

# Multi-excitation Raman spectroscopy technique for fluorescence rejection

Scott T. McCain, Rebecca M. Willett, and David J. Brady

Fitzpatrick Institute for Photonics, Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, 27708

[dbrady@duke.edu](mailto:dbrady@duke.edu)

**Abstract:** Multi-excitation Raman spectroscopy filters out Raman signals from a fluorescent background by sequentially using multiple excitation frequencies. The filtering method exploits the shift of the Raman spectra with excitation frequency and the static response of the fluorescent background. This technique builds upon previous work which used two slightly shifted excitations, Shifted Excitation Raman Difference Spectroscopy (SERDS), in order to filter the Raman signal. An Expectation-Maximization algorithm is used to estimate the Raman and fluorescence signals from multiple spectra acquired with slightly shifted excitation frequencies. In both simulation and experiment, the efficacy of the algorithm increases with the number of excitation frequencies even when holding the total excitation energy constant, such that the signal to noise ratio is inversely proportional to the number of excitation frequencies. In situations where the intense fluorescence causes significant shot noise compared to the weak Raman signals, the multi-excitation approach is more effective than non-iterative techniques such as polynomial background subtraction.

© 2008 Optical Society of America

**OCIS codes:** (300.6450) Spectroscopy, Raman; (100.3190) Inverse problems; (170.6510) Spectroscopy, tissue diagnostics.

---

## References and links

1. J. R. Ferraro, K. Nakamoto, and C. W. Brown, *Introductory Raman Spectroscopy*, 2nd Ed. (Academic Press, San Diego, Cal., 2003).
2. P. P. Yaney, "Reduction of fluorescence background in Raman spectra by the pulsed Raman technique," *J. Opt. Soc. Am.* **62**, 1297–1303 (1972).
3. A. P. Shreve, N. J. Cherepy, and R. A. Mathies, "Effective rejection of fluorescence interference in Raman spectroscopy using a shifted excitation difference technique," *Appl. Spectrosc.* **46**, 707–711 (1992).
4. S. H. Mosier-Boss, S. H. Lieberman, and R. Newbery, "Fluorescence rejection in Raman spectroscopy by shifted-spectra, edge detection, and FFT filtering techniques," *Appl. Spectrosc.* **49**, 630–638 (1995).
5. P. Matousek, M. Towrie, A. Stanley, and A. Parker, "Efficient rejection of fluorescence from Raman spectra using picosecond Kerr gating," *Appl. Spectrosc.* **53**, 1485–1489 (1999).
6. J. Zhao, M. M. Carrabba, and F. S. Allen, "Automated fluorescence rejection using shifted excitation Raman difference spectroscopy," *Appl. Spectrosc.* **56**, 834–845 (2002).
7. C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl. Spectrosc.* **57**, 1363–1367 (2003).
8. P. Matousek, M. Towrie, and A. W. Parker, "Simple reconstruction algorithm for shifted excitation Raman difference spectroscopy," *Appl. Spectrosc.* **59**, 848–851 (2005).
9. M. Kasha, "Characterization of electronic transitions in complex molecules," *Discuss. Faraday Soc.* **9**, 14–19 (1950).
10. S. E. J. Bell, E. Bourguignon, and A. Dennis, "Analysis of luminescent samples using subtracted shifted Raman spectroscopy," *Analyst* **123**, 1729–1734 (1998).

11. L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.* **79**, 745754 (1974).
12. W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Am.* **62**, 5559 (1972).
13. L. A. Shepp and Y. Vardi, "Maximum likelihood reconstructions for emission tomography," *IEEE Trans. Med. Imaging* **MI-1**, 113122 (1982).
14. S. E. Bialkowski, "Overcoming the multiplex disadvantage by using maximum-likelihood inversion," *Appl. Spectrosc.* **52**, 591–598 (1998).
15. D. R. Fuhrmann, C. Preza, J. A. O'Sullivan, D. L. Snyder, and W. Smith, "Spectrum estimation from quantum-limited interferograms," *IEEE Trans. Signal Process.* **52**, 950–961 (2004).
16. C. V. Raman, "A new radiation," *Indian J. Phys.* **2**, 387 (1928).
17. R. Nowak and E. Kolaczyk, "A multiscale statistical framework for Poisson inverse problems," *IEEE Trans. Inf. Theory* **46**, 1811–1825 (2000).
18. R. Willett, "Multiscale intensity estimation for marked Poisson processes," in *Proc. IEEE Conf. Acoust., Speech, Signal Processing - ICASSP* (2007).
19. E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Annals of Stat.* **32**, 500–527 (2004).
20. R. Willett and R. Nowak, "Multiscale Poisson intensity and density estimation," *IEEE Trans. Inf. Theory* **53**, 3171–3187 (2005).
21. M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Process. Lett.* **3**, 10–12 (1996).
22. R. Willett and R. Nowak, "Fast multiresolution photon-limited image reconstruction," in *Proc. IEEE Int. Sym. Biomedical Imaging — ISBI '04* (15–18 April, Arlington, VA, USA, 2004).
23. E. F. Y. Hom, F. Marchis, T. K. Lee, S. Haase, D. A. Agard, and J. W. Sedat, "AIDA: an adaptive image deconvolution algorithm with application to multi-frame and three-dimensional data," *J. Opt. Soc. Am. A* **24**, 1580–1600 (2007).
24. 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).
25. M. E. Gehm, S. T. McCain, N. P. Pitsianis, D. J. Brady, P. Potulri, and M. E. Sullivan, "Static 2D aperture coding for multimodal multiplex spectroscopy," *Appl. Opt.* **45**, 2965–2974 (2006).
26. S. L. Rudder, J. C. Connolly, and G. J. Steckman, "Hybrid ECL/DBR wavelength and spectrum stabilized lasers demonstrate high power and narrow spectral linewidth," in *Laser Beam Control and Applications*, A. V. Kudryashov, A. H. Paxton, V. S. Ilchenko, A. Giesen, D. Nickel, S. J. Davis, M. C. Heaven, and J. T. Schriempf, eds., *Proc. SPIE* 6101, pp. 112–119 (2006).
27. P. A. Jansson, *Deconvolution of Images and Spectra* (Academic Press, San Diego, Calif., 1997).
28. P. Matousek, M. Towrie, and A. W. Parker, "Fluorescence background suppression in Raman spectroscopy using combined Kerr gated and shifted excitation Raman difference techniques," *J. Raman Spectrosc.* **33**, 238–242 (2002).
29. M. Maiwald, G. Erbert, A. Klehr, H. D. Kronfeldt, H. Schmidt, B. Sumpf, and G. Trnkle, "Rapid shifted excitation Raman difference spectroscopy with a distributed feedback diode laser emitting at 785 nm," *Appl. Phys. B* **85**, 509–512 (2006).
30. I. Osticioli, A. Zoppi, and E. M. Castellucci, "Shift-excitation Raman difference spectroscopy difference deconvolution method for the luminescence background rejection from Raman spectra of solid samples," *Appl. Spectrosc.* **61**, 839–844 (2007).

