Journal of Molecular Spectroscopy 379 (2021) 111467

Contents lists available at ScienceDirect

Journal of Molecular Spectroscopy

journal homepage: www.elsevier.com/locate/jms

A high speed fitting program for rotational spectroscopy

P. Brandon Carroll, Kin Long Kelvin Lee, Michael C. McCarthy

Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA, United States

ARTICLE INFO

Article history:
Received 6 October 2020
In revised form 25 March 2021
Accepted 2 April 2021
Available online 03 May 2021

Keywords:
Rotational spectroscopy
Numerical methods

ABSTRACT

The advent of chirped-pulse Fourier transform microwave spectroscopy has dramatically increased the acquisition rate of high-resolution data. However, this data is only useful when it can be translated into molecule-specific information, and ultimately chemical knowledge. To adapt to this high data rate regime, new data analysis tools, which do not result in commensurate increases in time, are needed. Several methods, both computational and experimental, have been developed to simplify and automate the assignment of microwave spectra, and such tools can also make microwave spectroscopy far more approachable and usable for non-experts. Computationally efficient spectral assignment is a particularly promising avenue to enable rapid and nearly exhaustive spectral identification. Towards this end, we have recently developed an efficient, high-speed program for the prediction and fitting of rotational spectra. This program is built on a simple framework that is applicable to a wide variety of molecules. We present this program, discuss its performance, and demonstrate its use in fitting broadband spectra at centimeter wavelengths.

1. Introduction

The relentless improvement of high-speed electronics driven principally by the telecommunications industry has resulted in the widespread availability of extremely high bandwidth frequency sources, digitizers, and high power amplifiers[1]. The advent of chirped pulse Fourier transform microwave (CP-FTMW) spectroscopy [2] successfully leveraged these advances, with numerous instruments based on this approach now in operation throughout the radio band [3–7]. These spectrometers routinely supply large, information-rich data sets [8,9] at a rate far faster than the data can be analyzed, resulting in spectral assignment limiting both the rate and completeness of analysis.

Broadband rotational spectra present a particularly difficult set of challenges for spectral assignment. Among these are the nonlinear scaling of potential solutions with data set size, the large potential parameter space and sparse solution space, and the limitations inherent to experimental data, e.g., finite frequency coverage, intensity variations, etc. A further complication is that a spectrum almost invariably contains features from more than one carrier or variant; the high-sensitivity of CP-FTMW spectrometers allows for species of low abundance, for example isotopologues, to be readily observed. Taken together, an enormous number of potential solutions may exist for any data set. Confidently decomposing experimental spectra into individual molecular components is therefore a formidable undertaking.

This situation is further exacerbated for open-ended searches where few constraints may be available in advance. Because little information is encoded in a single rotational feature, as a practical matter it is nearly impossible to make definitive assignments without *a priori* rotational constants. However, the magnitude of observed rotational constants can span several orders of magnitude depending on the precursors and the type of experiment. As illustrated in Fig. 1, even for a rigid rotor, there are numerous local minima in search space, a phenomenon that often precludes use of standard least-squares fitting approaches. Furthermore, because solution space is often extremely narrow, different minima may be found depending on the search resolution, with the global minimum only obvious at high resolution.

As an example, for a spectrum containing 100 unknown features, there are $\sim 1.6 \times 10^5$ possible sets of three transitions, the minimum needed to fit all three rotational constants. The number of unique sets grows exponentially with the number of features. In practice using reasonable exclusion criteria far fewer sets will be plausible, however it should be noted that for larger data sets worst case scenarios can still reach several million sets [10]. This requires an input of initial rotational constants. When these are not known, a search of rotational constants scales with the cube of the search range and inverse cube of the search resolution. Practical values give millions or more unique sets of rotational constants. Together these two factors alone can easily result in several trillion unique sets of constants and features to be checked.

