4.1 Theoretical Raman Spectra Generation

A model that predicts the bandshape of the H$_2$ Stokes vibrational $O$, $Q$, and $S$ branches of the Raman spectra has been developed using the principles outlined in Chapter Two. The Raman spectral model accounts for the collisional energy shifting of each Raman line, natural linewidth and collisional broadening effects, and convolution of the lines due to instrument responses and resolution. First, the frequencies for the Raman transitions are generated by calculating $\nu_3$, from Eq. 2-17, for each vibrational-rotational transition using the corresponding selection rules for $O$, $Q$, and $S$ transitions. Next, the $J$-dependent collisional energy shifts are calculated, as previously described, for the appropriate H$_2$ collision partner pairs and added to the results of Eq. 2-17. For the H$_2$-N$_2$ mixtures and rich H$_2$-air flames examined in this work, the collision partners for H$_2$ are H$_2$, N$_2$, and H$_2$O with the relative concentrations of each based on the composition of the equilibrium gas mixture. The gas density, $\rho$, is calculated using Redlich-Kwong data (Redlich and Kwong 1949). Likewise for the rich LH$_2$/LO$_x$ hot-fire conditions examined the collision partner pairs for H$_2$ are H$_2$ and H$_2$O. The Raman frequencies and collisional energy shifts are determined for all initial energy levels from $\nu = 0, J = 0$ up to $\nu = 10, J = 25$.

After the Raman transition frequencies/line positions are determined, the intensity of each transition is calculated using Eq. 2-20 for a specified $T$. Values for the nuclear-rotational spin symmetry in Eq. 2-20 are determined through the evaluation of Eq. 2-25, or the equivalent expression based on proper nuclear spin symmetry, at the initial sample fluid temperature, e.g. if a laboratory flame is examined the initial fluid temperature is the temperature of the unburnt gas mixture and similarly for LH$_2$/LO$_x$ combustion in the MCTA the initial fluid temperature is the temperature of the LH$_2$ in cryogenic storage. The $J$-dependence of the Raman scattering cross-sections are calculated using the basic expressions for $P_J$ in Table 2-3 in conjunction with the geometry tables in Long (1977) to obtain the proper expressions for the polarization of the incident electromagnetic field. Since $\eta \approx 85\%$ for the laser used in the current work, the $P_J$ term in Eq. 2-20 is calculated assuming the laser light entering the sample volume is 85% vertically
polarized and 15% unpolarized (polarization is directly related to $\eta$, Grunefeld et al. 1996). Additionally, the resonance enhancement correction is included in the evaluation of Eq. 2-20 for completeness.

The natural (no instrumental effects) intensity distribution of each line is then determined, depending on $\rho$ and evaluation of the mean free path, from either a Gaussian lineshape with a linewidth given by Eq. 2-29 for Doppler broadening or a Lorentzian lineshape with linewidths given by Eqs. 2-31 and 2-33 for Dicke-narrowed and collisionally broadened lines, respectively. Chapman-Enskog and Wilke theories are used to obtain the molecular diffusion coefficient for H$_2$ diffusing into a gas mix. In addition to the natural physical processes that broaden spectral lines, the finite resolution of the instruments further convolute the shape of the spectral lines. Instrumental broadening occurs through $\Delta \nu$$_{spect}$ and $\Delta \nu$$_L$. These instrumental parameters can be convoluted together to form an instrumental lineshape with a linewidth of $(\Delta \nu$$_{spect}^2 + \Delta \nu$$_L^2)$^{1/2}. The overall linewidth is then determined by convoluting the total instrumental linewidth with the natural linewidth in a similar manner. The Raman lines are also convoluted by the spectrograph entrance slit width, which, when multiplied by the dispersion of the spectrograph/dispersive-grating system, produces an additive linewidth function (for the Vanderbilt Raman system the 37.5$\mu$m slit width and 56cm$^{-1}$/mm system dispersion produce a linewidth of $\sim$2.1cm$^{-1}$. Thus as shown in Fig. 3-6, the actual spectral linewidth is larger than the 6.75cm$^{-1}$ spectrograph linewidth due to the additional broadening mechanisms.