## 1. Introduction

Raman spectroscopy offers rich chemical information in the optical spectra of scattered light from molecules. The Raman peaks in the spectra are frequently difficult to distinguish, however, from the intense fluorescent background that occurs in the same spectral region. In applications such as the analysis of biological specimens, laser exposure limits and long exposure times are an issue, and the fluorescent background poses a significant problem to the routine use of Raman spectroscopy. Additionally, to reduce sample fluorescence and the fluorescence of impurities in chemical and materials analysis applications, near-infrared (NIR) excitation is used at the expense of the reduced Raman cross-section at longer excitation wavelengths. [1]

Various techniques have been developed to isolate the Raman features that involve both system design and data post-processing [2–8]. Time-gating takes advantage of the near-instantaneous response of Raman scattering and the longer lifetimes of fluorescence by employing pulse-gating [2] and optical Kerr-gating [5]. While effective, time-gating at fast enough intervals to avoid fluorescence from molecules with shorter lifetimes creates a difficult experi-

mental setup and has photon collection problems.

Since most Raman features are spikes on a smooth fluorescent background, digital post-processing can also provide signal separation. Fourier-domain filtering and edge detection algorithms have been applied to perform this filtering [4], however this can lead to artifacts and distortions of the Raman spectrum in noisy situations. A polynomial fit of the background is a standard technique which allows for minimal distortion of the Raman spectrum [7] and is straight-forward to implement, but again requires significant exposure time to ensure adequate an signal-to-noise ratio (SNR).

The shift response of the Raman effect to excitation wavelength shifts allows another mechanism for removing the fluorescent background. By Kasha's rule [9], small changes in excitation wavelength will have a minimal effect on the fluorescence emission. For Raman spectra, however, the entire spectrum will shift in energy by the amount of excitation shift. This property has been used by many groups under the heading of "Shifted Excitation Raman Difference Spectroscopy" (SERDS) [3, 6, 8]. In these implementations, two spectra of a sample are acquired with slightly different excitation wavelengths, and are then subtracted to estimate the Raman spectrum of the sample. Curve-fitting [3] and linear deconvolution [6, 8] are the principal signal processing used to estimate the Raman spectrum from the difference spectrum. Alternatively, Bell et al. has rotated the grating in a spectrometer to produce up to three shifted spectra to estimate a sample's Raman spectrum. [10] This approach is slightly different from SERDS, however, in that both the Raman and fluorescence spectra shift on the detector plane, and is most effective in regards to the pattern noise inherent in detector arrays rather than Kasha's rule.

In this paper, we investigate an extension of SERDS with a new approach with two key novel elements— (a) the use of more than two excitation frequencies and (b) Raman signal estimation using an Expectation-Maximization (EM) algorithm. The use of more excitation frequencies allows for more complex excitation patterns which, in our simulations, improve the performance of the signal estimation. The multiple excitation frequencies could come from either a single tunable laser, or a bank of lasers at slightly spaced wavelengths. The experimental setup used for this report uses eight separate lasers at fixed wavelengths. A new spectrum estimation framework which is robust to noisy data is presented which has been designed to estimate the Raman spectrum from multiple observed spectra (as opposed to the two used in SERDS). Similar EM algorithms [11–13] have been successful in other inverse problems in spectroscopy [14, 15]. We show that for situations where the background fluorescence completely overwhelms the Raman features, the EM algorithm is more effective than background subtraction techniques at extracting the Raman spectrum. In both simulation and experiment the efficacy of the algorithm increases with the number of excitation frequencies used, even when holding the total exposure energy deposited constant.

## 2. Theory

### 2.1. Signal model

Raman scattering is an inelastic photon scattering process which transfers a fraction of the energy from arriving photons to rotational/vibrational modes of a molecule [16]. The scattered photons thus experience an energy shift, called the "Raman shift", making the process shift-variant with respect to the excitation energy. Optical frequency will be used as the unit of spectrum since it is directly proportional to photon energy. The intensity of the total Raman signal measured,  $g_R$ , is

$$g_R = \int h_e(\nu') S_R(\nu - \nu') d\nu'. \quad (1)$$

where  $\nu$  is the measured optical frequency,  $\nu'$  is the excitation frequency,  $h_e$  is the spectrum of the excitation laser, and  $S_R$  is the Raman spectral impulse response (the spectrum if measured with an excitation source of negligible spectral width). In order to minimize the broadening of the measured Raman signal by the excitation spectrum  $h_e$ , narrow bandwidth lasers are used in most applications. The  $S_R$  term is composed of variable width peaks, each corresponding to a Raman-active transition of the molecule(s) being interrogated.

Many molecules also emit a spectrally broad fluorescence signal, which for small shifts in excitation frequency can be approximated by a spectral impulse response  $S_F$  with no dependence on excitation frequency  $\nu'$ . The measured fluorescence signal,  $g_F$ , is then approximated as

$$g_F = \int h_e(\nu') S_F(\nu) d\nu'. \quad (2)$$

For molecules with both Raman and fluorescence emission, the measured  $S_F$  and  $S_R$  signals occur in the same spectral bands for commonly used excitation wavelengths in the visible and near-infrared. The Raman signal carries more specific chemical information than the fluorescence signal, motivating the use of techniques for separating the two.

## 2.2. Raman signal estimation method

With only one measured spectrum the shift nature of the Raman signal can not be used in estimation; however, by making multiple measurements with different excitation frequencies the Raman and fluorescence signals can be separated. In SERDS systems, two measurements at slightly spaced excitation frequencies are taken. By subtracting the two measurements the fluorescence signal cancels out, forming a pseudo-derivative of the Raman signal. If the two excitation frequencies are narrow enough to be approximated as delta functions in frequency,  $\delta(\nu' - \nu_1)$  and  $\delta(\nu' - \nu_2)$ , and noise is ignored, the inverse problem becomes,

$$g_{R1} - g_{R2} = \int S_R(\nu - \nu') (\delta(\nu_1 - \nu') - \delta(\nu_2 - \nu')) d\nu'. \quad (3)$$

This pseudo-derivative is then used to estimate Raman signal by using curve-fitting [3], Fourier deconvolution, or discrete integration [6, 8]. These techniques are not always ideal, however, since the curve-fitting is limited to simple spectra, the deconvolution is poorly conditioned and can amplify noise, and discrete integration leads to a broadening of the Raman features and a sloping baseline. When Poisson noise is included in the two-excitation model, the measurement becomes

$$y_{SERDS} \sim \text{Poisson}(S_F + S_R(\nu - \nu_1)) - \text{Poisson}(S_F + S_R(\nu - \nu_2)). \quad (4)$$

