DOI: 10.1177/0003702819895299

An Optical Model for Quantitative Raman Microspectroscopy

Joseph Razzell Hollis*, David Rheingold, Rohit Bhartia, Luther Beegle

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Corresponding author email: hollis@jpl.nasa.gov

Abstract

Raman spectroscopy is an invaluable technique for identifying compounds by the unique pattern of their molecular vibrations, and is capable of quantifying the individual concentrations of those compounds provided that certain parameters about the sample and instrument are known. We demonstrate the development of an optical model to describe the intensity distribution of incident laser photons as they pass through the sample volume, determine the limitations of that volume that may be detected by the spectrometer optics, and account for light absorption by molecules within the sample, in order to predict the total Raman intensity that would be obtained from a given, uniform sample such as an aqueous solution. We show that the interplay between the shape and divergence of the laser beam, the position of the focal plane, and the dimensions of the spectrometer slit are essential to explaining experimentally observed trends in deep-UV (DUV) Raman intensities obtained from both planar and volumetric samples, including highly oriented pyrolytic graphite (HOPG) and binary mixtures of organic nucleotides. This model offers the capability to predict detection limits for organic compounds in different matrices based on the parameters of the spectrometer, and to define the upper/lower limits within which concentration can be reliably determined from Raman intensity for such samples. We discuss the potential to quantify more complex samples, including as solid phase mixtures of organics and minerals, that are investigated by the unique instrument parameters of the scanning habitable environments with Raman and luminescence for organics and chemicals (SHERLOC) investigation on the upcoming Mars 2020 mission.

Keywords: Raman spectroscopy, quantitative Raman, optical modelling, organic compound detection, detection limits, Mars 2020

Introduction

Raman spectroscopy is a useful technique for the identification of compounds through the analysis of laser light scattered from a sample. Three different Raman spectrometers have been selected as scientific instruments on upcoming missions to the surface on Mars in 2020,

DOI: 10.1177/0003702819895299

including NASA's Mars 2020 rover and ESA's Exomars rover.[1–3] The scanning habitable environments with Raman and luminescence for organics and chemicals (SHERLOC) instrument, onboard Mars 2020, consists of a microscopic imager used in conjunction with a deep ultraviolet (DUV) fluorescence/Raman spectrometer. SHERLOC's objectives are to identify potential biosignatures and assess the habitability of the environment that existed on the Martian surface at Jezero crater. In an effort to enable quantitative analysis, we are moving beyond the fluorescence and Raman optical models that were used to design and develop the instruments,[1,4] into models to deconvolve mixtures and quantitate the molecules detected on the surface of Mars at concentrations commensurate with the organics that have been detected on Mars to date.[5] To truly ascertain the significance of potential biosignatures, the simple identification of trace amounts of organic compounds needs to be more quantitative, including a determination of concentration that is independent of bulk mineral composition. This both enables increased science return from in situ instruments like SHERLOC, and will support the selection of critical and/or significant samples that should be cached for a return to Earth.

The combination of microscopy with Raman spectroscopy provides the tangible advantage of mapping how the Raman spectrum of a sample varies across its surface at the millimeter to micron scale.[6–8] In microspectroscopy, spectra are obtained at particular points on the surface but each spectrum will represent the total signal collected over a volume in and around that point based on how tightly the laser is focused.[9] Any molecules in the focal volume will be exposed to the laser and potentially generate Raman scattering that can be detected by the instrument, and the total intensity measured should be proportional to the illuminated volume. Certain compounds are more easily polarized than others and generate more Raman scattering per molecule, represented by its Raman scattering cross-section.[10–11]

The most basic interpretation of Raman scattering suggests that the Raman intensity obtained from any given compound (the analyte) is proportional to its concentration, most Raman measurements are purely qualitative, focusing instead on identifying compounds by the unique pattern of their Raman peaks while quantitative measurement of concentration from Raman scattering is limited to investigations of a well-studied analyte in a medium of known optical properties, using an instrument of well-defined parameters. As such, "quantitative" Raman is sometimes used to refer to the direct measurement of certain sample parameters, such as crystallinity or composition, through the numerical analysis of peak properties such as

DOI: 10.1177/0003702819895299

position, width or relative intensity.[12–14] The direct calculation of absolute concentration from absolute Raman intensity is challenging due to the difficulty of accurately determining the effective number of molecules being probed in the sample, $N_{scatterer}$, or the effective intensity of incident light as it passes through the sample, I_{Laser}. Both parameters may be modified by a number of confounding factors specific to the instrument or the sample itself, and it is generally inadvisable to compare absolute Raman intensities between any two measurements under anything less than identical experimental conditions.[15] Many approaches to quantifying an analyte's concentration from Raman intensity try to sidestep these problems by introducing an internal standard: a second Raman-scattering compound of known concentration that will be measured simultaneously, under identical conditions, so that the relative intensity becomes independent of any confounding factors (whether sample properties or instrumental parameters) and depend solely on the ratio of the concentrations of the compounds and their relative Raman scattering cross-sections.[11] However, this approach relies on there being no appreciable interaction or spectral overlap between the internal standard and the analyte, and precludes the quantitative measurement of samples whose composition cannot be easily controlled or modified.

For our purposes, the development of a more quantitative measurement will enable the SHERLOC investigation to better characterize a sample on the surface of Mars, with the end goal of reliably determining concentrations and defining the detection limit for organic compounds embedded in solid samples observed in situ.[1,4,16] The first step towards achieving this objective is the development of a model to describe how the parameters of the optics of the instrument interact with the sample. Rather than immediately considering complex samples of organic material embedded within mineral matrices, where the signal may be affected by variations in composition or scattered by structures within the sample, we have begun by considering simpler systems such as solutions, where any organic molecules are uniformly distributed within a transparent medium; and planar solid samples of well-known optical properties, such as highly oriented pyrolytic graphite (HOPG). Our model was developed with the SHERLOC instrument and its laboratory analog mineralogy and organic based investigations using ultraviolet (UV) spectroscopy (MOBIUS) in mind, and thus focuses on parameterizing the annular lasers employed by this class of DUV Raman microspectrometers, but can easily be applied to more conventional laser profiles by substituting the appropriate equations. In most

DOI: 10.1177/0003702819895299

cases, we expect that the model equations will be much simpler for more conventional, terrestrial applications.

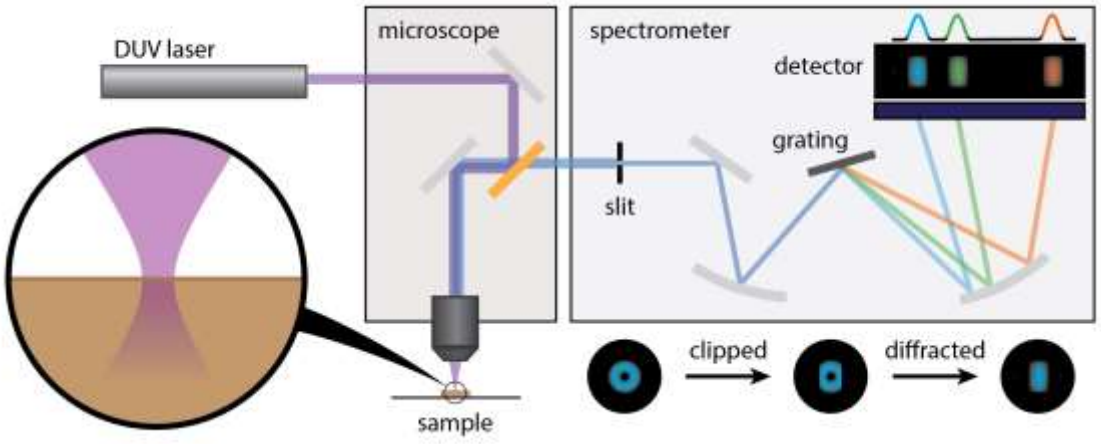


Figure 1. Illustration of the MOBIUS Raman microspectrometer configuration, depicting the hyperbolic cross-section of the focal volume at the sample, the collection of a 2D image representing the distribution of Raman scattered light, which as an annulus due to the profile of the 248.6 nm NeCu laser, subsequent clipping of the image by the spectrometer's rectangular slit, and spectral diffraction by the grating to create the detectable peaks of a Raman spectrum.

In this paper we will present a novel optical model for describing the measurable Raman intensity that will be obtained from a uniform sample using a microspectrometer. This model will account for the non-cylindrical distribution of the incident laser beam, which has a minimum diameter at the focal plane and broadens as it goes out of focus with increasing distance from that plane, as shown in Figure 1. We will also consider the implications of the spectrometer's optics in limiting the dimensions of the volume that can actually be detected, particularly the clipping of the sample image as it passes through the slit of the spectrometer (Figure 1). We must finally factor in how the absorption of light by the sample itself may reduce the measurable volume further, based on the absorption coefficients of the compounds present.[17–19] These factors will be essential to describing experimentally observed phenomena and trends in Raman intensity that cannot be adequately explained using a more basic interpretation. Once developed and tested against experimental data, the model may be used to predict trends in Raman intensities from a variety of uniform samples, examining how the measured intensity will be altered by composition, focusing, magnification, or the width of the spectrometer slit. Once these

DOI: 10.1177/0003702819895299

relationships have been established, we will be able to consider more structurally complex samples such as organics embedded in mineral matrices, and further develop the model towards the reliable estimation of concentration in Mars-relevant samples, and the prediction of detection limits for any given organic compound in any given matrix.

Methodology

Organic Sample Preparation

The four deoxyribonucleotides, deoxyribose adenosine triphosphate (dATP), deoxyribose cytidine triphosphate (dCTP), deoxyribose guanosine triphosphate (dGTP), and deoxyribose thymidine triphosphate (dTTP) were received as polymerase chain reaction (PCR)-grade 100 mM aqueous solutions (Sigma-Aldrich, DNTP100), containing a trace quantity of phosphate buffer. A 100 mM aqueous solution of Na_2SO_4 was obtained by dissolving sodium sulfate decahydrate ($Na_2SO_4.10H_2O$; Sigma-Aldrich) in MilliQ H_2O. Stock solutions containing 50 mM of a particular nucleotide with 50 mM of the internal standard Na_2SO_4 were prepared by mixing the PCR grade nucleotide solution and Na_2SO_4 solution at a volume ratio of 1:1. Lower concentrations were produced by serial dilution of the stock solution using the Na_2SO_4 stock, producing solutions with a composition generalized by N mM of nucleotide and 100 N mM of Na_2SO_4. Immediately prior to measurement, 50 ±0.05 μL of solution was deposited onto a clean Aluminum wafer beneath the objective lens, and the sample stage z-axis position adjusted to ensure the wafer was in focus through the droplet. Mixtures containing multiple nucleotides were produced by combining two or more stock solutions by pipetting them individually onto the same spot on the wafer at the appropriate volumes necessary to produce a single 50 μL droplet with the desired concentration ratio.

DUV Raman Measurements

MOBIUS, a custom DUV resonant Raman spectrometer at the NASA Jet Propulsion Laboratory was used for all measurements. MOBIUS uses a 248.579 nm NeCu pulsed laser (Photon Systems Inc.) reflected off of a 248.58 nm RazorEdge ultra-steep long-pass edge filter (Semrock Inc.) and focused onto the sample through a DUV chromatically corrected f/4 objective lens with a numerical aperture of 0.13 (ThorLabs LMU-5x-UVB). Raman-scattered photons were collected in a 180° backscatter geometry, passed through a Horiba 550i spectrometer, and recorded by a

DOI: 10.1177/0003702819895299

Horiba Symphony e2v 42–10 CCD liquid nitrogen cooled (−140 °C) detector. The collected light was passed through a slit with a width of 250 μm, and diffracted by a grating with a groove density of 1800 lines/mm. Raman spectra were obtained over 1024 points between ~570 cm^{-1} (252.2 nm) and ~4200 cm^{-1} (277.6 nm) with a spectral accuracy of 3.8 cm^{-1}. The output energy of each laser pulse was recorded using a photo diode at the exit aperture of the laser, previously calibrated to the value measured previously at the sample by an Ophir sensor. Prior to data collection, wavelength values were calibrated by validating the position of the primary and secondary laser lines at 248.58 and 252.93 nm respectively. Raman spectra were integrated over 1200 pulses per point with a frequency of 40 Hz, a pulse width of 40 μs, and an energy of 2.5–3.4 μJ/pulse. Spectra were recorded in a 5x5 array over a 500x500 μm square across the surface of each sample, with a 15 second delay between acquisitions, with the entire measurement completed within 20 minutes of deposition to minimize the impact of evaporation. Resulting spectra were corrected for pulse-to-pulse variation in laser output by normalizing the Raman signal based on recorded laser intensity during each acquisition. Cosmic rays were identified as outliers in the distribution of intensity values in each Raman shift channel and replaced by the value of adjacent points.[20] Further processing was done by in-house Python scripts utilizing NumPy, SciPy, and LMFIT packages.[21–23] Processing included subtraction of a linear baseline by least squares regression, a recalibration of Raman shift values based on the position of the atmospheric N_2 peak relative to its literature position of 2331 cm^{-1}, and subsequent removal of atmospheric N_2 and O_2 peaks by subtraction of a standard atmospheric Raman spectrum acquired on the same spectrometer. For spectra taken of samples in solution, the height of the H_2O peak at ~3300 cm^{-1} was measured and a purified water (MilliQ H_2O) spectrum was scaled and subtracted from the sample spectrum to remove the minor H_2O bending vibration that overlaps with any organic vibrational modes in the ~1600 cm^{-1} region. All steps of spectral processing are visualized in Supplementary Information Figure S1. Mixed nucleotide spectra were deconvoluted by scalable linear combination of Raman spectra from the individual nucleotides taken under standard conditions (50 μL of stock 50 mM solution) based on nonlinear least squares regression of all points between 800 and 1800 cm^{-1}. Peak intensities were obtained by trapezoidal integration over 900–1000 cm^{-1} for the sulfate peak, 1000–1800 cm^{-1} for the organic peaks, and 3000–4200 cm^{-1} for the water peak.