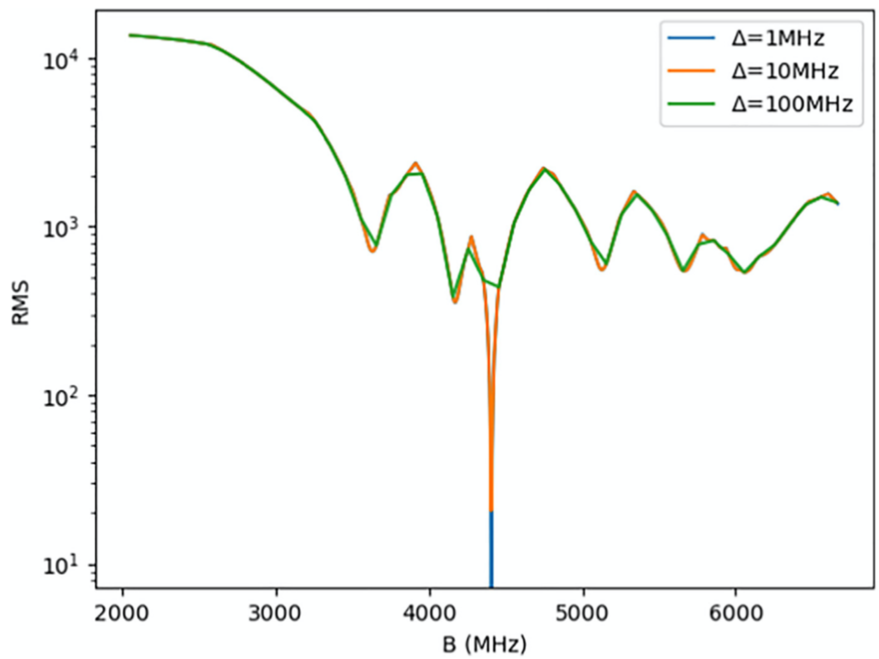


Fig. 1. RMS values as a function of the *B* rotational constant for a hypothetical rigid asymmetric top rotor. The surface is calculated for a reference molecule with A = 6666 MHz, B = 4408 MHz, C = 2056 MHz, using its ten most intense transitions at 2 K. RMS is calculated as a function of *B* with *A* and *C* fixed to their correct values, using a catalog of transitions with $J < 7$, $K_a < 3$. For each value of *B*, the nearest line predicted to each reference molecule transition is used to compute the chi squared value. Δ refers to the step size in *B*.

To address this multitude of factors, numerous computational and experimental methods have been developed to aid in rotational assignment[11,12,10,13–15]. None of these methods are complete solutions alone, and may struggle in various situations. Regardless of the specific approach, a constraint for many is simply the time required to calculate a rotational spectrum. As such, strongly nonlinear scaling can make it difficult or impossible to fully explore parameter space when potentially millions to billions of spectra must be considered. Alternative solutions such as machine learning may be able to address these issues, but at present struggle with real world data sets [13]. An attractive alternative is to simply to increase the sheer speed with which rotational spectra are calculated.

To that end, we have developed a method for high-speed calculation of rotational spectra of asymmetric top rigid rotors[16]. An archived version of the code discussed can be found with the digital object identifier doi:10.5281/zenodo.3950543[1]. By precalculating rigid rotor eigenvalues, the need for expensive matrix diagonalization is eliminated. This method is not only inherently much faster, but also eliminates the need for many extraneous calculations, thereby allowing us to realize a several of order of magnitude speed-up relative to programs commonly used by the microwave community. This method can be used as the basis for building search and fitting algorithms that previously would have been computationally intractable. This simple brute force algorithm has already been used with considerable success in studies to determine rotational constants for entirely new molecules based on a very small number of double resonance linkages[17,9]. The following sections describe the implementation and performance of this program.