Thus, the subtraction does not simplify the inverse problem as it does in the noise-free case, making the previous SERDS approaches less effective. The previously stated techniques are also limited to only two excitation frequencies. More robust techniques are particularly critical in the presence of large amounts of noise, since reduced exposure time and laser power could be useful in many applications. The multi-excitation approach described in this paper treats spectral measurements from different excitation frequencies as independent Poisson measurements, and then uses iterative solution techniques to estimate the most likely Raman and fluorescence signals. To estimate the Raman spectrum given  $K$  excitations ( $k = 1 \dots K$ ), the model for the Raman signal changes to

$$g_{Rk} = \int h_{ek}(\nu') S_R(\nu - \nu') d\nu', \quad (5)$$

where  $h_{ek}$  is the excitation spectrum of each of the  $K$  excitation lasers. Similarly the fluorescence emission becomes,

$$g_{Fk} = \int h_{ek}(\nu') S_F(\nu) d\nu'. \quad (6)$$

To simplify the analysis, we now assume that each laser has the same total energy in each excitation, so the measured  $g_{Fk}$  is approximately  $S_F$ . If this were not the case, a normalization of the measurements would be necessary. We can then define each measured spectrum as,

$$g_k = g_{Rk} + S_F. \quad (7)$$

The problem will now be discretized to introduce the application of matrix methods. The measured spectra  $g$  are the  $K$  recorded spectra by the measurement system, each of length  $N$ ,

$$g \equiv [g_1^T, g_2^T \dots g_K^T]^T. \quad (8)$$

The Raman and fluorescence signals are defined as  $S \equiv [S_F^T S_R^T]$ , with length  $2N$ , leading to the inverse problem

$$y \sim \text{Poisson}(HS), \quad (9)$$

where  $H$  is a matrix of size  $KN \times 2N$  such that  $G = HS$ .  $H$  can be formed with knowledge of the excitation spectrum of each laser used,  $h_{ek}$ , and the assumptions of the fluorescence invariance. In discrete form, Eq. (5) and Eq. (6) become matrix operators  $H_{Rk}$  and  $H_F$ , respectively. The total operator  $H$  is then

$$H = \begin{bmatrix} H_F & H_{R1} \\ \vdots & \vdots \\ H_F & H_{Rk} \end{bmatrix} \quad (10)$$

Since the uncertainty in the measurements are typically dominated by photon noise due to the strong nature of the fluorescence, the task of estimating the Raman signal  $S_R$  is a Poisson inverse problem. An Expectation-Maximization (EM) algorithm is used for signal estimation. The estimate for  $S$ , defined as  $\hat{S}$ , is iterated upon by a Lucy-Richardson formula,

$$\hat{x}_{t+1} = \hat{S}_t \cdot (H^T(y \cdot / (H\hat{S}_t))) \quad (11)$$

$$\hat{S}_{t+1} = \hat{x}_{t+1} \quad (12)$$

where  $t$  is the iteration number and the dots indicate element-wise multiplication and division of vectors. Equation (12) can be replaced by a regularization step, as described in the next section, to improve the accuracy of the estimate  $\hat{S}_{t+1}$  and convergence properties. [17, 18]

### 2.3. Multiscale regularization

The multiresolution Poisson denoising method of regularization employed in this paper is referred to as translation-invariant Haar tree pruning. Intensity estimates are calculated from noisy realizations in  $\hat{S}$  by determining the ideal partition of the domain of observations (assumed to be  $[0, 1]$  in some unit of measurement for both the Raman and the fluorescent spectra) and computing the sample average in each interval of the optimal partition. This approach and its theoretical properties are detailed in [19, 20] and briefly reviewed in this section.

The optimal partition is selected by searching over a large collection of partitions called *Recursive Dyadic Partitions* (RDPs). The RDPs are formed by recursively splitting the domain  $[0,1]$  in halves to generate a collection of intervals at varying resolutions, where the length of each interval is a dyadic (power of two) number. Partitions defined by a set of these intervals can be represented by a binary tree, which each leaf in the tree corresponds to an interval in the partition. RDPs give our estimators the capability of varying the resolution with the frequency  $\nu$  to automatically increase the smoothing in very regular regions of the spectra and to preserve detailed structure in less regular regions. Selection of the optimal pattern is made using a penalized likelihood criterion.

The RDP framework leads to a model selection problem that can be solved by a tree pruning process. Each of the terminal intervals in the pruned RDP could correspond to a region of homogeneous or smoothly varying spectral intensity. Such a partition can be obtained by merging neighboring intervals of (*i.e.* pruning) a complete RDP to form a data-adaptive RDP  $\mathcal{P}$  and fitting sample averages to the terminal intervals of  $\mathcal{P}$ . Thus the spectrum estimate,  $\hat{S}$ , is completely described by  $\mathcal{P}$ . Given a partition estimate  $\hat{\mathcal{P}}$ ,  $\hat{S} \equiv S(\hat{\mathcal{P}})$  can be calculated by computing the sample average of the observations over each interval in  $\hat{\mathcal{P}}$ .

This provides for a very simple framework for penalized likelihood estimation, wherein the penalization is based on the complexity of the underlying partition. The goal here is to find the partition which minimizes the penalized likelihood function, so that Eq. (12) above is replaced with the following:

$$\begin{aligned}\hat{\mathcal{P}}_{t+1} &\equiv \arg \min_{\mathcal{P}} \left[ -\log p(\hat{x}_{t+1} | S(\hat{\mathcal{P}})) + \text{pen}(\hat{\mathcal{P}}) \right] \\ \hat{S}_{t+1} &\equiv S(\hat{\mathcal{P}}_{t+1})\end{aligned}\tag{13}$$

where

$$p(x|S(\hat{\mathcal{P}})) = \prod_{i=0}^{N-1} \frac{e^{-S[i]} S[i]^{x[i]}}{x[i]!}$$

denotes the likelihood of  $x$  given the spectrum estimate  $S(\hat{\mathcal{P}})$  and where  $\text{pen}(\hat{\mathcal{P}})$  is the penalty associated with the estimate  $S(\hat{\mathcal{P}})$ . (We penalize the estimates according to a codelength required to uniquely describe each model with a prefix code; the penalties are discussed in detail in [20].) The resulting estimator  $\hat{S}$  is referred to as the maximum penalized likelihood estimator (MPLE).

The accuracy of these estimates can be augmented by a process called cycle-spinning, or averaging over shifts, resulting in translation-invariant (TI) estimates [21]. Cycle-spinning, as originally proposed, requires  $O(N \log N)$  operations, but was derived in the context of undecimated wavelet coefficient thresholding in the presence of Gaussian noise, and is difficult to implement efficiently in the case of Poisson noise. The above multiscale tree-pruning method can be modified to produce the same effect by averaging over shifts, but the increase in quality comes at a high computational cost; naïve algorithms require  $O(N^2)$  operations. Novel computational methods, as described in [22], however, can be used to yield TI-Haar tree pruning estimates in  $O(N \log N)$  time. Unlike traditional wavelet thresholding techniques, this method is near minimax optimal for Poisson noise distributions and is robust to noise due to the hereditary nature of the tree-pruning process.

#### 2.4. System forward model formation

The described technique is completely general in regards to the sequence of laser excitation spectra used. A thorough optimization of the excitation patterns is beyond the scope of this pa-



per and will be left for future work. In computer simulations, an excitation technique that proved effective was multiple delta function-like excitations each of slightly different frequency. This was done experimentally through the use of multiple lasers, each at a different operating frequency, and thus the system forward model formation will be covered in detail for this case. In order to implement the EM algorithm, the  $H$  operator and its adjoint  $H^T$  need to be determined from a calibration of the spectrometer and excitation lasers. An important detail of dispersive spectrometers is that typically the measured channel spacing is nearly constant in wavelength units and thus highly non-uniform in energy units. Since the Raman shift is constant in energy units over the entire spectrum, this makes interpolations rather than simple shifts necessary for the operators. In order to make the data processing more efficient, functions are used for the interpolations instead of actual matrices. By analyzing the forward model functions, equivalent transpose model functions can be determined.