DOI: 10.1177/0003702819895299

Laser Beam Profiling

The laser spot profile was imaged using an Ophir charge-coupled device (CCD) placed onto the sample stage beneath the objective. Images were recorded using the BeamGage software package, with an acquisition time of 419.47 ms and a gain of 0. To avoid saturation of the sensor (at >4000 counts per pixel), the laser was fired at 10 Hz, with a pulse width of 8 μsec and an energy of ~0.09 μJ/pulse to avoid saturating the sensor. The focal plane was initially determined by focusing onto the visible elements of the CCD using the focusing camera of the spectrometer, and the exact optical axis (*z*) position in μm was recorded using the absolute value of the stage z-motor (in steps) with a conversion factor of 0.09921875 μm/step. The beam profile was imaged at varying z-positions up to ±1500 μm from the focal plane, and 16 images were taken per plane (totaling ~67 pulses). Each grayscale image was saved as a 64-bit color JPEG, with a sensitivity of ~0.06 counts per step in brightness, and later imported into a custom Python script for averaging and fitting using a Gaussian-blurred ellipse function. Fitting was done using the LMFIT python package. The dimensions of the beam spot in each image were calculated based on a pixel size of 3.69 x 3.69 μm.

Results

Annular Elliptical Beam Parameters

Unlike solid state and conventional gas lasers, the hollow cathode NeCu laser within both SHERLOC and MOBIUS produces an annular beam rather than a Gaussian beam. In order to describe the distribution of incident light from the annular 248.6 nm laser evolves as it passes through the sample space, the laser spot was imaged using an Ophir CCD that was moved incrementally up to ±1.5 mm from the focal plane in the *z*-axis direction. As shown in Figure 2a, the laser spot at the focal plane appeared as two centered ellipses of similar size, roughly 60 μm (16 pixels) along the long axis, but at different angles. With increasing distance from the focal plane, the angle between the two ellipses is reduced until they overlap completely by ±1000 μm, while the general diameter of the beam increases monotonically as it goes out of focus, reaching ~160 μm at ±1000 μm. The thickness of the ellipse, from outer to inner edge, is approximately 7–11 μm (two to three pixels) at the focal plane, but increases roughly in proportion to overall diameter, reaching 18–21 μm (five to six pixels) at ±1000 μm. We also observed that the angle of the ellipse was very different above and below the focal plane, shifting clockwise by ~1.9

DOI: 10.1177/0003702819895299

radians, which is fairly typical of an astigmatic beam undergoing a Gouy phase shift as it passes from near-field to far-field propagation.[24]

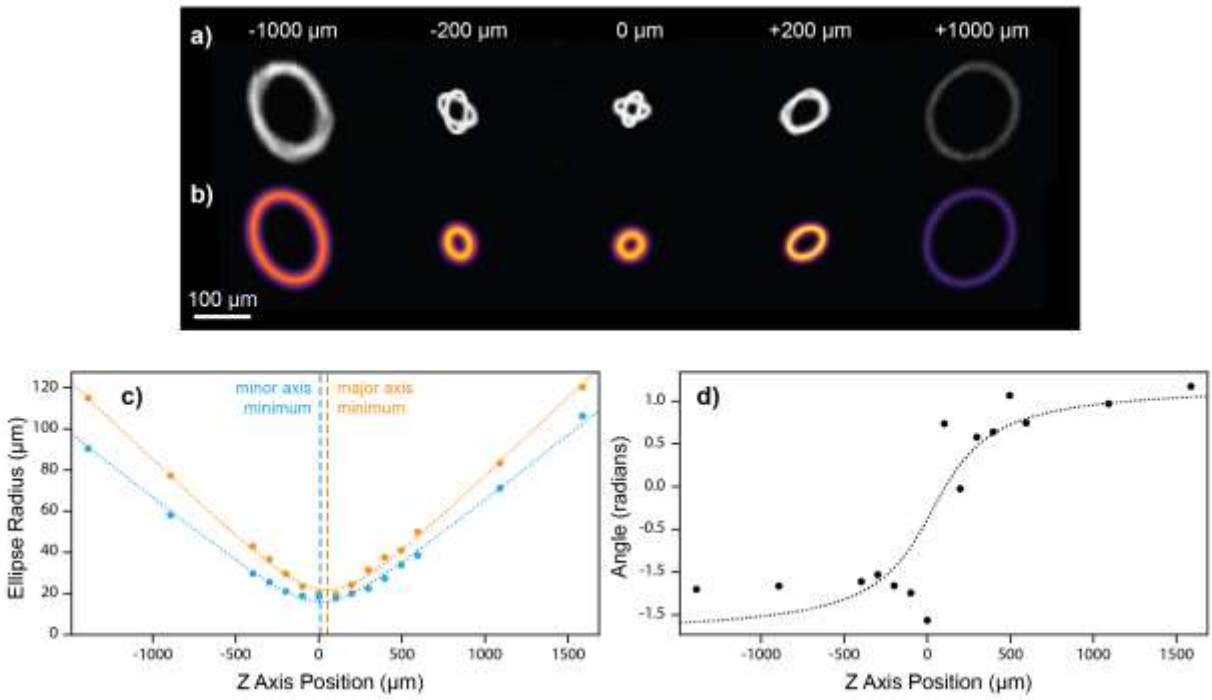


Figure 2. (a) The elliptical beam spot imaged at several positions along the *z*-axis, and b) intensity distributions obtained by fitting using a Gaussian-blurred ellipse function. (c) semi-major and semi-minor radii versus z-position, fitted with hyperbolae. (d) Angle of the ellipse versus z-position, fitted with an arctan function to approximate the Gouy phase shift of the astigmatic beam.

$$L(x,y) = I_0 \exp\left(- \frac{\left(x - x_0 - r_m \cos(\theta - \phi)\right)^2}{2\sigma^2} - \frac{\left(y - y_0 - r_m \sin(\theta - \phi)\right)^2}{2\sigma^2} \right) \quad (1)$$

Each image of the beam spot was fitted using a Gaussian-blurred ellipse function (Eq. 1). This function has seven parameters: the center of the ellipse (x_0, y_0), the ellipse angle φ, the semi-major radius r_M, the semi-minor radius r_m, the total incident intensity I_0, and the standard deviation of the Gaussian blur, σ, which were all optimized by using nonlinear least squares fitting to minimize the residual between the real image and an image generated using the function (shown in Figure 2b). The appearance of a second ellipse was not accounted for as it

DOI: 10.1177/0003702819895299

proved impossible to obtain realistic fits for all *z*-axis positions when a second function was included, and we note that in images where there are two very distinct ellipses the single-function fit tends to obtain the best possible overlap with both by selecting an intermediate angle and compensating by overestimating the standard deviation. Despite this, we obtain fairly reliable results in terms of semi-major and semi-minor radii, as shown in Figure 2c, and both are consistent with the hyperbolic relationship between beam diameter and *z*-axis position for a confocal laser.

$$r(z) = R_0\sqrt{1+(z - z_0)^2} / z_R^2 \quad (2)$$

By fitting the trends in both major and minor radii using a hyperbolic function (as defined in Eq. 2), we are able to determine the peak-to-peak semi-major and semi-minor axis diameters of the ellipse reached minima of 44 μm and 32 μm respectively, and that the laser's focal plane (z_0) was ~30 μm above the spectrometer's focal plane. The divergence of the beam as it goes out of focus is represented by the Rayleigh length (z_R, the distance from the focal plane at which the beam's diameter increases by a factor of $\sqrt{2}$), which had an average value of 264 μm. This corresponds to an angular spread of 0.071 rad, markedly smaller than the 0.13 rad angle of the *f*/4 objective because the incident laser beam only fills a fraction of the lens. The beam was also not perfectly centered on the lens, as indicated by a lateral drift of 32 μm per mm of propagation, equivalent to an angle ~32 mrad off normal (see Figure S2, Supplemental Material).

Highly Oriented Pyrolytic Graphite

Highly oriented pyrolytic graphite provides a very distinctive Raman spectrum, shown in Figure 3a, dominated by a peak at ~1570 cm^{-1} (the in-plane coordinated C=C stretching mode of graphite, known as the G-band). The rising signal at lower Raman shifts is typical of highly reflective samples like HOPG, and all other observed peaks are consistent with contributions by atmospheric gases along the beam path and in the illuminated volume directly above the sample. The N_2 Raman peak has a reported position of 2331 cm^{-1} but was observed to be slightly downshifted with respect to that value, and was used to recalibrate Raman shift values in all spectra before subtraction of a scaled atmospheric reference spectrum.[25] Subtraction is essential

DOI: 10.1177/0003702819895299

to reliably determining the HOPG G-band parameters as the atmospheric O_2 peak at ~1525 cm^{-1} overlaps with the edge of the G-band, apparent as a shoulder in the original data that disappears after subtraction (Figure S3, Supplemental Material). After subtraction, the isolated HOPG peak was fitted using a Gaussian function to measure its intensity, position and full width half-maximum (FWHM).

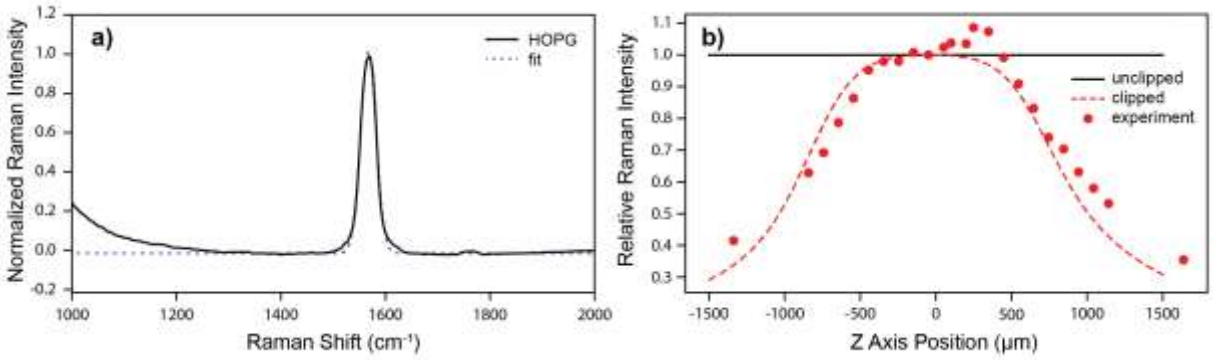


Figure 3. (a) The DUV Raman spectrum of HOPG under 248.6 nm excitation. (b) The evolution of HOPG's integrated G band intensity with varying position along the optical (z) axis, described using the optical model with and without image clipping by the spectrometer slit.

When the HOPG sample is moved incrementally along the optical (*z*-axis) direction, the position of the G-band remains fairly constant within 1.5 mm of the focal plane, with only a small but consistent variation that we attributed to inexact alignment of the laser, but there is a marked drop in intensity once the HOPG is more than 500 μm out of focus. The intensity continues to drop with increasing distance from the focal plane, following a gradually flattening curve, with some 35% of the max intensity still obtained at the furthest z-positions measured (±1.5 mm). The shape of the G-band was observed to evolve over the *z*-axis as well, appearing narrower (a FWHM of ~42 cm^{-1}) near the focal plane and gradually broadening as went out of focus, before saturating at a FWHM of ~52 cm^{-1} once 500 μm from the focal plane (Figure S3, Supplemental Material). This is consistent with the expectation that the width of the peak is dependent on the image of the illuminated sample as it is projected onto the detector, and that as the sample goes out of focus, the image gets wider and thus the peak broadens. No such trend in peak shape or intensity was observed for the atmospheric N_2 peak, indicating that the phenomenon is specific to the G band and its dependence on the z-position of the HOPG relative to the focal plane. The z-positions at which the width of the peak saturates appears to coincide

DOI: 10.1177/0003702819895299

with the sudden drop in intensity, suggesting that the loss of signal is a direct result of the edges of the peak being suppressed despite the image being significantly defocused.

Nucleotide Solutions

The deoxyribose nucleotide triphosphates of adenine, guanine, cytosine, and thymine are known as dATP, dCTP, dGTP, and dTTP respectively. As shown in Figure 4, each molecule exhibits a Raman spectrum with a unique pattern of Raman peaks between 1300 and 1650 cm^{-1}, and each has dominant mode at a characteristic Raman shift, all of which have been previously assigned in the literature to vibrational stretching and bending modes of specific bonds located on the aromatic moiety (see Table I). The combination of high frequency excitation and molecular resonance due to the nucleotides having strong absorption bands in the UV, leads to surprisingly large Raman scattering cross-sections for these molecules, on the order of 10^{-25} cm^2 $molecule^{-1}$ sr^{-1},[26] versus more typical cross-section values of 10^{-30} cm^2 $molecule^{-1}$ sr^{-1} under nonresonant conditions or using visible or infrared excitation wavelengths.[27,28] Another consequence of molecular resonance is the appearance of a weak pattern of overtones (excitation of higher order vibrational energy levels) around 2000–3000 cm^{-1} (Figure S4, Supplemental Material).

