Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

ESI: Teaching Vibrational Spectra to Assign Themselves

Paul L. Houston,*[a,b] Brian L. Van Hoozen, Jr.,[c] Chen Qu,[c] Qi Yu,[c] and Joel M. Bowman[c]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

S1. Methods

S1.1 Algorithms

S1.1.1 Guess the Hamiltonian and diagonalize - the MCD algorithm

This algorithm uses a Monte Carlo search to find Hamiltonians with a set of target eigenvalues. The diagonal elements of the Hamiltonian are assumed to be known while the off-diagonal elements are obtained from the directed Monte Carlo search. This method relies on an diagonally dominant initial guess for the Hamiltonian.

Two different initial guesses for the Hamiltonian are explored and discussed. One initial guess assumes that off-diagonal elements near the diagonals are small (relative to the diagonals) random numbers, denoted ξ_i. For example, a Hamiltonian, **H**, that has a basis of five 1D harmonic wavefunctions with $\hbar = \omega = 1$ would have a guess of

$$\mathbf{H} = \begin{pmatrix} 0.5 & \xi_1 & 0 & 0 & 0 \\ \xi_1 & 1.5 & \xi_2 & 0 & 0 \\ 0 & \xi_2 & 2.5 & \xi_3 & 0 \\ 0 & 0 & \xi_3 & 3.5 & \xi_4 \\ 0 & 0 & 0 & \xi_4 & 4.5 \end{pmatrix} \quad (1)$$

In other examples in this report, all of the off-diagonal elements of the initial guess are zero. For some potentials including random numbers for the "near-off diagonal" elements of the "zero initial guess" significantly improves the results.

The Monte Carlo search for the off-diagonal components is performed by adding small random perturbations (which are constrained to maintain a real and symmetric Hamiltonian) to the off-diagonal elements of the Hamiltonian. An example of a perturbation for a 5×5 **H** is displayed in the equation below

$$\Delta\mathbf{H} = \begin{pmatrix} 0 & \xi_{12} & \xi_{13} & \xi_{14} & \xi_{15} \\ \xi_{21} & 0 & \xi_{23} & \xi_{24} & \xi_{25} \\ \xi_{31} & \xi_{32} & 0 & \xi_{34} & \xi_{35} \\ \xi_{41} & \xi_{42} & \xi_{43} & 0 & \xi_{45} \\ \xi_{51} & \xi_{52} & \xi_{53} & \xi_{54} & 0 \end{pmatrix} \quad (2)$$

The random numbers in the perturbation are typically three orders of magnitude smaller than the diagonal components to which they are connected. If adding the perturbation results in a **H** with eigenvalues that are closer to the target eigenvalues, then the perturbation is accepted, otherwise not. The figure of merit (FOM) used to determine if a step should be accepted is the sum of the squares of the differences between the n_{dim} eigenvalues of the current matrix (λ_M) and the target eigenvalues (λ_T), given by Eq. 3 below; it is proportional to the variance. If this number decreases by adding $\Delta\mathbf{H}$ to the Hamiltonian, the step is accepted. This procedure minimizes the least squares between the eigenvalues of the matrix and the target eigenvalues.

$$FOM_{MCD} = \sum_{j=1}^{n_{dim}} \left(\lambda_M^{(j)} - \lambda_T^{(j)}\right)^2 \quad (3)$$

For each trial, this process is repeated for up to thousands of iterations in order to find a Hamiltonian matrix with the target eigenvalues.

S1.1.2 Guess the eigenvectors and perform Inverse Diagonalization - the MCID algorithm

This algorithm takes the diagonal matrix of eigenvalues, Λ, develops Monte Carlo guesses of the initial values of the unconstrained eigenvector components in **C**, and minimizes a figure of merit (FOM), subject to the constraint that $\mathbf{C}^T\mathbf{C} = \mathbf{I}$. The figure of merit in this case is given by sum of the squared deviations between the basis energies, $E_M^{(i)}$, calculated from the minimization

[a] Department of Chemistry and Chemical Biology, Cornell Universtiy, Ithaca, NY 14853, USA. Fax: 01 607 255 4137; Tel: 01 607 592 2713; E-mail: plh2@cornell.edu
[b] School of Chemistry and Biochemistry, Georgia Inst. of Technology, Atlanta, GA 30332-0400, USA.
[c] Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University Atlanta, Georgia 30322, USA.

and the target ones, $E_T^{(i)}$, presumed known:

$$FOM_{MCID} = \sum_{j=1}^{n_{dim}} \left(E_M^{(j)} - E_T^{(j)} \right)^2, \quad (4)$$

where $E_M^{(j)}$, with $j = 1,...n_{dim}$, are calculated as the diagonal elements of $\mathbf{C}^T \Lambda \mathbf{C}$. Minimization of FOM_{MCID} is calculated using the "FindMinimum" function of MATHEMATICA. In practice, trials are discarded if the average of the sum of the elements of $\mathbf{C}^T\mathbf{C} - \mathbf{I}$ is greater than 10^{-5}, or if FOM_{MCID} is greater than $F \sum_{j=1}^{n_{dim}} \left(\lambda_T^{(j)} - E_T^j \right)^2$, where F is an adjustable parameter generally in the range from 0.001 to 0.005. The logic of this last choice is that this summation would be equal to the FOM if $\mathbf{H}$ were assumed to have zeros for all off-diagonal elements and the basis energies as the diagonal elements; i.e., if there were no improvement on the original guess. The minimization is rapid for n_{dim} up to about 30, and the algorithm generally converges if the trace of Λ is within a few percent of the sum of the basis energies, as discussed next.

S1.1.3 When will the algorithms work and when not?

A general property of matrices and their eigenvectors is that the trace of a matrix, $\mathbf{H}$, and the trace of its eigenvalue matrix, Λ, are equal; thus, $Tr(\mathbf{H}) = Tr(\Lambda)$. If the Hamiltonian matrix has the basis energies as its diagonals, then any completely converged solution provided by the algorithms should maintain this conservation of trace. This conservation would hold exactly if all perturbations to the basis functions were proportional to odd powers of the coordinates, because the diagonal elements of $\mathbf{H}$ would then be given by the basis energies plus an integral that, by symmetry, is required to be zero. For perturbations proportional to even powers of the coordinates, however, the diagonal elements of $\mathbf{H}$ and the basis energies will not generally be equal, although the difference will typically be small. In this case, the target of the algorithm, either the eigenvalues or the basis energies, will not be met exactly. For example, in the Inverse Diagonalization (MCID) method, the target will be the basis energies, but the result of the algorithm will give the basis energies each adjusted by an equal amount to make the trace of the resulting Hamiltonian equal to the sum of the eigenvalues. If the even perturbations are too large, then the results will not fall within the limit set by the factor F described in the previous section, and thus the algorithm will not provide useful results.

Similarly, if the sum of the basis energies is substantially different from the sum of the eigenvalues to begin with, perhaps because of a shift or scaling, it may be impossible for the algorithm to converge to within the limit set by F. If there is no convergence even with $F = 1$, it is clear that the functions that give the basis energies cannot be basis functions for the eigenvalues, no matter what off-diagonal elements of $\mathbf{H}$ are selected.

S1.1 Analysis

Because our primary objective is to determine the best eigenvectors describing the connection between the basis energies and the eigenvalues, we focused on comparing the calculated eigenvector component magnitudes with the exact magnitudes. A convenient method for comparison is to plot the n_{dim}^2 exact component magnitudes as a function of the calculated component magnitudes. If the data fall on a straight line of unit slope, then the algorithm has done a perfect job. The degree of fit can be characterized by R^2, where $R^2 = 1$ designates a perfect fit.