The theoretical lineshapes of spectroscopic data can be predicted using numerous lineshape models, some of which range from the basic Gaussian or Lorentzian functions to complex combinations of these basic functions that include the superposition of multiple lineshapes, combinations of superpositions of multiple lineshape functions with varying linewidths, and lineshape functions including asymmetry factors. A comprehensive discussion on lineshape models and their use in fitting experimental spectra is presented by Conny et al. (1998) and Conny and Powell (2000a and 2000b). In the present work the experimental Raman spectra are fitted with a combined Gaussian-Lorentzian lineshape function. This lineshape, as described by Cignoli et al. (1991) and Conny and Powell (2000b), accounts for the Gaussian-like central peak and the broad low intensity base that is often observed in experimental spectra, which results from spatial smearing by the finite resolution of the CCD detector. Use of the Gaussian-Lorentzian function results in smaller bias/errors in the measurements than the use of a
simple Gaussian function alone (Conny et al. 1998). The detector response and hence the fitted lineshape is \(\sim 80\%\) Gaussian and \(\sim 20\%\) Lorentzian (as determined in the analysis of the results).

Since the CCD is not a sampling instrument but is rather an integrating detector with each pixel acquiring signal over a finite frequency range, the theoretical spectra should represent an integrated spectral segment instead of simply sampling the mid-point of each pixel (Tellinghuisen 2001). Therefore, the convoluted theoretical H\(_2\) Stokes Raman spectra are generated on a uniform grid of 0.045cm\(^{-1}\) (0.00035nm) increments between 36496-36036 cm\(^{-1}\). This fine mesh divides each pixel, which covers \(\sim 2\text{cm}^{-1}\), into 50 integrated grid points. Additionally, an absorption scan is performed to determine the frequency dependence of the dielectric thin-film notch filter, as shown in Fig. 4-1. The absorbance/transmission frequency dependence of the dielectric thin-film notch filter must be used to adjust the intensities of the generated spectra since it affects the intensities of the experimentally measured signals, \(\sim 4\%\) across the H\(_2\) Raman signal. The resulting Raman spectrum is generated by plotting the intensity profiles of all the transitions on the frequency dependent grid and constructively summing the overlapped intensities.

The actual theoretical Raman spectra used to determine the contour-fit Raman-derived temperatures are the combination of two spectra: one spectrum using \(\nu_L\) and \(\Delta \nu_L\) of the narrowband component of the laser emission and the other using \(\nu_B\) and \(\Delta \nu_B\) of the broadband component of the laser emission. This method is required for two reasons: 1) the residual broadband component has a much broader spectral linewidth, which slightly broadens the spectral profiles, and 2) the peak emission frequency of the broadband component produces a spectrum at slightly different absolute frequencies. The laser cannot be tuned to emit at the frequency of peak broadband intensity because \(\nu_B\) lies on the strong \(P_{3}(8)\) OH(3←0) absorption transition (as can be seen in Fig. 3-3). The narrowband and broadband spectrums are scaled by \(\eta\) and 1-\(\eta\), respectively and then summed to produce the final H\(_2\) Stokes Raman spectrum as shown in Fig. 4-2. The H\(_2\) Raman spectra in Fig. 4-2 are for the following parameter set: \(T = 293\text{K}, \rho = 0.933\text{amagats}, [H_2] = 0.02, [N_2] = 0.98, \Delta \nu_{spect} = 6.75\text{cm}^{-1}, \Delta \nu_{slit} = 2.1\text{cm}^{-1}, \nu_L = 40257.65\text{cm}^{-1}, \Delta \nu_L = 0.8\text{cm}^{-1}, \nu_B = 40249.55\text{cm}^{-1}, \Delta \nu_B = 50\text{cm}^{-1}\), and \(\eta = 0.85\). Thus the complete H\(_2\) Stokes vibrational Raman spectrum is obtained through the aforementioned method using a specified \(T\), \(\rho\), and species concentrations and estimated instrument parameter set (\(\Delta \nu_{spect}, \Delta \nu_{slit}, \nu_L, \Delta \nu_L, \nu_B, \Delta \nu_B,\) and \(\eta\)).
Fig. 4-1  Wavelength dependent absorbance scan of the dielectric thin-film notch filter. The figure shows how the filter blocks the Rayleigh/laser line and due to the dielectric film has a wave-like response. Transmission across the H₂ Raman signal varies as much as ~4%.
Spectral synthesis showing the convolution method used to generate the total H$_2$ Raman spectrum and match the spectral response of the CCD. The spectral parameters are $T = 293K$, $\rho = 0.933$ amagats, $\Delta \nu_{spect} = 6.75 \text{cm}^{-1}$, $\Delta \nu_{slit} = 2.1 \text{cm}^{-1}$, $\nu_L = 40257.65 \text{cm}^{-1}$, $\Delta \nu_L = 0.8 \text{cm}^{-1}$, $\nu_B = 40249.55 \text{cm}^{-1}$, $\Delta \nu_B = 50 \text{cm}^{-1}$, and $\eta = 0.85$. The narrowband spectrum is convoluted with the broadband spectrum according to $\eta$. At $T = 293K$ only $Q_0(1)$ through $Q_0(4)$ transitions are visible.
4.2 Contour Fitting Procedure and Parameter Estimation