Each measured spectrum  $g_k$  contains  $N$  channels, corresponding to frequency values which are non-uniformly spaced in  $N$ . This relationship can be determined by recording the spectrum of a known source such as a gas discharge lamp. The peaks in the measured spectrum are fit to the frequency values of the emission lines, providing a vector containing the frequency value of each channel,  $w(n)$ . In order to determine the frequency value of each laser, a sample of a molecule with a large Raman cross section is measured with each excitation laser, and a reference peak at a Raman-shifted frequency value  $w_p$  can be used to determine the frequency of each laser. This peak location in each spectrum,  $p_k$  for  $k = 1 \dots K$  number of lasers, can then be used to determine each laser frequency,  $l_k = w(p_k) - w_p$ .

The Raman signal estimate,  $S_R$ , will then be referenced to the spectrum from the first excitation laser. The relative shift of each laser compared to the first is then  $\Delta(k) = l_k - l_1$ , allowing for the Raman signal for the rest of the lasers to be generated by shifting them  $\Delta(k)$ . An interpolation is used to perform the shift. With this calibration information, the forward function  $H$  can be written as,

$$H(S_F, S_R) = [S_F + S_R^w, S_F + S_R^w(w(n) + \Delta_2), \dots, S_F + S_R^w(w(n) + \Delta_K)] \quad (14)$$

where  $S_R^w$  is the  $S_R$  signal in frequency units,

$$S_R^w(w(n)) = S_R(n). \quad (15)$$

Thus, the  $S_R$  and  $S_F$  signals have the same channel spacing and number of channels as each of the original measurements. The adjoint operator  $H^T$  operates on  $K$  spectra, and results in two vectors of length  $N$  for the Raman fluorescence signal corrections. The fluorescence correction operation is an averaging of the  $K$  spectra, while the Raman correction operation involves shifting the spectra from each laser the opposite direction as in the  $H$  operator. These “unshifted” spectra become

$$z_k = g_k^w(w(n) - \Delta_k). \quad (16)$$

This shifting in frequency is again carried out by a spline interpolation. The adjoint function can then be written as

$$H^T(g_1, \dots, g_K) = \left[ \frac{1}{K} \sum_{k=1}^K g_{kn}, \frac{1}{K} \sum_{k=1}^K z_{kn} \right]. \quad (17)$$

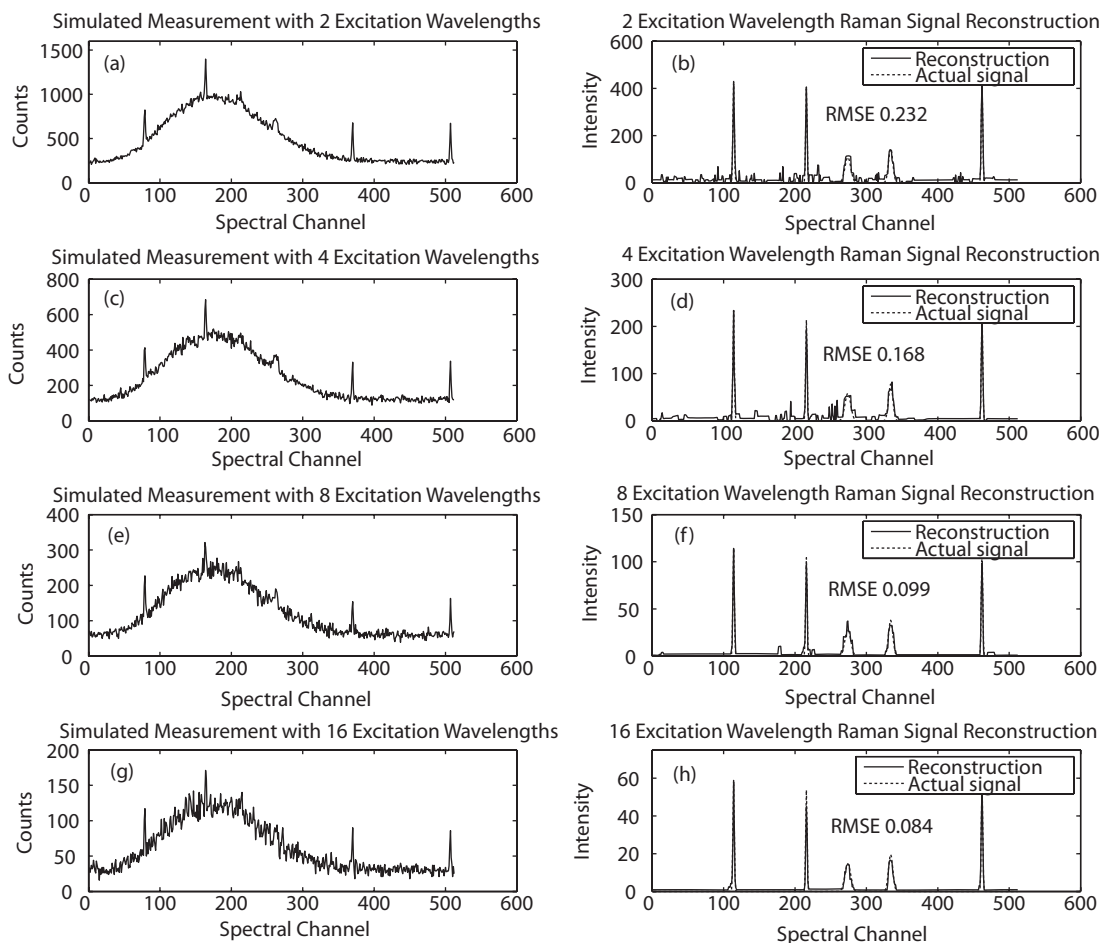


Fig. 1. One simulated measurement from each of the 2, 4, 8, and 16 frequency simulations (aceg) and corresponding EM reconstructions of the Raman component (bdfg).

### 3. Simulations

Simulations were performed to investigate the performance of the EM algorithm under different signal conditions. The relationship between the number of excitation frequencies used and reconstruction performance was another goal of the simulations.

For the modeling of the fluorescence, a Beta probability distribution function produced a similar shape to experimental fluorescence profiles. The Raman signals contained five Gaussian peaks with random locations, heights, and widths. The relative intensity of the fluorescence and Raman signals were varied to judge its impact on the reconstruction performance.

The number of excitation frequencies tested was 2, 4, 8 and 16, since these were feasible for experimental implementation. The spacing of the excitations was varied to keep the total span of the excitations constant. As the number of excitations increased, the intensity of each observation was decreased to keep the total excitation energy constant. This is apparent in the decreasing SNR in Fig. 1(a,c,e,g). If the laser power were held constant, this would be equivalent to having the same total exposure time for the entire collection of excitations.