We focus on the absolute *magnitudes* of the components because the algorithms produce some solutions where the value of a particular component is positive and some where the absolute value is nearly the same but the sign is negative. It is actually sufficient to know the magnitudes of the components, since, if the magnitudes are known and the vectors are normalized, then there are several possible sign choices for the components that can be made to ensure that the vectors are orthogonal.

Focusing on the magnitude, however, raises a question of analysis. The simplest approaches would be to either 1) use the mean over all trials of the absolute value of each component, or 2) use the absolute value of the mean over all trials of the (signed) value of the component. The former method is more sensible if the distribution of the components is centered about some non-zero value and falls off substantially between that value and zero. The latter method is more sensible if the distribution of the components is centered on zero. Of course, many of the components will be larger than, for example, 0.1, and with a narrow distribution. For these the first method will work well. However using this method for very small values risks assigning a non-zero value to the component when the proper component might actually be much closer to zero. We decided to investigate two methods that might correct this situation.

The first method uses the average of the absolute values if the mean of the absolute value of a component is greater than two standard deviations of its distribution. If the mean of the absolute value is smaller than two standard deviations, then it uses the absolute value of the mean. The second method fits a Gaussian to a histogram of the positive values (including those that have zero weight) and another to that of negative values, using numerical methods to give the best center, amplitude and width. If the Gaussian center for the positive values is positive and that for the negative values is negative, then it uses the mean of the absolute values. If the Gaussian center for the positive values is negative and that for the negative values is positive, it uses the absolute value of the mean. In the rare case where the results are mixed, it uses the method indicated for the larger number of values. Comparison R^2 values for $n_{dim} = 5$ and $n_{dim} = 7$ cases shows that for either method that R^2 is within 0.01 of simply taking the mean of the absolute values. The reason for this outcome is that the smaller components have much less influence on R^2 than the larger ones, and only the smallest components are changed by the corrective methods. Thus, we will simply use the metric of taking the mean of the absolute value of each component.

S2. Dipole transition moments

We provide more detail here on steps b) and c) of Procedure 1. In Eq. (1) of the main text, ν_f is the frequency of the transition and the remaining part of the equation is the squared dipole transition element. There are three fundamentals that carry os-

cillator strength in the FAD spectral region we consider: ν_{20}, the C=O stretch, ν_{22}, the C-H stretch, and ν_{24}, the O-H stretch. If there were only one bright state for any eigenstate, then from Eq. (4) of the main text, the amplitude of the Gaussian centered on each eigenstate would be proportional to the eigenstate energy times the squared dipole transition element times the squared probability amplitude that the eigenstate has the character of the relevant normal mode. When there are more than one normal modes that carry oscillator strength, say ν_{24} and ν_{22}, the situation is more complex. The ground-state wavefunction for any eigenstate will be $|0\rangle = \left|\chi_{24,0}\,\chi_{22,0}\,\Psi_0^{(i)}\right\rangle$, where $\chi_{k,0}$ are the ground-state harmonic oscillator functions for the bright states, and $\Psi_0^{(i)}$ represents the parts of the wavefunction due to the ground-state dark modes in eigenvalue (i), modes that are unchanged by the excitation. The excited state will be $\langle f| = \left\langle c_{24}^{(i)}\,\chi_{24,1}\,\chi_{22,0}\,\Psi_0^{(i)} + c_{22}^{(i)}\chi_{24,0}\,\chi_{22,1}\,\Psi_0^{(i)}\right|$, where $c_k^{(i)}$ are the probability amplitudes for the bright states in eigenvalue (i) and $\chi_{k,1}$ are the excited-state harmonic oscillator functions for the bright states. Finally, the intensity will be proportional to the frequency corresponding to the eigenstate (i) times the squared dipole transition element (Eq. (1) repeated here from main text, slightly modified):

$$I \propto \nu_f^{(i)} \quad \left|\langle f|\,\vec{\mu}\,|0\rangle\right|^2, \tag{5}$$

Substituting, we have

$$I \propto \nu_f^{(i)} \times \left|\left\langle c_{24}^{(i)}\,\chi_{24,1}\,\chi_{22,0}\,\Psi^{(i)} + c_{22}^{(i)}\chi_{24,0}\,\chi_{22,1}\,\Psi^{(i)}\right|\vec{\mu}\left|\,\chi_{24,0}\,\chi_{22,0}\,\Psi^{(i)}\right\rangle\right|^2, \tag{6}$$

$$I \propto \nu_f^{(i)} \times \left|\,c_{24}^{(i)}\,\langle\chi_{24,1}\,\chi_{22,0}|\,\vec{\mu}\,|\,\chi_{24,0}\,\chi_{22,0}\rangle + c_{22}^{(i)}\,\langle\chi_{24,0}\,\chi_{22,1}|\,\vec{\mu}\,|\,\chi_{24,0}\,\chi_{22,0}\rangle\right|^2, \tag{7}$$

$$I \propto \nu_f^{(i)} \times \left[\,(c_{24}^{(i)})^2\,\langle\chi_{24,1}|\,\vec{\mu}\,|\,\chi_{24,0}\rangle^2 + (c_{22}^{(i)})^2\,\langle\,\chi_{22,1}|\,\vec{\mu}\,|\,\chi_{22,0}\rangle^2 + Xterm\right], \tag{8}$$

where

$$Xterm = 2(c_{24}^{(i)})(c_{22}^{(i)})\langle\chi_{24,1}|\,\vec{\mu}\,|\,\chi_{24,0}\rangle\cdot\langle\,\chi_{22,1}|\,\vec{\mu}\,|\,\chi_{22,0}\rangle. \tag{9}$$

The first two terms in Eq. (8) are just the probability of $k = 24$ or $k = 22$ character in the eigenvalue (i) times the squared transition dipole moment for that fundamental. Note that the cross term, *Xterm*, can be positive or negative depending on the relative signs of the probability amplitudes and the matrix elements. In addition, because $\vec{\mu}$ is a vector, there is a dot product between the two matrix elements. For any given direction, $\alpha = x, y, z$, the dipole matrix element will go as

$$\langle\chi_{j,1}\,|\mu_\alpha|\,\chi_{j,0}\rangle = \left\langle\chi_{j,1}|\mu_{\alpha 0} + \frac{\partial\mu_\alpha}{\partial Q}Q|\chi_{j,0}\right\rangle \tag{10}$$

$$= \frac{\partial\mu_\alpha}{\partial Q}\langle\chi_{j,1}|Q|\chi_{j,0}\rangle \tag{11}$$

$$\propto \frac{1}{\sqrt{\omega}}\frac{\partial\mu_\alpha}{\partial Q} \tag{12}$$

Calculated fundamental modes, frequencies and dipole derivatives for the formic acid dimer are given in Table S4 (see Supplementary Tables section below). The amplitude for each calculated Gaussian peak is then given by Eqs (8), (9), (12), and Table S4 using the values for $c_k^{(i)}$ provided by diagonalization of the trial Hamiltonian. The cross term is important for the ν_{24} and ν_{22} modes, but it has been neglected in the case of ν_{20} for two reasons. The probability amplitudes for ν_{20} are quite small relative to the other two in the important 2500-3300 cm^{-1} region, and the direction of the ν_{20} transition dipole moment is nearly orthogonal to the other two, making the cross term very small. Finally, because experimental spectra are often presented in arbitrary absorption units, the calculated spectrum is multiplied by a constant chosen to give the best match between experiment and calculation.