The simplest and most basic approach to analyzing spectroscopic data consists of the method of linear least squares, where experimentally measured data with a fully known parameter set are compared with theoretically predicted data until the sum of the squares, the error between the two data sets, is minimized. Linear least squares analysis has several major advantage: 1) analysis of normally distributed (Gaussian) data yields unbiased normally distributed parameter estimates with standard errors known exactly if the error structure of the data is known; and 2) the solution is obtainable in a single-step and yields a single, minimum-variance solution. This allows the measurement errors to be calculated prior to the experiment for use in the design of the experiment. However in many applications, the theoretical expressions that characterize the experimentally measured data/variables are not linear in response to the desired parameters, i.e. the parameters are adjustable, not independent of one another, and parameter errors are unknown, thus necessitating a nonlinear least squares analysis where the parameter errors are estimated, assuming a weighted error structure, in an iterative solution. As in linear least squares methods, nonlinear least squares analysis still provides an estimation of the statistical errors in the parameters, but with considerably more computational effort and attention given to the error structure since the parameter errors are unknown. In addition to the inherent parameter dependence of the physical models (e.g. $T$, $\rho$...) in both linear and nonlinear least squares methods, the lineshape functions and parameters, which are inherently nonlinear in response, further complicate the analysis of spectroscopic data. Thus the resulting spectroscopic models are extremely complex and almost always nonlinear in nature.

Temperature measurement based on the contour-fit method is a classic inverse problem, in that temperature is to be determined directly from theory through analysis/prediction/interpretation of experimental data. Inverse problems are best solved using least-squares methods since the solution to the problem minimizes the error between experiment and theory by estimating/optimizing the unknown/desired parameters. Therefore, a nonlinear least-squares model is used to derive the contour-fit Raman-derived temperatures and parameter estimates because the Raman spectra are inherently nonlinear in response with respect to the underlying physical parameters, i.e. $f(e^{E/T})$ where $E$ is a $f(\rho, [x], \nu_L, \nu_B)$, and lineshape parameters ($\Delta \nu_{spec}, \Delta \nu_{slit}, \nu_L, \Delta \nu_L, \nu_B, \Delta \nu_B, T, \rho, [x], \eta$). Hence, a nonlinear computer model
based on a modification to Gauss’s method is used to obtain the Raman-derived temperatures and estimated parameters by minimizing the signal-to-noise (S/N) weighted least-squares fit between the experimental spectra and theoretical $\text{H}_2$ Stokes $Q$-branch Raman spectra.

The overall experimental error, or weighting factor, in the measurements is taken as the variance in the measured signal intensities, or more simply the inverse square of the S/N ratios calculated using Poisson statistics. Poisson statistics accurately describe the errors in CCD measurements (Tellenghuisen 2000). Additionally, the frequency of each pixel associated with the experimentally measured spectra is allowed to vary within $\pm \sigma^2$ of the CCD pixel to frequency calibration fit of the experimental spectra to the theoretical spectra. This frequency adjustment is needed to account for the uncertainty in the frequency calibration fit of the CCD.

Several assumptions can be made to minimize the number of parameters to estimate and hence greatly simplifying the computation of the Raman spectra. Species concentrations are assumed constant and are calculated based on the composition of the equilibrium gas mixture, which also reduces $\rho$ to only a function of $T$. The narrowband and broadband laser frequencies are fixed since they can be accurately determined experimentally from OH scans and Rayleigh spectra. The narrowband laser linewidth is extremely small, $\sim0.8\text{cm}^{-1}$ as deduced in Chapter III and measured by the manufacturer. Thus the narrowband linewidth parameter can be fixed to a constant value since it has little effect on the linewidth of the resulting spectra (as will be shown in the results). As previously described, the CCD response can be approximated by a combined Gaussian-Lorentzian lineshape where the Gaussian shape simulates the spectra induced by the narrowband laser component and the Lorentzian shape simulates the broadband laser component. The linewidth parameter dependence is simplified by keeping the spectrograph entrance slit-width and grating dispersion constant during the calculation since they are accurately determined hardware specifications. This results in a slightly simplified lineshape function, but the Gaussian and Lorentzian components are still coupled with a highly nonlinear response to $\Delta \nu_{\text{spect}}$, $\nu_L$, $\nu_B$, $\Delta \nu_B$, and $\eta$. The coupled response can be envisioned via