Table I. The Raman shift and differential Raman scattering cross-section for the dominant Raman peak of each nucleotide under 248.6 nm excitation,[26] and their assignments to the vibrational modes of specific atomic bonds, defined as either stretching (s) or bending (b).[26,29]

Nucleotide	dATP	dCTP	dGTP	dTTP
Dominant Peak Position (cm^{-1})	1322	1523	1474	1651
Cross-section (10^{-26} cm^2 $molecule^{-1}$ sr^{-1})	16.2	7.2	23.7	4.7
Assignment	C8H (b), C8N7 (s)	N3C4 (s)	C8H (b), N9C8 (s)	C4=O(s), C5C7(s)

DOI: 10.1177/0003702819895299

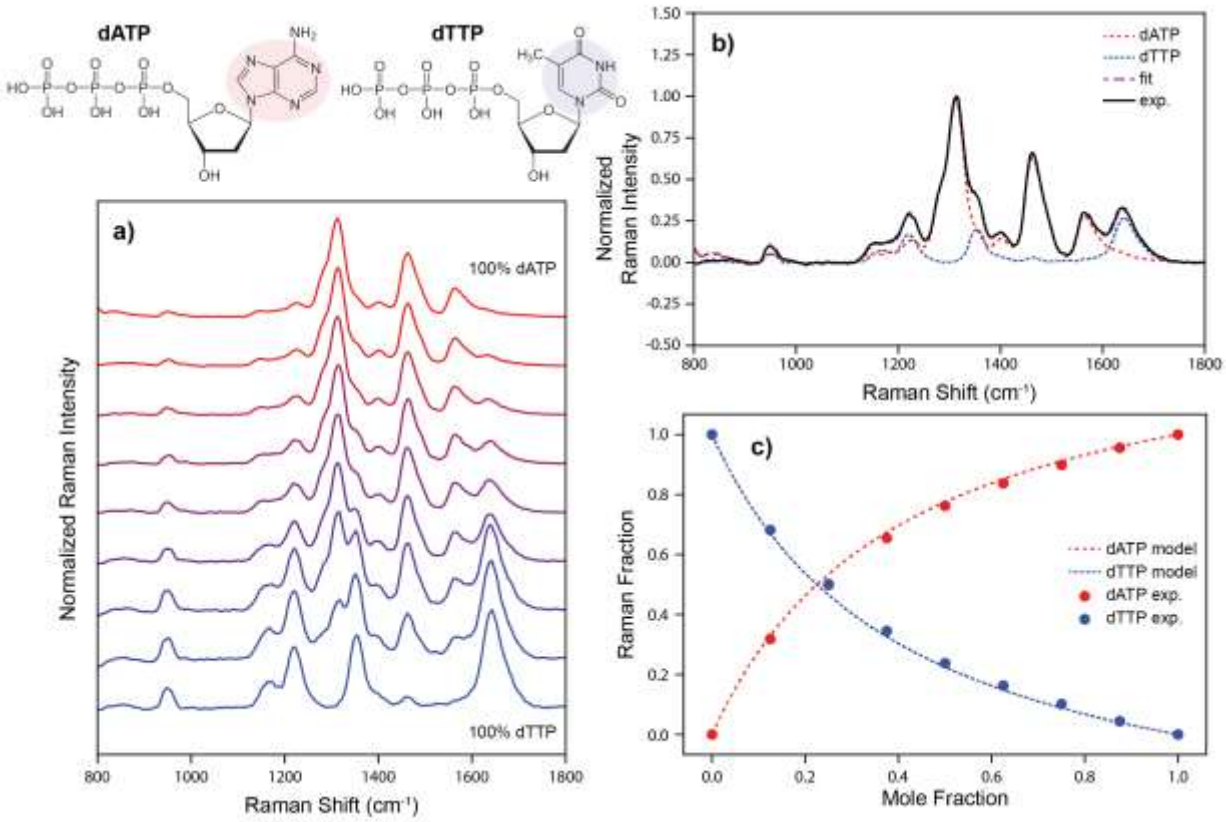


Figure 4. (a) The normalized DUV Raman spectra of mixed dATP and dTTP solutions (total concentration 1 mM) with varying mole fraction of dATP, with spectra offset for clarity. (c) Deconvolution of a 1:1 mixture of dATP and dTTP using their individual standard spectra. (d) Fractions of Raman intensity from dATP and dTTP, as obtained from deconvolution versus the Raman intensity fractions predicted by the optical model.

Figure 4a demonstrates that when dATP and dTTP are mixed, the aggregate Raman spectrum is a combination of the individual molecule spectra. The exact intensity of the peaks from a particular compound will depend on its relative concentration in the mixture, and the ratio of their Raman scattering cross-sections. This is evident in the Raman spectrum of a 1:1 by mole mixture (Figure 4b), which shows that the peaks from dATP are significantly stronger than those of dTTP despite being at the same concentration. Each mixed spectrum was deconvoluted using the standard spectra for the individual compounds, in order to reveal the exact contribution of each molecule to the overall Raman spectrum, varying their relative intensities until the residual between the fitted total and the measured spectrum was minimized. Their separate intensities can be calculated by integrating the deconvoluted (weighted) spectra between 800 and 1800 cm^{-1}.

DOI: 10.1177/0003702819895299

For a 1:1 mixture containing equal number of dATP and dTTP molecules, we found that dATP contributes roughly 76% of Raman intensity while dTTP contributes only 24%, due to the disparity in their Raman scattering cross-sections. This equates to a dATP:dTTP intensity ratio of 3.2:1, similar to the 3.4:1 ratio of their reported Raman scattering cross-sections at 248 nm Figure 4c demonstrates how the fraction of Raman intensity contributed by each component varies with the composition of the mixture from 100% dTTP to 100% dATP. It is worth noting that there is a peak at ~970 cm^{-1} apparent in all mixtures, which cannot be attributed to either nucleotide but is in fact the sulfate ion SO_4^{2-}. This was added to the stock solutions as an internal standard and despite a constant concentration of 99 mM, the intensity of the SO_4^{2-} was observed to decrease with increasing dATP content.

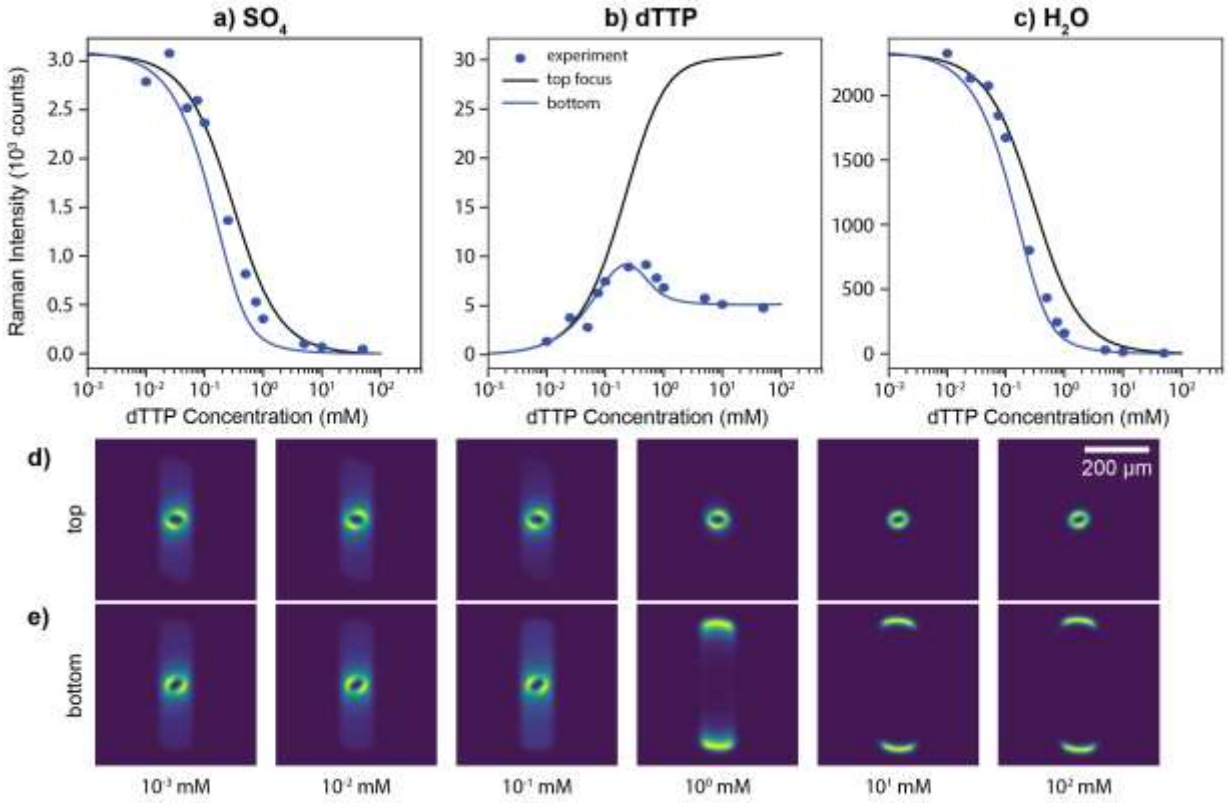


Figure 5. (Top panel) Experimentally measured DUV Raman intensities for (a) SO_4, (b) dTTP, and (c) H_2O for aqueous solutions of dTTP of varying concentration, measured under 248.6 nm excitation, versus Raman intensities predicted by the optical model with the laser focused either at the top or the bottom of a 2.5 mm thick aqueous sample. (Bottom panel) simulated images of Raman scattering projected through the slit at various concentrations when the laser is focused onto (d) the top, or (e) the bottom of the sample.

DOI: 10.1177/0003702819895299

The nucleotide dTTP was also measured in solution over a range of concentrations from 0.01 mM to 50 mM, prepared such that any given solution contains 3 Raman active components of known concentrations: the nucleotide dTTP with a concentration of *N* mM, the internal standard SO_4^{2-} with a concentration of 100 – *N* mM, and the solvent H_2O at an effectively constant concentration of 55560 mM. Figure 5 plots the integrated Raman intensity of each component (SO_4^{2-} at ~970 cm^{-1}, the nucleotide peaks at 800–1800 cm^{-1} and H_2O at ~3300 cm^{-1}) against dTTP concentration in order to observe how each is influenced by the UV-absorbing nucleotide. At the lowest dTTP concentration measured, 0.01 mM, the spectrum is dominated by H_2O (at ~$2.4x10^7$ counts) and SO_4^{2-} (at $3x10^4$ counts), but with increasing dTTP concentration the intensity of both components decreased dramatically, effectively reaching zero at >10 mM dTTP. The dTTP intensity increased in proportion to its concentration before reaching maximum of $1x10^5$ counts at ~0.75 mM. At higher concentrations it decreased, before levelling out around $6x10^4$ counts, or ~60% of the maximum, for concentrations >5 mM. Similar trends were observed for the other nucleotides (Figure S5, Supplemental Material), though there were differences in the magnitude of the intensities measured, the exact concentration at which the inflection point occurs, and the relative drop in intensity at higher concentrations.

Discussion

In order to predict and interpret the Raman intensity obtained from any given sample, we must understand how the Raman microspectrometer interrogates the sample. The simplest description of how the intensity of Raman scattered light obtained for a particular vibrational mode from a certain molecular species in a sample is generally summarized by the following equation:

$$I_{Raman} \propto I_{Laser} \times J_{scatterer} \times N_{scatterer} \quad (3)$$

where I_{Laser} is the incident laser intensity, $J_{scatterer}$ is the Raman scattering cross-section of the molecular vibration (the likelihood that scattering occurs), and $N_{scatterer}$ is the number of molecules being illuminated. Because cross-sections are specific to certain vibrations, the quoted cross-section of a molecule is usually assumed to refer to its dominant vibrational mode, e.g., the most intense peak in its spectrum. The innate broadening of Raman peaks leads to their intensity

DOI: 10.1177/0003702819895299

being distributed over a range of Raman shifts, typically ~10–100 cm^{-1}, and thus the true intensity of a peak is represented by its integrated area rather than its maximum value. The two major challenges in using Eq. 3 to predict Raman intensities, or calculate concentrations from Raman intensities, are (i) accurately estimating the Raman cross-section of a molecule in a given environment, and (ii) ascertaining how many molecules are actually being probed during the measurement. Regarding the first challenge, it is not the focus of this work to provide a detailed discussion of how to define and measure the Raman scattering cross-section of even a single molecule across a variety of environments, the issues of which have been addressed elsewhere.[10] Instead, we consider how to determine the total intensity that would be obtained from a given sample by addressing the second challenge, how to define the number of molecules probed by the instrument. We will demonstrate the development of a detailed model to describe how the optical parameters of a DUV Raman spectrometer, the profile of the annular excitation laser and the properties of the sample all influence the overall signal measured.

Using Eq. 3 as a starting point, we can develop a more accurate equation to describe the Raman scattered intensity that would be detected from a molecule in any part of the sample volume, and then integrate over the entire volume to obtain the overall signal. Firstly, we must account for how the intensity of the incident laser varies across the sample, based on the size and shape of the beam, to find the illuminated volume within which molecules may be exposed to the incident laser. Second, we must account for the limitations of the instrument's optics, which will define the detectable volume within which Raman-scattered light may actually reach the detector. Finally, we must consider how both incident light and Raman-scattered light will be attenuated by passing through the sample itself, which can be significant when measuring samples under resonant conditions (e.g., when the excitation wavelength is within the absorption band of the molecule being probed).