S3. Supplementary Results

S3.1 Overview

We consider several tests of the algorithms, starting from very simple perturbations to a one-dimensional harmonic oscillator, moving to multidimensional systems that exhibit Fermi Resonance, and ending with a study of randomly generated matrices of increasing dimension from $n_{dim} = 3$ to 20. In each of these cases we first generate a Hamiltonian matrix from basis functions and a model potential, then diagonalize it, and finally perform the algorithms, using the basis energies and the eigenvalues but neither the potential nor the off-diagonal elements of the Hamiltonian. We compare the actual with the calculated eigenvector component magnitudes as well as the actual and calculated off-diagonal elements. These examples show that the algorithms provide useful information.

S3.2 One-dimensional harmonic oscillator with a cubic perturbation

In this section we will investigate a simple cubic perturbation to the harmonic oscillator potential energy function (Eq. 13) to determine the effectiveness of the algorithms:

$$V(q) = 0.5q^2 - 0.02q^3. \tag{13}$$

We computed the Hamiltonian matrix for a basis set of five vibrational levels using raising and lowering operators. The matrix is shown below for $\hbar = \omega = 1$:

$$\mathbf{H} = \begin{pmatrix} 0.5 & -0.0212 & 0 & -0.0173 & 0 \\ -0.0212 & 1.5 & -0.06 & 0 & -0.0346 \\ 0 & -0.06 & 2.5 & -0.1102 & 0 \\ -0.0173 & 0 & -0.1102 & 3.5 & -0.1697 \\ 0 & -0.0346 & 0 & -0.1697 & 4.5 \end{pmatrix}. \tag{14}$$

The eigenvalues, given by the set (0.4994, 1.4964, 2.4914, 3.4841, 4.5286), are very close to the harmonic (diagonal) values.

The eigenvalue magnitudes were determined using the MCI algorithm. They fit the actual values to an R^2 of 0.99993, as shown by the red dots in the plot of Fig. S1. Thus, the algorithm does an

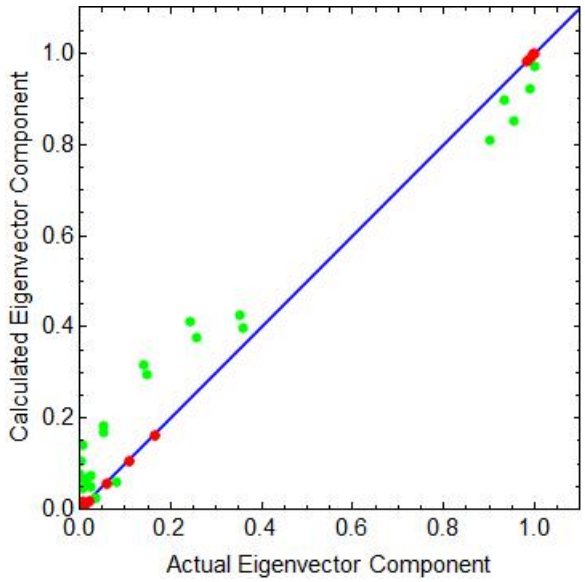

Figure S1. Plot of actual eigenvector component magnitude as a function of calculated eigenvector component magnitude. The red dots for a cubic perturbation to a harmonic oscillator ($R^2 = 0.99993$, Sect. S3.2). The green dots are for cubic and quartic perturbations ($R^2 = 0.9352$, Sect. S3.3). The blue line is of unit slope.

excellent job of finding the eigenvector magnitudes.

S3.3 One-dimensional harmonic oscillator with cubic and quartic perturbations

In the previous section we examined a slightly anharmonic cubic potential. Although the Monte Carlo search method was successful in describing the eigenfunctions for this potential, the limited anharmonicity of the potential may not have presented it with a significant challenge. In addition, the presence of even terms in the perturbation will affect the diagonal elements of **H**, perhaps affecting the algorithms. Thus, we now examine the following potential:

$$V(q) = 0.5q^2 - 0.053q^3 + 0.0075q^4. \qquad (15)$$

For a basis set of 5 harmonic wavefunctions (lowest three eigenvalues converged to within 0.1%), the Hamiltonian is

$$\mathbf{H} = \begin{pmatrix} 0.5056 & -0.0562 & 0.0159 & -0.0459 & 0.0092 \\ -0.0562 & 1.5281 & -0.1590 & 0.0459 & -0.0918 \\ 0.0159 & -0.1590 & 2.5731 & -0.2921 & 0.0909 \\ -0.0459 & 0.0459 & -0.2921 & 3.6406 & -0.4497 \\ 0.0092 & -0.0918 & 0.0909 & -0.4497 & 4.7306 \end{pmatrix}. \qquad (16)$$

The eigenvalues are (0.5019, 1.5054, 2.5188, 3.5409, 4.9111). Note that the diagonal elements of the Hamiltonian are not the basis energies (0.5, 1.5, 2.5, 3.5, 4.5), but have contributions from the quartic term in the perturbation. Also, some of the off-diagonal elements are quite a bit larger, relative to the diagonal elements, than in the cubic example. The difference between the sum of the basis energies and the sum of the eigenvalues is 3.7%.

Nonetheless, the algorithms do well in predicting the eigenvectors. Table S1 shows the results for starting both from the harmonic basis energies and from the actual Hamiltonian diagonals (although the latter are not known without information about the potential energy surface). The two algorithms give similar results. A plot of the actual vs. calculated eigenvector magnitudes for the MCD calculation using the harmonic basis diagonals is shown by the green dots in Fig. S1.

Table S1. R^2 results of algorithms for an harmonic oscillator perturbed by cubic and quartic potential terms.

MCD actual diag.	MCD HO diag.	MCID actual diag.	MCID HO diag.
0.9975	0.9352	0.9961	0.9345

Table S2. R^2 results of algorithms for the symmetric stretching mode of CO_2 using a quartic force field

MCD actual diag.	MCD HO diag.	MCID actual diag.	MCID HO diag.
0.9960	0.9949	0.9995	0.9816

S3.4 Further tests with the CO_2 Quartic Potential

In this section we investigate further tests with the CO_2 quartic potential given by Eg. (5) in the main text.

S3.4.1 The symmetric stretch frequency

As a simple test of our algorithms, we used this force field to calculate the frequencies of the fundamental symmetric stretching mode. Using the constants they determined, we used the first five terms in Eq. (5) of the main text to calculate the Hamiltonian matrix for a basis set of five vibrational levels, as shown in Eq. 17:

$$\mathbf{H} = \begin{pmatrix} 677 & -28 & 4 & 39 & 2 \\ -28 & 2025 & 27 & 11 & 77 \\ 4 & 27 & 3378 & 115 & 22 \\ 39 & 11 & 115 & 4738 & 228 \\ 2 & 77 & 22 & 228 & 6102 \end{pmatrix}. \qquad (17)$$

Note that because of the quartic terms in the potential, the diagonal elements of the Hamiltonian are not the harmonic basis energies, which are (672, 2015, 3358, 4701, 6044). The eigenvalues are (676, 2023, 3369, 4710, 6141). All energies are in cm^{-1}. The sum of the basis energies and the sum of the eigenvalues differ by only 0.8%, so both algorithms work well even when using the harmonic basis energies as the diagonal elements for the algorithms.

