applied optics

Multicomponent analysis using a confocal Raman microscope

Zhengyuan Tang,[1] Sinead J. Barton,[1] Tomas E. Ward,[2] John P. Lowry,[3] Michelle M. Doran,[3] Hugh J. Byrne,[4] and Bryan M. Hennelly[1,5,*]

[1]Department of Electronic Engineering, Maynooth University, Maynooth, Co. Kildare, Ireland
[2]School of Computing, Dublin City University, Dublin, Ireland
[3]Department of Chemistry, Maynooth University, Maynooth, Co. Kildare, Ireland
[4]FOCAS Research Institute, Dublin Institute of Technology, Kevin St., Dublin, Ireland
[5]Department of Computer Science, Maynooth University, Maynooth, Co. Kildare, Ireland
*Corresponding author: bryanh@cs.nuim.ie

Received 13 April 2018; revised 30 May 2018; accepted 31 May 2018; posted 1 June 2018 (Doc. ID 328080); published 5 July 2018

Measuring the concentration of multiple chemical components in a low-volume aqueous mixture by Raman spectroscopy has received significant interest in the literature. All of the contributions to date focus on the design of optical systems that facilitate the recording of spectra with high signal-to-noise ratio by collecting as many Raman scattered photons as possible. In this study, the confocal Raman microscope setup is investigated for multicomponent analysis. Partial least-squares regression is used to quantify physiologically relevant aqueous mixtures of glucose, lactic acid, and urea. The predicted error is 17.81 mg/dL for glucose, 10.6 mg/dL for lactic acid, and 7.6 mg/dL for urea, although this can be improved with increased acquisition times. A theoretical analysis of the method is proposed, which relates the numerical aperture and the magnification of the microscope objective, as well as the confocal pinhole size, to the performance of the technique.

OCIS codes: (170.1790) Confocal microscopy; (170.5660) Raman spectroscopy.

https://doi.org/10.1364/AO.57.00E118

1. INTRODUCTION

Raman spectroscopy is a non-invasive optical technique that can be used to identify the presence of, and quantify the concentration of, chemical substances by detecting the vibrations of molecules within the sample. Raman spectroscopy is based on the inelastic scattering of light and occurs when photons from a monochromatic laser source are incident on, and interact with, these molecular vibrations. This results in a change in the energy of the incident photons, or more specifically a shift in wavelength. The Raman scattered photons' wavelength and magnitude contain information relating to the identity and concentration of a specific chemical, respectively.

Well-established methods for measuring analytes in blood and urine typically require large volumes of fluid or lengthy processing time. In order to overcome these drawbacks, Raman spectroscopy has been proposed to quantify multiple components simultaneously and in real time with the advantage of small volume sampling and less sample contact [1–12]. A key advantage of Raman spectroscopy is that it is non-destructive; the sample can be reused for further analysis following inspection with Raman spectroscopy. Multivariate statistical analysis of the recorded spectra is central to the approach; most commonly, partial least-squares regression (PLSR) [13] is used to provide a predictive model that can estimate the relationship between a set of independent variables (the Raman spectrum) and dependent variables (chemical concentrations).

Measuring the concentration of chemical metabolites in body fluids is important in clinical and biological analysis. Three examples that are highlighted in this paper are urea, glucose, and lactic acid. Urea, a common metabolite existing in urine, reflects information on the condition of the body in terms of nutrition, and provides information about renal disorder [1]. The measurement of glucose is of obvious importance in the context of diabetes; diabetic patients must measure blood glucose concentration in order to avoid the possible complication of kidney failure, blindness, and heart disease [2]. The concentration of lactic acid in blood provides information regarding the degree of fatigue, especially for athletes [3]. The investigation presented in this paper is focused only on these three chemicals in aqueous mixture; however, it should be noted that many other blood analytes have also been measured by Raman spectroscopy including bicarbonate, triactin, ethanol, acetaminophen, creatinine, triglyceride, albumin, protein, globulin, cholesterol, and hemoglobin [4–7].

Over the past two decades, a number of independent research groups have investigated the potential of Raman spectroscopy to measure the concentration of multiple chemical components in aqueous mixture. In 1995, Goetz *et al.* [14] used an Argon-ion laser to measure the concentration of urea, glucose, and lactic acid simultaneously in an aqueous mixture solution. Following on from this initial experiment, Berger *et al.* [11,15,16], Qu and Shao [6], Enejder *et al.* [17], Rohleder *et al.* [18], and Qi and Berger [19] all attempted to further exploit Raman spectroscopy in order to predict the concentration of multiple components in greater number and with greater accuracy in terms of the smallest measurable concentration. In Section 2, the background research is briefly reviewed, as are the principles and experimented methods that underpin Raman spectroscopy. Particular attention is given to the various optical architectures that have been proposed for multicomponent analysis to date. The motivation in all of these designs is to maximize the number of Raman scattered photons that can be collected by the detector from the sample container, thereby maximizing the signal-to-noise ratio (SNR) of the recorded spectra.

In Section 3, the conventional confocal Raman microscope is discussed in the context of analyzing an aqueous solution. Variants of this type of microscope are produced by Horiba, Renishaw, and Princeton Instruments, and are commonly found in research laboratories. The optical architecture of a confocal Raman microscope is ubiquitous in the sense that it is not optimized for a specific application and is as likely to be found in a material science laboratory as it is to be found in a clinical setting. The optical system is designed to reduce the unwanted background signal from the microscope objective/substrate by using a confocal aperture [20]. However, this results in a limited depth resolution that reduces the number of collected Raman scattered photons and will, therefore, reduce the SNR of the spectrum recorded from a bulk aqueous solution; in this regard, a theoretical analysis of the performance of a confocal Raman microscope is presented. In Section 4 we extend this model to facilitate a comparison of the confocal Raman system with the bespoke designs that are reviewed in Section 2.A, in terms of collection efficiency.

In Section 5, an experiment is outlined that repeats the initial investigation of Goetz *et al.* [14] using a conventional confocal Raman microscope to measure the concentrations of glucose, lactic acid, and urea in aqueous mixtures. The results of this experiment are provided in Section 6, and a brief conclusion is offered in Section 7.

2. BACKGROUND

A. Optical Systems for Multicomponent Analysis with Raman Spectroscopy

Over the past two decades, a number of research groups have proposed different optical systems and methods for multicomponent analysis using Raman spectroscopy. In 1995, Goetz *et al.* [14] proposed the application of Raman spectroscopy to identify and quantify the concentration of three different chemicals (glucose, urea, and lactic acid) in aqueous mixture. The experimental setup that was used is shown in Fig. 1(a) and used an Argon-ion laser source with wavelength

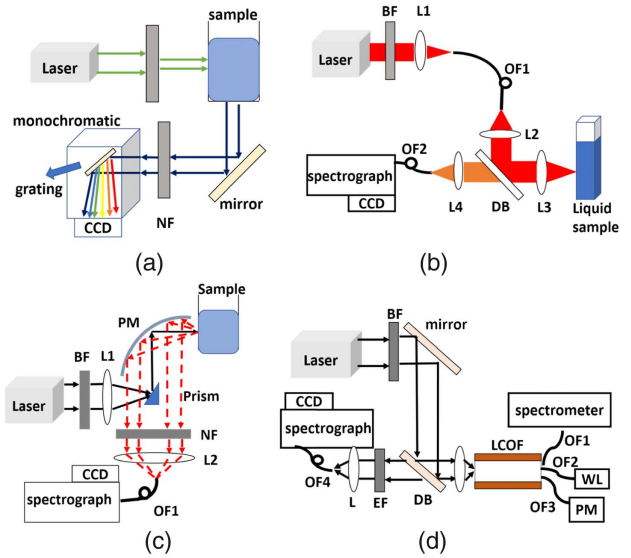

Fig. 1. Raman spectroscopy systems that have previously been proposed for multicomponent analysis. (a) Optical system similar with basic system used by Goetz *et al.* in 1995 [14]: LF, line pass filter; NF, notch filter. (b) Raman setup using optical fiber bundle used by Berger *et al.* [11] in 1996: BF, bandpass filter; L, lens; OF1, optical fiber; OF2, a bundle of seven optical fibers; DB, dichoric beam splitter filter. (c) Optical system with parabolic mirror used by Enejder *et al.* [17] in 2002: BF, bandpass filter; L, lens; NF, notch filter; PM, parabolic mirror. (d) LCOF Raman setup used by Qi and Berger [21]: BF, bandpass filter; DB, dichoic beam splitter; LCOF; liquid-core optical fiber; OF, optical fiber; EF, edge filter.

of 514.5 nm, 10 mW power. The scattered Raman irradiance was collected at 90° with respect to the source delivery path and focused onto the spectrograph slit using a lens (not shown). A total acquisition time of 40 s was used.