$$\Gamma_{\text{total}} = \sqrt{(\Delta \nu_{\text{spect}}^2 \cdot \eta + (\Delta \nu_{\text{spect}}^2 + \Delta \nu_B^2)(1-\eta))},$$

where the first and second terms represent the narrowband and broadband spectral components, respectively. The advantage of the aforementioned assumptions is a reduction in the number of variable parameters, 4 vs. the >10
total parameters, required to generate the theoretical Raman spectra, with the minimal estimated parameter set defined as $T$, $\Delta \nu_{\text{spec}}$, $\Delta \nu_B$, and $\eta$.

The method of least squares and parameter estimation are described in great detail in numerous texts and journal publications. In the proceeding discussion a brief description of nonlinear least squares using a modified version of Gauss’s method with parameter estimation will be given. This explanation is not intended to be a comprehensive derivation but rather to simply show how the spectroscopic problem is cast mathematically and indicate the methods used to obtain a solution. For a complete discussion concerning the mathematics and underlying theory of least squares methods with parameter estimation see Beck and Arnold (1977) and various journal articles such as Tellinghuisen (1994, 1996, 2000, 2001).

The equations for nonlinear least squares fitting and parameter estimation are obtained by casting the problem to be solved in a mathematical representation and minimizing the sum of the squared residuals $S$,

$$
S = \sum_{all \ i} \delta_i^T W \delta_i \quad (4-1)
$$

where $\delta$ is the residual vector (measured-predicted) and $W$ is the weight vector containing the corresponding weights for all the measurement points $i$. From statistics, the measured values, $Y$, are equal to the estimated or expected values, $\eta$, plus random values, $\varepsilon$, which accounts for error in the measurement

$$
Y = \eta(X, \beta) + \varepsilon \quad (4-2)
$$

with the bold syntax indicating matrix notation. The estimated values are a function of the estimated/dependent parameters, $\beta$, and the independent variables $X$. In linear least squares methods it is important to separate these parameters; however, in nonlinear least squares methods both the dependent and independent variables are considered to be completely uncertain and thus $\beta$ and $X$ can both contain elements of the same parameter set. It is important to note that this is where the nonlinear response with respect to the parameters in the predicted values is modeled. In most problems, and as in this work, $\varepsilon = 0$ if the measurement errors are assumed normal. Thus the matrix definitions for Eq. 4-2 are
where the vector $Y$ contains the number of measured values $n$, the vector $\beta$ contains the number of variables in the parameter set $p$, the vector $X$ contains the parameter variables as a function of $n$ with size $n \times p$, and the vector $\eta$ contains the predicted values.

Gauss’s linearization procedure is a relatively simple method that indicates the direction and size of the correction estimates to the parameters and effectively minimizes $S$ provided that the initial estimates are well-defined and in the region of the minimum. For this method to provide a unique solution, the matrix derivative of $S$ with respect to $\beta$ must be equal to zero at the minimum of $S$. It can be shown that the derivative of $S$ with respect to $\beta$ is

$$
\nabla_\beta S = 2\left[ -\nabla_\beta \eta^T(\beta) \right] W [Y - \eta(\beta)]
$$

(4-4)

where

$$
\nabla_\beta = \begin{bmatrix}
\frac{\partial}{\partial \beta_1} \\
\vdots \\
\frac{\partial}{\partial \beta_p}
\end{bmatrix} \quad \quad \nabla_\beta \eta^T(\beta) = \begin{bmatrix}
\frac{\partial \eta_1}{\partial \beta_1} & \ldots & \frac{\partial \eta_n}{\partial \beta_1} \\
\vdots & \ddots & \vdots \\
\frac{\partial \eta_1}{\partial \beta_p} & \ldots & \frac{\partial \eta_n}{\partial \beta_p}
\end{bmatrix}
$$

