decreasing the percentage of preserved signal counts after filtering out dark counts. Improvements in the detection electronics can thus further increase the fraction of dark counts removed by this technique. For instance, without FPGA noise and reaching the SPAD resolution of 35 ps would give another ~10-fold improvement.

We presented a scheme here of modulated lasers for the example of CARS. Typically laser modulation and lock-in detection is also used for stimulated Raman scattering (SRS). Since SRS requires detection of small changes in the pump laser intensity, the detected powers are far above the single-photon regime. In an optimised configuration, our FPGA setup can only record up to 1 Mcts/s count rates (pW) and is therefore not suitable for SRS without upgrading electronics to deal with much higher count rates.

To highlight the potential of the presented technology, Fig. 9 compares different photon detection techniques. Depending on the photon count rates, different photodetectors must be selected. For single-photon experiments, SPADs are suitable while traditional photodiodes (PD) are needed for spectroscopies delivering higher light intensities (> nW). In between, avalanche photodiodes (APD) and photomultiplier tubes (PMT) provide detection of photocurrents with high gain.

The crucial experiment signal-to-noise ratio strongly depends on signal detection and amplification. A SPAD delivers one voltage pulse for every detected photon and thus the only

362 source of noise is photon shot noise and dark events. For photodiodes, where an electrical
363 current induced by light is produced, additional noise sources include the electronic shot noise
364 due to the diode dark current and thermal detector noise as well as noise from amplifiers
365 necessary to record the small output currents.

366

367
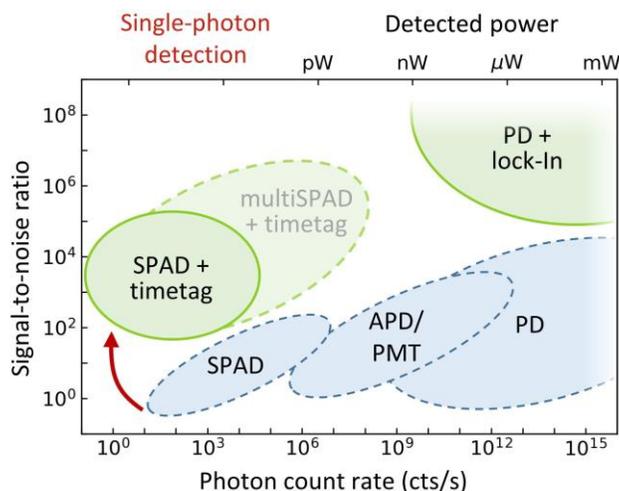
Fig. 9. Comparison of different photon detection techniques. Single-photon avalanche diodes (SPAD), avalanche photo
diodes (APD), and photomultiplier tubes (PMT) can detect single-photon count rates whereas photodiodes (PD) operate
at higher light intensities. The signal-to-noise ratio of the detectors can either be improved by lock-in amplifiers in the
high signal regime, or with the photon time stamping setup (SPAD + Timetag) presented here for single-photon
detection, which can be extended to higher count rates using a combination of several SPADs (multiSPAD + Timetag).
A detailed description of the comparison is provided in the SI.

374

375 For photodiodes, a typical way to improve this signal-to-noise is to modulate the excitation
376 light source (or sample perturbation) at a fixed frequency and use a lock-in amplifier to record
377 the amplitude of the modulated detected photocurrent. Analogously, our FPGA system can
378 enhance the signal-to-noise ratio in single-photon experiments by more than two orders of
379 magnitudes. By resolving this modulation of the signal, it is possible to filter out unwanted
380 background photons and those from other contributions. In the experiment presented here, we
381 demonstrate reliable detection of 3-pulse CARS at count rates of 1 cts/s within a background
382 of more than 300 cts/s. This ratio of $>10^2$ is remarkable for a single-photon experiment, but can
383 increase even further for applications with high backgrounds such as stray light.
384   Improving the signal-to-noise even further would be possible by enhancing the timing
385 precision to narrow the electronic pulse as discussed above. Increasing the detector count rates
386 beyond the saturation of a single SPAD can be handled by splitting the light intensity over
387 multiple SPADs and connecting them to different channels of the FPGA. In combination with
388 a spectrometer grating, a SPAD array could even then resolve the spectral dependence of the
389 signal.
390   Here, we have demonstrated the working principle of our setup using laser pulses, laser
391 modulation and the delay stage sweep as reference clocks. However, any periodic signal can
392 act as a reference, making our setup attractive for a wide range of experimental research fields.
393 In particular the additional time resolution of the signal during a reference period can enable a
394 plethora of new applications. For example, the light emission from an optoelectronic device
395 induced by an alternating voltage can be tracked in time to characterize the response time of
396 the device. In scanning near-field optical microscopy, an oscillating tip above the sample can
397 provide a reference frequency, both drastically increasing the signal-to-noise and recording the

398 signal as a function of tip-sample distance. Moreover with a dispersive fibre, photons can be
399 delayed depending on their colour enabling optical time-of-flight spectroscopy [33] with the
400 80 MHz lock-in frequency. Finally, the switching of photoactive molecules can be resolved in
401 time with modulated lasers. We suggest this is of particular interest for single-molecule
402 fluorescence spectroscopy and microscopy.