In 1996 and 1997, Berger *et al.* applied a similar Raman system with a near-infrared (NIR) source in order to measure the concentrations of glucose, lactic acid, and creatinine in saline solutions, and in a second experiment, to measure the concentrations of glucose in blood [11,15]; Fig. 1(b) illustrates the Raman system used in these experiments. NIR wavelengths have been demonstrated to be optimal for tissue and biosamples due to the significantly reduced unwanted background signal [11,12]. The source excitation in these experiments was produced by an Argon-ion pump laser and an 830 nm dye laser with 200 mW power, and for the second experiment a NIR diode laser with 150 mW power. In order to collect photons from as large an area as possible and couple these photons into a narrow spectrograph slit (to ensure high spectral resolution), in the first experiment the scattered photons were coupled into a fiber array bundle that was subsequently separated into individual fibers, which were input to the spectrograph along the linear slit (only one fiber was used to input to the spectrograph in the second experiment). The integration time for each spectrum was 100 s for the first experiment and 5 min for the second experiment.

Qu and Shao [6] and Rohleder *et al.* [18] subsequently proposed an ultrafiltration technique in order to remove macromolecules from their samples, thereby enhancing the predictive

accuracy for small molecules in the multicomponent mixture. Qu *et al.* [6] also proposed a waveguide capillary cell that guided the source laser irradiance over an extended volume of the sample. Enejder *et al.* [17] designed and optimized a Raman system [Fig. 1(c)] using a Monte Carlo model that estimated the brightness and the spatial support of the scattered light. The resulting design is based on the trade-off between solid angle and area and uses a parabolic mirror with high f-number and numerical aperture in order to collect a large number of Raman scattered photons.

Increasing collection solid angles, collection area, integration time, and laser power can effectively increase the number of Raman scattered photons that can be collected and will, therefore, increase the SNR of the recorded spectrum. Liquid-core optical fibers (LCOF) currently offer the gold standard for Raman multicomponent analysis by providing the highest SNRs from aqueous solutions to date. This approach relies on a significantly increased collection volume by guiding the laser over the entire length of an LCOF often many meters in length that is filled with the solution. LCOFs greatly enhance the number of scattered photons that can be collected over a given integration time, with the additional advantage of requiring a small sample volume in the order of 1 μL [22]. Building on the work of Altkorn *et al.* [23,24], the experiments of Qi and Berger [19,21,22,25] demonstrated the application of LCOFs for multicomponent analysis. An illustration of the LCOF Raman system used by Qi and Berger in 2007 [21] for quantifying different analytes in blood serum and urine samples is given in Fig. 1(d). The experiment is complicated by the effect of wavelength-dependent absorption as the scattered photons are guided back to the fiber input from where these photons are coupled to the spectrograph using a fiber bundle. This absorption is both wavelength- and distance-dependent, and is modelled using the Beer–Lambert law. In order to compensate for this effect, a calibrated white light source must be included in the setup, as well as additional spectrometer.

B. Limitation of Measurement

Experimental noise places an unavoidable limitation on the capacity of multicomponent Raman analysis both in terms of the number of analytes that can be simultaneously measured for a given mixture, as well as the smallest concentrations that can be measured for each analyte. Noise sources result from (i) shot noise, the random arrival of photons collected by detector, (ii) dark current noise, generated by thermally excited electrons within the detector pixels, and (iii) read noise, resulting from the imperfect conversion of electrons into a digital voltage in the camera. The shot noise and dark current noise are both modelled by a Poisson distribution and are both linear functions of the acquisition time. The SNR of the Raman spectrum can be defined in term of the spectral irradiance and the standard deviation of the various noise terms as follows:

$$\mathrm{SNR} = \frac{it}{\sqrt{it + ct + \sigma^2}}, \tag{1}$$

where i denotes the mean irradiance in electrons/seconds, which also takes into account the quantum efficiency of the detector; c is mean dark current (electrons/second); σ is the standard deviation of the read noise; and t is the integration time. In summary, the numerator in Eq. (1) represents the signal power at a given detector pixel, and the denominator represents the standard deviation of the total noise term. The SNR increases nonlinearly with respect to exposure time. The SNR increases rapidly in the beginning of the exposure, and this rate of increase gradually reduces over time. For practical reasons, the integration time, t, has been rarely increased beyond five minutes in any of the multicomponent experiments to date.

It has been noted that the noise present in the spectrum places a hard limit on the accuracy of measurement [26]. In particular, the Raman spectrum from water, a weak but abundant Raman scatterer, will generate shot noise that remains after the Raman spectrum of the water is numerically subtracted. Since a biofluid sample, such as urine or blood serum, will inevitably contain water in high concentration, the shot noise will be appreciably large when compared with the Raman spectrum from a chemical with a low concentration, even one with a relative large Raman cross section. The raw spectrum may also contain other undesired background signals for which the same argument can be made, including Raman spectra from the optical elements and the sample's container, as well as the unwanted baseline signal [20], which is less pronounced for NIR excitation. These unwanted background signals can all be subtracted or reduced, but the shot noise associated with these signals can never be removed. Many of the optical systems discussed in Section 2.A include design features that reduce unwanted signal; for example, the 90° collection geometry in Fig. 1(a) will reduce signals emanating from the optical elements in the illumination path and from the sample container. As will be discussed in the following section, the confocal Raman setup permits only signal originating from within a small three-dimensional volume in the sample to contribute to the Raman spectrum, and will, therefore, optically filter much of the unwanted background contaminants from the recorded spectrum. However, the cost of this is a reduction in the SNR, when compared to many of the systems reviewed in Section 2.A. In Section 4, the collection efficiency of each of the systems illustrated in Fig. 1 is compared with that of a confocal Raman microscope.

3. APPLICATION OF CONFOCAL RAMAN MICROSCOPY TO AQUEOUS SOLUTIONS

A. Confocal Raman Microscopy

All of the optical systems mentioned in Section 2.A are designed with the goal of enhancing the amount of Raman backscattered photons that can be collected from an aqueous mixture, thereby reducing the impact of shot noise. The resulting complexity of these systems limits their range of application. The most obvious example is the LCOF technique, for which the spectrum is collected from an aqueous sample that has been pumped into an optical fiber. While this setup has demonstrated the best results for multicomponent analysis to date, the system is not readily available for the analysis of cell or tissue samples.

The confocal Raman microscope has been widely applied in biology [27–30] as well as in material science [31–33]. The objective of this paper is to investigate the application of this

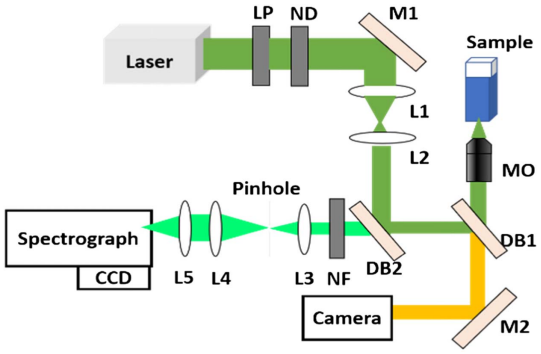

Fig. 2. Typical confocal Raman microscopy system, similar to that used in Section 5: LP, line pass filter; ND, neutral density filter; L, lens; DB, dichroic beam splitter; M, mirror; MO, microscope objective.

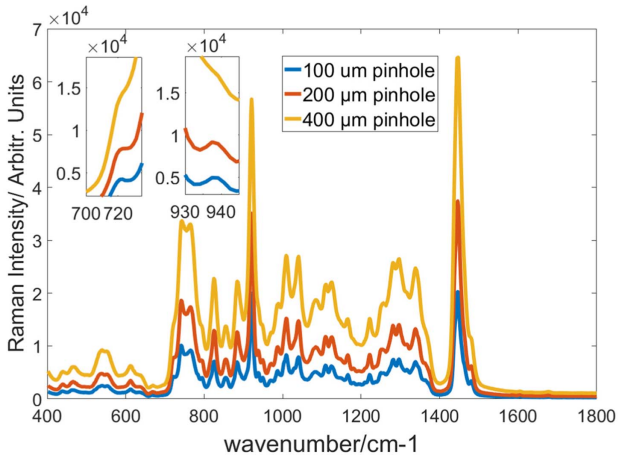

Fig. 3. Comparison of Raman spectra recorded from a transparent polymer sample using 100, 200, and 400 μm pinholes.

ubiquitous and common instrument for multicomponent Raman analysis. The optical setup for a confocal Raman microscope is shown in Fig. 2; more details on the design considerations for such a system can be found in Ref. [34]. When compared with the optical systems discussed in Section 2.A, the design is relatively simple and involves the inclusion of a pinhole in an intermediate image plane in the collection path. In confocal Raman microscopy, only photons that pass through this pinhole can contribute to the recorded spectrum; the spectrum is, therefore, composed of a contribution from a three-dimensional, spatially resolved volume. In this way, the scattering from the microscope objective, which can generate a strong unwanted background, particularly for NIR excitation, can be reduced significantly. However, it must be noted that in the context of analyzing bulk samples such an aqueous multicomponent mixture, this advantage comes at the expense of a reduced sampling volume. In particular, the reduced depth of field will result in a weaker Raman spectrum and, therefore, a lower SNR. The laser will propagate over an extended depth within the sample, generating Raman scattering at each point along that path. The confocal aperture will limit the range of depth from which this scattering can be collected. In the following sections, this limitation is explored, and a theoretical analysis of the performance is proposed.