(4-5)

and $W$ is the diagonal weight matrix with values $w_n$ of size $n \times n$. The first bracketed term in Eq. 4-4 is referred to as the sensitivity matrix, $X(\beta)$, which describes the change in the predicted values as the parameters are changed.

$$
X(\beta) = \left[ \nabla_\beta \eta^T(\beta) \right]^T
$$

(4-6)

The importance of the sensitivity matrix in examining the parameter dependence will be discussed in the results. Setting Eq. 4-4 equal to zero for the case when the estimated parameters are optimized and substitution of $X(\beta)$ yields

$$
X^T(\beta) W [Y - \eta(\beta)] = 0
$$

(4-7)
which is the set of equations that provide the solution to the minimization of $S$ in the nonlinear case.

The set of equations in Eq. 4-7 must be solved iteratively since $\beta$ appears in both $\eta$ and $X$. A Taylor series is used to express $\eta(\beta)$ about $b$, where $b$ are estimates of $\beta$.

$$\eta(\beta) = \eta(b) + \left[\nabla_\beta \eta^T(b)\right](\beta - b)$$

(4-8)

Substitution into Eq. 4-7 yields

$$X^T(b)W\left[Y - \eta(b) - X(b)(\beta - b)\right] = 0$$

(4-9)

Switching to an iterative notation, $k$, and solving Eq. 4-9 for $b$ yields the solution to the equation set (Eq. 4-7) that provides the estimated parameters that minimizes Eq. 4-1

$$b^{(k+1)} = b^{(k)} + \Delta h \cdot P^{(k)}\left[X^{T(k)}W(Y - \eta^{(k)})\right]$$

(4-10a)

where

$$b^{(k)} = b \quad b^{(k+1)} = \beta \quad \eta^{(k)} = \eta(b) \quad X^{(k)} = X(b) \quad P^{-1(k)} = X^{T(k)}WX^{(k)}$$

(4-10b)

and $\Delta h$ is a scalar interpolation factor used to find the $(k+1)$st iterate for the $b$ vector. This modified form of Gauss’s linearization method provides more reliable convergence in nonlinear applications and is known as the Box-Kanemasu interpolation method. The iteration is stopped when the values in $b$ do not change significantly with successive iterations according to

$$\frac{|b^{(k+1)} - b^{(k)}|}{|b^{(k)}| + \delta_x} < \delta$$

(4-11)

$\delta_x$ is a small value ($10^{-100}$) to ensure that the left hand side of Eq. 4-11 is not undefined in the case that $b^{(k)}$ equal zero. Typically, $\delta$ has values on the order of $10^{-4}$.

The step-by-step procedure for using Eq. 4-10a to obtain the solution to nonlinear problems is:

1) Initial estimates of $\beta$ are given and designated $b^{(0)}$.
2) $\eta$ (the Raman spectrum) is calculated using the $b^{(0)}$ parameter set, or the $b^{(k)}$ parameter set for successive iterations.
3) $S$ is calculated using Eq. 4-1 and stored for comparison with future values of $S$ calculated using the newly estimated $b^{(k)}$.
4) $X(b)$ is determined using Eq. 4-6. $\nabla_\beta \eta^T$ can not be evaluated directly and therefore must be estimated numerically using a forward differencing method.
\[
\frac{\partial \eta}{\partial \beta} = \frac{\eta(\beta + \Delta) - \eta(\beta)}{\Delta} = \frac{\eta(\beta_1 + \Delta) - \eta(\beta_1)}{\Delta} \eta(\beta_2 + \Delta) - \eta(\beta_2) \ldots \eta(\beta_p + \Delta) - \eta(\beta_p)
\]

Such calculations require partial derivatives of the spectrum with respect to the adjustable parameters, which are taken by incrementing each parameter by \( \Delta \) and then subtracting the current reference spectrum for each iteration cycle. The value of \( \Delta \) must be determined such that the derivative is accurately estimated. If \( \Delta \) approaches zero the derivative is reached, however rounding errors associated with computation can sometimes prevent distinction between values in the numerator and lead to convergence errors. At the same time, the value of \( \Delta \) must not be too large or the derivative may not be well approximated. Proper estimation of \( \Delta \) is the most difficult aspect of parameter estimation.

5) \( P^{1(k)} \) is evaluated with the inverse yielding \( P^{(k)} \).