2. Implementation

calculated set of eigenvalues for all states and geometries to be used, as described in Section 2.1; second, a catalog listing transitions as connectivity between states, and an associated dipole axis; third, a dictionary containing quantum numbers associated with each state. For speed and simplicity the archived version of the catalog contains only first order ($\Delta K_{a,b,c} = 0, \pm 1$) transitions, however it is straightforward for the user to add rotational transitions to the file. Energies of rotational levels are then calculated by passing rotational constants and desired transitions to the program's energy calculation subroutine. For each state, the energy is found by locating the state's nearest eigenvalues in the program's precalculated eigenvalues and linearly interpolating between them.

2.1. Calculation the Rigid Rotor Eigenvalues

The asymmetric top rigid rotor Hamiltonian and its solutions have been thoroughly described [18]. The Hamiltonian is given in Eq. 1, where $A/B/C$ are the rotational constants about the *a*-, *b*-, and *c*-inertial axes, respectively, and are related to the principal moments of inertia by $A/B/C = h^2/(8\pi^2 I_{a/b/c})$. For an asymmetric top, where $A > B > C$, analytic expressions for the eigenvalues only exist for a handful of states. General solutions therefore require the Hamiltonian to be solved numerically.

$$H = A\widehat{J}_a^2 + B\widehat{J}_b^2 + C\widehat{J}_c^2 \tag{1}$$

As suggested by Ray [19], the Hamiltonian can be refactored as a combination of total angular momentum and an asymmetric contribution (Eq. 2) using a unitless asymmetry parameter $\kappa = (2B - A - C)/(A - C)$, where κ takes on values from -1 to $+1$, which represent the prolate and oblate limits, respectively. This formulation greatly simplifies the calculation by allowing eigenvalues only be solved in terms of κ.

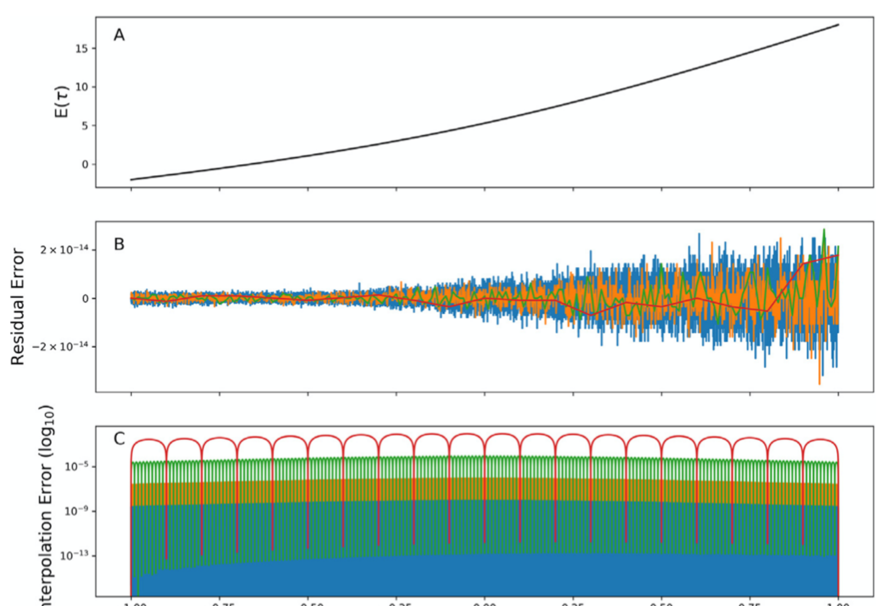
To derive numerical solutions, a basis set must be chosen, and a convenient choice are those of a symmetric rotor. The matrix elements of symmetric rotor functions are well known, and are diagonal in J[18]. This feature allows eigenvalues to be solved for a single J, rather than for the entire Hamiltonian, without any loss in generality. Matrices are constructed and solved one quantum of J at a time. Matrix elements are determined as a function of κ in the I^r representation, although solutions are independent of representation [18]. The eigenvalues are then calculated in units of $\hbar$ and stored for all values of κ from -1 to $+1$ at a resolution $\Delta\kappa$ using a Python implementation. The accuracy of the eigenvalues can be roughly estimated by comparing the handful of states with analytic solutions to their numerical values. In all cases these agree to within one to two orders of magnitude of machine epsilon. An example is given in Fig. 2, panel B. The chosen resolution will depend on the required accuracy, but is generally of order 10^{-2}-10^{-4} for most high-resolution applications. Eigenvalues are mapped to asymmetric rotor states by noting energies and states, as labeled by τ = (K_a-K_c) within a given J manifold, are both in ascending order. The values are then saved for use by the fitting routine. With these values calculated, the eigenvalues are stored as an external file that is read by the program at run time. With this suite of functions built, the basic functions of a spectral prediction and fitting engine can be performed.