The Raman intensity I that is generated from a particular compound at any given position in the sample will depend on the scattering of the cross-section of the molecule, $J_{scatterer}$, the number of molecules at that position, $N_{scatterer}(x,y,z)$, the laser power at that position, $L(x,y,z)$, and the degree of light absorption based on its depth beneath the sample surface, $A(z)$. Eq. 3 thus becomes Eq. 4.

$$I(x,y,z) = J_{scatterer} \times N_{scatterer} \times L(x,y,z) \times A(z) \quad (4)$$

DOI: 10.1177/0003702819895299

For the purpose of this study, we have assumed that the scattering molecule is uniformly distributed through the illuminated volume, e.g., $N_{scatterer}$ is a constant. This is representative of samples in the gas phase or in solution but many solid samples, such as thin films or stratified minerals, can display heterogeneity at scales smaller than the spot size of microspectroscopy techniques, and therefore the total intensity measured during a spectrum acquisition will necessarily be a convolution of signal across any variations present. To analyze the illuminated volume using our model, we assumed a realistic sample volume of 500 x 500 x 2500 μm represented by a 200 x 200 x 25000 array containing $1x10^9$ points, equating to a sampling resolution of 2.5 μm in the *x*- and *y*-axes and 0.1 μm in the *z*-axis. The Raman intensity $I(x, y, z)$ at any given point can then be calculated based on Eq. 4, and the scattered intensity that is projected onto the detector can be obtained by flattening the Raman intensity array along the *z*-axis, producing a 2D image that is scaled by 2.5x in the *x*- and *y*-axes to reflect the magnification of the MOBIUS instrument's *f*/4 objective lens. The total Raman intensity that reaches the detector can then be estimated by the sum of the intensity image, provided that the terms in Eq. 4 are well defined.

Defining the Illuminated Volume

The volume of the sample that is illuminated by the laser is often approximated for conventional Gaussian lasers by assuming the illuminated volume is a cylinder, defined by the width of the focused laser spot and the thickness of the sample. The limitations of this approximation are clear: in reality a focused laser will trace out a hyperboloid rather than a cylinder, its diameter changing as it goes in and out of focus (as illustrated in the inset of Figure 1). The length scale for this divergence can range from micrometers to centimeters depending on the properties of the laser and the focusing lens, with lasers of higher M^2 value and lenses with higher *f*-numbers tending to diverge more rapidly, which means that the cylindrical approximation often applies to only the thinnest of samples.

This is complicated further by using lasers with less conventional beam profiles, such as the DUV lasers used by SHERLOC and MOBIUS. These lasers are 40 microsecond pulse, hollow cathode NeCu lasers, where lasing occurs near the inner wall of the cylindrical cathode during discharge.[30] When projected through the aperture and following optics, the cross-section

DOI: 10.1177/0003702819895299

of the resulting beam is annular, i.e., ring-shaped, resulting in a very different distribution of illumination versus a conventional Gaussian laser. The annular profile is formally defined as a Laguerre–Gaussian LG_{0l} function with a high azimuthal index *l*, versus the fundamental TEM_{00} Gaussian function of a conventional laser.[31] When imaged at the sample, as shown in Figure 1, the annular beam appears as an ellipse rather than a rotationally symmetric ring (as would be expected from a cylindrical source) due to distortion by an astigmatic aberration in the MOBIUS instrument's laser illumination path.

Fitting an image of the elliptical laser beam using a Gaussian-blurred ellipse function provides an approximation of the distribution of incident intensity in *x* and *y* dimensions. The fitted parameters can then be used to estimate the intensity at any given in-plane coordinate (x, y) using Eq. 1, where r_M and r_m are the semi-major and semi-minor radii of the ellipse, φ is the angle of the ellipse, σ is the standard deviation of the Gaussian blur, and θ is the angle from the coordinate (x, y) to the center of the ellipse (x_0, y_0). It is worth noting that this representation of the beam assumes an even distribution of laser intensity around the ellipse's circumference. At the spatial resolutions being considered (as imaged in Figure 2), there doesn't appear to be any detectable variation, but we acknowledge that periodic variations may occur in certain Laguerre–Gaussian beams depending on their azimuth index.[31]

However, the diameter of the beam is not constant and will vary with distance, along the *z*-axis, from the focal plane. This is true for all confocal laser spectrometers, Gaussian or annular, where the laser is focused at a specific position along the *z*-axis and will go out of focus with increasing distance from this plane.[32] The relationship between radius r and *z* is a hyperbola, given in Eq. 2, and when hyperbolic functions are substituted for the radii parameters r_M and r_m in Eq. 1, the intensity at any given point (x, y, z) can be determined from the minimum radii R_M and R_m as measured as the beam waist. We must also account for the fact that as the beam goes out of focus, the incident intensity is spread over an increasingly larger area and the maximum intensity passing through any given point will be reduced proportionately. Integrating Eq. 1 over *x* and *y* is not trivial, and so instead we approximate the area of the beam as a ring with outer and inner radii defined by the average radius of the ellipse ± twice the standard deviation σ. This definition of the area accounts for ~95% of the total intensity, and is expressed by the first term in Eq. 5.

DOI: 10.1177/0003702819895299

$$L(x,y,z) = \frac{I_0}{4\pi\sigma\left(R_M + R_m\right)\sqrt{1+\left(z - z_0\right)^2 / z_R^2}} \cdot \exp\left(- \frac{\left(x - x_0 - R_m\sqrt{1+\left(z - z_0\right)^2 / z_R^2}\cos\left(\theta - \phi\right)\right)^2}{2\sigma^2} - \frac{\left(y - y_0 - R_M\sqrt{1+\left(z - z_0\right)^2 / z_R^2}\sin\left(\theta - \phi\right)\right)^2}{2\sigma^2} \right)$$

(5)

Equation 5 describes the incident light intensity (ignoring sample absorption effects) at any point in the sample volume, provided that the parameters, I_0, R_M and R_m, φ, c, z_R, and the hyperboloid's center (x_0, y_0, z_0) are known for that instrument and laser. By measuring the shape of the laser beam spot at multiple positions along the *z*-axis, these parameters can be determined experimentally, and the volume illuminated by the MOBIUS instrument can be modelled. Figure 2a presents actual images of the elliptical laser beam measured at multiple positions up to ±1500 μm from the focal plane, and their best fits using the two-dimensional Gaussian-blurred ellipse function. The fits produce major and minor radii that closely follow hyperbolas along the *z*-axis, as expected (Figure 2b), despite the appearance of a second ellipse at a different angle close to the focal plane. This phenomenon is known as a Gouy phase anomaly, a consequence of the laser cavity randomly generating clockwise and counter-clockwise laser pulses, which then interact with the astigmatic aberration in the MOBIUS optics to produce ellipses with different Gouy phase angles.[24] This anomaly is only apparent when the beam passes from near to far-field propagation, e.g., near the focal plane, and beyond ±500 μm both clockwise and counter-clockwise laser pulses exhibit the same ellipse angle. Although the model fails to account for the presence of a second ellipse, the fitting was able to reproduce the hyperbolic curves for both major and minor radii and the evolution of the fitted angle followed the approximate arctan curve expected for a Gouy phase shift. We found that σ tended to be the most variable parameter, likely compensating for the appearance of the second ellipse, with an average value of 7.5 ± 1.4 μm.

Using Eq. 5, with the parameters obtained for the MOBIUS laser focus, we can simulate the incident laser intensity across a sample volume of pre-defined dimensions. The shape of the illuminated volume is represented by the red surface in Figure 6, given by the inner/outer edges (±2σ, representing 95% of the overall intensity) for the ellipse, which appear in three dimensions as two nested hyperboloids. For a more conventional Gaussian beam, there would only be a

DOI: 10.1177/0003702819895299

single hyperboloid, defined by 2σ of the single, centered Gaussian distribution.

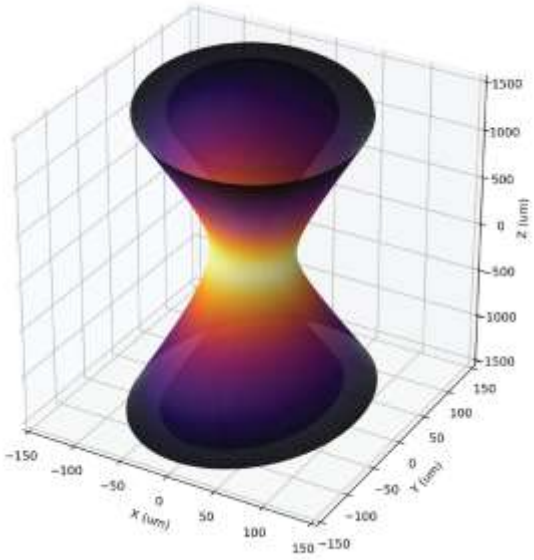


Figure 6. An illustration of the sample volume illuminated by the elliptical laser, colored according to the logarithmic flux density of the beam in that (x, y) plane.

It is interesting to note the implications of this illumination profile, namely that the focal point of the measurement (i.e., the center of the hyperboloid) receives effectively zero incident light. Of course, this means that the actual signal obtained from a measurement at a given position will actually come from the volume immediately surrounding that point. The total volume that is appreciably illuminated by the laser (based on points that meet a minimum threshold of 5% of the maximum intensity) can be calculated using the model at ~3.3×10^7 cubic micrometers (~33 nL). This represents only 4.0% of the simulated sample volume, but encompasses 94.5% of the volume weighted by the local intensity of the incident laser. When considering an actual measurement of a sample where the laser is focused onto the top surface, only half of the volume (beneath the focal plane) will overlap with the sample and a mere ~16 nL of sample will be illuminated, which would contain only ~16 picomoles (~9.9×10^{12} molecules) of a uniformly distributed compound at a concentration of 1 mM. However, any estimation of the total volume fails to account for variations in the intensity received throughout that volume, and so it is best to proceed with point-by-point modelling of the sample to determine the overall Raman intensity measured.

DOI: 10.1177/0003702819895299

Defining the Detectable Volume

Once the sample volume has been illuminated by the focused laser, we must consider what portion of that volume will produce Raman scattering that is measurable by the collection optics of the spectrometer. As illustrated in Figure 1, any scattered light from the sample that is collected by the objective lens and transmitted (with some magnification factor) through a narrow slit and onto the detector. The slit serves to control the width of the beam, balancing the desire for maximize spectral resolution (after the grating disperses the beam according to wavelength) against the need for sufficient light to ensure a detectable signal. The beam is in fact an image of the illuminated volume (in terms of its scattered light) and the edges of that image may be clipped as it passes through the slit, or otherwise fall outside the region of the detector that is actually recorded during acquisition. Figure 6 demonstrates these spatial limitations visually, displaying the borders of the slit in the *x*-axis and the detector in the *y*-axis as projections onto the sample space, with only the portion of the illuminated volume within these limits generating Raman scattering that will reach the detector and be measured. This phenomenon is exploited in confocal Raman spectroscopy, where the slit is replaced by a narrow pinhole in order to restrict the detectable volume in all three dimensions and ensure Raman spectrum comes from a highly localized region, enabling micron resolution imaging at the price of a significant loss in signal.[33]

The impact of these limitations on the measurable signal can be observed experimentally by deliberately enlarging the image of the sample by measuring it out of focus. Figure 3b shows how the signal from HOPG varies with the sample's position along the *z*-axis, relative to the focal plane. HOPG is an effectively planar source of Raman scattering: it has an extremely large (18.44 μm^{-1}) absorption coefficient at 248 nm,[34] which means that 95% of Raman intensity will come from within 0.16 μm of the top surface. This means that the distribution of detectable Raman scattering will be solely dependent on the size and shape of the laser spot at that z-position, and once the (magnified) image of the illuminated area exceeds the width of the slit, we will see a decrease in signal. For the MOBIUS instrument, where the objective lens has a magnification of 2.5x and the slit is a window 250 μm across, we would expect the image to be clipped at z-positions wherever the laser spot is >100 μm in diameter along the *x* axis. Based solely on the hyperbolic radii functions plotted in Figure 2c, the onset of clipping should occur at ±500–700 μm from the focal plane, depending on the exact angle of the ellipse relative to the slit.

DOI: 10.1177/0003702819895299

And as Figure 3b shows, that is exactly what is observed experimentally: the HOPG signal is fairly constant within ~500 μm of the focal plane – as all the scattered light is being collected – but once the sample is sufficiently out of focus there is a significant drop in measured intensity because the illuminated area is being increasingly clipped by the slit.

In order to simulate the impact of clipping by the slit, we treated the slit as a bandpass function that is applied to the modelled Raman intensity along the x axis, represented by a symmetric Fermi–Dirac function (Eq. 6) where x_s is the center of the slit, W is the half-width half-maximum of the slit function in sample space (50 μm for MOBIUS) and R is a rounding factor (5 μm^{-1}). This equation was chosen for its effectiveness in describing the shape of Raman peaks that are obtained when the instrument response is slit-limited, e.g., when the image of the laser spot is so large that it fills the slit and the response becomes solely dependent on its bandpass. Examples of this include the atmospheric O_2 and N_2 peaks, which are best fit using the Fermi–Dirac function (Eq. 6) than the more typical Gaussian, Lorentzian or Voigt functions.

$$S(x) = 1/(1 + \exp\left(\frac{-|x-x_s|-W}{R}\right)) \quad (6)$$

We simulated the HOPG Raman intensity across a range of z-positions from +1500 μm to −1500 μm (assuming a focal plane at 0 μm) by solving Eq. 5 for every point in a sample volume with dimensions 500 x 500 x 3000 μm. Each cross-section of the sample volume in (x, y) equates to the image that would be ideally obtained from the HOPG at that z-position without clipping, with the evolution in total intensity plotted by the black line in Figure 4b. Unsurprisingly, it is constant across the entire z-range considered, as there are no factors included in Eq. 5 that reduce the total intensity I_0 summed across all points. But when Eq. 6 is applied to each point, we introduce clipping by the simulated slit and there is now a marked decrease in signal at large distances from the focal plane, plotted by the red line in Figure 4b. As expected, there is no difference between the two lines within 500 μm of the focal plane, but with increasing distance from the focal plane there is a significant decrease in intensity due to clipping. The exact onset of the drop in simulated intensity (defined by a decrease of 5% versus the true total) is at ±486 μm, or roughly 1.5x the Rayleigh length. The simulated clipping function does not describe the trend exactly, which appears slightly lopsided versus the modelled curve, but the deviation is likely a consequence of astigmatism or some optical misalignment in

DOI: 10.1177/0003702819895299

the MOBIUS system, such as the beam not being entirely centered on the slit, or may be because the model does not account for the ~32 mrad drift of the laser as it propagates along the *z*-axis.