Table S2 shows the results of the algorithms. For each algorithm we list the results for starting either with the actual diagonal elements of the Hamiltonian or with the harmonic basis energies, noting, of course, that only the latter would be available if we did not have a potential energy function. The degree of quartic perturbation is evidently small, so that the difference between using the two different diagonals is barely perceptible in the R^2 value. Similarly, the results from the two methods are nearly equivalent.

Although the results will not be presented here, we note that the algorithms also do well predicting the eigenvector component magnitudes for the bending and antisymmetric stretching modes for this model of CO_2.

Table S3. R^2 results of algorithms for the coupled symmetric stretching, bending and antisymmetric stretching modes of CO_2 using a quartic force field and the harmonic oscillator basis elements.

MCD 9×9	MCD 14×14	MCID 9×9	MCID 14×14
0.9602	0.8463	0.9578	0.8751

3.4.2 Addition of the third vibrational mode

A two-mode model of CO_2 has been discussed in the main text. We now consider a more complete potential of Eq. (5) of the main text, including the symmetric stretching, bending, and antisymmetric stretching coordinates, but not a degenerate bending mode. We used two basis sets, one consisting of 9 states, $(v_{ss}, v_b, v_{as}) = (0,0,0)$, $(0,1,0)$, $(0,2,0)$, $(1,0,0)$, $(1,1,0)$, $(1,2,0)$, $(0,0,1)$, $(0,1,1)$, and $(0,2,1)$, and a second consisting of 14 states, where the additional states were $(0,3,0)$, $(1,3,0)$, $(0,3,1)$, $(1,0,1)$, and $(1,1,1)$. The Hamiltonian matrices, the eigenvalues, and the basis energies for both the 9×9 and 14×14 cases are provided in Sect. 3 of the Supporting Information.

All calculations were performed using the harmonic basis energies as diagonals. Table S3 shows the results for the two basis sets and for each of the algorithms.

S3.5 Tests of algorithms on randomly generated Hamiltonian matrices of dimensions from 3 to 20

In order to further test the algorithms and to see how their success depends on the dimension of the Hamiltonian, we randomly generated Hamiltonians, diagonalized them, executed the algorithms using only the eigenvalues and the diagonal Hamiltonian elements, and compared the calculated results to the exact answers. Each random Hamiltonian was generated by selecting diagonal elements randomly on the region (1-5), selecting off-diagonal elements randomly on the region (0-0.2), and ensuring symmetry. Four different Hamiltonians were tested for each dimension, and, for each Hamiltonian, results were obtained using both algorithms. For each test, 100 trials were calculated. In the case of the Inverse Diagonalization (MCID) algorithm, we used a value of $F = 0.001$. In the case of the MCD algorithm, we choose 2000 iterations for each trial and we assumed initially that the near off-diagonal elements were zero. We also performed the above tests with 1000 trials but found that the R^2 values were changed on average by less than 0.001 from those for 100 trials. For the trials at each dimension, we determined the mean value of R^2, the standard deviation, and the mean value of the time required per trial using a common laptop computer.

Results of the tests are shown in Fig. S2. The top panel gives R^2 for the magnitudes of the vector components as a function of n_{dim}, for $n_{dim} = 3, 5, 9, 14,$ and 20. The results for the MICD algorithm are in blue, while those for the MCD algorithm are in red. The two algorithms, as expected, show similar results, and both exhibit a fall-off in R^2 with an increase in n_{dim}. It should be noted that these random matrices described above pose a particularly stringent test for the algorithms. Most real applications will have Hamiltonian matrices with smaller and sparser off-diagonal elements, as seen from the examples above. We note from the top panel that the accuracy of the algorithms, even for these difficult matrices, is good up until somewhere between $n_{dim} = 14$ and $n_{dim} = 20$.

The bottom panel of Fig. S2 gives the log_{10} of the time per trail in seconds. The log_{10} of the time increases with the dimensionality, but since only about 100 steps are necessary, the calculations for all dimensions shown are relatively rapid, each on the order of minutes.

S4. Supplementary Tables

Table S4. Formic acid dimer normal mode harmonic frequencies (in cm^{-1}), symmetries, and dipole derivatives

Mode	Freq.	Symmetry	$\frac{\partial\mu_x}{\partial Q}$	$\frac{\partial\mu_y}{\partial Q}$	$\frac{\partial\mu_z}{\partial Q}$
1	70	A_u	0.000000	-0.001080	0.000000
2	167	A_g	0.000000	0.000000	0.000000
3	170	A_u	0.000000	0.002289	0.000000
4	209	A_g	0.000000	0.000000	0.000000
5	254	B_g	0.000000	0.000000	0.000000
6	275	B_u	-0.000913	0.000000	-0.005996
7	693	A_g	0.000000	0.000000	0.000000
8	716	B_u	0.002764	0.000000	0.003737
9	956	B_g	0.000000	0.000000	0.000000
10	970	A_u	0.000000	0.009038	0.000000
11	1084	B_g	0.000000	0.000000	0.000000
12	1100	A_u	0.000000	-0.004781	0.000000
13	1255	A_g	0.000000	0.000000	0.000000
14	1258	B_u	0.013354	0.000000	0.004738
15	1406	B_u	-0.000134	0.000000	-0.004417
16	1408	A_g	0.000000	0.000000	0.000000
17	1448	B_u	-0.001229	0.000000	-0.000026
18	1481	A_g	0.000000	0.000000	0.000000
19	1715	A_g	0.000000	0.000000	0.000000
20	1780	B_u	-0.019076	0.000000	0.007760
21	3095	A_g	0.000000	0.000000	0.000000
22	3097	B_u	0.000250	0.000000	0.010534
23	3232	A_g	0.000000	0.000000	0.000000
24	3326	B_u	-0.003997	0.000000	-0.034996

Table S5. 9×9 Hamiltonian for CO_2 example of main text. Note that because of the quartic term in the perturbation, the diagonal elements of the Hamiltonian are different than the basis function energies.

$$\mathbf{H} = \begin{pmatrix} 1014 & 0 & -2 & -54 & 0 & -36 & -3 & 0 & -10 \\ 0 & 1692 & 0 & 0 & -105 & 0 & 0 & -18 & 0 \\ -2 & 0 & 2378 & -36 & 0 & -157 & -10 & 0 & -33 \\ -54 & 0 & -36 & 2352 & 0 & -16 & -9 & 0 & -51 \\ 0 & -105 & 0 & 0 & 3009 & 0 & 0 & -82 & 0 \\ -36 & 0 & -157 & -16 & 0 & 3675 & -51 & 0 & -155 \\ -3 & 0 & -10 & -9 & 0 & -51 & 3695 & 0 & -31 \\ 0 & -18 & 0 & 0 & -82 & 0 & 0 & 4331 & 0 \\ -10 & 0 & -33 & -51 & 0 & -155 & -31 & 0 & 4976 \end{pmatrix}.$$