The simulation process involved generating the test data by adding together the fluorescence



signal with shifted versions of the Raman signal. The proper amount of Poisson noise was then added to this test data to form the simulated observations. Readout noise was ignored since shot noise is usually the dominant source of noise in Raman systems with fluorescent samples.

Instead of starting the signal estimates for the EM algorithm from arbitrary locations, an approach was developed to form the starting point. An initial estimate for the fluorescence signal was found by taking the minimum value of all of the observations at each spectral channel. By subtracting this initial fluorescence estimate from each of the observations, a rough estimate for the Raman signal from each observation is obtained. These rough estimates are unshifted as in Eq. (16) and then averaged to form the first estimate of the Raman signal. While similar to the  $H^T$  operator, this approach made the simulation converge much faster than using the  $H^T$  operator alone on the simulated measurements to form the initial estimate.

This rough estimate  $\hat{S}(0)$  is then regularized and then Eq. (11) is used to form  $\hat{S}(1)$ . This process is repeated for 100 iterations in the simulations done for this report. Alternatively, a stopping condition can be used to stop further iterations. For the simulations performed, however, there was a negligible change in the estimates past  $\simeq 50$  iterations, so 100 was chosen for simplicity. More sophisticated strategies for stop conditions are an active research area [23] and were not explored thoroughly in this work.

There were three parameters that were varied in the simulations— the ratio of the peak fluorescence intensity to peak Raman intensity, the SNR of the observations, and the number of excitation frequencies used. One of the simulations with a fluorescence to Raman ratio of 1 and an SNR of 30 dB is shown in Fig. 1. The root-mean-squared error (RMSE) is calculated for each number of excitation frequencies by comparing each value of the reconstructed Raman signal to the known Raman signal. As seen in the figure, the RMSE of the reconstructed Raman signal decreases as the number of exposures increases, even though the total energy deposited remains constant.

In order to more thoroughly investigate the performance of the reconstruction process, a series of simulations was performed. The fluorescence to Raman ratio was set at 1 and 1000, and the SNR of the observations was varied over 3 orders of magnitude. At each of these settings, 20 simulations were performed at each number of excitation frequencies. The results of the simulation are shown in Fig. 2 with error bars indicating the mean plus and minus the standard deviation of the RMSE of the 20 simulations at each setting. As in the case of the one simulation discussed earlier, the trend is for the RMSE of the Raman signal reconstruction to decrease as the number of excitation lasers increases. When the RMSE is very low,  $< 0.05$ , there is little difference between the reconstructions using a different number of lasers. This is thought to be so since the noise is so low it hardly impacts the reconstruction performance. For the low SNR cases, however, there is a clear advantage to using more excitation lasers. This suggests that the conditioning of the inverse problem becomes better as the number of excitation frequencies increase. An increase in SNR of the observations decreases the RMSE, as expected, however this comes at a cost of either increased exposure time or increased laser power.

#### 4. Experiments

All the experiments were performed on a custom-built Raman spectrometer described in an earlier work. [24] The spectrograph is a grating-based instrument implementing static two-dimensional aperture coding. [25] The spectrometer has a spectral range of 800-900 nm with a resolution of 0.5 nm. This allows for the Raman shift region of  $500\text{-}2000\text{ cm}^{-1}$  to be measured at a resolution of  $8\text{-}12\text{ cm}^{-1}$ . The system was modified by adding eight excitation lasers and these will be described here, along with the excitation and collection optics.

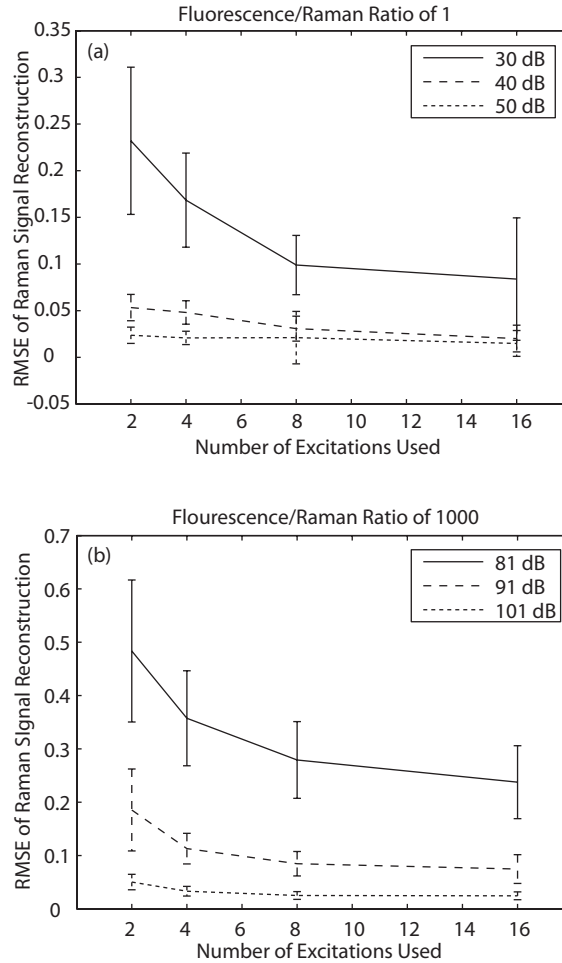


Fig. 2. (a) Root-mean-squared error of simulations as a function of number of excitation frequencies used and SNR of observations. The fluorescence signal to Raman signal ratio was 1 and 20 simulations were performed for each case. (b) Root-mean-squared error of simulations as a function of number of excitation frequencies used and SNR of observations. The fluorescence signal to Raman signal ratio was 1000 and 20 simulations were performed for each case. These results highlight the improvement associated with multiple excitation frequencies.

#### 4.1. Laser system

The laser system contained 8 wavelength stabilized diode lasers with center wavelengths of 782.6, 784.1, 784.4, 786.8, 788.6, 790.7, 793.6 and 794.3 nm (Innovative Photonic Solutions). This choice of wavelengths required the custom manufacturing of only 3 lasers. Integrated volume-phase holographic gratings provided stabilization of each laser's wavelength which resulted in a narrow linewidth,  $< 0.2$  nm, and high power operation,  $> 300$  mW per laser. [26]

The lasers were in 14-pin butterfly packages and fiber-coupled to  $105\text{ }\mu\text{m}$  core multi-mode fibers. A Newport Model 8008 Laser Diode Driver controlled the laser currents and temperatures. The temperature was set such that the side-mode suppression ratio of the laser emission was  $> 40$  dB. To set the drive current for the lasers, a power meter was set in the sample position and the laser current varied to reach an equivalent power level for each laser.

#### 4.2. Excitation and collection optics

A custom fiber bundle (RoMack Fiber Optics) composed of eight  $100\text{ }\mu\text{m}$  core fibers delivered the excitation light to sample. On one end of the bundle the fibers were closely packed into one SMA connector, and on the other end the fibers fanned out into eight FC connectors for the laser sources. The closely packed end then attached via a SMA connector to a  $400\text{ }\mu\text{m}$  core fiber, sized appropriately to allow virtually all the light from the eight fibers in the bundle. This larger diameter fiber was  $\approx 30$  m long, and wound into a tight loop inside a box in order to randomize the distribution of light from all the different excitation lasers. This "mode mixer" had an output SMA connector to couple the uniform excitation light onto a  $400\text{ }\mu\text{m}$  core fiber to the sample. A diagram of the setup is shown in Fig. 3.