As Figure 4b shows, the loss of signal due to the limits of the detectable volume can be change dramatically over a relatively short distance along the *z*-axis. For an annular laser such as the one in MOBIUS or SHERLOC, this effectively defines the limit of the detectable volume in the *z*-axis: at a sufficient distance from the focal plane, the image of the laser spot is so large that it exceeds the limits in either *x* and *y* axes such that only a tiny fraction of the total signal will be detectable, essentially defining the limits of the detectable volume in z-position also known as the depth of field. Depth of field is traditionally defined in photography by the distance at which the defocused image exceeds a certain apparent size (the circle of confusion), and is sometimes approximated by the Rayleigh length of the laser's divergence, but considering that the primary output of a spectroscopic instrument is a total intensity at a given wavelength, it is perhaps better to define the key threshold in terms of the loss of signal: for a 50% reduction in intensity from planar HOPG, the model predicts that the MOBIUS instrument has an effective depth of field of ±1000 μm.

It is worth noting that the decrease in signal due to clipping will be far less rapid for more conventional Gaussian beams, as the majority of illumination will be much closer to the beam center (further from the edge of the slit), but the exact onset and rate of the signal decrease for any given instrument will also depend significantly on the divergence of the laser beam, the magnification of the optics and the width of the slit. Although this work has focused on a single instrument, with a particular laser profile, we argue that it is important to include the ramifications of the spectrometer design and laser profile of all Raman instruments when considering the limits of the volume that they will interrogate.

Accounting for Absorption

The laser profile, optical depth of field and the dimensions of the spectrometer slit are not the only parameters that may alter the incident light intensity or reduce the detectable Raman signal, our model must also consider the direct absorption of light as it passes through the sample.[17–19] This will apply to both the incident laser light (at 248.6 nm for MOBIUS and SHERLOC) as it passes through the sample, and the Raman scattered light (at a Raman-shifted wavelength) that returns back through the sample to the collecting objective lens. The intensity of the attenuated

DOI: 10.1177/0003702819895299

laser that will be received by any point in the sample volume can be predicted using the Beer–Lambert Law based on how far the light has travelled through the sample to reach that point (e.g., its depth beneath the top surface), the concentration of the compound and its absorption coefficient at the excitation wavelength. Because the absorbing compound may not be the same as the scattering compound, so we must sum the absorption from each and every compound present in the sample to determine the overall attenuation of the incident laser intensity. As the incident light intensity has been reduced, the Raman scattering will be reduced proportionately, but will then also be reduced further as it passes back through the sample if there is absorption at the Raman-scattered wavelength as well. The exact wavelength of Raman scattering will depend on the molecular vibration involved, but for DUV Raman the majority of Raman peaks occur within 25 nm (~3800 cm^{-1}) of the excitation wavelength, and for the purposes of this work they can be generalized into 3 bands: inorganic peaks at ~1000 cm^{-1} (255 nm), organic C=C peaks at ~1500 cm^{-1} (258 nm), and the water peak at ~3300 cm^{-1} (271 nm). Thus, our modified Beer–Lambert equation (Eq. 7) must contain two absorption coefficients for each compound present in the sample, one for the excitation wavelength and one for the Raman-scattered wavelength, in order to calculate the overall impact of absorption on the measured Raman intensity.

$$A(z) = \exp\left(-dz \sum c_i \left(\varepsilon_i^0 + \varepsilon_i^v\right)\right) \quad (7)$$

where dz is the penetration depth (distance in z from the top of the sample), and for each compound i present in the sample, c_i is its concentration, ε_i^0 is its absorption coefficient at the excitation wavelength, and ε_i^v is its absorption coefficient at the Raman-scattered wavelength. Because Eq. 7 is expressed as an exponential rather than a power of 10, care must be taken to ensure the absorption coefficients are expressed as their Napierian value rather than the more conventional decadic value – easily converted by multiplying the decadic value by $\ln(10)$.

The overall Raman intensity obtained from a given molecule at any given point in the sample volume, $I(x, y, z)$, can then be determined by combining Eqs. 5, 6, and 7 in order to account for the intensity of the laser at that point, $L(x, y, z)$, potential clipping by the slit along the x axis, $S(x)$, and the impact of absorption (on both incident and scattered light) at that depth, $A(z)$. The result is given by Eq. 8, which also includes $J_{scatterer}$, the scattering cross-section of the molecule, and $c_{scatterer}$, its concentration.

DOI: 10.1177/0003702819895299

$$I(x,y,z) = J_{scatterer} \cdot c_{scatterer} \cdot L(x,y,z) \cdot S(x) \cdot A(z) = J_{scatterer} \cdot c_{scatterer} \cdot \frac{I_0}{4\pi\sigma(R_M+R_m)\sqrt{1+(z-z_0)^2/z_R^2}} \cdot \exp\left(-\frac{\left(x-R_m\sqrt{1+(z-z_0)^2/z_R^2}\cos(\theta-\varphi)\right)^2}{2\sigma^2} - \frac{\left(y-R_M\sqrt{1+(z-z_0)^2/z_R^2}\sin(\theta-\varphi)\right)^2}{2\sigma^2} \right) \cdot \left(1+\exp\left(\frac{-|x-x_s|-W}{R}\right)\right)^{-1} \cdot \exp(-dz\sum c_i(\varepsilon_i^0+\varepsilon_i^v)) \quad (8)$$

Equation 8 can then be used to simulate the detectable Raman intensity obtained from any given point in a uniform sample, provided that the corresponding optical parameters of the sample (e.g., its UV absorption) are known and the instrument and laser are well defined. Obviously the $L(x,y,z)$ terms expressed in full in Eq. 8 are specific to an annular laser beam, but can be easily substituted by the corresponding functions that describe the intensity distribution of a Gaussian laser beam, or any other. Although Eq. 8 is far too complex to integrate in three dimensions, which would provide a single value for the total intensity that would be obtained, it can be solved point by point throughout the sample volume to generate a 3D map of Raman intensities, which can then be summed to determine the total intensity for that simulated sample volume.

In order to confirm the reliability of the overall model, we tested how well it could predict the trends in Raman intensity for a series of solutions containing a strongly absorbing organic compound of varying concentration. With the addition of Na_2SO_4 as an internal standard, we obtain three regions of characteristic Raman scattering from the three components of each solution: SO_4^{2-} at ~1000 cm^{-1}, the organic C=C peaks at 1200–1600 cm^{-1}, and H_2O at ~3300 cm^{-1}. The organic standards chosen were the four nucleotides dATP, dCTP, dGTP, and dTTP, due to their intense Raman scattering under DUV excitation and their significance as potential biosignature markers, as they represent the four bases of DNA and are present in all known terrestrial organisms. The absorption coefficients of each nucleotide at the relevant wavelengths are presented in Table II, and demonstrate how even for chemically similar compounds there can be significant variations in absorption between compounds and across a relatively narrow range of wavelengths. For an arbitrary sample containing 1 mM of each nucleotide, we should expect that for every 100 μm of sample the laser passes through, the incident light will be reduced 21%

DOI: 10.1177/0003702819895299

by dATP, 5% by dCTP, 23% by dGTP, and 11% by dTTP.

Table II. Molar absorption coefficients in M^{-1} cm^{-1} for the deoxyribonucleotide triphosphates, within a nanometer of four wavelengths corresponding to excitation (248.6 nm) and the Raman shifts for the characteristic peaks of SO_4^{2-} (~1000 cm^{-1}), C=C (~1500 cm^{-1}), and H_2O (~3300 cm^{-1}).[35]

Wavelength (nm)	Raman Shift (cm^{-1})	dATP	dCTP	dGTP	dTTP
249	0	10520	2183	11350	5093
255	1000	13932	4018	13092	6223
258	1500	14588	5495	13397	6839
271	3300	9638	9817	10023	6668

To demonstrate the predicted impact of absorption on the measurable Raman intensity, we used Eq. 8 to simulate the Raman intensities that would be obtained from a 1.5 mm-thick sample of a transparent, nonscattering medium (e.g., water) containing a single nucleotide of varying concentration. Figure 7a reveals how the intensity of each nucleotide is initially proportional to concentration when the nucleotide is sufficiently dilute, but the intensity starts to saturate as concentration increases to the point where absorption becomes significant, eventually reaching a point at which the intensity no longer varies with increasing concentration, and is absorption limited. The point at which the intensity saturates is indicative of how strongly they absorb in the UV: dATP and dGTP have significantly higher absorption coefficients than dCTP or dTTP, and saturate at lower concentrations. The reason the intensity becomes constant can be understood in terms of the interplay between the concentration, $c_{scatterer}$, and the effective detection depth. When absorption is minimal, Raman scattering can be obtained from the full detection volume, limited in the z-direction by either the sample thickness or the instrument's depth of field, and at a certain concentration absorption will become the limiting factor in determining the depth at which Raman scattering can still be detected. At that point, further increases in concentration are counteracted by a proportional decrease in the effective detection

DOI: 10.1177/0003702819895299

volume, essentially rendering the total number of detectable molecules a constant.

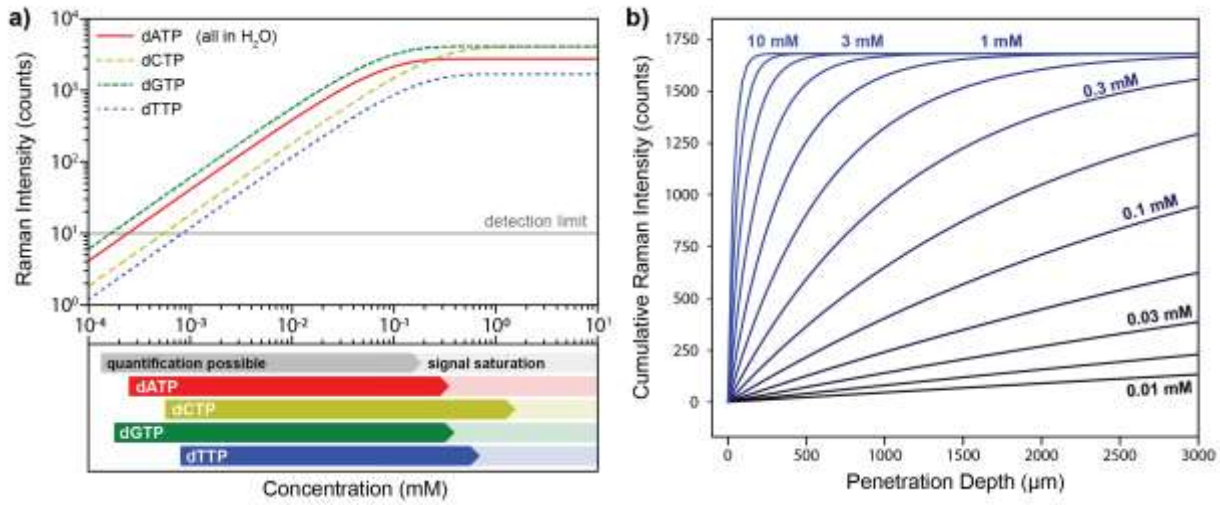


Figure 7. (a) The predicted intensity/concentration curves for the four nucleotides in a transparent medium, as simulated for the MOBIUS instrument, with the limits of detection and quantification. (b) the cumulative Raman intensity versus depth as predicted for an aqueous solution of dTTP at varying concentrations, demonstrating a varying depth of signal saturation.

Introducing absorption into the equation can complicate matters further, as it means the position of the focal plane relative to the sample becomes significant. This is not immediately apparent from Figure 7, but when a series of aqueous solutions containing absorbing organic compounds were measured experimentally (Figure 5) there were some unexpected trends that did not fit the curves predicted in Figure 7, as the curves were simulated for a sample measured with the focal plane at the top surface of the sample, rather than the bottom. The experimental measurements were done on droplets on an Al wafer, and the metal substrate provided a more reliable plane to focus on than the air/water interface at the top surface. Figure 5 demonstrates the various trends in Raman intensity from the sulfate, organic and water components of the series containing dTTP: the impact of absorption by the aromatic organic is immediately apparent as a decrease in intensity for SO_4^{2-} and H_2O despite their concentrations remaining relatively constant, this can be understood as an effective decrease in the illuminated volume with increasing absorption by the organic. The deviation of the experimental data from the initial simulations presented in Figure 7 is that of the intensity from the organic itself. It initially increases in proportion with concentration, as one would expect from a basic interpretation of

DOI: 10.1177/0003702819895299

Raman intensity, but quickly reaches a maximum at ~0.75 mM before decreasing again. It does eventually saturate at a constant Raman intensity for concentrations >5 mM, as would be expected for an absorption-limited measurement, but at a value significantly lower than expected. This trend was apparent in the other three nucleotides as well (Supplemental Material).