2.2. Additional Functions

Several ease of use functions are available, including those to calculate individual state energies and approximate intensities, predict full catalogs, select subsets of catalogs based on parameters such as J, K_a, and intensity, as well as sorting of catalogs. Additionally, a subroutine allows rotational constants to be fit to a set of experimental frequencies and their assigned transitions. This subroutine uses the GSL implementation of the Levenberg–Marquardt nonlinear least-squares fitting algorithm [20] which is also used in CALPGM [12].

To improve calculation speed, individual state energies are not retained by the program during frequency calculations. Once the absolute energy of the two states in a transition are calculated, only the difference is retained as a transition frequency. For functions that require individual state energies, it is inefficient to recalculate all energies, and they are instead calculated using the existing frequency information. For this function, individual lower and upper state energies are calculated by assuming the $0_{0,0}$ state is at a zero energy ($0\,\mathrm{cm}^{-1}$) and counting up iteratively. Intensity calculations are at present somewhat imprecisely implemented. The program calculates intensities according to Eq. (1) of Ref. [12], with two modifications. First, normalization by the partition function is not performed. Second, S_{ij} values are omitted since they are both state and geometry dependent, and would therefore require a dedicated and perhaps substantial subroutine to calculate correctly.The explicit form used is $\mathrm{I(T)} = (8\pi^3/3\mathrm{hc})\nu\,[e^{E_{lower}/kT}\text{-}e^{E_{upper}/kT}]$. However, for most practical applications S_{ij} values vary by at most an order of magnitude across the majority of molecular transitions, and are therefore omitted in the present implementation. The degree to which this affects intensity accuracy will be geometry dependent, but as an example, the S_{ij} of the 4_{04}-4_{14} varies from 4.5 to 7.2 from $\kappa = -1$ to $+1$ respectively, while the 4_{04}-3_{13} varies from 1.5 to 3.5, and the 7_{17}-6_{06} varies from 4.0 to 6.5. More extreme cases like the 4_{41}-5_{32} vary by several orders of magnitude due to the near zero values of S_{ij} near $|\kappa| = 1$. For more exact intensities, it should be relatively straightforward to implement S_{ij} lookup tables and partition function calculations in a future version of this code.

With respect to sorting, the initial ordering of the program's catalog is simply that of its reference files. When the user calculates catalogs with values of κ that differ from the input catalog, the ordering of states by frequency or energy may change. Many applications or algorithms require sorted catalogs, so re-sorting efficiently is a secondary but necessary consideration. Many sorting algorithms show improved performance for more ordered inputs, so the choice of input catalog is non-trivial[21]. Generally,



the more a user's kappa value differs from that of the reference file, the more disordered the catalog will be. For this reason the input reference catalog is chosen to be ordered as calculated for κ = 0 to ensure it is on average close to the user's value, reducing the time required for sorting.

3. Discussion

This program is potentially applicable to numerous problems in microwave spectroscopy. Its restriction to rigid rotors certainly precludes its use in more complex problems, for which programs such as CALPGM are far more suitable. However, in cases where rigid rotor approximations are sufficient it may be highly effective provided it meets two criteria.