403 In conclusion, we developed a method to separate different contributions to a signal by
404 resolving the variation of the detected single-photons over the period of a reference signal. The
405 technique relies on continuously recording the arrival time of each photon at a single-photon
406 detector with an FPGA board and comparing it to reference clock time stamps. With this setup,
407 we reconstructed the periodic modulation of a sub-ps excitation laser allowing 98% of stray
408 light and dark counts to be filtered out. The capability of this method was demonstrated in a
409 CARS experiment, where single-photon per second count rates of a nonlinear signal were
410 detected. Due to the high flexibility for different reference clock signals from Hz to MHz
411 frequencies this concept can be applied universally to all single-photon experiments, drastically
412 increasing their signal-to-noise.

413

420 **Disclosures.** The authors declare no conflict of interest.

421 **Data availability.** Data underlying the results presented in this paper are available at DOI: (will be provided on
422 acceptance of the paper).

423 **Supplemental document.** See Supplement 1 for supporting content.

424

References

426 1. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, 1984).
427 2. A. Lombardi, M. K. Schmidt, L. Weller, W. M. Deacon, F. Benz, B. De Nijs, J. Aizpurua, and J. J.
428   Baumberg, "Pulsed Molecular Optomechanics in Plasmonic Nanocavities: From Nonlinear Vibrational
429   Instabilities to Bond-Breaking," Phys. Rev. X **8**, 1–17 (2018).
430 3. K. T. Crampton, A. Zeytunyan, A. S. Fast, F. T. Ladani, A. Alfonso-Garcia, M. Banik, S. Yampolsky, D.
431   A. Fishman, E. O. Potma, and V. A. Apkarian, "Ultrafast Coherent Raman Scattering at Plasmonic
432   Nanojunctions," J. Phys. Chem. C **120**, 20943–20953 (2016).
433 4. C. J. DeSantis, D. Huang, H. Zhang, N. J. Hogan, H. Zhao, Y. Zhang, A. Manjavacas, Y. Zhang, W. S.
434   Chang, P. Nordlander, S. Link, and N. J. Halas, "Laser-Induced Spectral Hole-Burning through a
435   Broadband Distribution of Au Nanorods," J. Phys. Chem. C **120**, 20518–20524 (2016).
436 5. M. Huang, F. Zhao, Y. Cheng, N. Xu, and Z. Xu, "Origin of laser-induced near-subwavelength ripples:
437   Interference between surface plasmons and incident laser," ACS Nano **3**, 4062–4070 (2009).
438 6. M. L. Meade, *Lock-in Amplifiers: Principles and Applications* (IEE Electrical Measurement Series, 1983).
439 7. P. Borri, J. Mørk, and J. M. Hvam, "Heterodyne pump-probe and four-wave mixing in semiconductor
440   optical amplifiers using balanced lock-in detection," Opt. Commun. **169**, 317–324 (1999).
441 8. L. Piatkowski, N. Accanto, G. Calbris, S. Christodoulou, I. Moreels, and N. F. van Hulst, "Ultrafast
442   stimulated emission microscopy of single nanocrystals," Science **366**, 1240–1243 (2019).
443 9. Y. Inouye and S. Kawata, "Near-field scanning optical microscope with a metallic probe tip," Opt. Lett. **19**,
444   159–161 (1994).
445 10. G. Marriott, S. Mao, T. Sakata, J. Ran, D. K. Jackson, C. Petchprayoon, T. J. Gomez, E. Warp, O.
446   Tulyathan, H. L. Aaron, E. Y. Isacoff, and Y. Yan, "Optical lock-in detection imaging microscopy for
447   contrast-enhanced imaging in living cells," Proc. Natl. Acad. Sci. U. S. A. **105**, 17789–17794 (2008).
448 11. A. Lenhard, S. Zaske, and C. Becher, "Lock-in detection of single photons after two-step frequency
449   conversion," Opt. Lett. **37**, 4254–4256 (2012).
450 12. F. T. Arecchi, E. Gatti, and A. Sona, "Measurement of Low Light Intensities by Synchronous Single
451   Photon Counting," Rev. Sci. Instrum. **37**, 942–948 (1966).

13. D. Braun and A. Libchaber, "Computer-based photon-counting lock-in for phase detection at the shot-noise limit," Opt. Lett. **27**, 1418–1420 (2002).

14. A. Restelli, R. Abbiati, and A. Geraci, "Digital field programmable gate array-based lock-in amplifier for high-performance photon counting applications," Rev. Sci. Instrum. **76**, 093112 (2005).

15. X. Wang, G. Zhang, R. Chen, Y. Gao, L. Xiao, and S. Jia, "Detection efficiency enhancement of single-photon detector at 1.55-μm by using of single photons lock-in and optimal threshold," Opt. Laser Technol. **44**, 1773–1775 (2012).