The concentration series samples were simulated based on a height of 2.5 mm, assuming a hemispherical droplet with a volume of 50 μL and a diameter of ~6 mm. The sample volume was considered to contain three components: nucleotide, water, and sulfate, each with its own distinct Raman scattering cross-section ($3.7\text{x}10^{-28}$ and $1.0\text{x}10^{-31}$ cm^2 molecule^{-1} steradian^{-1} for SO_4^{2-} and H_2O respectively), absorption coefficients (zero for SO_4^{2-} and H_2O) and appropriate concentration for that sample (as described in the results section). We ran two simulations, one with the laser focused onto the top of the sample (z_0 = 0 μm) and one with the laser focused onto the bottom of the sample (z_0 = 2500 μm), with the latter being equivalent to the experimental set up. Figures 5a and 5c show that both simulations predict an absorption-driven loss of Raman intensity from the SO_4^{2-} and H_2O components consistent with the trend in experimental data, though the decrease in signal occurs at a slightly lower concentration for bottom-focus measurements. For the organic dTTP component shown in Figure 5b, the top-focus simulations predict an intensity curve that follows the expected trend of Raman intensity tending towards absorption-limited saturation, but the bottom-focus simulations predict an intensity curve that is significantly lower in intensity but is an excellent match to the experimental data. Based on the simulations, we find that the inflection point occurs at ~0.2 mM for dTTP, with a bottom-focused Raman measurement providing only ~50% of the signal that would be obtained by a top-focused measurement, and that the signal saturates at ~20% of the maximum that would be obtained with top focus. The deviation can be explained by looking at the simulated Raman intensity images (after clipping) generated by the model for the two focal positions, shown in Figures 5d and 5e. At low concentrations, the images are identical, but at a certain point the contribution from the focused part of the second image, e.g., the bottom of the sample, starts to disappear due to the impact of absorption. By the time the concentration is 100 mM, the only detectable Raman intensity comes from the very top of the sample (95% from within ~9 μm of the surface for dTTP) and the bottom-focus image becomes entirely out of focus, with a much lower intensity due to losses from clipping by the slit. The inflection point observed in Figure 5 is essentially an artifact of the transition from predominantly unclipped image to a clipped one. It is worth noting

DOI: 10.1177/0003702819895299

that the experimental data in Figure 5 represents a fairly unusual measurement set up, for most samples the instrument would be focused onto the readily observable top surface rather than the bottom, but it does demonstrate a significant variation in measurable signal that could not be explained by a more simplistic model that fails to consider the impact of laser divergence or clipping.

The model can also be applied to samples containing multiple absorbing compounds, such as mixtures of two or more aromatic organics. Figure 4a shows how the Raman spectrum of a mixture of dATP and dTTP varies with composition from 100% dATP to 100% dTTP. The spectrum of the mixture is a simple linear combination of the individual compounds' spectra, and can be deconvoluted as such (Figure 4b) to isolate their contributions and determine their separate Raman intensities. When expressed as a fraction of the total intensity, the contributions from dATP and dTTP were found to follow distinctive curves with respect to the fraction of molecules they represent (see Figure 4c). When these mixtures were simulated using the model, it accurately predicts the trend in both compounds' contribution to the total Raman intensity. This raises an important point regarding the application of the model: when Raman intensities are not expressed in absolute terms but as in relation to another compound present within the sample, or to the total as in Figure 4c, it is possible to simplify the calculation considerably. Instead of solving Eq. 8 for each Raman-scattering compound, by taking the ratio of two results for the same sample we effectively cancel out the terms relating to the detection volume, $L(x,y,z) \cdot S(x) \cdot A(z)$ – these apply equally to all compounds present within that sample, including the total impact of their absorption – such that the only terms that change between each simulation are $c_{scatterer}$ and $J_{scatterer}$, the concentration and cross-section of the compound in question. Thus, the Raman intensity ratio becomes a product of their concentration ratio and their cross-section ratio, and for an equal (by mole) mixture of dATP:dTTP we find that the predicted intensity ratio is exactly equal to the cross-section ratio of 3.4:1, and is confirmed experimentally by a measured intensity ratio of 3.2:1. The nonlinearity of the trend in Figure 4c indicates the magnitude of the cross-section ratio for that pair, compounds that have a larger disparity in cross-section (like dATP:dTTP) exhibit a less linear trend, and pairs with a smaller disparity appearing more linear (like dATP:dGTP, as shown in the Supplemental Material). This follows the pre-existent understanding of Raman that if you include a compound of known cross-section (such as an internal standard) and known concentration relative to another compound, you can determine

DOI: 10.1177/0003702819895299

the second compound's relative Raman scattering cross-section from the intensity ratio. In fact, it shows that having multiple mixtures of varying concentration ratio can give a more reliable result than a single data point. And perhaps more importantly, we have shown that it is not necessary to account for the detection volume when measuring relative cross-sections, assuming the two compounds are identically distributed throughout the sample volume.

Although there will always be a minimum signal required for accurate detection (depending on the noise of the instrument), Figure 7 indicates that there is also an upper limit to the measurable signal from an absorption-limited sample. The linear trend in Raman intensity with respect to concentration between these two points suggests that concentration can indeed be quantified based on Raman intensity, provided that it falls between the limit of detection and the saturated signal as determined by the absorption of the sample. The latter can be considered to be the upper limit of quantification, at which there is still a very detectable signal but one that is essentially decoupled from concentration. The exact range for quantification will depend on the instrumental parameters and the properties of the sample, particularly its absorption, but taking the samples simulated in Figure 7 as examples, we should be able to quantify the concentrations of those four nucleotides over approximately two to three orders of magnitude, between ~10^{-3} and ~10^{-1} mM – assuming a detection threshold of 100 counts. For context, 10^{-3} mM equates to ~$5x10^{-4}$ mg/mL (for dATP), or ~20 parts per billion (in water), and based on the previously calculated interrogation volume of 33 nL, we would be getting a detectable signal from a total of only ~$2x10^{10}$ molecules. In reality, we can expect confounding factors such as overlap with other Raman peaks, transmittance losses due to elastic scattering, or simply absorption by other compounds to complicate this calculation and alter the thresholds of quantification. Therefore, it will be important to consider the exact parameters of any given sample before attempting to quantify concentration from Raman intensities. However, it is worth noting that for DUV resonant Raman spectrometers such as SHERLOC, only those compounds that have significant absorption in the UV will need to be considered, this limits the list of potential compounds to aromatic organics and certain minerals, with aliphatic organics having negligible absorption in the UV. As that same absorption typically results in resonance enhancement of Raman scattering under the UV laser, the problem is simplified: we will only have to consider those compounds that appear in the resonant Raman spectrum, and the general properties of the sample matrix itself.

DOI: 10.1177/0003702819895299

Conclusion

We have developed a relatively simple model to describe the Raman intensities that will be obtained from a homogenous sample of aqueous solution containing multiple chemical compounds, based on their optical properties and the physical parameters of the spectrometer being used to conduct the measurement. The model predicts the 3D distribution of incident intensity from a known laser profile as it passes through the sample volume, accounts for projection of the image of that volume through the spectrometer slit, and the impact of the sample's absorption at both incident and Raman-scattered wavelengths. The inclusion of the slit function is essential to describe certain phenomena observed experimentally, such as the loss of signal from a planar source of Raman scattering (HOPG) when sufficiently out of focus, and how highly absorbing organics produce a non-monotonic trend in Raman intensity with increasing concentration, depending on the position of the focal plane with respect to the sample surface and the extent of absorption by compounds present within the sample.

This represents a critical step towards accurately describing the Raman intensities that would be obtained from more physically complex samples that will be encountered by the SHERLOC instrument aboard NASA's Mars 2020 mission as it explores the chemical composition of the Martian surface. Most importantly, it provides a means to estimate the concentration of an organic compound based on its Raman intensity, provided that the matrix is accounted for, and to predict the detection limit for that compound in given any set of conditions. To ensure that these calculations are reliable across the broad range of samples that will be encountered by SHERLOC and other instruments on Mars, we intend to further develop this model to account for the impact of more physically complex matrices such as minerals, where metallic composition and physical parameters like grain size may affect the Raman signal obtained from any organics present. While the specific model equations described here were developed for annular lasers used by SHERLOC and similar instruments, it can be easily modified to describe more conventional Gaussian laser profiles, and can form the basis of a more advanced model to systematically interpret and quantify the composition of solid samples measured using this instrument and others.

Acknowledgments

DOI: 10.1177/0003702819895299

The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This work was funded by a NASA Postdoctoral Program fellowship awarded to Joseph Razzell Hollis, administered by the Universities Space Research Association on behalf of NASA. We would like to thank Dr. Haley Sapers, Dr. Evan Eshelman, and Dr. Michael Malaska for insightful and informative discussions.

References

1. W.J. Abbey, R. Bhartia, L.W. L. Beegle Deflores, et al. “Deep UV Raman Spectroscopy for Planetary Exploration: The Search for in Situ Organics". Icarus. 2017. 290: 201–214.
2. K.H. Williford, K.A. Farley, K.M. Stack, A.C. Allwood, et al. “The NASA Mars 2020 Rover Mission and the Search for Extraterrestrial Life". In: N.A. Cabrol, E.A. Grin, editors. From Habitability to Life on Mars. Amsterdam: Elsevier, 2018.
3. H.G.M. Edwards, I. Hutchinson, R. Ingley. “The Exomars Raman Spectrometer and the Identification of Biogeological Spectroscopic Signatures Using a Flight-Like Prototype". Anal. Bioanal. Chem. 2012. 404(6–7): 1723–1731.
4. R. Bhartia, Hug, W., Reid, R., L. Beegle “Explosives Detection and Analysis by Fusing Deep Ultraviolet Native Fluorescence and Resonance Raman Spectroscopy". In: P.M. Pellegrino, E.L. Holthoff, M.E. Farrell, editors. Laser-Based Optical Detection of Explosives. Boca Raton, Fl: CRC Press, 2015. Pp. 67–97.
5. J.L. Eigenbrode, R.E. Summons, A. Steele, C. Freissinet, et al. “Organic Matter Preserved in 3-Billion-Year-Old Mudstones at Gale Crater, Mars". Science. 2018. 80(360): 1096–1101.
6. M. Fries, A. Steele. “Raman Spectroscopy and Confocal Raman Imaging in Mineralogy and Petrography". In: T. Dieing, O. Hollricher, J. Toporski, editors. Confocal Raman Microscopy. Berlin: Springer, 2003. Chap. 6, Pp. 111–135.
7. C.E. Amri, M.C. Maurel, G. Sagon, M.H. Baron. “The Micro-Distribution of Carbonaceous Matter in the Murchison Meteorite as Investigated by Raman Imaging". Spectrochim. Acta, Part A. 2005. 61(9): 2049–2056.

DOI: 10.1177/0003702819895299

8. Y. Kumamoto, A. Taguchi, N.I. Smith, S. Kawata. “Deep Ultraviolet Resonant Raman Imaging of a Cell". J. Biomed. Opt. 2012. 17(7): 0760011.

9. L. Schermelleh, R. Heintzmann, H. Leonhardt. “A Guide to Super-Resolution Fluorescence Microscopy". J. Cell Biol. 2010. 190(2): 165–175.

10. D.A. Long. Raman Spectroscopy. London: McGraw–Hill, 1977.

11. M.J. Pelletier. “Quantitative Analysis Using Raman Spectrometry". Appl. Spectrosc. 2003. 57(1): 20A–42A.

12. J. Razzell-Hollis, S. Limbu, J.-S. Kim. “Spectroscopic Investigations of Three-Phase Morphology Evolution in Polymer: Fullerene Solar Cell Blends". J. Phys. Chem. C. 2016. 120(20): 10806–10814.

13. H.J. Butler, L. Ashton, B. Bird, G. Cinque, et al. “Using Raman Spectroscopy to Characterize Biological Materials". Nat. Protoc. 2016. 11(4): 664–687.

14. H.M. Sapers, J. Razzell-Hollis, R. Bhartia, L.W. Beegle, et al. “The Cell and the Sum of Its Parts: Patterns of Complexity in Biosignatures as Revealed by Deep UV Raman Spectroscopy". Front. Microbiol. 2019. 10: 1–15.

15. X. Rodríguez-Martínez, M.S. Vezie, X. Shi, I. McCulloch, et al. “Quantifying Local Thickness and Composition in Thin Films of Organic Photovoltaic Blends by Raman Scattering". J. Mater. Chem. C. 2017. 5(29): 7270–7282.

16. L. Beegle, R. Bhartia, M. White, L. Deflores, et al. “SHERLOC: Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals". Paper presented at: 2015 IEEE Aerospace Conference. Big Sky, Montana; 7–14 March 2015.

17. S.A. Asher. “Ultraviolet Resonance Raman Spectrometry for Detection and Speciation of Trace Polycyclic Aromatic Hydrocarbons". Anal. Chem. 1984. 56(4): 720–724.

18. S.A. Asher. “UV Resonance Raman Spectroscopy for Analytical, Physical, and Biophysical Chemistry ". Anal. Chem. 1993. 65(4): 201A–210A.

19. A.J. Sedlacek III, S.D. Christesen, T. Chyba, P. Ponsardin. “Application of UV-Raman Spectroscopy to the Detection of Chemical and Biological Threats". Proc. SPIE Proc. Chemical and Biological Point Sensors for Homeland Defense. 2004. 5269: 23.

20. K. Uckert, R. Bhartia, J. Michel. “A Semi-Autonomous Method to Detect Cosmic Rays in Raman Hyperspectral Data Sets". Appl. Spectrosc. 2019. DOI: 10.1177/0003702819850584.

DOI: 10.1177/0003702819895299

21. S. van der Walt, S.C. Colbert, G. Varoquaux. “The NumPy Array: A Structure for Efficient Numerical Computation". Comput. Sci. Eng. 2011. 13(2): 22-30.

22. E. Jones, E. Oliphant, P. Peterson, et al. “SciPy: Scientific Tools for Python". 2001. Available at: http://www.scipy.org/ [accessed 20 July 2018].

23. M. Newville, T. Stensitzki, D.B. Allen, A. Ingargiola. “LMFIT: Non- Linear Least-Square Minimization and Curve-Fitting for Python". https://lmfit.github.io/lmfit-py/index.html [accessed 10 March 2018].