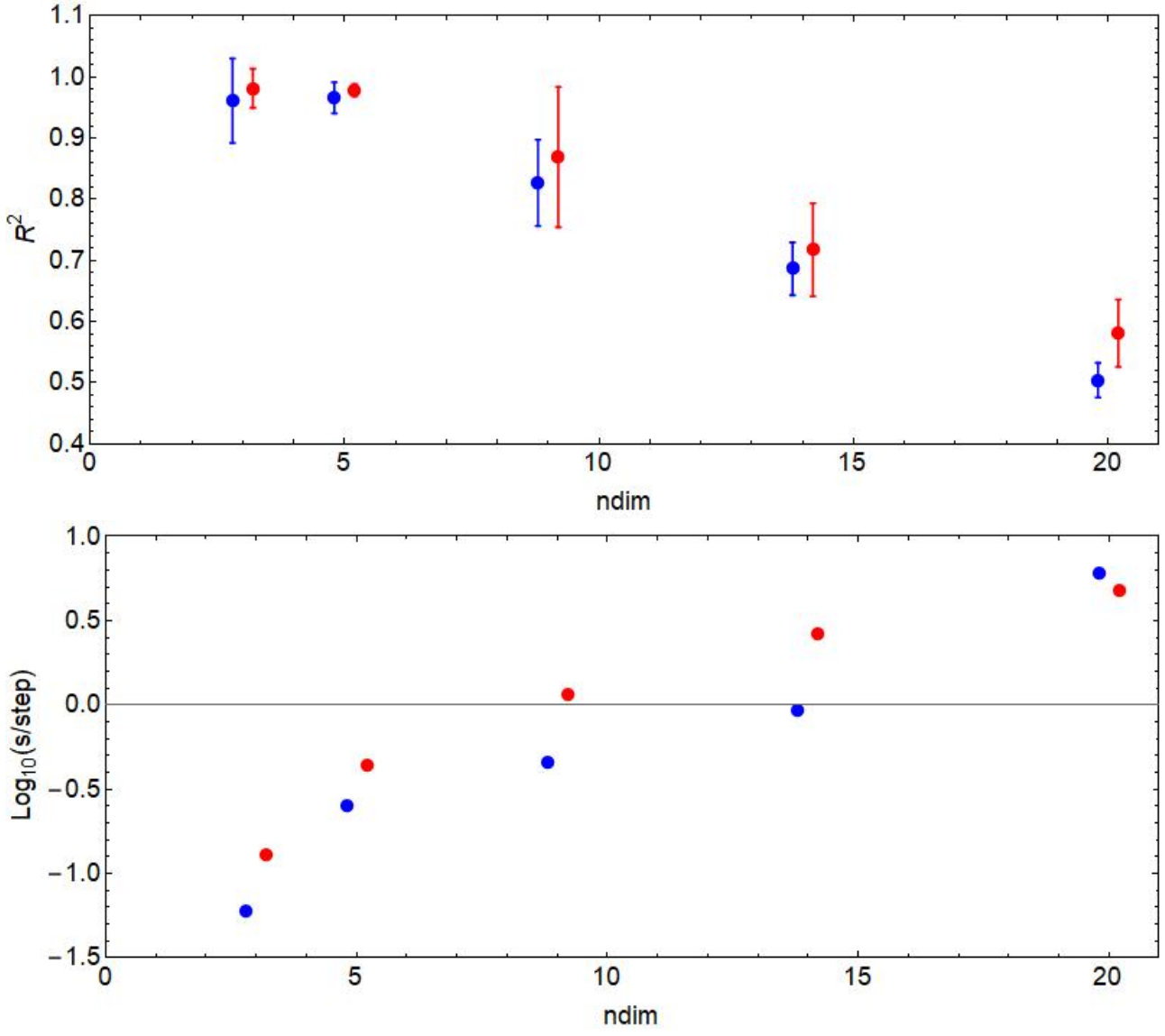

Figure S2. Results of tests with random Hamiltonian matrices. Red data are for the MCD algorithm, whereas blue data are for the MCID algorithm. The top panel gives R^2 as a function of n_{dim}, whereas the bottom panel gives the log_{10} of the time per trial in seconds. Red and blue data are displaced slightly along the abscissa for clarity.

Table S6. Properties for the fit of the jet cooled formic acid dimer spectrum. The first three columns give the harmonic basis energies and identities in order of increasing energy. Column 4 gives the eigenvalues from the fit in order of increasing energy. Columns 5-10 give the three most important components of the expansion $\psi^{(i)} = \sum_j c_j^{(i)} \chi_j$, where i is the eigenvalue, j is the basis state, and χ_j are harmonic basis functions. The last column gives the sum of the squares of the three components listed $S = \sum_{j=1}^{3} |c_j^{(i)}|^2$. For example, the first eigenvector is identified with basis state 1 ($2\nu_2 + \nu_{15}$) with a probability amplitude of 0.945, basis state 3 (ν_{20}) with a probability amplitude of 0.2324, and basis state 2 ($\nu_6 + \nu_{18}$) with a probability amplitude of 0.032. These three components comprise 0.999 of the total probability.

j/i	Basis E	State	Eigenval.	j	$\lvert c_j^{(i)}\rvert$	j	$\lvert c_j^{(i)}\rvert$	j	$\lvert c_j^{(i)}\rvert$	S
1	1740	$2\nu_2 + \nu_{15}$	1732.98	1	0.945	3	0.324	2	0.032	0.999
2	1756	$\nu_6 + \nu_{18}$	1754.36	2	0.998	3	0.055	36	0.024	0.999
3	1780	ν_{20}	1784.17	3	0.944	1	0.325	2	0.048	0.999
4	2540	$3\nu_6 + \nu_{19}$	2529.72	4	0.940	5	0.288	10	0.100	0.976
5	2585	$\nu_4 + \nu_6 + \nu_7 + \nu_{16}$	2580.93	5	0.890	6	0.274	4	0.270	0.941
6	2606	$\nu_4 + \nu_6 + \nu_8 + \nu_{15}$	2595.86	6	0.809	7	0.419	5	0.312	0.927
7	2615	$2\nu_4 + \nu_8 + \nu_{18}$	2610.6	7	0.721	6	0.432	8	0.408	0.873
8	2644	$2\nu_7 + \nu_{14}$	2630.15	9	0.537	7	0.472	11	0.470	0.732
9	2649	$2\nu_6 + \nu_7 + \nu_{15}$	2644.4	8	0.721	9	0.589	7	0.195	0.905
10	2661	$\nu_{13} + \nu_{15}$	2655.59	10	0.889	9	0.266	8	0.190	0.897
11	2666	$\nu_{14} + \nu_{16}$	2665.85	11	0.696	12	0.469	9	0.405	0.869
12	2683	$\nu_6 + \nu_7 + \nu_{19}$	2675.55	12	0.789	11	0.414	13	0.290	0.879
13	2703	$\nu_{13} + \nu_{17}$	2702.19	13	0.909	12	0.277	11	0.127	0.919
14	2739	$\nu_{14} + \nu_{18}$	2730.59	14	0.864	15	0.356	20	0.142	0.893
15	2771	$\nu_6 + \nu_8 + \nu_{20}$	2763.93	15	0.792	14	0.408	18	0.194	0.831
16	2785	$\nu_6 + 2\nu_{13}$	2769.15	16	0.499	17	0.451	20	0.374	0.592
17	2791	$\nu_6 + 2\nu_{14}$	2779.28	18	0.691	16	0.638	25	0.136	0.903
18	2792	$2\nu_7 + \nu_{15}$	2794.37	17	0.661	18	0.376	16	0.363	0.710
19	2814	$2\nu_8 + \nu_{15}$	2811.39	22	0.569	20	0.524	17	0.434	0.786
20	2817	$\nu_{15} + \nu_{16}$	2817.91	19	0.691	21	0.318	22	0.314	0.677
21	2834	$\nu_7 + \nu_8 + \nu_{16}$	2822.52	19	0.602	21	0.513	20	0.489	0.865
22	2838	$2\nu_7 + \nu_{17}$	2844.36	21	0.549	23	0.485	22	0.438	0.729
23	2856	$\nu_{16} + \nu_{17}$	2857.24	23	0.770	24	0.266	20	0.263	0.733
24	2880	$2\nu_8 + \nu_{17}$	2879.55	25	0.708	24	0.330	27	0.321	0.713
25	2887	$\nu_{15} + \nu_{18}$	2891.88	24	0.745	22	0.425	25	0.342	0.853
26	2916	$\nu_6 + 2\nu_7 + \nu_{13}$	2914.45	26	0.736	29	0.379	28	0.336	0.799
27	2929	$\nu_{17} + \nu_{18}$	2932.81	27	0.693	26	0.443	25	0.322	0.780
28	2938	$\nu_6 + \nu_{13} + \nu_{16}$	2938.72	28	0.775	29	0.335	30	0.308	0.808
29	2939	$\nu_6 + \nu_{14} + \nu_{15}$	2946.7	29	0.773	27	0.432	26	0.294	0.870
30	2973	$\nu_{14} + \nu_{19}$	2970.72	30	0.769	31	0.480	28	0.290	0.905
31	2981	$\nu_6 + \nu_{14} + \nu_{17}$	2995.67	31	0.802	30	0.438	27	0.242	0.893
32	3011	$\nu_6 + \nu_{13} + \nu_{18}$	3013.29	32	0.963	33	0.103	30	0.101	0.948
33	3035	$\nu_{13} + \nu_{20}$	3030.35	33	0.852	34	0.392	28	0.175	0.910
34	3063	$2\nu_6 + \nu_{13} + \nu_{14}$	3060.14	34	0.757	36	0.477	33	0.276	0.876
35	3087	$\nu_6 + 2\nu_{15}$	3083.98	35	0.809	37	0.363	34	0.277	0.863
36	3097	ν_{22}	3105.18	36	0.716	35	0.368	34	0.352	0.772
37	3121	$\nu_{15} + \nu_{19}$	3130.67	37	0.893	35	0.235	39	0.228	0.905
38	3163	$\nu_{17} + \nu_{19}$	3168.74	38	0.928	39	0.256	36	0.186	0.962
39	3171	$\nu_6 + 2\nu_{17}$	3178.1	39	0.885	38	0.266	35	0.229	0.906
40	3237	$\nu_6 + 2\nu_{18}$	3248.2	40	0.965	39	0.212	38	0.073	0.981
41	3326	ν_{24}	3372.88	41	0.929	36	0.264	42	0.086	0.940
42	3507	$\nu_6 + \nu_{23}$	3510.86	42	0.994	41	0.084	38	0.033	0.995