B. Throughput of Confocal Raman Microscopy

1. Confocal Pinhole Size

Although the confocal aperture greatly reduces unwanted background signals, the depth selectivity limits the number of Raman scattered photons that can be collected from the sample. The result is a relatively low SNR when compared with the systems reviewed in Section 2.A, which will limit the capability of the confocal setup for Raman multicomponent analysis. Increasing the pinhole size to mitigate this effect is not an attractive option; it must be noted that the pinhole diameter can also affect the wavenumber (spectral) resolution in the recorded spectrum; increasing the size of the pinhole diameter beyond the width of the pixel size will blur the spectrum, and some spectral features may be obscured (here, it is assumed that the spectrograph slit is at least as wide as the pinhole diameter). In Fig. 3, the loss in resolution associated with increasing the pinhole diameter is demonstrated; three Raman spectra recorded from a polymer sample (μ-Slide I Luer, Ibidi GmbH, Munich, Germany) are shown using the Raman system described in Section 5.A with an acquisition time of 15 s and using pinhole sizes of 100, 200, and 400 μm. Two areas of the spectrum are magnified and shown in the left upper corner of Fig. 3, which clearly demonstrates that increasing pinhole size results in a loss of spectral resolution, even though the intensity is of the spectrum is increased. In the experiment outlined in Section 5, a 200 μm confocal aperture was selected as the best compromise between spectral intensity and resolution, providing an overall resolution of approximately 10 cm^{-1}.

In the absence of a confocal aperture, the width of the spectrograph slit will control the spectral resolution and, therefore, will limit the volume of scattering from within the sample that contributes to the spectrum. In order to couple as many photons as possible into the spectrograph, various solutions have been proposed; one method is to disentangle a fiber bundle and place each fiber along the slit entrance [11]; another technique involves using coded two-dimensional masks at the entrance to spectrograph, which facilitates the use of a deconvolution algorithm to recover the spectrum [35]. In this way, a high spectral resolution can be obtained from a wide slit.

2. Microscope Objective

The confocal aperture cannot be considered in isolation. The spectral intensity will also depend on the properties of the microscope objective (MO). This dependence is considered here, specifically in terms of numerical aperture (NA) and magnification. NA is defined in terms of the refractive index of the immersion medium, and the maximum angle of light that can be collected from the center of the sample. The NA is related to the minimum spatial resolution of an image as well as the image brightness and the depth of field. The optical sectioning properties of a confocal microscope are related to the NA and the pinhole aperture size [36]. This relationship can be defined in terms of the full width half-maximum (FWHM), which is defined as the width between the axial points where the intensity of an image defocuses to 50% of

its peak value in the image plane. The FWHM of a confocal microscope has been defined by Wilson [36] as follows:

$$\mathrm{FWHM} = 0.67\frac{\lambda}{(n - \sqrt{n^2 - \mathrm{NA}^2})} \times \sqrt{1 + \mathrm{AU}^2}, \quad (2)$$

where λ is the wavelength of the laser excitation, n denotes the refractive index of the immersion medium, NA is the numerical aperture, and AU is the pinhole size in Airy units, which are defined as follows:

$$\mathrm{AU} = (D \times \mathrm{NA})/(1.22\lambda \times M), \quad (3)$$

where D is the pinhole aperture size and M is the magnification of the MO. Raman scattering can be collected from each axial point along the full range of the FWHM if the laser power remains constant over this range, and it can, therefore, be assumed that the irradiance of the Raman scattering at the detector, I_r, will be directly proportional to the FWHM, i.e. $I_r \propto \mathrm{FWHM}$. Raman scattering that occurs at axial points outside of the FWHM will not contribute to the value of I_r, even though the laser may remain focused over an extended range of depth. The laser power will spread out at increasing distance from the focal plane of the MO, and the assumption that the laser power remains constant throughout the range of the FWHM requires that the full laser power remains focused over this range. Gaussian optics can be used to determine the width of the beam at various depths, which confirms that the laser beam will remain focused within the area of the pinhole aperture over the full range of the FWHM; for the sake of brevity, this analysis is not presented here.

The number of Raman scattered photons from a single point in a solution is assumed to be isotropic, and, consequently, the number of photons that contribute to the spectrum is determined by the solid angle of light that can be collected by the MO. Therefore, the irradiance of the Raman scattering at the detector is proportional to the square of the NA, i.e., $I_r \propto \mathrm{NA}^2$. The intensity of the Raman spectrum will also be dependent on the transmittance of the MO, which represents the fraction of light that is transmitted by the MO. Since the same MO both delivers the source laser and collects the spectrum, the irradiance is also dependent on this quantity as follows: $I_r \propto T_{\mathrm{MO}}^2$. The transmittance of the MO is wavelength dependent; here, for the sake of simplicity, only a constant value, T_{MO}, is used to represent the transmittance of the MO for the laser wavelength as well as for all of the wavelengths that make up the Raman spectrum, T_{MO}. Taking into account these various dependencies altogether, the following relationship can be written:

$$I_r \propto \mathrm{FWHM}(T_{\mathrm{MO}}\,\mathrm{NA})^2. \quad (4)$$

Equation (4) must be amended in order to account for the attenuation of the laser as it propagates through the FWHM, due to absorption and scattering by the molecules within the solution. In simple terms, a solution containing a high concentration of molecules will absorb and diffuse the laser light as it propagates over a relatively short distance. This attenuation can be modelled as follows:

$$T = \exp(-\alpha L), \quad (5)$$

where T denotes the transmittance of a sample of thickness L, with attenuation coefficient α, which is given by the sum of the absorption coefficient μ_a, and the reduced scattering coefficient μ_s' [37] of the sample, $\alpha = \mu_a + \mu_s'$. Both of these quantities are typically measured in terms of cm^{-1}, i.e., the attenuation resulting from a thickness of 1 cm, which should not be confused with the unit of wavenumber used elsewhere in this paper. The reduced scattering coefficient is a lumped property that takes into account both the traditional scattering coefficient μ_s as well as the anisotropy of the sample g. The purpose of the reduced scattering coefficient, which is defined as $\mu_s' = \mu_s(1 - g)$, is to describe the diffusion of photons in a random walk of step size $1/\mu_s'$ cm, where each step involves isotropic scattering [37]. Both the reduced scattering coefficient and the absorption coefficient are wavelength dependent. For the sake of simplicity, a constant value is assumed here for both quantities for all of the wavelengths that make up the Raman spectrum as well as for the source laser wavelength. Taking attenuation into account, Eq. (4) is rewritten as follows:

$$I_r \propto E = (T_{\mathrm{MO}}\,\mathrm{NA})^2 \int_0^{\mathrm{FWHM}} \exp(-2\alpha L)\mathrm{d}L, \quad (6)$$

where the factor 2 appears in the exponential function in order to account for the attenuation of the backscattered Raman photons along the same length of the sample, as well as for the source laser. The term on the right-hand side of Eq. (6) is named the "collection efficiency" in the discussion that follows and denoted by E. The value of α for a particular solution is proportional to the concentration of the analyte in solution. Specific values of α for a given mixture can be determined by experiment [37–40]; however, an examination of the literature revealed no information on the value of the scattering and absorptivity coefficients for glucose, lactic acid, or urea solutions. For the purpose of this discussion, a qualitative analysis of Eq. (6) is given below, for a wide range of α, without focusing on specific values of α for mixtures containing physiologically relevant concentrations of various analytes. For the purpose of providing context, the values of the absorption and reduced scattering coefficients of a number of clinically relevant sample types are as follows [37,38]: skin at 500 nm, $\mu_a \approx 1.3\ \mathrm{cm}^{-1}$, $\mu_s' \approx 30.6$–$68.7\ \mathrm{cm}^{-1}$, $\alpha \approx 31.9$–$70\ \mathrm{cm}^{-1}$; whole blood at 500 nm, $\mu_a \approx 100\ \mathrm{cm}^{-1}$, $\mu_s \approx 32\ \mathrm{cm}^{-1}$, $\alpha \approx 132\ \mathrm{cm}^{-1}$; water at 500 nm, $\mu_a \approx 0.0001\ \mathrm{cm}^{-1}$, $\mu_s' < 0.003\ \mathrm{cm}^{-1}$, $\alpha \approx 0.003\ \mathrm{cm}^{-1}$.