3.1. Accuracy

The first criteria is the program's accuracy; the accuracy of a calculated frequency depends on several factors. From Eq. 2, it is clear that the absolute accuracy will decrease with J^2 and $A + C$. The magnitude of any error will then depend on the accuracy with which Eq. 3 is calculated. In the current implementation the dominant source of error results from interpolating between supplied eigenvalues. The importance of the interpolation error will vary depending on the eigenstate, resolution of numerical eigenvalue solution, and value of κ. This error will be periodic, with a magnitude that depends on the second derivative of the eigenvalue and the square of the numerical resolution, and reaches a maximum halfway between calculated points. In practice this difference amounts to $\sim \pm 10^{-3}\hbar$ for $\Delta\kappa = 0.1$ for J<5, and is the dominant source of error in nearly all calculations. A similar plot displaying the variation of this error with κ is given in panel C of Fig. 2. To determine the impact of this error on line frequencies, spectra for a set of molecules was calculated and compared with those derived from the CALPGM program [12]. 10000 sets of rotational constants were sampled from a normal distribution, with the magnitude of the constants between 1000–10000 MHz, for first-order one-quanta transitions ($\Delta K_{a,b,c}$ = 0,±1) from J = 0 to J = 25. The same rotational levels were predicted using our fitting program, using various values of $\Delta\kappa$. The average (and standard deviation) of the largest difference between CALPGM's and this work's catalogs were $8.41\times10^{-2}(1.434\times10^{-1})$ MHz for $\Delta\kappa$ = 10^{-3}, $8.38\times10^{-4}(1.45\times10^{-3})$ MHz for $\Delta\kappa$ = 10^{-4}, and 4.9×10^{-5} (1.20×10^{-5}) MHz for $\Delta\kappa 10^{-5}$. From this comparison we conclude the present method achieves the accuracy needed for essentially any microwave spectroscopy experiment, and furthermore can be selected by the user based on their specific application.

An additional obvious limitation of the program's accuracy is the rigid rotor model used to calculate energies. For molecules that substantially deviate from a rigid rotor, the accuracy achieved will be proportionately reduced. As an example, we compare the predicted frequencies of the fifty strongest transitions at 2 K for glycolaldehyde, a moderately large asymmetric top, neglecting varying levels of distortion with their experimental Hamiltonian counterparts[22]. These transitions, which span J = 0 to J = 7 and K_a=0 to K_a=4, have an RMS difference of 0.2 kHz and a maximum difference of 1 kHz when only quartic distortion constants are included. When only D_J is included, the RMS becomes 1.85 MHz, with the worst predicted transition 5.63 MHz off. If all distortion constants are neglected, the RMS is 2.72 MHz, with the worst predicted tran-

code to combat this. This would however have to be done for each effect that would be added, and may be challenging to implement with state splitting, or multiple effects of similar magnitude. Alternatively, future versions could include simple effective corrections. While not necessarily complete or physically meaningful, these could improve accuracy with minimal performance cost. Thus while the program is highly accurate, it is limited by the degree to which a molecule deviates from a rigid rotor. Practically this limits the use cases for this program, and will depend on the balance between a user's need for accuracy or speed, with programs like CALPGM more appropriate in the former case.

3.2. Performance

The second criteria is the program's performance, specifically the raw speed with which spectral line catalogs can be calculated or fit. To benchmark performance, we repeatedly calculated or fit a single catalog using the present method and CALPGM, taking care to ensure that both programs used the same number of transitions. Catalog prediction and sorting was performed using the same constants, A = 3 GHz, B = 2 GHz, and C = 1 GHz. For fitting, the same set of ten or one hundred lines based on the same catalog were fit using a guess of A = 3.333 GHz, B = 2.222 GHz, and C = 1.111 GHz. For the current implementation, the time to load external data is not included in calculation time, as it is minimal compared to calculations performed, as would be expected for most uses of this code. As indicated in Table 1, we achieve enhancements in speed relative to CALPGM by factors of 114 to 897 depending on the specific type of calculation. For CALPGM, calculating spectra, intensities, and sorting are all necessarily done at once, however for the current program these are separable. Therefore the catalog calculations do not include any sorting or calculation of energy or intensity. The speed of these operations is given separately in Table 1. For all speed calculations, the program is compiled using the CLANG compiler v11.0.0, and GCC version 2.4, with compiler flags -O3, -lGSL -lm, -lcblas, and -lgslcblas. The program is run using an Intel I9-9880H processor.