The output fiber from the mode-mixer was collimated by a lens and spectrally filtered with a short-pass filter. Most notch filters did not have enough bandwidth needed to transmit all 8 of the lasers, thus requiring a short-pass filter to remove the fluorescence emission of the optical fiber. A  $10 \times 10$  mm, gold-coated, right-angle prism attached to the center of a 50 mm BK7 window with epoxy reflected the excitation light through three fused silica collection lenses onto the sample. The large core of the excitation fiber,  $400\text{ }\mu\text{m}$ , made the choice of collimating lens important, since an extremely short focal length lens would create a large beam due to the off-axis region of the fiber. Having a longer focal length lens, however, would create a large beam due to the large cone angle of the light from the fiber. Both these situations would cause the obscuration by the prism to be substantial. As a compromise, a 8 mm diameter, 16 mm focal length allowed for almost all the beam to be reflected, and only have a 5 percent obscuration by the mirror.

#### 4.3. Calibration of spectrometer

To calibrate the spectrometer, first an argon pen lamp spectrum was acquired to determine the wave-number values of the spectrometer channels. The chemical naphthalene was used to calibrate the laser frequency values. The 8 collected spectra are shown in Fig. 4. The strong peak at  $1382.2\text{ cm}^{-1}$  was used to determine the relative shift of the lasers in frequency, and thus was needed to implement the  $H^T$  operator described earlier. The effectiveness of the calibration can be checked by unshifting the spectra using the proper interpolation. These unshifted spectra from the 8 lasers are shown in Fig. 5.

### 5. Results

To investigate the performance of the EM algorithm with the experimental system, a trace amount of the laser dye HITC perchlorate was dissolved in pure ethanol at a concentration of  $\approx 1 \times 10^{-8} M$  to form a fluorescent sample with a known Raman spectrum. Since the concentration

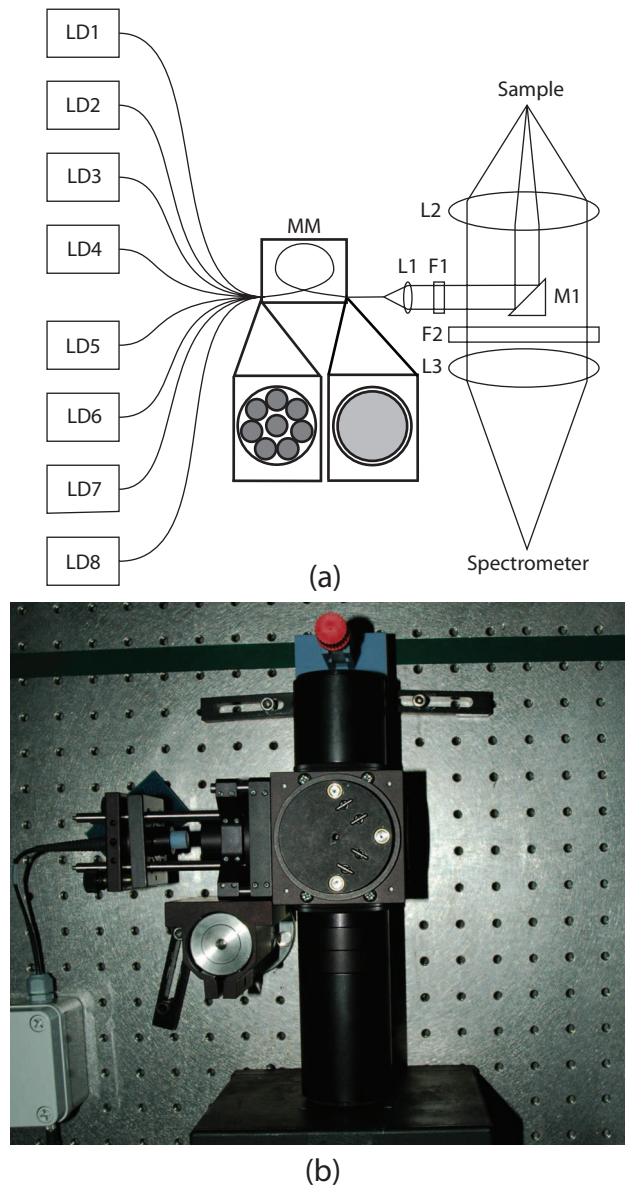


Fig. 3. (a) Laser diodes 1-8 (LD1-8) are fiber-coupled, then pass through mode-mixer (MM) to randomize their spatial distribution. A lens (L1) collimates the light from the fiber, which is then spectrally filtered by a short-pass thin-film filter (F1). A mirror (M1) mounted on a BK7 window reflects the beam towards the sample. A lens (L2) focuses the beam onto the sample, and two lens systems (L2,L3) image the scattered light onto the entrance aperture of the spectrometer. A long-pass filter (F2) blocks the laser light. (b) Photograph of excitation/collection setup.

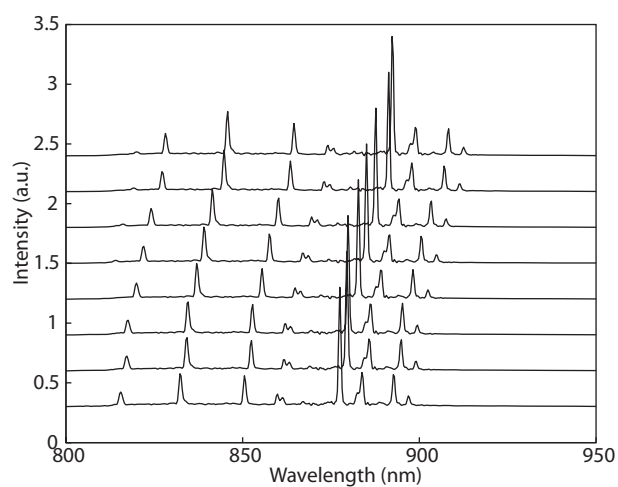


Fig. 4. Naphthalene Raman spectra collected with each excitation laser.

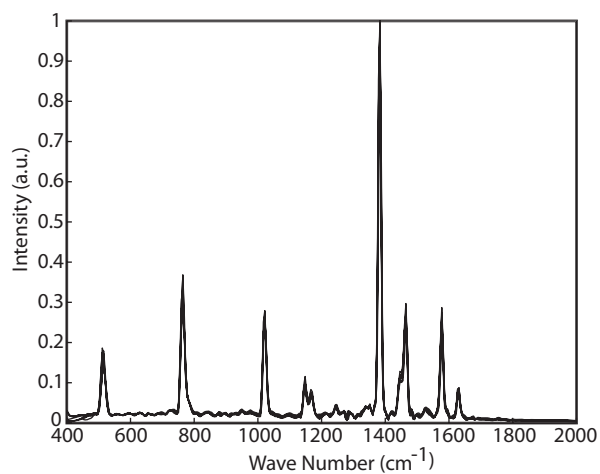


Fig. 5. Naphthalene Raman spectra corrected for relative shift of each excitation laser.