In order to elucidate the variation in the collection efficiency [defined in Eq. (6) above] as a function of magnification, NA, and α, a number of simulations are presented in Figs. 4 and 5. In Figs. 4(a)–4(c), the collection efficiency E is plotted as a function of NA for $\alpha = 0.00001\ \mathrm{cm}^{-1}$, $\alpha = 10\ \mathrm{cm}^{-1}$, and $\alpha = 100\ \mathrm{cm}^{-1}$, respectively. For all cases, the results are shown for a range of different magnifications that are commonly found in research microscopes: 2×, 4×, 10×, 20×, 40×, 50×, and 100×. The confocal pinhole diameter used in all of these calculations is 200 μm. Although the value of E is plotted as a function of the full range of NA for the various magnifications, the NA of low magnification objectives, such as 2× and 4×, is practically limited to approximately 0.05–0.1, while high magnification objectives such as 50× and 100× may have a range of NA from 0.8 to 1.4. MOs with a NA that is >1 require the use of an

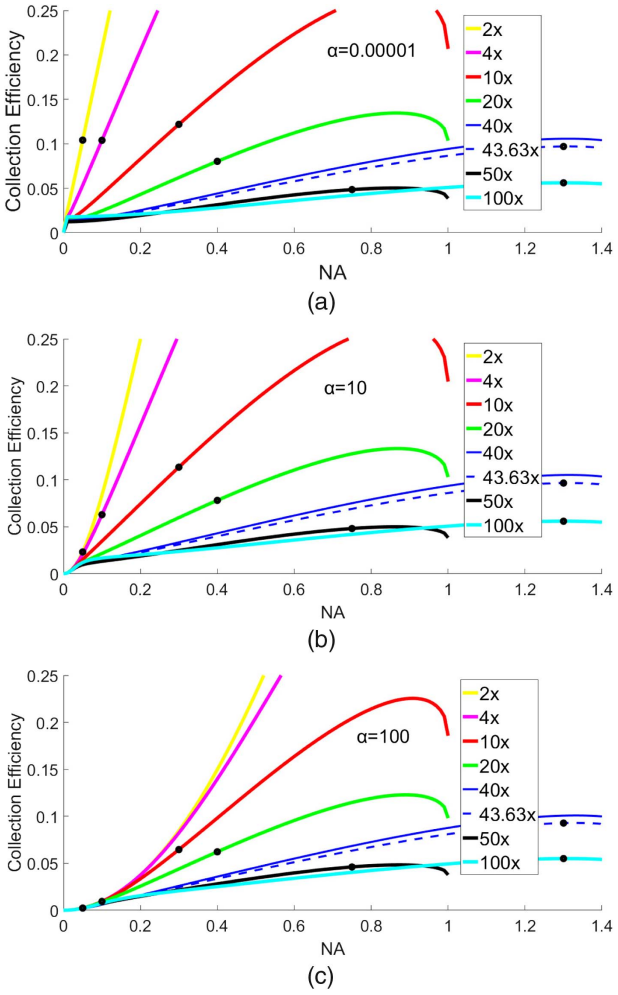

Fig. 4. Collection efficiency for various MOs with the different magnification for (a) $\alpha = 0.00001$, (b) 10 cm^{-1}, and (c) 100 cm^{-1}.

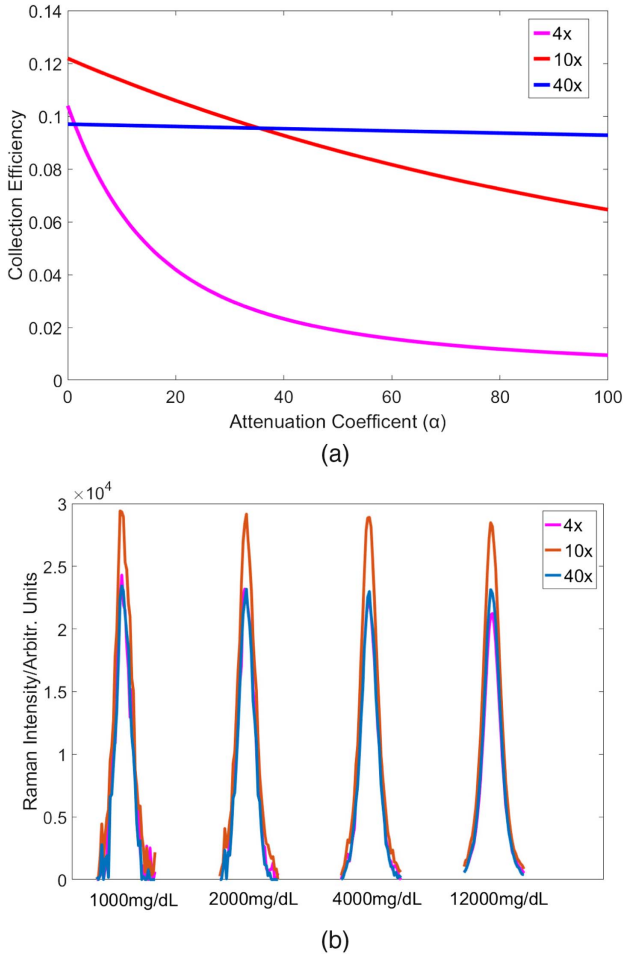

Fig. 5. (a) Collection efficiency for the 4×/0.1, 10×/0.3, and 40×/1.3 MOs listed in Table 1 as a function of the attenuation coefficient, α; (b) the intensity of the most prominent peak (in the range 801–861 cm^{-1}) recorded using the 4×, 10×, and 40× MOs, from a solution with increasing concentrations of lactic acid. For each concentration, the peaks for 4× and 10× are normalized to the peak recorded using the 40× MO. The same exposure time was used in all cases.

immersion medium such as water or oil; the results in Fig. 4 are based on a refractive index of $n = 1$ (air) for the 2×, 4×, 10×, 20×, and 50× magnifications, and a value of $n = 1.51$ (oil) for the 40× and 100× cases.

A set of high-quality commercial MOs that match the magnifications used in these calculations are listed in Table 1. The properties of each MO are shown in the figure based on the manufacturer specifications; the values of the FWHM for each case are calculated based on Eq. (2) for $\lambda = 532$ nm and $D = 200$ μm. For each of the MOs, the value of E is also listed corresponding to the three values of α used in Fig. 4, and these values are indicated with thick black dots in both figures.

For the case of the Zeiss 40×/0.13 MO, the effective magnification of this objective when used in an Olympus microscope, such as that in the Horiba system described in Section 5, it will be greater than 40×. Since Olympus and Zeiss microscopes use tube lenses with focal lengths of 180 mm and 165 mm, respectively, the effective magnification the Zeiss objective is 40× (180/165) = 43.636. A dashed blue line is included in each part of Fig. 4, representing this magnification. The transmittance of the 40× varies between 0.8 and 0.9 in the visible range of wavelengths; a value of 0.8 is selected to calculate the collection efficiency in the results presented here.

Figure 4(a) indicates that the collection efficiency is maximized for high NA and low magnification; however, these two properties are, in general, mutually exclusive for conventional MOs. It is clear that for low levels of scattering and absorption ($\alpha = 0.00001$ cm^{-1}) similar to the case of pure water, the lower magnification MOs, 2×, 4×, and 10× provide superior performance when compared with the higher magnification MOs, 20×, 50×, and 100× (with the 10× providing the optimal performance), which is due to the significantly higher values of FWHM for these objectives. This was confirmed experimentally by testing the performance of the various MOs listed in Table 1. The 40×/1.3 MO also provides high performance due to the relatively high NA for this level of magnification, which increases the number of photons that are collected from the relatively narrow FWHM.

Table 1. Specifications (Where Available) for Several MOs That Are Indicated in Fig. 4 Using Black Dots

Microscope Objective	Olympus Plan 2×/0.05	Olympus PlanN 4×/0.1	Olympus UplanFLN 10×/0.3	Olympus PlanN 20×/0.4	Zeiss Plan-Neofluar 40×/1.3 oil	Olympus MPlanN 50×/0.75	Olympus UPLFLN 100×/1.3 Oil
Magnification	2×	4×	10×	20×	43.63×	50×	100×
Numerical aperture	0.05	0.1	0.3	0.4	1.3	0.75	1.3
Immersion medium	air	air	air	air	oil	air	oil
T_{MO}	0.97	0.97	0.97	0.97	0.8–0.9	0.93	0.91
FWHM (cm)	0.22	5.52×10^{-2}	7.2×10^{-3}	2.7×10^{-3}	4.83×10^{-4}	4.97×10^{-4}	1.98×10^{-4}
$E_{(\alpha=0.00001)}$	0.104	0.104	0.121	0.08	0.097	0.049	0.056
$E_{(\alpha=10)}$	0.023	0.063	0.114	0.078	0.096	0.048	0.056
$E_{(\alpha=100)}$	0.002	0.009	0.065	0.062	0.093	0.046	0.055

The equivalent results are shown in Fig. 4(b) for a significantly higher attenuation ($\alpha = 10$ cm^{-1}). The performance of the lower-magnification objectives, 2× and 4×, deteriorates for this case; this is due to the increased attenuation over the relatively large FWHM for these objectives. In contrast, the MOs with a narrow FWHM remain almost unchanged. The 10×/0.3 MO provide optimal performance for all cases for the range of $0 < \alpha < 10$ cm^{-1} due to its competitive values for both FWHM and NA when compared with the other objectives.

The case of $\alpha = 100$ cm^{-1} is shown in Fig. 4(c), where the attenuation is similar to that for whole blood. The low-magnification MOs suffer a significant drop in performance, and the collection efficiency of the 10× MO reduces by ∼45% compared with the previous cases. The performance of the higher-NA lenses remains relatively stable. One can extend this argument to the case where the concentration becomes so high that the sample becomes solid powder. In this case, the high-NA MOs will continue to provide high performance while MOs with low NA will provide poor results due to the limited depth that the source can penetrate into the sample.