24. J.P. Rolland, K.P. Thompson, K. Lee, J. Tamkin, et al. “Observation of the Gouy Phase Anomaly in Astigmatic Beams". Appl. Opt. 2012. 51(15): 2902–2908.

25. W.R. Fenner, H.A. Hyatt, J.M. Kellam, S.P.S. Porto. “Raman Cross Section of Some Simple Gases". J. Opt. Soc. Am. 1973. 63(1): 73–77.

26. Z.Q. Wen, G.J. Thomas. “UV Resonance Raman Spectroscopy of DNA and Protein Constituents of Viruses: Assignments and Cross Sections for Excitations at 257, 244, 238, and 229 nm". Biopolymers. 1998. 45(3): 247–256.

27. J.M. Dudik, C.R. Johnson, S.A. Asher. “Wavelength Dependence of the Preresonance Raman Cross Sections of CH_3CN, SO_4^{2-}, ClO^{4-}, and NO^{3-}". J. Chem. Phys. 1985. 82(4): 1732–1740.

28. S.A. Asher, M. Ludwig, C.R. Johnson “UV Resonance Raman Excitation Profiles of the Aromatic Amino Acids". J. Am. Chem. Soc. 1986. 108(12): 3186–3197.

29. Z.Q. Wen, S.A. Overman, G.J. Thomas. “Structure and Interactions of the Single-Stranded DNA Genome of Filamentous Virus fd: Investigation by Ultraviolet Resonance Raman Spectroscopy". Biochem. 1997. 36(25) 7810–7820.

30. M.C. Storrie-Lombardi, W.F. Hug, G.D. McDonald, A.I. Tsapin, et al. “Hollow Cathode Ion Lasers for Deep Ultraviolet Raman Spectroscopy and Fluorescence Imaging". Rev. Sci. Instrum. 2011. 72(12): 4452–4459.

31. L. Allen, M.W. Beijersbergen, R.J.C. Spreeuw, J.P. Woerdman. “Orbital Angular Momentum of Light and the Transformation of Laguerre-Gaussian Laser Modes". Phys. Rev. A. 1992. 45: 8185–8189.

32. Y. Suzaki, A. Tachibana. “Measurement of the Micrometer-Sized Radius of Gaussian Laser Beam Using the Scanning Knife-Edge". Appl. Opt. 1975. 14(12) 2809–2810.

DOI: 10.1177/0003702819895299

33. R. Tabaksblat, R.J. Meier, B.J. Kip. “Confocal Raman Microspectroscopy: Theory and Application to Thin Polymer Samples". Appl. Spectrosc. 1992. 46(1): 60–68.

34. T. Smausz, B. Kondasz, T. Gera, T. Ajtai, et al. “Determination of UV–Visible–NIR Absorption Coefficient of Graphite Bulk Using Direct and Indirect Methods". Appl. Phys. A 2017. 123: 633.

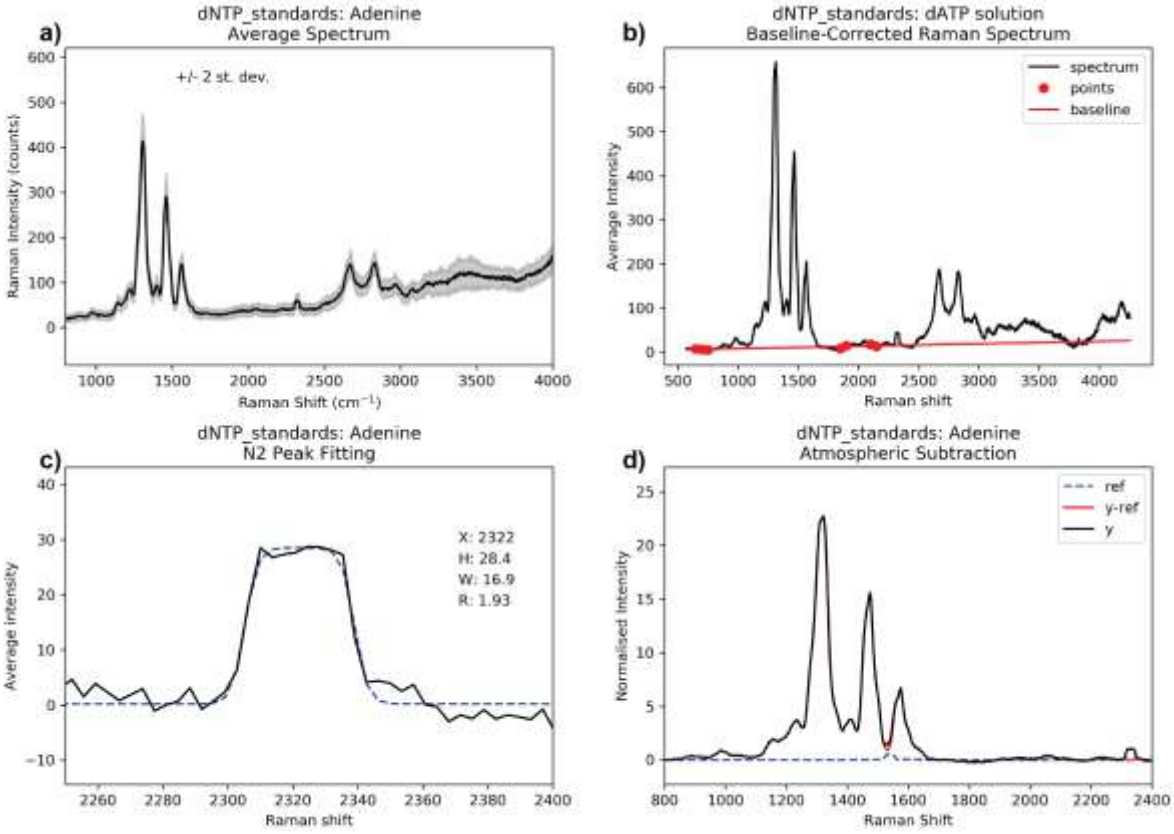
35. M.J. Cavaluzzi. “Revised UV Extinction Coefficients for Nucleoside-5’-Monophosphates and Unpaired DNA and RNA". Nucleic Acids Res. 2004. 32(1): e13.

Supplemental Material

An Optical Model for Quantitative Raman Microspectroscopy

Joseph Razzell Hollis, David Rheingold, Rohit Bhartia, Luther Beegle

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA



DOI: 10.1177/0003702819895299

Figure S1. Data processing pipeline for DUV Raman spectra, showing (a) spectral averaging, (b) baseline subtraction, (c) N_2 peak fitting, and (d) atmospheric spectrum subtraction, for the nucleotide dATP. The atmospheric contribution could not be removed by placing the sample under vacuum, as around 30-40% of the total N_2 intensity is an aggregate signal from all the atmospheric gas along the beam path between laser and sample.

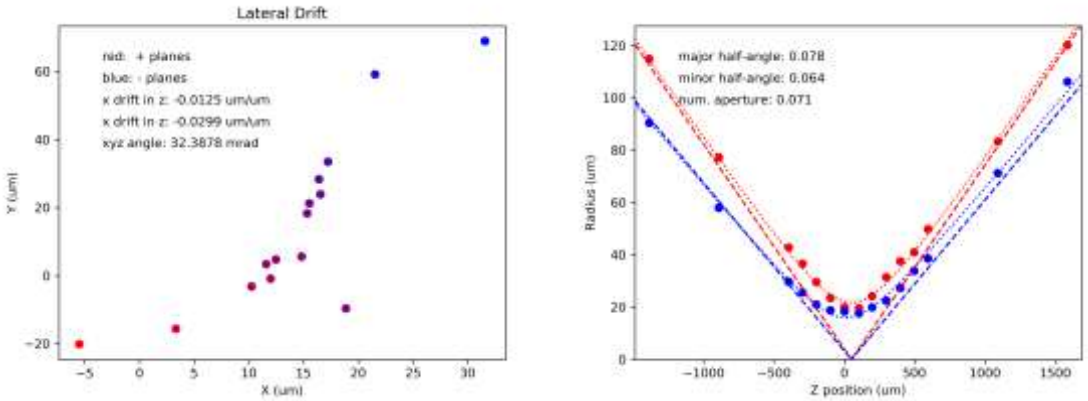


Figure S2. (a) The coordinates of the annular laser beam's center, and (b) major and minor radii, as the laser beam propagates along the *z* axis from −1500 to +1500 μm, obtained by fitting each beam image with a Gaussian-blurred ellipse function. Drift was calculated based on change in *x*,*y* positions between z = −1500 to +1500 μm, angular spread of the beam was calculated based on the ratio between beam waist and Rayleigh length parameters from fitting major and minor ellipse radii with the hyperbolic function.

DOI: 10.1177/0003702819895299

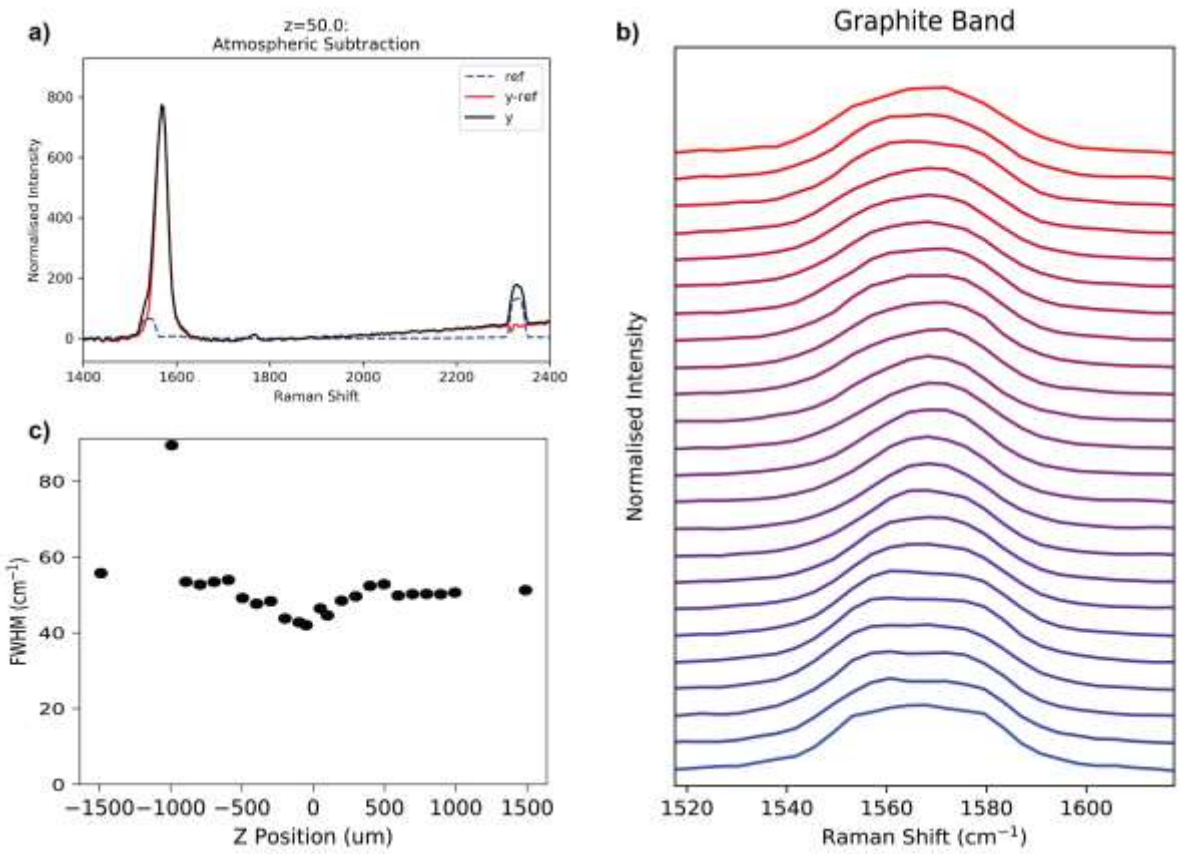


Figure S3. The DUV Raman spectrum for highly oriented pyrolytic graphite (HOPG), (a) before and after atmospheric subtraction, (b) the HOPG G band at ~1570 cm^{-1}, recorded at various *z*-axis positions (spectra offset for clarity), and (c) the measured full width half-maximum (FWHM) of the HOPG G band, versus *z* position.

DOI: 10.1177/0003702819895299

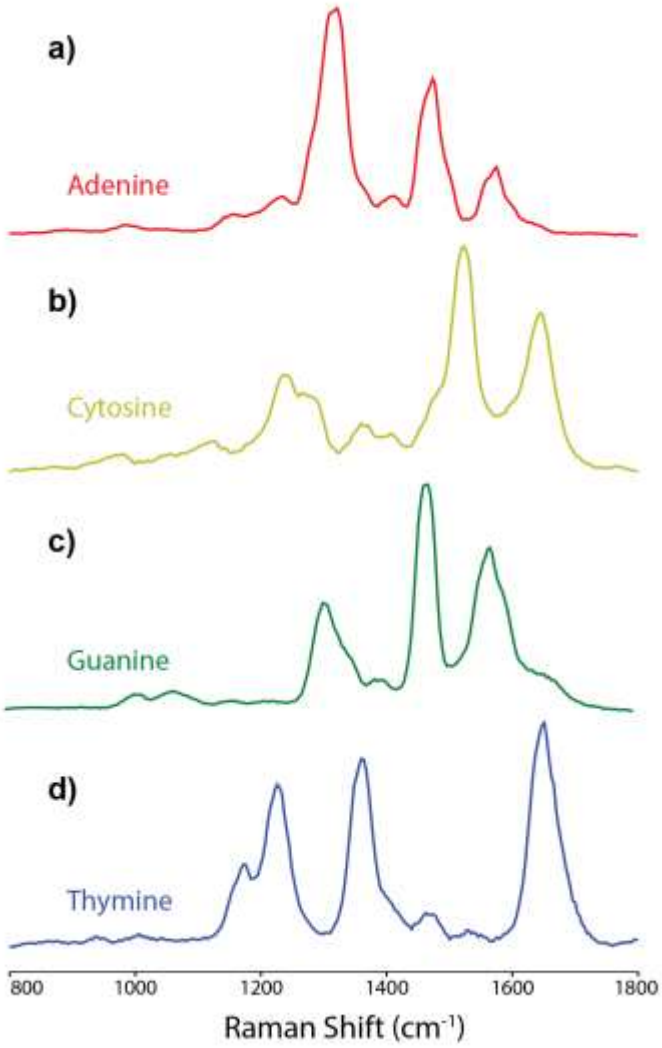


Figure S4. The DUV Raman spectra for (a) dATP, (b) dCTP, (c) dGTP, and (d) dTTP.