Table S7. Properties for the fit of the room-temperature formic acid dimer spectrum. The first three columns give the harmonic basis energies and identities in order of increasing energy. Column 4 gives the eigenvalues from the fit in order of increasing energy. Columns 5-10 give the three most important components of the expansion $\psi^{(i)} = \sum_j c_j^{(i)} \chi_j$, where i is the eigenvalue, j is the basis state, and χ_j are harmonic basis functions. The last column gives the sum of the squares of the three components listed $S = \sum_{j=1}^{3} |c_j^{(i)}|^2$. For example, the first eigenvector is identified with basis state 1 ($2\nu_2 + \nu_{15}$) with a probability amplitude of 0.873, basis state 3 (ν_{20}) with a probability amplitude of 0.442, and basis state 2 ($\nu_6 + \nu_{18}$) with a probability amplitude of 0.174. These three components comprise 0.988 of the total probability.

j/i	Basis E	Mode	Eigenval.	j	$\lvert c_j^{(i)}\rvert$	j	$\lvert c_j^{(i)}\rvert$	j	$\lvert c_j^{(i)}\rvert$	S
1	1740	$2\nu_2 + \nu_{15}$	1732.54	1	0.873	3	0.442	2	0.174	0.988
2	1756	$\nu_6 + \nu_{18}$	1741.78	3	0.777	1	0.483	2	0.378	0.979
3	1780	ν_{20}	1754.14	2	0.907	3	0.413	36	0.049	0.995
4	2540	$3\nu_6 + \nu_{19}$	2513.43	4	0.831	36	0.260	11	0.221	0.807
5	2585	$\nu_4 + \nu_6 + \nu_7 + \nu_{16}$	2533.68	7	0.539	5	0.501	4	0.364	0.674
6	2606	$\nu_4 + \nu_6 + \nu_8 + \nu_{15}$	2552.12	5	0.637	6	0.380	8	0.317	0.651
7	2615	$2\nu_4 + \nu_8 + \nu_{18}$	2586.	6	0.789	5	0.317	7	0.279	0.801
8	2644	$2\nu_7 + \nu_{14}$	2610.2	9	0.599	8	0.464	7	0.345	0.693
9	2649	$2\nu_6 + \nu_7 + \nu_{15}$	2624.51	7	0.603	11	0.539	5	0.314	0.752
10	2661	$\nu_{13} + \nu_{15}$	2641.76	10	0.849	11	0.220	13	0.181	0.801
11	2666	$\nu_{14} + \nu_{16}$	2647.69	8	0.670	12	0.449	9	0.365	0.784
12	2683	$\nu_6 + \nu_7 + \nu_{19}$	2676.41	12	0.765	8	0.281	9	0.267	0.736
13	2703	$\nu_{13} + \nu_{17}$	2686.65	13	0.781	14	0.365	10	0.268	0.815
14	2739	$\nu_{14} + \nu_{18}$	2722.78	14	0.574	13	0.419	11	0.308	0.600
15	2771	$\nu_6 + \nu_8 + \nu_{20}$	2743.51	15	0.593	14	0.435	17	0.386	0.690
16	2785	$\nu_6 + 2\nu_{13}$	2756.05	18	0.644	17	0.311	14	0.248	0.574
17	2791	$\nu_6 + 2\nu_{14}$	2778.98	19	0.616	17	0.372	16	0.313	0.616
18	2792	$2\nu_7 + \nu_{15}$	2791.91	20	0.667	16	0.329	17	0.298	0.642
19	2814	$2\nu_8 + \nu_{15}$	2795.37	16	0.698	14	0.325	17	0.325	0.698
20	2817	$\nu_{15} + \nu_{16}$	2802.21	21	0.476	23	0.409	22	0.401	0.555
21	2834	$\nu_7 + \nu_8 + \nu_{16}$	2826.39	21	0.630	20	0.364	22	0.317	0.629
22	2838	$2\nu_7 + \nu_{17}$	2841.07	17	0.464	24	0.384	19	0.377	0.505
23	2856	$\nu_{16} + \nu_{17}$	2847.03	15	0.435	22	0.385	21	0.305	0.431
24	2880	$2\nu_8 + \nu_{17}$	2866.57	23	0.514	22	0.435	24	0.362	0.584
25	2887	$\nu_{15} + \nu_{18}$	2880.86	25	0.722	27	0.294	22	0.245	0.668
26	2916	$\nu_6 + 2\nu_7 + \nu_{13}$	2903.59	26	0.510	29	0.491	24	0.400	0.661
27	2929	$\nu_{17} + \nu_{18}$	2925.16	28	0.639	27	0.419	30	0.344	0.701
28	2938	$\nu_6 + \nu_{13} + \nu_{16}$	2942.21	26	0.521	30	0.447	25	0.309	0.567
29	2939	$\nu_6 + \nu_{14} + \nu_{15}$	2958.78	31	0.590	27	0.438	24	0.355	0.666
30	2973	$\nu_{14} + \nu_{19}$	2973.79	27	0.535	26	0.374	30	0.346	0.545
31	2981	$\nu_6 + \nu_{14} + \nu_{17}$	2989.9	29	0.556	28	0.493	31	0.369	0.689
32	3011	$\nu_6 + \nu_{13} + \nu_{18}$	3008.26	32	0.736	31	0.363	24	0.284	0.754
33	3035	$\nu_{13} + \nu_{20}$	3022.28	33	0.506	32	0.422	31	0.356	0.561
34	3063	$2\nu_6 + \nu_{13} + \nu_{14}$	3049.95	33	0.667	34	0.364	30	0.283	0.657
35	3087	$\nu_6 + 2\nu_{15}$	3075.34	35	0.730	37	0.266	38	0.245	0.664
36	3097	ν_{22}	3105.79	34	0.624	37	0.435	33	0.287	0.662
37	3121	$\nu_{15} + \nu_{19}$	3138.12	37	0.658	40	0.415	34	0.336	0.719
38	3163	$\nu_{17} + \nu_{19}$	3169.04	38	0.664	39	0.612	37	0.194	0.853
39	3171	$\nu_6 + 2\nu_{17}$	3196.24	39	0.569	38	0.546	37	0.374	0.762
40	3237	$\nu_6 + 2\nu_{18}$	3292.24	40	0.804	39	0.336	38	0.286	0.841
41	3326	ν_{24}	3509.03	42	0.982	41	0.110	36	0.079	0.982
42	3507	$\nu_6 + \nu_{23}$	3712.65	41	0.687	36	0.542	35	0.252	0.830