Three of the objectives (4×/0.1, 10×/0.3, and 40×/1.3) were selected for further investigation. In Fig. 5(a), the collection efficiency of these three objectives is plotted as a function of the attenuation coefficient in the range $0 < \alpha < 100$ cm^{-1}. The 10× MO is predicted to provide the best performance over the range $0 < \alpha < 35.43$ cm^{-1}, and beyond this range the 40× MO will provide optimal performance. The performance of the 10× MO is predicted to reduce by 45% over the full range, while the performance of the 4× is predicted to decrease more rapidly to ∼10% of its initial value. An experiment was carried out to investigate this behavior qualitatively by recording spectra from a solution of lactic acid with increasing concentration using these three MOs, and the results are shown in Fig. 5(b). The most prominent peak, located in the region of 801–861 cm^{-1}, is shown for concentrations of 1000, 2000, 4000, and 12,000 mg/dL. An attempt to quantitatively relate the value of α to a specific concentration of lactic acid is beyond the scope of this paper; however, it is possible to make a number of qualitative observations that support the theory presented here. For all four cases, the peak intensities are normalized to the intensity of the peak recorded using the 40× MO, since this objective is predicted to provide the most stable performance over all values of α. For the four different concentrations, the 10× MO provides the best performance, which is predicted by the theory for the range of $0 < \alpha < 35.43$ cm^{-1}. It is clear that as the concentration of lactic acid is increased, the performance of both the 4× and 10× MOs reduce relative to the 40× case, with the 4× MO deteriorating more rapidly, which is also predicted by the theory. Although these observations support the theory proposed here, a more comprehensive investigation is required for a full validation.

This section is concluded by defining the wavelength-dependent version of Eq. (6). In this case, $I_r(\lambda)$ refers to the irradiance of a specific wavelength component λ of the Raman spectrum on the detector,

$$I_r(\lambda) \propto T_{MO}^{\lambda} T_{MO}^{\lambda_s} \mathrm{NA}^2 \int_0^Q \exp[-L(\alpha_\lambda + \alpha_{\lambda_s})]\mathrm{d}L$$
$$Q = \min(\mathrm{FWHM}_{\lambda_s}, \mathrm{FWHM}_{\lambda}), \tag{7}$$

where λ_s denotes the source laser wavelength. The transmittance of the MO is now considered separately for λ_s and λ. The attenuation is also considered separately for both wavelengths; the value of the FWHM is wavelength dependent as are the absorption and scattering coefficients.

4. COMPARISON OF COLLECTION EFFICIENCIES OF DIFFERENT RAMAN SYSTEMS

In this section, a comparison of the collection efficiencies of the four systems illustrated in Fig. 1, as well as the confocal architecture, is provided. This comparison is approximate and is not intended to provide a rigorous analysis of each of these systems. In order to compare the various architectures, a more general definition of the collection efficiency is presented here:

$$E = T_{fb} \mathrm{NA}^2 \int_0^Q e^{-2\alpha L}\mathrm{d}L. \tag{8}$$

This equation is similar in form to that derived earlier in the paper for confocal microscopy [see Eq. (6)]; however, in this case the term T_{MO}^2 is dropped for the sake of simplicity, as the transmission loss associated with the delivery and collection paths will be similar for all cases and will be approximately equal to 1. An additional term, T_{fb}, is included to take into account loss in coupling the Raman scattered photons into the spectrograph for the case of using a fiber bundle, which has a limited collection area due to the physical separation of fibers.

In addition, the term FWHM is replaced with Q to more generally represent the depth into the sample that the laser can be delivered to and the scattered photons can be collected from. For the sake of convenience, it is assumed that the same microscope objective (10×/0.3) is used to collect the scattered photons from the sample. For cases in which a fiber bundle is used, the value of T_{fb} is taken to be 0.68, based on the design of commercially available bundles (Thorlabs; BFL200HS02); in all other cases, T_{fb} is taken to be 1. The collection efficiency is examined for the three cases of $\alpha = 0.5$, 10, and 100 cm^{-1}. For the sake of simplicity, it is assumed that all values of transmission, absorption, and scattering are constant for all wavelengths.

The setup in Fig. 1(a) employs a 90° collection geometry, which can significantly reduce the unwanted background signal by separating the delivery and collection paths. For the ideal case, it can be assumed that the source laser remains focused over an extended length within the sample, which can be imaged to the full length of spectrograph slit. Therefore, the value of Q in Eq. (8) is given by the spectrograph slit length divided by the magnification, which for the spectrograph/charge-coupled device (CCD) used in our experiment, this gives a value of $Q = 0.64$ mm.

Equation (8) takes into account the propagation of the Raman scattered photons back along the same path from which the laser was delivered. In this case, however, if the scattering takes place close to the edge of the container, it can be assumed that the path length of the Raman scattered photons within the sample is approximately zero. Rather than adapt Eq. (8) for this special case, the value of Q is halved to give $Q = 0.32$ mm.

The setup in Fig. 1(b) employs a fiber bundle. The fibers are arranged linearly in the terminal end and coupled with the slit of spectrograph. The finite diameter of the collection end of the fiber bundle has the same effect as a confocal aperture. This diameter is taken to be 640 μm (Thorlabs; BFL200HS02). This system can be considered to be the same as the confocal Raman microscope considered in this paper; however, it has a significantly larger confocal aperture, which will result in greater collection efficiency. In this case, the large confocal aperture will not reduce the spectral resolution, due to the linear arrangement of the fibers at the output. However, the large confocal aperture will capture a more intense background signal, which will limit the performance of the system, as discussed in Section 3.A. For this system, the calculation of Q is based on the FWHM in Eq. (2) and is calculated to be $Q = 0.23$ mm.

The setup in Fig. 1(c) employs a non-imaging paraboloidal mirror to maximize the signal that can be collected from a 1 mm^2 spot area with a large collection solid angle, and is designed to achieve optimal collection efficiency for a whole blood sample. The optical properties of this system, in terms of collecting photons from an extended depth in a sample, cannot be discerned, and it is, therefore, not possible to describe this system using Eq. (8), so it is not included in Table 2.

The LCOF system illustrated in Fig. 1(d) enhances the Raman spectrum by guiding the laser over the length of a waveguide and collecting the Raman photons that are guided back to the end of the fiber. A comprehensive analysis of the collection efficiency of this system is provided by Qi [19]. Here we provide a simplified model, once again using Eq. (8), where Q is taken to be the length of the LCOF, which is assumed to be 2 m.

Table 2. Comparison of Setups Illustrated in Fig. 1

	Fig. 1(a) [14]	Fig. 1(b) [11]	Fig. 1(d) [21]	Confocal
Q	320 μm	230 μm	2 m	72 μm
Back-ground	low	high	low	low
$E_{(\alpha=0.5)}$	0.567	0.275	12.30	0.129
$E_{(\alpha=10)}$	0.425	0.226	0.615	0.121
$E_{(\alpha=100)}$	0.089	0.061	0.061	0.068

In Table 2, the four systems illustrated in Fig. 1 are compared using the definition of collection efficiency given in Eq. (8). The value of Q relates to the depth in the sample from which scattering can be collected, and it can be seen that the confocal architecture has the worst performance in this regard. For samples with low attenuation ($\alpha = 0.5$ cm^{-1}), the LCOF system is the best performer with a collection efficiency that is between 25 and 100 times better than the other systems, while the confocal system has approximately half the collection efficiency of the traditional architecture for Raman multicomponent analysis proposed in Ref. [11]. For highly attenuating samples, such as whole blood, the confocal architectures are predicted to perform as well as the other systems.

5. MATERIALS AND METHODS

In this section, we outline an experiment to examine the performance of the confocal Raman microscope for multicomponent analysis. In terms of chemicals, concentrations, mixtures, and processing, the experiment is similar to that used in Ref. [14].

A. Confocal Raman Microscope

All spectra were recorded using a commercial confocal Raman microscope (Horiba Jobin Yvon LabRam 800 HR). All of the experiments used a 600 lines/mm diffraction grating in a spectrograph with focal length of 800 mm. This system uses a 100 mW single-mode diode-pumped solid-state (DPSS) laser with a wavelength of 532 nm (Torus; Laser Quantum, UK). A back-illuminated, cooled, CCD detector (Synapse; Horiba, Japan) operating at −80°C, was used to record all spectra. This camera has a typical read noise of 5 electrons and a dark current of 0.002 electrons/pixel/s. A 10× microscope objective (UMPlanFl 10 × /0.3; Olympus, Japan) and a confocal aperture of 200 μm were used for all experiments unless otherwise indicated. The MO focuses the laser onto the sample, which is contained within a cuvette, with a base made from a Raman-grade calcium fluoride coverslip with a thickness of ∼200 μm (Crystran, UK), which provides a spectral resolution of approximately ∼10 cm^{-1} using the spectrograph and grating mentioned above. In order to reduce the effect of wavenumber shifting due to temperature variation, all experiments were conducted in a temperature-controlled laboratory. The wavenumber range of all Raman spectra recorded was 402–2048 cm^{-1}.