A notable performance difference between this method and direct solutions of a Hamiltonian is for the former, many extraneous calculations can be ignored versus the latter where these are unavoidable. As noted above, several operations, for example sorting a set of transitions, are necessary for methods that diagonalize a Hamiltonian. These are not required for the program's core functions, and can be ignored if the user's application does not require it. Similarly, for applications that only require a small number of strong transitions, the user can only calculate these states, rather than diagonalize a full set of states. This constraint can reduce the number of calculations by an order of magnitude and yield a commensurate increase in speed.

Additional performance improvements could be achieved by native parallelization of common tasks. Because the program uses minimal memory, it is relatively simple to create a large number of

Table 1
Performance comparison of the present method versus CALPGM. Independent energy and intensity calculations are not possible with CALPGM, and are therefore not included.

Operation	This work	CALPGM	Speed
	(kHz)	(Hz)	Up
Fitting (100 Lines)	17.9	155.9	114.8
Fitting (10 Lines)	150.4	167.5	897.9

parallel threads to perform tasks. Similarly, many of the common searches, e.g. triples fitting [10], are embarrassingly parallel—meaning that individual tasks within a search have no dependence on each other— and can be trivially split among as many threads as available. Such parallelization would allow even very large tasks to be completed quickly on large compute clusters.

This performance may be affected by the ability to load the required data, particularly into processor cache. By far the largest data block used is the eigenvalues. These are stored in a memory contiguous block as double precision floating point numbers. The size required scales with the precision with which the eigenvalues are solved, and the number of J states used. For a J of twenty-five and $\Delta\kappa$ of 10^{-3} only approximately 10 MB is required. This is equivalent to or slightly below the L3 cache size of many modern processors. Unfortunately the current implementation does not allocate the dictionary or catalog files in the same memory block as the eigenvalues, limiting the speed up of loading eigenvalues into cache. Future implementations could potentially achieve a large speed up by rectifying this. For larger J and higher precision additional memory constraints may apply. For a J of one hundred and a precision of $\Delta\kappa$ of 10^{-5}, approximately 16 GB is required. This represents a worst case scenario, but should fit in the RAM of a modern computer. For parallel computing, loading eigenvalues for each thread may prevent slowdowns due to resource sharing, with the caveat that memory requirements scale with the number of threads. In such edge cases, it may be necessary to sacrifice either speed or precision.

4. Conclusion

We have built and demonstrated a program capable of predicting and fitting rigid rotor rotational spectra at very high speed. The program substantially outperforms existing programs as measured by clock speed, while retaining sufficient accuracy for use with high precision rotational data. The potential uses for this program are varied. Simple use cases employing double resonance data have been demonstrated [17,9], although far more complex and varied methods can almost certainly be implemented using this program as the kernel. By doing so, the potential to perform analysis of rotational spectra at speeds commensurate with the ever increasing rate of data acquisition may be feasible.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

M.C.M., K.L.K.L., and P.B.C. acknowledge financial support from NSF grant AST-1908576 and NASA grant 80NSSC18K0396. P.B.C. is funded by the Simons Collaboration on the Origins of Life. We thank Zachary Nicolaou for helpful discussions.

References

[1] P.J. Pupalaikis, B. Yamrone, R. Delbue, A.S. Khanna, K. Doshi, B. Bhat, A. Sureka, Technologies for very high bandwidth real-time oscilloscopes, 2014 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), IEEE (2014) 128–135.