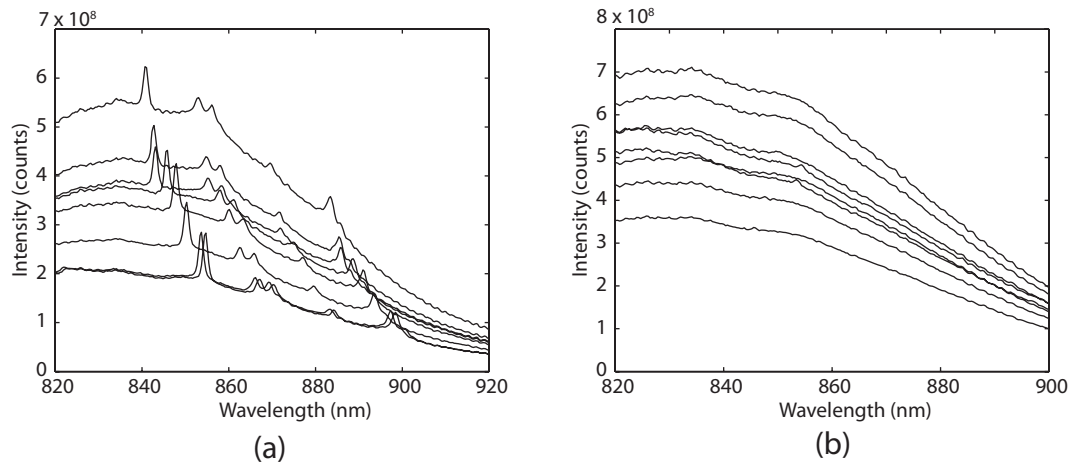


Fig. 6. (a) Measured spectra with all 8 excitation lasers for solution of  $1 \times 10^{-8}$  M dye dissolved in ethanol, 50 mW excitation power per laser, 1 second per excitation (b) Measured spectra with all 8 excitation lasers for solution of  $1 \times 10^{-6}$  M dye dissolved in ethanol, 5 mW excitation power per laser, 1 second per excitation

of the ethanol is many orders of magnitude larger than the dye, it is assumed that the dye's contribution to the Raman spectrum is negligible. A cuvette filled with the dye mixture was then placed into the sample holder of the spectrometer and measured at various exposure times and laser excitations. At 50 mW of excitation power, the peak counts on the CCD were  $\approx 25,000$  for a 1 second exposure. This was chosen as the exposure time to use since the maximum counts on the detector were 65,536. Exposures at each of the 8 lasers were taken 16 times to allow for different signal-to-noise ratios by summing different numbers of spectra taken with the same laser. The 2D CCD data was formed into an 800 channel spectrum by custom software written in Matlab, as described earlier. One of the measured spectra for each laser are shown in Fig. 6(a). While the incident power on the sample is approximately equal for each sample, the dye's absorption changes over the excitation wavelength range leading to the difference in measured intensities. The spectra are normalized to a unit sum before processing to estimate the Raman features. Raman features of ethanol are clearly visible at this concentration and exposure time. In addition, a sample with a higher concentration of dye,  $1 \times 10^{-6}$  M, was measured to provide a more obscured Raman signal. At this higher dye concentration the CCD saturated at the 50 mW power level. A 5 mW power level was used to achieve the same count level as with the lower dye concentration. Spectra were acquired again at one second exposures 16 times with each laser, and one spectrum from each laser are shown in Fig. 6(b). In these spectra the higher dye concentration and lower laser power makes the Raman features of ethanol no longer visible.

### 5.1. Comparison between processing methods

To better evaluate the multi-wavelength EM algorithm, the data was processed three ways—by a polynomial fit background subtraction, SERDS using Fourier deconvolution, and the previously described EM algorithm approach. The polynomial fit background subtraction is implemented by fitting a 5th order polynomial to the regions of the spectrum from one of the lasers that have no Raman features. This smooth background is then subtracted from the original spectrum to recover a Raman spectrum free of the fluorescent background, shown in Fig. 7(a)



and (e) for the low and high fluorescent dye concentrations, respectively. For the low dye concentration this works well, resulting in a Raman spectrum similar to the reference spectrum of ethanol taken with one excitation laser on the same instrument in Fig. 7(d). At the higher dye concentration, however, the polynomial background fit is not as successful, as shown in the Fig. 7(e). The low frequency ripple is probably due to a combination of the filter responses of the laser blocking filters and the non-uniformity of the CCD detector response. Even with the smoothing that could be done to minimize the effect of these features in the spectrum, not even the largest Raman feature of ethanol is prominent. The previously mentioned SERDS technique was also used to process spectra from 2 of the excitation lasers. By subtracting spectra taken with two different excitation wavelengths, a derivative-like signal is obtained as shown in the Fig. 7(b). The derivative signal can be estimated by solving the equation

$$y_d = H_d * S_R, \quad (18)$$

where  $y_d$  is the difference spectrum,  $H_d$  is a positive and negative impulse separated by the frequency difference of the two lasers, and the  $*$  indicates a convolution. [6] Before being subtracted, the spectra are linearized in wave-number and intensity corrected in order to make the  $H_d$  function be a constant function of frequency. The Raman signal  $S_R$  can then be found by solving Eq. (18) with Fourier analysis,

$$S_R = F^{-1}(F(y_D)/F(H_d)), \quad (19)$$

with  $F$  and  $F^{-1}$  defining the forward and inverse discrete Fourier transform, respectively. A further cosine apodization step helps reduce the Gibbs ringing in the reconstructed spectrum, and results in a low-pass filtering of the Raman estimate as well. [27] As shown in Fig. 7(b-c), the derivative estimate has a high SNR, and the deconvolution estimate is very similar to the measured pure ethanol. There is a broadening of the features due to the wide spacing of the laser frequencies used. For the high concentration dye in Fig. 7(e-h), the SERDS method is more effective at eliminating the low frequency ripples in the spectrum that were evident in the polynomial subtraction. The shot noise of the fluorescence is much more apparent in the derivative estimate in Fig. 7(f). The primary peak of ethanol is easily identified after the deconvolution shown in Fig. 7(g), however the other ethanol features are not as apparent as in the low dye concentration case.

## 5.2. EM processing analysis

For the EM estimation with the fluorescent dye data, two analyses were pursued—a comparison of the EM approach with SERDS and polynomial background subtraction, and to see how the number of excitation lasers affected the estimation performance. In order to keep the total excitation energy constant, as the number of excitation frequencies doubled the number of one 1 exposure to use would be reduced by a factor of 2. Thus, the 3 cases used were using two lasers with a 4 second exposure time each, 4 lasers with a 2 second exposure time each, and 8 lasers with a 1 second exposure each. The EM algorithm was applied to each set of spectra using lasers 1 and 2 for the 2 laser case, lasers 1-4 for the 4 laser case, and all 8 lasers for the 8 laser case. There weren't large differences in the reconstructions as long as the spacing of the lasers was not very large, such as using lasers 1 and 8 for the two laser case. This is probably due to the breaking down of the assumption of the measured fluorescence being completely insensitive to excitation wavelength for larger wavelength ranges. This is an issue with using all 8 excitation lasers since the lasers used spanned a large wavelength range, 12 nm, nonetheless the algorithm performed well with 8 lasers even with this span.