DOI: 10.1177/0003702819895299

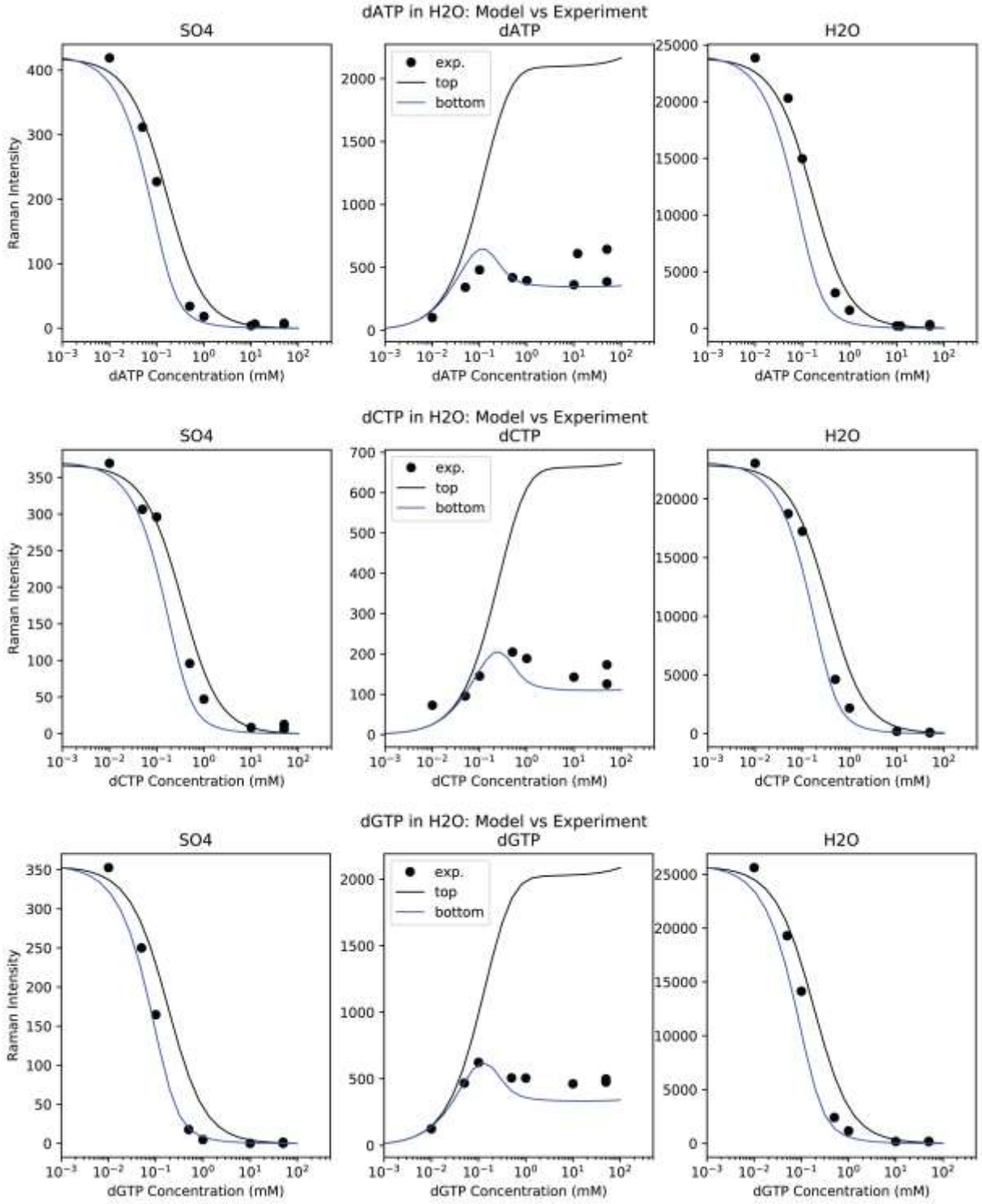


Figure S5. The integrated Raman intensities obtained for the sulfate, nucleotide and water bands of separate solutions with varying concentrations of (a) dCTP, (b) dGTP, or (c) dTTP. All

DOI: 10.1177/0003702819895299

measurements were done of 50 μL droplets on Al wafers, with the instrument focused on the wafer beneath the droplet.

Derivation of Equation 3

The annular laser beam profile is summarized in Eq. 3 using a combination of functions that represent a Gaussian-blurred ellipse, such that the intensity at any point (x, y) is determined by a Gaussian distribution centered on the nearest point of the ellipse. First, the angle and distance from the center of the ellipse (x_0, y_0) to the point (x, y) must be determined.

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

$$\theta = \tan^{-1}\left(\frac{y - y_0}{x - x_0}\right)$$

The next step is to calculate the expected radius of the ellipse at angle θ, where the semi-major axis is at angle φ, and the ellipse's semi-major and semi-minor radii are r_M and r_m respectively.

$$r_{ellipse}(\theta) = \frac{r_M r_m}{\sqrt{(r_m \cos(\theta - \varphi))^2 + (r_M \sin(\theta - \varphi))^2}}$$

The position of the point on the ellipse at angle θ can be determined correspondingly:

$$x_{ellipse}(\theta) = r_m \cos(\theta - \varphi)$$

$$y_{ellipse}(\theta) = r_M \sin(\theta - \varphi)$$

The result at point (x, y) for a Gaussian distribution originating at $\left(x_{ellipse}, y_{ellipse}\right)$ is given by the following equation, where A is a normalization factor and σ is the standard deviation.

$$G(x, y) = A \exp\left(-\frac{\left(x - x_{ellipse}\right)^2}{2\sigma^2} - \frac{\left(y - y_{ellipse}\right)^2}{2\sigma^2}\right)$$

Substituting for $x_{ellipse}$ and $y_{ellipse}$ means the intensity equation becomes the following, as expressed in Eq. 3 in the main text.

$$I(x, y) = A \exp\left(-\frac{(x - r_m \cos(\theta - \varphi))^2}{2\sigma^2} - \frac{(y - r_M \sin(\theta - \varphi))^2}{2\sigma^2}\right)$$

Derivation of Equation 4

DOI: 10.1177/0003702819895299

In Equation 4, we express the radius of the beam at any position along the optical (z) axis as a hyperbolic function. The canonical form of a hyperbola is as follows, where x and y are the variables, a is the semi-axis in the y direction and b is the semi-axis in the x direction.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

To produce the appropriate curve, we need to evaluate for x.

$$x^2 = a^2\left(1 + \frac{y^2}{b^2}\right)$$

$$x = \sqrt{\frac{a^2b^2}{b^2} + \frac{a^2y^2}{b^2}}$$

$$x = \sqrt{\frac{a^2}{b^2}(b^2 + y^2)}$$

$$x = \pm\frac{a}{b}\sqrt{y^2 + b^2}$$

This is a highly generalized expression of a hyperbola, and to use it we need to substitute these terms for the appropriate parameters of our physical system, e.g., x becomes the beam radius r, and y becomes the z position. Thus, the equation becomes:

$$r(z) = \frac{R_0}{z_R}\sqrt{(z - z_0)^2 + z_R^2}$$

where r is the (positive) radius, R_0 is the radius at the beam waist (e.g., the minimum radius), z is the position being evaluated, z_0 is the position of the beam waist (the focal plane) and z_R is the Rayleigh length (the distance from the focal plane at which r increases by $\sqrt{2}$). At large absolute values of $(z - z_0)$, the equation tends towards an asymptote with a gradient of R_0/z_R. We can then rearrange this equation, to simplify it and produce Eq. d 4.

$$r(z) = \frac{R_0}{z_R}z_R\sqrt{\frac{(z - z_0)^2}{z_R^2} + \frac{z_R^2}{z_R^2}} = R_0\sqrt{\frac{(z - z_0)^2}{z_R^2} + 1}$$

Derivation of Equation 6

Equation 6 describes clipping by the spectrometer slit as a band-pass function evaluated over the x axis, approximating it as a symmetric Fermi–Dirac function. This is based on the Fermi–Dirac

DOI: 10.1177/0003702819895299

distribution employed to describe the energetic statistics of fermions at thermodynamic equilibrium, expressed as follows.

$$F(\epsilon) = \frac{1}{1 + \exp\big((\epsilon - \mu)/k\big)}$$

Where ϵ is the dependent variable, μ is the value at which the equation evaluates to half its maximum, and *k* is the roundness of the curve. We have adapted this equation to describe the shape exhibited by Raman peaks that are slit-limited with our instrument. Therefore, it becomes:

$$S(x) = \frac{1}{1 + \exp\big((x - W)/R\big)}$$

where x is the position on the *x* axis, *W* is the half width half maximum (defined by half the maximum), and *R* is the rounding factor. This represents a single-sided function, but can be made to express both edges of the slit by introducing the modulus of the distance from x to the center of the slit, x_s.

$$S(x) = \frac{1}{1 + \exp\big((|x - x_s| - W)/R\big)}$$

Derivation of Equation 7

Equation 5 describes the absolute intensity obtained at any given point (x, y, z) in the sample volume, based on Eq. 4 and introducing a new term to replace the normalization factor *A*, which accounts for the reduction in absolute intensity as the total beam intensity I_0 is dispersed over a larger area as it goes out of focus. To approximate the area of the annular beam at any given *z* position, we treat it as a circular ring with an outer radius and an inner radius given by ± twice the standard deviation σ.

$$area = \pi(r + 2\sigma)^2 - \pi(r - 2\sigma)^2 = \pi\big(r^2 + 4\sigma r + 2^2\sigma^2 - (r^2 - 4\sigma r + 2^2\sigma^2)\big) = 8\pi\sigma r$$

where *r* is the average radius of the ellipse, approximated by the average of the semi-major and semi-minor radii.

$$area = \frac{8\pi\sigma(r_M + r_m)}{2} = 4\pi\sigma(r_M + r_m)$$

The next parameter to account for is the fact that both radii of the elliptical beam follow the hyperbolic function given in Eq. 2, based on the *z* position, the beam's focal plane z_0 its Rayleigh length z_R, and its semi-major and semi-minor radii at the minimum (R_M and R_m).

DOI: 10.1177/0003702819895299

$$area = 4\pi\sigma(R_M + R_m)\sqrt{1 + (z - z_0)^2/z_R^2}$$

and thus, when we account for the spread of the total intensity I_0 over the full beam area, the normalization factor *A* becomes the first term in Eq. 5:

$$A = \frac{I_0}{4\pi\sigma(R_M + R_m)\sqrt{1 + (z - z_0)^2/z_R^2}}$$

Derivation of Equation 7

Light that is absorbed as it passes through a sample will obey the Beer–Lambert Law, given below. The total fraction of light that is transmitted by the sample will depend on the absorbing compound's absorption coefficient, ε, its concentration, *c*, and the optical path length through the sample, *L*.

$$\frac{T}{T_0} = 10^{-\varepsilon cL}$$

This is expressed in powers of 10, but in this paper the absorption is handled as a power of *e*, which means that the equation is as follows, provided that ε_{10} is multiplied by $\ln(10)$.

$$\frac{T}{T_0} = \exp(-\varepsilon cL)$$

For the model, *L* becomes the total distance between the top of the sample and the current *z* position, *dz*. Furthermore, there is absorption of the incident light T_0 as it passes through the sample, and then the Raman-scattered light T' that returns back to the detector. The total light that returns from this depth, $A(z)$, will be the product of both these absorption processes, with separate absorption coefficients reflecting the possibility of different degrees of absorption at incident (ε^0) and Raman-scattered (ε^{v}) wavelengths.

$$A(z) = \frac{T}{T_0} \cdot \frac{T'}{T} = \exp(-\varepsilon^0 c \cdot dz) \cdot \exp(-\varepsilon^{\mathrm{v}} c \cdot dz) = \exp\left(-dz(\varepsilon^0 c + \varepsilon^{\mathrm{v}} c)\right)$$

When accounting for samples that contain multiple absorbing species, numbered 1, 2, 3, etc., their cumulative contributions to the total absorption are represented by a series of terms within the exponent, each with its own absorption coefficients and concentration.

$$A(z) = \exp\left(-dz\left((\varepsilon_1^0 c_1 + \varepsilon_1^{\mathrm{v}} c_1) + (\varepsilon_2^0 c_2 + \varepsilon_2^{\mathrm{v}} c_2) + (\varepsilon_3^0 c_3 + \varepsilon_3^{\mathrm{v}} c_3) + \cdots\right)\right)$$

which can be simplified to a single summation over each compound i that is present.

DOI: 10.1177/0003702819895299

$$A(z) = \exp\left(-dz \sum_{i=1}^{N} c_i(\varepsilon_i^0 + \varepsilon_i^{\mathrm{v}})\right)$$