16. W. Chen, X. Wang, B. Wang, Y. Wang, Y. Zhang, H. Zhao, and F. Gao, "Lock-in-photon-counting-based highly-sensitive and large-dynamic imaging system for diffuse optical tomography," Biomed. Opt. Express **7**, 499–511 (2016).

17. J. Wu, Z. Shi, and I. Y. Wang, "Firmware-only Implementation of Time-to-Digital Converter (TDC) in Field-Programmable Gate Array (FPGA)," Proc. IEEE Nucl. Sci. Symp. Conf. Rec. **1**, 177–181 (2003).

18. S. Tancock, E. Arabul, and N. Dahnoun, "A Review of New Time-to-Digital Conversion Techniques," IEEE Trans. Instrum. Meas. **68**, 3406–3417 (2019).

19. E. Bayer and M. Traxler, "A High-Resolution (<10ps RMS) 48-ChannelTime-to-Digital Converter (TDC) Implementedin a Field Programmable Gate Array (FPGA)," IEEE Trans. Nucl. Sci. **58**, 1574–1552 (2011).

20. A. Neiser, J. Adamczewski-Musch, M. Hoek, W. Koenig, G. Korcyl, S. Linev, L. Maier, J. Michel, M. Palka, M. Penschuck, M. Traxler, C. Uğur, and A. Zink, "TRB3: a 264 channel high precision TDC platform and its applications," J. Instrum. **8**, C112043 (2013).

21. D. O'Connor, *Time-Correlated Single Photon Counting* (Academic Press, 2012).

22. S. Yampolsky, D. A. Fishman, S. Dey, E. Hulkko, M. Banik, E. O. Potma, and V. A. Apkarian, "Seeing a single molecule vibrate through time-resolved coherent anti-Stokes Raman scattering," Nat. Photonics **8**, 650–656 (2014).

23. T. Ichimura, N. Hayazawa, M. Hashimoto, Y. Inouye, and S. Kawata, "Tip-enhanced coherent anti-Stokes Raman scattering for vibrational nanoimaging," Phys. Rev. Lett. **92**, 220801 (2004).

24. J. J. Baumberg, J. Aizpurua, M. H. Mikkelsen, and D. R. Smith, "Extreme nanophotonics from ultrathin metallic gaps," Nat. Mater. **18**, 668–678 (2019).

25. F. Benz, C. Tserkezis, L. O. Herrmann, B. De Nijs, A. Sanders, D. O. Sigle, L. Pukenas, S. D. Evans, J. Aizpurua, and J. J. Baumberg, "Nanooptics of Molecular-Shunted Plasmonic Nanojunctions," Nano Lett. **15**, 669–674 (2015).

26. M. Danckwerts and L. Novotny, "Optical frequency mixing at coupled gold nanoparticles," Phys. Rev. Lett. **98**, 026104 (2007).

27. A. Tosi, A. D. Mora, F. Zappa, A. Gulinatti, D. Contini, A. Pifferi, L. Spinelli, A. Torricelli, and R. Cubeddu, "Fast-gated single-photon counting technique widens dynamic range and speeds up acquisition time in time-resolved measurements," Opt. Express **19**, 10735 (2011).

28. M. Buttafava, G. Boso, A. Ruggeri, A. Dalla Mora, and A. Tosi, "Time-gated single-photon detection module with 110 ps transition time and up to 80 MHz repetition rate," Rev. Sci. Instrum. **85**, 083114 (2014).

29. I. Nissinen, J. Nissinen, A. K. Länsman, L. Hallman, A. Kilpelä, J. Kostamovaara, M. Kögler, M. Aikio, and J. Tenhunen, "A sub-ns time-gated CMOS single photon avalanche diode detector for Raman spectroscopy," Proc. Eur. Solid-State Device Res. Conf. 375–378 (2011).

30. S. Ly, G. McNerney, S. Fore, J. Chan, and T. Huser, "Time-gated single photon counting enables separation of CARS microscopy data from multiphoton-excited tissue autofluorescence," Opt. Express **15**, 16839 (2007).

31. F. Benz, M. K. Schmidt, A. Dreismann, R. Chikkaraddy, Y. Zhang, A. Demetriadou, C. Carnegie, H. Ohadi, B. De Nijs, R. Esteban, J. Aizpurua, and J. J. Baumberg, "Single-molecule optomechanics in "picocavities,"" Science **354**, 726–729 (2016).

32. C. Carnegie, M. Urbieta, R. Chikkaraddy, B. de Nijs, J. Griffiths, W. M. Deacon, M. Kamp, N. Zabala, J. Aizpurua, and J. J. Baumberg, "Flickering nanometre-scale disorder in a crystal lattice tracked by plasmonic flare light emission," Nat. Commun. **11**, (2020).

33. M. Sidorova, S. G. Pavlov, A. D. Semenov, M. Gensch, and H.-W. Hübers, "Fiber-dispersive Raman spectrometer with single-photon sensitivity," Opt. Express **29**, 20941 (2021).