Table S8. Probability amplitude for the basis functions ν_{24}, ν_{22}, and ν_{20} for the fit to the jet-cooled formic acid dimer spectrum

i	Eigenval.	$c_{24}^{(i)}$	$c_{22}^{(i)}$	$c_{20}^{(i)}$
1	1732.98	-0.0097	-0.0100	0.3241
2	1754.36	-0.0022	-0.0238	-0.0552
3	1784.17	0.0075	0.0055	0.9437
4	2529.72	0.0265	0.0197	0.0109
5	2580.93	-0.0300	0.0256	0.0143
6	2595.86	0.0095	-0.0344	0.0051
7	2610.6	-0.0277	-0.0229	0.0019
8	2630.15	0.0301	0.0115	-0.0010
9	2644.4	-0.0359	-0.0127	-0.0035
10	2655.59	-0.0139	-0.0134	-0.0104
11	2665.85	-0.0307	-0.0182	0.0013
12	2675.55	-0.0432	0.0041	-0.0037
13	2702.19	-0.0071	-0.0156	-0.0080
14	2730.59	0.0038	-0.0790	-0.0007
15	2763.93	0.0008	-0.1023	0.0008
16	2769.15	0.0209	-0.0040	0.0098
17	2779.28	-0.0357	-0.0010	-0.0071
18	2794.37	-0.0641	-0.0314	-0.0075
19	2811.39	0.0258	0.0195	0.0022
20	2817.91	-0.0134	-0.0185	0.0069
21	2822.52	0.0492	0.0013	0.0048
22	2844.36	0.0080	-0.1249	0.0086
23	2857.24	0.0485	0.0546	-0.0000
24	2879.55	-0.0463	0.0698	-0.0031
25	2891.88	-0.0492	-0.0122	-0.0016
26	2914.45	0.0330	-0.0968	0.0060
27	2932.81	0.0594	0.0421	0.0048
28	2938.72	-0.0485	0.0297	0.0040
29	2946.7	0.0665	0.0096	-0.0081
30	2970.72	0.0563	0.0292	0.0034
31	2995.67	0.0792	0.0431	-0.0047
32	3013.29	-0.0297	0.0695	0.0010
33	3030.35	-0.0064	0.1408	0.0001
34	3060.14	0.2025	0.4766	-0.0075
35	3083.98	-0.1073	-0.1564	0.0095
36	3105.18	-0.1641	-0.7156	-0.0014
37	3130.67	-0.0143	-0.2077	-0.0042
38	3168.74	0.0354	0.1856	0.0003
39	3178.1	0.0562	0.0515	-0.0014
40	3248.2	-0.0047	0.0005	0.0047
41	3372.88	-0.9291	0.2641	0.0031
42	3510.86	-0.0835	0.0178	0.0007

Table S9. Probability amplitude for the basis functions ν_{24}, ν_{22}, and ν_{20} for the fit to the room-temperature FAD spectrum

i	Eigenval.	$c_{24}^{(i)}$	$c_{22}^{(i)}$	$c_{20}^{(i)}$
1	1732.54	0.0415	0.0850	-0.4416
2	1741.78	0.0024	-0.1255	0.7765
3	1754.14	0.0196	0.0491	-0.4126
4	2513.43	-0.0157	-0.2604	-0.0605
5	2533.68	0.0523	-0.2465	-0.0419
6	2552.12	0.0090	0.2020	0.0762
7	2586.	0.0771	0.0113	0.0168
8	2610.2	0.0665	-0.0534	-0.0007
9	2624.51	0.0200	0.0951	0.0275
10	2641.76	-0.0353	-0.0507	-0.0138
11	2647.69	0.0187	0.0710	0.0257
12	2676.41	-0.0735	0.0269	0.0079
13	2686.65	-0.0116	0.1490	0.0366
14	2722.78	-0.1175	0.1577	0.0161
15	2743.51	0.0851	-0.0679	-0.0067
16	2756.05	0.0730	-0.1610	-0.0189
17	2778.98	-0.0972	-0.0728	-0.0279
18	2791.91	-0.1380	0.0249	-0.0100
19	2795.37	0.0117	-0.0616	-0.0073
20	2802.21	0.0576	-0.1112	-0.0162
21	2826.39	0.1175	-0.0841	-0.0242
22	2841.07	-0.0682	0.0372	-0.0143
23	2847.03	-0.1333	0.0963	0.0165
24	2866.57	0.1005	-0.1636	-0.0268
25	2880.86	0.1683	-0.0340	0.0073
26	2903.59	0.1514	-0.0745	-0.0125
27	2925.16	-0.1330	0.1095	0.0351
28	2942.21	-0.2482	0.0939	-0.0091
29	2958.78	0.1079	-0.1627	-0.0274
30	2973.79	-0.0482	0.0965	0.0045
31	2989.9	0.1390	-0.1135	-0.0004
32	3008.26	0.1109	-0.1426	-0.0257
33	3022.28	0.1301	-0.1978	-0.0358
34	3049.95	-0.1354	0.0513	-0.0021
35	3075.34	-0.2063	-0.0189	-0.0167
36	3105.79	0.1899	-0.2077	-0.0289
37	3138.12	0.1841	-0.1757	-0.0108
38	3169.04	-0.1760	0.1682	0.0177
39	3196.24	0.1386	-0.2177	-0.0244
40	3292.24	-0.1170	0.2287	0.0196
41	3509.03	-0.1095	-0.0791	-0.0184
42	3712.65	0.6874	0.5420	0.0803

S5. Supplementary Figures

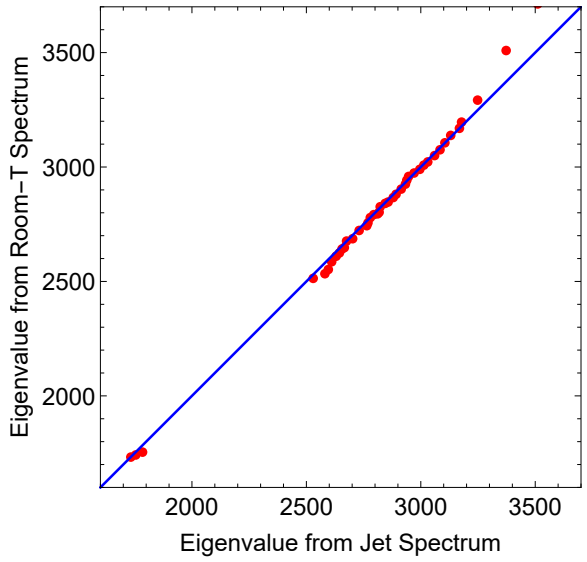

Figure S3. Correlation plot showing the eigenvalues obtained from the fit to the room-temperature spectrum as a function of the eigenvalues obtained from the jet-cooled spectrum. The blue line is of unit slope. The value of R^2 for the correlation is 0.987.