Raman spectra are recorded from all solutions using the UplanFLN10 × /0.3 MO; for all cases, three spectra with 20 s integration time were recorded and averaged together

to produce a single spectrum with an effective acquisition time of 60 s.

B. Sample Preparation

Powders of glucose (≥99.5%, G7528; Sigma-Aldrich, Ireland), urea (≥98%, U5378; Sigma-Aldrich, Ireland), and lactic acid (≥98%, L1750; Sigma-Aldrich, Ireland) were mixed based on the specific weight and diluted in deionized water into 19 mixtures with varying concentrations. The concentration of each chemical used in all 19 mixtures is shown in Table 3. Since lactic acid in solution can dissociate depending on different pH environments, the measurable concentration of lactic acid will change in solution [14]. In this experiment, the pH of each solution was measured using a pH meter (Eu Tech Instrument). Based on the measured concentration of lactic acid and the measured pH, the actual weight of lactic acid that was added to each solution can be calculated using the formula below:

$$C_{\mathrm{la}} = (1 - 1.38 \times 10^{(-4+\mathrm{pH})}) \times C_{\mathrm{ex}}, \tag{9}$$

where C_{ex} is the actual concentration (in the sense of the mass of lactic acid that was added to the volume of water), and 1.38×10^{-4} is the dissociation constant of lactic acid. Table 3 shows the list of concentrations and pH values for each solution.

In addition to the 19 mixtures listed in Table 3, three Raman spectra were recorded from "pure" solutions of glucose, lactic acid, and urea in isolation; these spectra are used in the final pre-processing step discussed below. A concentration of 5 g/dL was used for each case.

C. Numerical Pre-processing

The raw spectra that were recorded cannot be immediately input to a multivariate statistical analysis for the purpose of estimating component concentration due to the presence of noise and the unwanted baseline signal that varies across the set of recorded spectra. Therefore, it is necessary to remove or reduce the impact of these interferences before performing multicomponent analysis.

Cosmic ray artefacts appear in the spectrum in the form of spurious, narrow-band peaks with high intensity. The three raw 20 s spectra are used to remove cosmic rays; corresponding pixels are compared across the three spectra, and an intensity difference that is greater than the expected noise amplitude identifies the presence of a cosmic ray. Regions of the spectra that contain cosmic rays are omitted from the averaging process.

Following cosmic ray removal, smoothing is performed in order to reduce the impact of shot noise, using an algorithm that combines maximum likelihood estimation and Savitzky–Golay (SG) smoothing [41].

This algorithm has been demonstrated to perform better than other smoothing algorithms such as traditional SG smoothing filter and can improve the SNR of the input spectrum by >100% while also preserving the underlying spectral peaks [41].

The final pre-processing step is to subtract the unwanted baseline that varies for each recording. In order to estimate this baseline, a classical least-squares algorithm is applied to the raw spectrum that calculates the best fit of a set of component background spectra that make up the raw spectrum, as well as an N-order polynomial. This algorithm is similar to the well-known extended multiplicative signal correction (EMSC) algorithm that is commonly used to remove the unwanted baseline, including background signals such as the spectrum from glass components [42]. The least-squares algorithm used here assumes that the raw spectrum is composed of a linear weighted sum of reference spectra recorded from samples of (i) glucose, (ii) lactic acid, and (iii) urea, shown in Fig. 6(a) (for each of these three reference spectra, an automated background subtraction algorithm was applied [42]); (iv) a spectrum recorded from a pure water sample; this spectrum contains contributions from the water, the calcium fluoride window, the MO, and other optical elements (no baseline subtraction is applied to this spectrum), and finally (v) an N-order polynomial to account for the varying baseline [27]. The least-squares algorithm determines the weight of each of these five components (this is implemented separately for each term in the N-order polynomial) to optimally fit their sum to the raw spectrum. The final spectrum is obtained by subtracting each of the weighted background components from the raw spectrum. A polynomial of order $N = 7$ is used in this algorithm.

D. Partial Least-Squares Regression

PLSR is a multivariate statistical method that is commonly used in Raman multicomponent analysis. All the experiments reviewed in Section 2.A employed PLSR. The principle of this technique is to decompose a set of independent variables (in this case, the concentrations of the various analytes) and a set of corresponding dependent variables (a matrix of Raman spectra related to these components) into sets of scores and sets of corresponding loadings, and to find the maximum covariance between the scores of the independent variables and the scores of the dependent variables. In the process of establishing the PLS predictive model, choosing an optimal number

Table 3. List of the Concentrations (mg/dL) of Glucose, Urea, and Lactic Acid in the 19 Mixtures Used in This Experiment

	Glucose (mg/dL)	Lactic Acid (mg/dL)	Urea (mg/dL)	pH
1	415.32	191.89	88.4	2.69
2	606.52	133.13	108.6	2.78
3	675.24	54.87	172.2	2.98
4	271.24	155.11	189.16	2.78
5	255.78	84.56	253.94	2.89
6	208.06	141.31	70	2.76
7	183.6	183.52	126.25	2.63
8	255.36	191.09	51.9	2.68
9	546.46	55.25	132.16	2.96
10	569.82	200.21	201.74	2.66
11	497.82	133.05	244.6	2.77
12	233.74	89.22	252.36	2.83
13	363.98	212.20	56.54	2.64
14	160.18	172.70	228.76	2.76
15	682.92	225.40	227.5	2.69
16	785.1	91.13	101.9	2.93
17	203.38	144.27	75.16	2.76
18	180.08	184.12	131.89	2.72
19	302.7	215.46	54.79	2.68

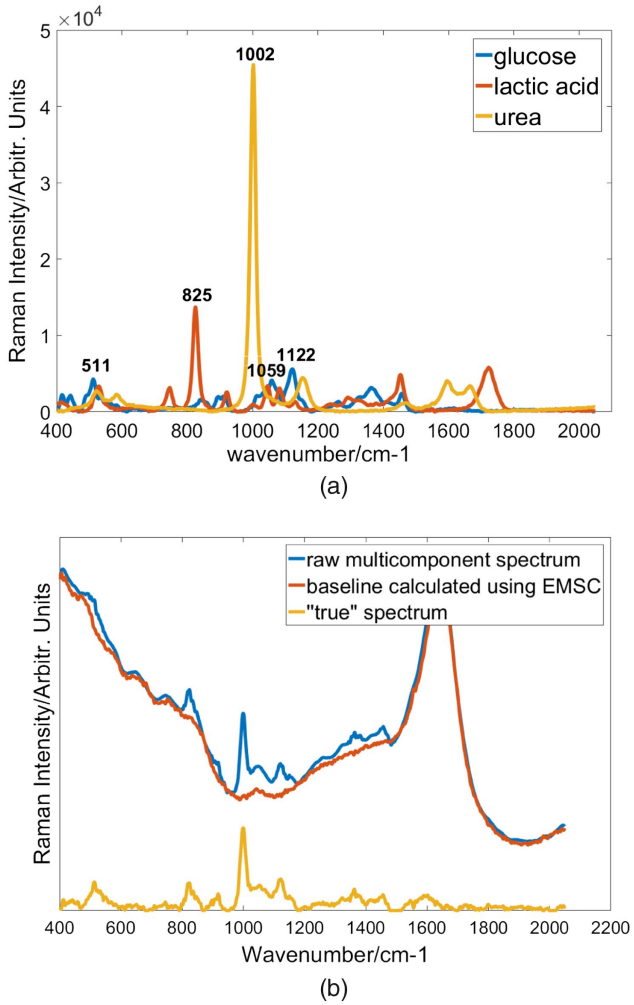

Fig. 6. (a) Reference spectra of glucose, lactic acid, and urea after processing. These spectra are used in the EMSC algorithm for removing the background from each component spectrum. (b) A raw spectrum from one of the 19 mixtures following cosmic ray removal and smoothing (blue), the baseline that is calculated using the EMSC algorithm (red), and the corrected spectrum (green).

of components is important in order to reduce the presence of noise.

The root mean square error (RMSE; the error between the reference concentrations and the predicted concentration) and the coefficient of determination (R^2; the correlation between the predicted and reference values), are two important metrics that are commonly employed to determine the appropriate number of components to use and to estimate the error of the model in terms of predicting the concentrations of an unknown mixture [1–19]. In the experiment presented here, RMSE and R^2 are used to evaluate the performance of the model based on leave-one-out cross-validation.

6. RESULTS

Raw spectra were recorded from individual solutions of glucose, lactic acid, and urea, in deionised water, as well as from a pure water sample; these spectra were processed for cosmic ray removal, smoothing, and background subtraction, as described in Section 5, to create the set of reference spectra for input to the EMSC algorithm. The reference spectra for glucose, lactic acid, and urea are shown in Fig. 6(a). As described in Section 5.C and Fig. 6(b), this algorithm estimates (and subtracts) the unwanted baseline from each of the spectra recorded from the 19 mixtures. These 19 pure spectra are independent variables for use in the PLS model for the three chemicals. The RMSE of the predictions vary as a function of the number of PLS components that are used in the model for each of the three chemicals, and this variation is shown in Fig. 7(a).