6) The new estimates of \( b \), designated \( b^{(k)} \) where \( k = 1 \) for the first iteration, are then determined by solving Eq. 4-10a. Since the initial parameter estimates are well-defined, a check is performed to prevent the parameter estimates from diverging too far from their true values. This is termed bounded parameter estimation and allows the parameters to vary within a user-defined limit, usually a fixed percentage of the parameter value or assumed error in the parameter measurement. If a new parameter estimate is outside the bounded limit then the specific parameter value is reset to the initial estimate and the procedure jumps to step 8.

7) If Eq. 4-11 is satisfied then the iteration is terminated and the parameters in \( b \) are optimized.

8) If the solution is not achieved, then the process is repeated starting at step 2 for successive iterations; \( k = k + 1 \).

4.3 Statistical Analysis Methods

The method of least squares provides built in procedures for estimating the statistical errors in the parameters. The variances in the parameters are proportional to the diagonal elements of the variance-covariance matrix \( P \), which is proportional to \( P^{-1} \). In order to be minimum-variance estimates, the statistical weights must be proportional to the inverse variances of the measurements, \( w_i \propto \sigma_i^{-2} \). Least squares is also a maximum likelihood method for this
condition, which is the more fundamental statistical principle behind data analysis (Tellinghuisen 2000). The preceding arguments are based on having complete prior knowledge of the statistics of the measurements, but this is almost never true from the experimental aspect. Fortunately, data obtained using counting instruments, as in the case of intensity measurements using a CCD, follow Poisson statistics with variances \( \sigma_{yi}^2 \) taken as \( y_i \), which is the measured signal strength. Consequently, the S/N ratios of measured spectra can be reliably used as statistical weights as follows

\[
\frac{w_i}{(S/N)^2} \quad \text{with} \quad S/N \propto \sqrt{y_i} \quad (4-12)
\]

An additional benefit of this physical property is that for large \( y_i \) Poisson data follow near Gaussian statistics, thus permitting the assumption of a normal parent distribution (Tellinghuisen 2000).

Confidence intervals can be constructed to estimate the uncertainty for each parameter \( \beta_i \) from the \( i \)th diagonal term in \( P \) and the normal \( t \)-distribution. The estimated standard error of each parameter \( b_i \) is

\[
est.e.(b_i) = \sqrt{P_{ii}} \cdot s
\]

where \( s^2 \) is the estimated sampling variance taken as

\[
s^2 = \frac{\delta^TW\delta}{\nu} \quad (4-14)
\]

The denominator, \( \nu \), in Eq. 4-14 is the number of degrees of freedom in the calculation \( n-p \). \( s^2 \) is essentially a chi-squared estimate used to scale the optimized parameters, which establishes a goodness-of-fit estimate. Thus, the \((100(1-\alpha))\%\) confidence intervals for the parameter estimates can be expressed as

\[
b_i - \est.e.(b_i) \cdot t_{1-\alpha/2}(n-p) < \beta_i < b_i + \est.e.(b_i) \cdot t_{1-\alpha/2}(n-p) \quad (4-15)
\]

where \( t_{1-\alpha/2}(n-p) \) is the student’s \( t \)-distribution statistic for \( \nu \) degrees of freedom. As \( \nu \) increases or becomes large (>100), the central-limit theorem is realized and the student’s \( t \)-distribution becomes approximately the standard normal distribution.

It is important to note that confidence intervals constructed in the manner previously described for nonlinear least squares are prone to error due to assumptions regarding the data’s error distribution. In linear methods (generally with normal error distributions) the errors given
by Eq. 4-13 are exact; however, in nonlinear methods the errors are estimates since the values in \( \mathbf{X} \) depend on \( y_i \) and the actual values of \( \beta \) (Tellinghuisen 2001). The values of \( y_i \) most likely may have normal errors but in general the parameter values in \( \beta \) are not normally distributed. Despite this limitation, estimates obtained from Eq. 4-13 for nonlinear methods can still be used to assess the parameter errors/\textit{goodness-of-fit} with reasonable confidence. For nonlinear least squares, Monte Carlo calculations can provide more exact estimates of the parameter errors. Since Eq. 4-13 is more or less a \textit{goodness-of-fit} indicator for the parameters, the total statistical uncertainty in nonlinear least squares can then be determined by combining the Monte Carlo and variance-covariance estimates in a propagation of errors analysis (as will be discussed in the results).