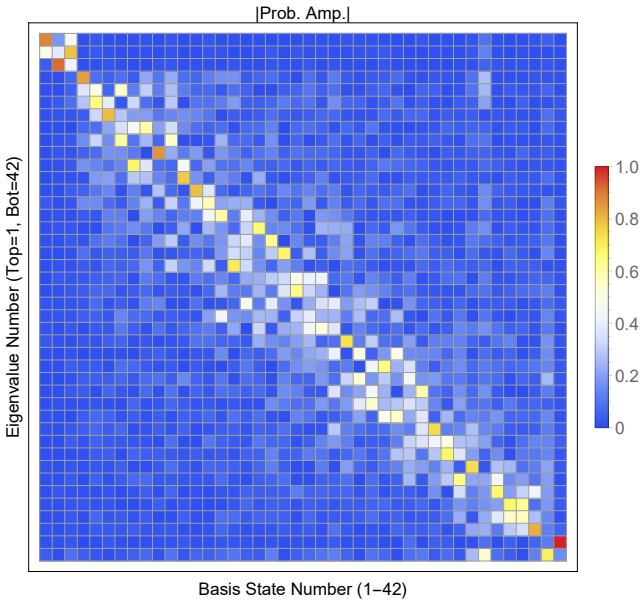

Figure S5. Array Plot of the eigenvectors corresponding to the best fit to the room-temperature formic acid dimer spectrum. Each cell depicts the absolute magnitude of the corresponding eigenvector component (probability amplitude). Color coding of the elements is given by the bar at the right. Eigenvectors run from the top row (number 1) to the bottom row, whereas basis states run from the left column (number 1) to the right column.

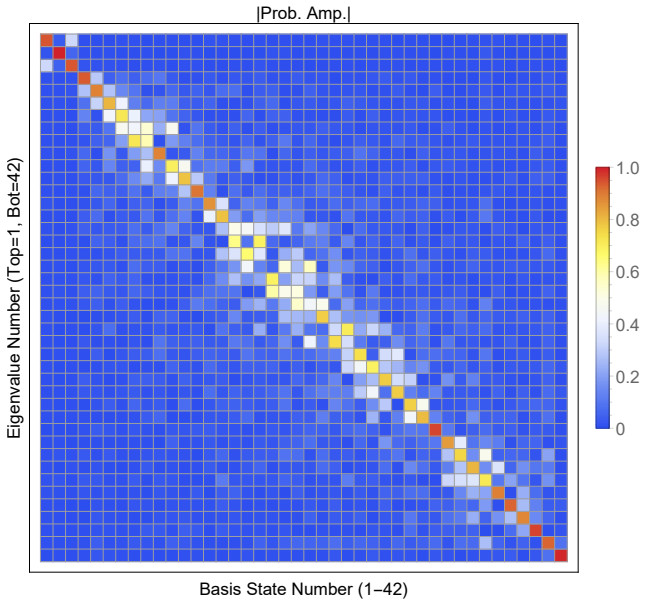

Figure S4. Array Plot of the eigenvectors corresponding to the best fit to the jet-cooled formic acid dimer spectrum. Each cell depicts the absolute magnitude of the corresponding eigenvector component (probability amplitude). Color coding of the elements is given by the bar at the right. Eigenvectors run from the top row (number 1) to the bottom row, whereas basis states run from the left column (number 1) to the right column.

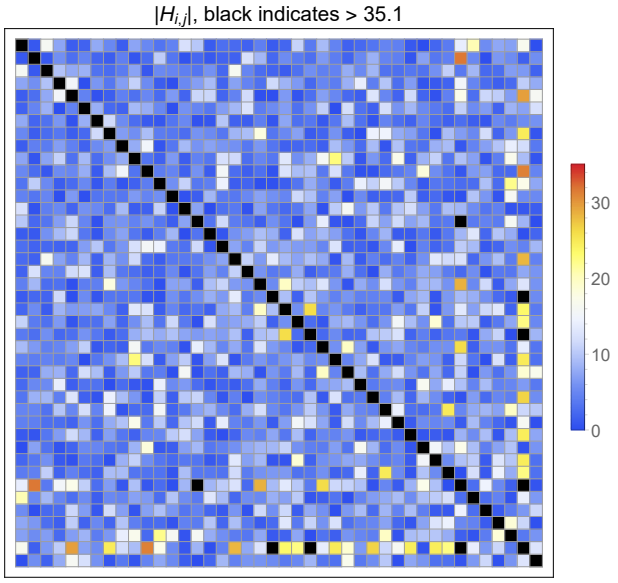

Figure S6. Array Plot of the Hamiltonian giving the best fit to the jet-cooled spectrum of the formic acid dimer. Each cell depicts the absolute magnitude of the corresponding $H_{i,j}$ element. Color coding of the elements is given by the bar at the right. Black indicates that the absolute values of the elements are larger than 35.1 cm^{-1}. The identities and energies of the basis states are given in Table S1. Note that basis states 36 (ν_{22}) and 41 (ν_{24}) have mixing with a large number of other states.

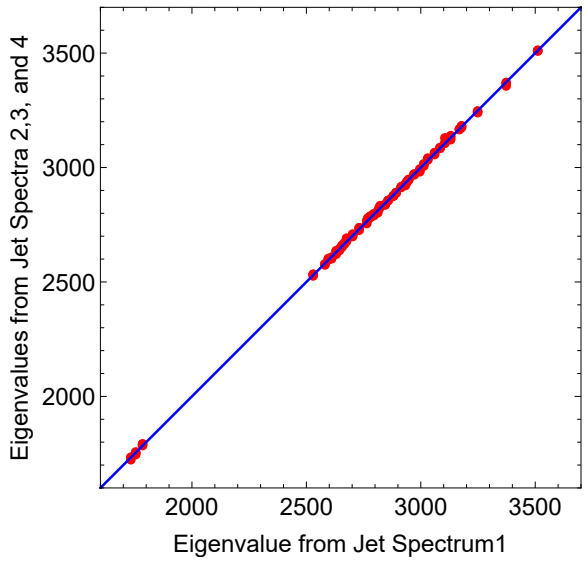

Figure S7. Correlation plot showing the eigenvalues obtained from fits 2, 3, and 4 as a function of the eigenvalues obtained from fit 1 to the jet-cooled spectrum of the formic acid dimer. The blue line is of unit slope. The value of R^2 for the correlation is 0.9997.

Acknowledgements

Support from the National Science Foundation (CHE-1463552) is gratefully acknowledged.