The optimal number of components to be used by the PLS algorithm is the number that results in the smallest RMSE. A crucial property of these components is that they should accurately represent the characteristics of the signal. The number of components that are selected should include many of the key spectral features of the analyte spectra, while omitting components with low SNRs. The coefficient of determinant (R^2) can be used to represent the robustness of the PLS model; in this experiment, both RMSE and R^2 are used to determine that the optimal number of components for glucose and urea is five, and for lactic acid the optimal number of components is four.

The predictions of PLS cross-validation are plotted for the three chemicals in Figs. 7(b)–7(d). The RMSE of the model for glucose is calculated to be 17.81 mg/dL with $R^2 = 0.93$; the RMSE for lactic acid is 10.6 mg/dL with $R^2 = 0.91$; and the RMSE for urea is 7.6 mg/dL with $R^2 = 0.96$. The accuracy of these predictions and the accuracy of similar predictions found in previous studies are compared in Table 4. The results presented here are comparable with those found in Goetz *et al.* [14] and Berger *et al.* [11], which were also based on three analyte mixtures. Also included in the table are the results of Enejder *et al.* [17] and Qi and Berger [19], both of which are based on body fluid (whole blood, blood serum, and urine), and, therefore, include several more analytes. Disregarding the increased complexity in these models due to the inclusion of more analytes, the results of the predictions presented here

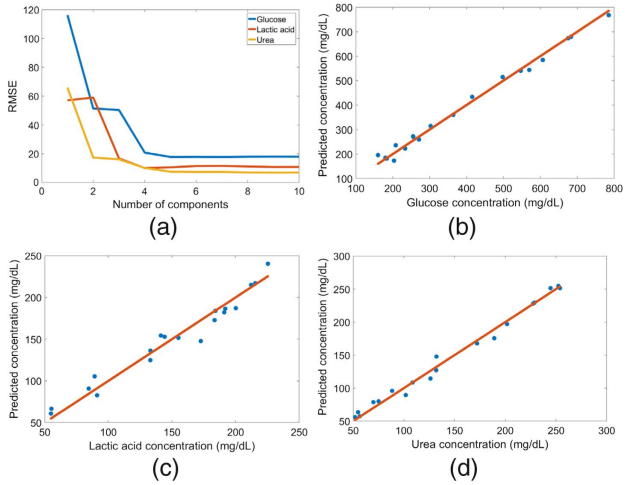

Fig. 7. (a) Number of components used in the PLS model versus RMSE for glucose, lactic acid, and urea; (b) result of PLS for glucose using five components; (c) prediction result of PLS for lactic acid using four components; (d) result of PLS for urea using five components.

Table 4. Comparison of Results with Other Raman Setups

Results Using Confocal Raman Microscopy (60 s Acquisition Time)		
Analytes	**RMSE (mg/dL)**	R^2
Glucose	17.81	0.93
Lactic acid	10.6	0.91
Urea	7.6	0.96
Goetz *et al.* (1995) [14] (40 s Acquisition Time)		
Analytes	**RMSE (mg/dL)**	R^2
Glucose	20.17	0.99
Lactic acid	12.92	0.94
Urea	19.07	0.92
Berger *et al.* (1996) [11] (100 s Acquisition Time)		
Analytes	**RMSE (mg/dL)**	R^2
Glucose	21.62	N/A
Creatinine	13.57	N/A
Lactic acid	11.71	N/A
Enejder *et al.* Using Parabolic Mirror [17] (5 min Acquisition Time)		
Analytes	**RMSE (mg/dL)**	R^2
Glucose	21	0.97
Urea	4.9	0.94
Qi and Berger (2007) [21] Using LCOFs (64 s Acquisition Time for Creatinine, 150 s for Glucose)		
Analytes	**RMSE (mg/dL)**	R^2
Glucose	8.8	N/A
Creatinine	4.3	N/A

for the three chemical mixtures are similar to the prediction results in these studies. The acquisition time used in the experiment presented here is 60 s, which is also consistent with the acquisition times used in the previous experiments, which vary significantly from 40 s to 5 min.

7. CONCLUSION

In this paper, the conventional confocal Raman microscope has been investigated to simultaneously measure the concentration of multiple chemicals in aqueous solution, which is commonly referred to as multicomponent analysis. Previous investigations have focused on developing optical systems that maximize the number of Raman scattered photons that can be collected from blood and urine samples, thereby maximizing the SNR of the recorded spectrum and reducing the necessary acquisition times, and these systems have been reviewed in Section 2.A. The best performing system that has been proposed to date, in terms of acquisition time, the number of analytes that can be simultaneously measured, and the minimum concentration that can be detected for each analyte, is based on the use of liquid-core optical fibers. These fibers can deliver the source laser over an increased volume of the solution and guide the scattered photons to the detector. While the confocal Raman microscope cannot match the same level of performance for multicomponent analysis, it has the advantage of being readily applicable to a multitude of different sample types and can be found in many research laboratories. Here, we have investigated the performance that can be expected from a confocal Raman microscope for multicomponent analysis both theoretically and experimentally.

In Section 3, the confocal Raman microscope is discussed. In particular, the efficiency of this system is analyzed in terms of the collection of Raman scattered photons from an aqueous solution. The role of the confocal pinhole is to (i) to provide a depth resolution and (ii) to reduce unwanted scattering that originates in optical elements that are common to the (laser) delivery and collection paths, such as the microscope objective. While the reduction of the unwanted background signal is especially important in Raman multicomponent analysis, the limited depth resolution is counterproductive, as this significantly reduces the scattering that can be obtained from a homogenous scatterer such as an aqueous solution. Increasing the pinhole size is an obvious solution to this problem; however, as demonstrated in Section 3.B.1, an increase in pinhole size has the unwanted effect of increasing the spectral resolution. A pinhole size of 200 μm was found to be a good compromise between these two effects, providing a resolution of ~10 cm^{-1}.

The interaction between the confocal aperture and the microscope objective was also investigated in terms of collection efficiency. Objectives with low NA can collect scattering over a significantly larger depth of field when compared to high-NA lenses for a given confocal pinhole size; however, high-NA lenses have the advantage of a larger solid angle over which scattered photons are collected. A theoretical investigation of these properties reveals that, in general, a high ratio of NA to magnification is desirable. The theory proposed here also takes into account scattering and absorption within the sample and the resulting attenuation of the source laser as it propagates (and diffuses) through the sample. For highly absorbing/scattering samples, the depth of field over which Raman scattering can be collected is reduced, and thus the advantage of low-magnification, low-NA objectives is negated. The experimental results presented in the paper are consistent with the theory, and it was found that a 10 × /0.3 microscope objective provided the best results for all mixtures that were investigated in this paper.

In Section 4 we extended the model for collection efficiency to estimate the performance of the various customized systems reviewed in Section 2.A. This analysis predicts that the LCOF system provides significantly superior performance for samples with low absorption and scattering coefficients for equivalent exposure times; however, as sample absorption/scattering increases, the confocal Raman system offers similar performance to the other systems.

Multicomponent analysis was performed using 19 mixtures of glucose, urea, and lactic acid in an experiment that closely emulates the initial work of Goetz *et al.* [14]. The experiment is described in detail in Section 5 and used a commercial high-resolution confocal microscope with a 60 s acquisition time. A rigorous pre-processing protocol is also proposed in Section 5 to reduce the noise in the recorded spectra and to remove the unwanted baseline that varies across the recordings.

The latter technique is based on the extended multiplicative scattering correction algorithm and is similar to that used in previous investigations. The results of PSLR applied to the 19 pre-processed spectra are presented in Section 6 and evaluated using the commonly used metrics of RMSE and the coefficient of determination. The results are shown to be similar to those obtained in previous studies, with only the liquid-core optical fiber approach showing markedly better measurements. It must be noted that the confocal Raman microscope can match the performance of any of the systems described in Section 2.A, including the liquid-core optical fiber system, if a sufficiently long acquisition time is used. Modern detectors have dark current values that are almost negligible (<0.001 electrons/pixel/second), and, therefore, shot noise is the only noise source that needs to be considered. Increasing the acquisition time will, therefore increase the SNR; however, the rate of increase over time will slow down. Recording times in the order of several minutes may be required to fully match the performance of the liquid-core systems.

It has been previously suggested that the shot noise associated with the water in the sample poses a fundamental limit for Raman multicomponent analysis in terms of the minimum concentration that can be measured. Although water is a relatively weak Raman scatterer, the abundance of water molecules relative to the molecules of interest results in an appreciably large spectrum from the water. The mean water spectrum can be subtracted, but the shot noise from this spectrum remains, the amplitude of which places a limit on the smallest peak that can be detected. The same argument can be applied to any of the background spectra that are subtracted in the pre-processing step described in Section 5. Since the confocal aperture optically filters much of this unwanted background from reaching the detector, the confocal Raman microscope is well suited to reducing this noise source (shot noise from background signals). Noise from the water spectrum cannot be removed in the same way; however, recently it has been proposed that removing water from the sample can help improve the spectroscopic measurement of human serum [43]. This approach could significantly reduce the minimum measurable concentrations using confocal Raman multicomponent analysis and may offer an exciting avenue for future research.