The EM algorithm reconstructions performs similar to the linear approaches with the low

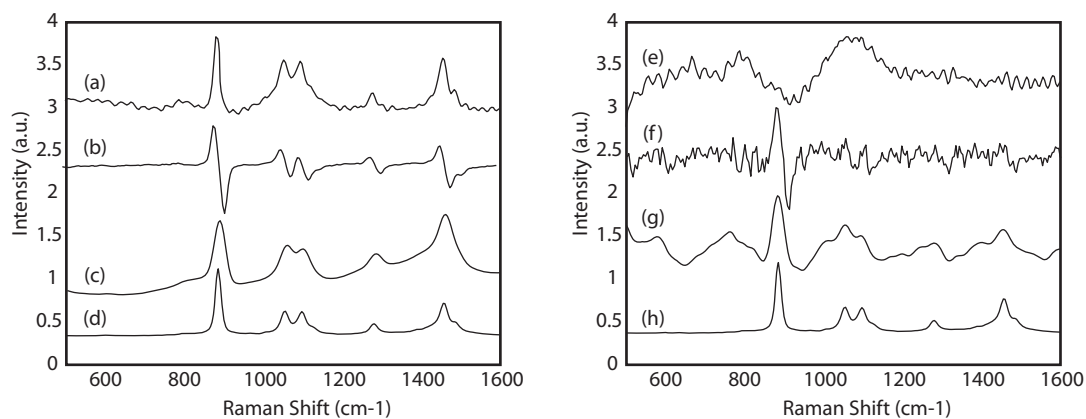


Fig. 7. (a-d) Low concentration dye results, (a)- Polynomial fit background subtraction of spectra from laser 1, (b)- Subtraction of spectra from lasers 1 and 2, (c)- Estimated Raman spectrum from Fourier deconvolution of b, (d)- Pure ethanol measured with one of the excitation lasers (e-h) High concentration dye results, (e)- Polynomial fit background subtraction of spectra from laser 1, (f)- Subtraction of spectra from lasers 1 and 2, (g)- Estimated Raman spectrum from Fourier deconvolution of f, (h)- Pure ethanol measured with one of the excitation lasers

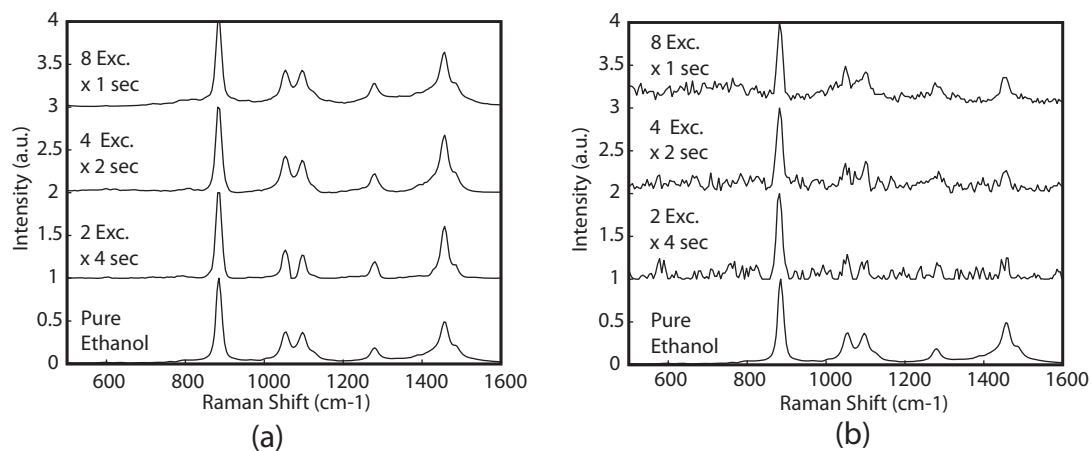


Fig. 8. (a) Low concentration dye EM reconstructions for 2, 4, and 8 excitation lasers, (b) High concentration dye EM reconstructions for 2, 4 and 8 excitation lasers

concentration dye as expected. As seen in Fig. 8(a), for the 2 laser case the two ethanol peaks near  $1100\text{ cm}^{-1}$  do not blend together like in the pure ethanol case. For the 4 and 8 laser case the reconstruction more resembles pure ethanol.

For the high concentration dye, the more lasers used, the better the reconstruction was, as shown in Fig. 8. The five ethanol peaks are clearly visible in the 8 laser reconstruction, whereas in the 2 and 4 laser reconstructions small artifacts in the reconstruction are of the same height as some of the ethanol peaks. Some of these smaller artifacts present in all three reconstructions could be a result of Raman bands from the fluorescent dye as seen in other studies [28], however without further investigation it is unclear at this point. In comparing the multi-wavelength

reconstructions to the linear processing techniques, at the high concentration of dye the EM algorithm processing seems to perform better. This would agree with the idea that for noisy observations, the EM technique is more robust than linear methods at extracting relevant information.

In agreement with the simulations, the multi-excitation performance increases with the number of excitation lasers used. The RMSE of the observations in Fig. 8 was calculated by comparing the reconstructed Raman signal to the pure ethanol signal acquired by the same instrument. For the reconstructions with 2, 4, and 8 lasers, the RMSE was 0.12, 0.09, and 0.08, respectively. Thus, the improvement from 2 to 8 lasers,  $\simeq 30\%$ , was less than in the simulated data,  $\simeq 50\%$ .

## 6. Conclusion

Multi-excitation Raman spectroscopy allows for the shift-variant Raman signal to be isolated from the broad fluorescent background present in many samples. An EM algorithm has been shown to be effective at solving this inverse problem even in situations where the Poisson noise of the observations is high compared to the weak Raman signal in both simulation and experiment. Once the spectrometer and laser calibration is performed, minimal user intervention is required making the technique less cumbersome than manual polynomial fitting of the background fluorescence. The regularization done by the EM algorithm does require some tuning to avoid over-smoothing the estimate for the Raman signal, and the automation of this process is something to be looked at in the future.

The use of more than 2 excitation frequencies provided an improvement in the reconstruction performance in both simulation and experiment. This suggests the conditioning of the inverse problem becomes less ill-posed as the excitation frequencies increase. This does, however, make the experimental setup more challenging, since more excitation frequencies puts more of a demand on the excitation sources and the filters used in the spectrometer. Thus, a trade-off must be made between more excitation frequencies and experimental complexity. As tunable diode lasers become more reliable for shifted excitation Raman applications [29, 30], the use of the EM approach with these lasers could help the approach become more practical.

## Acknowledgments

This work was supported by the National Institute on Alcohol Abuse and Alcoholism through the Integrated Alcohol Sensing and Data Analysis program under contract N01-AA-23013 and by NSF Award No. CCF-06-43947.