Monte Carlo analysis is an effective tool to analytically investigate the effects of nonlinearity in estimation methods. The basic procedure for building a Monte Carlo analysis model for a set of experiments is as follows:

1) The experiment is described mathematically, i.e. as in the previous discussion for nonlinear least squares parameter estimation.
2) True values, or values used to establish a fixed reference, are assigned to the dependent/independent parameters in \( \beta \) and \( \mathbf{X} \).
3) Using the variable set defined in 2, the true/reference value of \( \eta \) is calculated.
4) The errors in the measurement are identified and given an assumed/known distribution function with specified variances.
5) Next, random number generation is used to produce random errors within the limits of the specified variances for the assumed distribution function.
6) The simulated measurement is then obtained by combining the estimated errors with those of the true/reference values.
7) The nonlinear least squares procedure, or user defined method, is then used to fit the simulated measurement and provide estimates of \( \beta \) (assuming no prior knowledge of the values of \( \beta \)).
8) Steps 5, 6, and 7 are repeated for the desired number of simulated experiments \( N \), with new errors generated for each iteration.
9) Standard statistical methods, e.g. the measurement mean and standard deviation, can then be used to assess the uncertainty of the fitting procedure/parameter estimates for the simulated measurement data set.
The Monte Carlo simulations for the current investigation are carried out using the Raman spectrum generation and parameter estimation sub-models. True/reference values for the parameters are the experimentally measured instrumentation specifications given in Chapter Three. To accurately represent the spectral coverage of the CCD, the frequency grid upon which the simulated spectra are generated is $\sim 2.1\text{cm}^{-1}$ intervals or equivalent to the integrated spectral segment (i.e., a pixel) of the CCD. The simulated lineshape is $80\%$ Gaussian to represent the narrowband laser component and $20\%$ Lorentzian for the broadband laser component with $\eta \approx 85\%$ or the estimated value from the average fits. A constant baseline background, roughly 1000-2000 counts or the observed level of the CCD, is assumed. Randomly distributed noise is then generated and applied to the background to simulate the noise level of the CCD, e.g. typically between 1-2% of the constant background. The true experimentally measured peak signal intensity is used as the peak signal strength for calculation of the reference spectrum. S/N ratios are then determined according to Poisson statistics and used as the estimated variances for the simulated signal strengths. In the limit of large $y_i$ (either the number of photons or CCD counts, >50 photons) normal statistics are valid, thus a Gaussian distribution random number generation function with S/N estimated variances is used to generate normally distributed random noise in the spectra. The overall laser bandwidth, through variation in the locking efficiency $\eta$, affects the spectral shape; and therefore is allowed to vary within $\sim \pm 5\%$ of the reference $\eta$ for each generated spectrum. An additional noise, estimated from the standard deviation of the measured signal intensities from experimental data ($\sim 3.5\%$), was assumed to account for further sources of noise from spectrum-to-spectrum (e.g. from laser energy fluctuations that cannot be correctly accounted for). The free fit parameters for the Monte-Carlo simulations are $T$ and $\eta$, and the peak signal strength which is an inherent variable in the fit. Since the statistical variance scales as $N^{1/2}$, a minimum of $N = 1000$ spectra are simulated. This simulated spectral data set can then be used to provide reliable uncertainty estimates for the nonlinear least squares fitting procedure.

The randomly distributed noise in the background is generated using a Tausworthe random number generator while the normally distributed noise in the simulated spectrum is generated by a Gaussian distribution random number generation function. Both random number generation functions are in the numerical computation GNU Scientific Library (GSL) (GLS 2002). The random number generators are properly ‘seeded’ (seeding refers to a warm-up
procedure to initialize the random number generation sequence) to keep the errors truly random and prevent error/number patterns from re-occurring. An example of the Monte Carlo simulated H$_2$ Stokes Raman spectra are shown in Fig. 4-3 for $T = 295$K, a constant background level of 2000 counts with a noise level of 2%, peak signal strength of ~135,000 counts (~4500 photons, 30counts/e$^-$) or S/N = 67:1, and instrument parameter values as previously defined.

The Raman scattering simulation package (the Raman spectrum generation, parameter estimation functions, and Monte Carlo statistical analysis sub-programs) is written in C++: Microsoft Visual C++ Version 6.0 with extensive use of GSL (GSL, 2002).
Fig. 4-3 Monte Carlo simulated H$_2$ Stokes Raman spectra for $T = 295$K.