Funding. Science Foundation Ireland (SFI) (15/CDA/3667, 15/IA/3167).

REFERENCES

1. J. A. M. Bispo, E. E. de Sousa Vieira, L. Silveira, and A. B. Fernandes, "Correlating the amount of urea, creatinine, and glucose in urine from patients with diabetes mellitus and hypertension with the risk of developing renal lesions by means of Raman spectroscopy and principal component analysis," J. Biomed. Opt. **18**, 087004 (2013).
2. R. Pandey, S. K. Paidi, T. A. Valdez, C. Zhang, N. Spegazzini, R. R. Dasari, and I. Barman, "Noninvasive monitoring of blood glucose with Raman spectroscopy," Acc. Chem. Res. **50**, 264–272 (2017).
3. S. Pilotto, M. Pacheco, L. Silveira, Jr., A. B. Villaverde, and R. Zangaro, "Analysis of near-infrared Raman spectroscopy as a new technique for a transcutaneous non-invasive diagnosis of blood components," Laser Med. Sci. **16**, 2–9 (2001).
4. K. Virkler and I. K. Lednev, "Raman spectroscopic signature of blood and its potential application to forensic body fluid identification," Anal. Bioanal. Chem. **396**, 525–534 (2010).
5. J. Y. Qu, B. C. Wilson, and D. Suria, "Concentration measurements of multiple analytes in human sera by near-infrared laser Raman spectroscopy," Appl. Opt. **38**, 5491–5498 (1999).
6. J. Y. Qu and L. Shao, "Near-infrared Raman instrument for rapid and quantitative measurements of clinically important analytes," Rev. Sci. Instrum. **72**, 2717–2723 (2001).
7. A. M. Enejder, T. G. Scecina, J. Oh, M. Hunter, W. Shih, S. Sasic, G. L. Horowitz, and M. S. Feld, "Raman spectroscopy for noninvasive glucose measurements," J. Biomed. Opt. **10**, 031114 (2005).
8. C. J. Saatkamp, M. L. de Almeida, J. A. M. Bispo, A. L. B. Pinheiro, A. B. Fernandes, and L. Silveira, "Quantifying creatinine and urea in human urine through Raman spectroscopy aiming at diagnosis of kidney disease," J. Biomed. Opt. **21**, 037001 (2016).
9. S. J. Barton, B. M. Hennelly, T. Ward, K. Domijan, and J. Lowry, "A review of Raman for multicomponent analysis," in *Biophotonics: Photonic Solutions for Better Health Care IV* (International Society for Optics and Photonics, 2014), Vol. **9129**, paper 91290C.
10. J. W. McMurdy III and A. J. Berger, "Raman spectroscopy-based creatinine measurement in urine samples from a multipatient population," Appl. Spectrosc. **57**, 522–525 (2003).
11. A. J. Berger, Y. Wang, and M. S. Feld, "Rapid, noninvasive concentration measurements of aqueous biological analytes by near-infrared Raman spectroscopy," Appl. Opt. **35**, 209–212 (1996).
12. H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, and M. J. Walsh, "Using Raman spectroscopy to characterize biological materials," Nat. Protocols **11**, 664–687 (2016).
13. I. R. Lewis and H. Edwards, *Handbook of Raman Spectroscopy: from the Research Laboratory to the Process Line* (CRC Press, 2001).
14. M. J. Goetz, G. L. Cote, R. Erckens, W. March, and M. Motamedi, "Application of a multivariate technique to Raman spectra for quantification of body chemicals," IEEE Trans. Biomed. Eng. **42**, 728–731 (1995).
15. A. J. Berger, I. Itzkan, and M. S. Feld, "Feasibility of measuring blood glucose concentration by near-infrared Raman spectroscopy," Spectrochim. Acta A **53**, 287–292 (1997).
16. A. J. Berger, T.-W. Koo, I. Itzkan, G. Horowitz, and M. S. Feld, "Multicomponent blood analysis by near-infrared Raman spectroscopy," Appl. Opt. **38**, 2916–2926 (1999).
17. A. M. Enejder, T.-W. Koo, J. Oh, M. Hunter, S. Sasic, M. S. Feld, and G. L. Horowitz, "Blood analysis by Raman spectroscopy," Opt. Lett. **27**, 2004–2006 (2002).
18. D. Rohleder, W. Kiefer, and W. Petrich, "Quantitative analysis of serum and serum ultrafiltrate by means of Raman spectroscopy," Analyst **129**, 906–911 (2004).
19. D. Qi and A. J. Berger, "Quantitative analysis of Raman signal enhancement from aqueous samples in liquid core optical fibers," Appl. Spectrosc. **58**, 1165–1171 (2004).
20. L. T. Kerr, H. J. Byrne, and B. M. Hennelly, "Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen," Anal. Methods **7**, 5041–5052 (2015).
21. D. Qi and A. J. Berger, "Chemical concentration measurement in blood serum and urine samples using liquid-core optical fiber Raman spectroscopy," Appl. Opt. **46**, 1726–1734 (2007).
22. D. Qi and A. J. Berger, "Quantitative concentration measurements of creatinine dissolved in water and urine using Raman spectroscopy and a liquid core optical fiber," J. Biomed. Opt. **10**, 031115 (2005).
23. R. Altkorn, I. Koev, and M. J. Pelletier, "Raman performance characteristics of teflon-af 2400 liquid-core optical-fiber sample cells," Appl. Spectrosc. **53**, 1169–1176 (1999).
24. R. Altkorn, M. D. Malinsky, R. P. Van Duyne, and I. Koev, "Intensity considerations in liquid core optical fiber Raman spectroscopy," Appl. Spectrosc. **55**, 373–381 (2001).
25. D. Qi and A. J. Berger, "Correction method for absorption-dependent signal enhancement by a liquid-core optical fiber," Appl. Opt. **45**, 489–494 (2006).
26. A. J. Berger, "Raman spectroscopy of blood and urine specimens," in *Emerging Raman Applications and Techniques in Biomedical and Pharmaceutical Fields* (Springer, 2010), pp. 385–404.
27. F. Bonnier, S. M. Ali, P. Knief, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. C. Lee, F. M. Lyng, and H. J. Byrne, "Analysis of human skin

tissue by Raman microspectroscopy: dealing with the background," Vib. Spectrosc. **61**, 124–132 (2012).
28. C. Hammond, *Cellular and Molecular Neurophysiology* (Academic, 2014).
29. S. P. Reddy, P. Ramani, and P. Nainani, "Confocal microscopy and exfoliative cytology," J. Oral Maxillofac. Pathol. **17**, 217–221 (2013).
30. B. Matsumoto, *Cell Biological Applications of Confocal Microscopy* (Academic, 2003), Vol. **70**.
31. W. Hoheisel, W. Jacobsen, B. Lüttge, and W. Weiner, "Confocal microscopy: applications in materials science," Macromol. Mater. Eng. **286**, 663–668 (2001).
32. D. Hovis and A. Heuer, "The use of laser scanning confocal microscopy (LSCM) in materials science," J. Microsc. **240**, 173–180 (2010).
33. C. Franck, S. Hong, S. Maskarinec, D. Tirrell, and G. Ravichandran, "Three-dimensional full-field measurements of large deformations in soft materials using confocal microscopy and digital volume correlation," Exp. Mech. **47**, 427–438 (2007).
34. L. T. Kerr, T. M. Lynn, I. M. Cullen, P. J. Daly, N. Shah, S. O'Dea, A. Malkin, and B. M. Hennelly, "Methodologies for bladder cancer detection with Raman based urine cytology," Anal. Methods **8**, 4991–5000 (2016).
35. S. T. McCain, M. E. Gehm, Y. Wang, N. P. Pitsianis, and D. J. Brady, "Coded aperture Raman spectroscopy for quantitative measurements of ethanol in a tissue phantom," Appl. Spectrosc. **60**, 663–671 (2006).
36. T. Wilson, "Resolution and optical sectioning in the confocal microscope," J. Microsc. **244**, 113–121 (2011).
37. S. L. Jacques, "Optical properties of biological tissues: a review," Phys. Med. Biol. **58**, R37–R61 (2013).
38. I. S. Saidi, "Transcutaneous optical measurement of hyperbilirubinemia in neonates," Ph.D. thesis (Rice University, 1992).
39. L. Wang and S. L. Jacques, "Use of a laser beam with an oblique angle of incidence to measure the reduced scattering coefficient of a turbid medium," Appl. Opt. **34**, 2362–2366 (1995).
40. M. Johns, C. A. Giller, D. C. German, and H. Liu, "Determination of reduced scattering coefficient of biological tissue from a needle-like probe," Opt. Express **13**, 4828–4842 (2005).
41. S. Barton, T. Ward, and B. M. Hennelly, "Algorithm for optimal denoising of Raman spectra," Analyst (submitted).
42. B. D. Beier and A. J. Berger, "Method for automated background subtraction from Raman spectra containing known contaminants," Analyst **134**, 1198–1202 (2009).
43. F. Bonnier, F. Petitjean, M. J. Baker, and H. J. Byrne, "Improved protocols for vibrational spectroscopic analysis of body fluids," J. Biophoton. **7**, 167–179 (2014).