[2] G.G. Brown, B.C. Dian, K.O. Douglass, S.M. Geyer, S.T. Shipman, B.H. Pate, A broadband fourier transform microwave spectrometer based on chirped pulse excitation, Rev. Sci. Instrum. 79 (5) (2008) 053103.

[3] J.P. Porterfield, L. Satterthwaite, S. Eibenberger, D. Patterson, M.C. McCarthy, High sensitivity microwave spectroscopy in a cryogenic buffer gas cell, Rev. Sci. Instrum. 90 (5) (2019) 053104.

[4] E. Alonso, L. Kolesniková, I. Peña, S. Shipman, B. Tercero, J. Cernicharo, J. Alonso, Waveguide cp-ftmw and millimeter wave spectra of s-cis-and s-trans-acrylic acid, J. Mol. Spectrosc. 316 (2015) 84–89.

[5] I.A. Finneran, D.B. Holland, P.B. Carroll, G.A. Blake, A direct digital synthesis chirped pulse fourier transform microwave spectrometer, Rev. Sci. Instrum. 84 (8) (2013) 083104.

[6] D.P. Zaleski, J.L. Neill, M.T. Muckle, N.A. Seifert, P.B. Carroll, S.L.W. Weaver, B.H. Pate, A ka-band chirped-pulse fourier transform microwave spectrometer, J. Mol. Spectrosc. 280 (2012) 68–76.

[7] K.N. Crabtree, M.-A. Martin-Drumel, G.G. Brown, S.A. Gaster, T.M. Hall, M.C. McCarthy, Microwave spectral taxonomy: A semi-automated combination of chirped-pulse and cavity fourier-transform microwave spectroscopy, J. Chem. Phys. 144 (12) (2016) 124201.

[8] K.L.K. Lee, M. McCarthy, Study of benzene fragmentation, isomerization, and growth using microwave spectroscopy, The journal of physical chemistry letters 10 (10) (2019) 2408–2413.

[9] M.C. McCarthy, K.L.K. Lee, P.B. Carroll, J.P. Porterfield, P.B. Changala, J.H. Thorpe, J.F. Stanton, Exhaustive product analysis of three benzene discharges by microwave spectroscopy, J. Phys. Chem. A XX (33) (2020) XXXX.

[10] N.A. Seifert, I.A. Finneran, C. Perez, D.P. Zaleski, J.L. Neill, A.L. Steber, R.D. Suenram, A. Lesarri, S.T. Shipman, B.H. Pate, Autofit, an automated fitting tool for broadband rotational spectra, and applications to 1-hexanal, J. Mol. Spectrosc. 312 (2015) 13–21.

[11] B.P. Winnewisser, J. Reinstädtler, K.M. Yamada, J. Behrend, Interactive loomis-wood assignment programs, J. Mol. Spectrosc. 136 (1) (1989) 12–16.

[12] H. Pickett, R. Poynter, E. Cohen, M. Delitsky, J. Pearson, H. Müller, Submillimeter, millimeter, and microwave spectral line catalog, J. Quant. Spectrosc. Radiat. Transf. 60 (5) (1998) 883–890.

[13] D.P. Zaleski, K. Prozument, Automated assignment of rotational spectra using artificial neural networks, The Journal of chemical physics 149 (10) (2018) 104106.

[14] L. Yeh, L. Satterthwaite, D. Patterson, Automated, context-free assignment of asymmetric rotor microwave spectra, The Journal of chemical physics 150 (20) (2019) 204122.

[15] M.-A. Martin-Drumel, M.C. McCarthy, D. Patterson, B.A. McGuire, K.N. Crabtree, Automated microwave double resonance spectroscopy: A tool to identify and characterize chemical compounds, The Journal of chemical physics 144 (12) (2016) 124202.

[16] doi:10.5281/zenodo.3950543.

[17] J.P. Porterfield, K.L.K. Lee, V. Dell'Isola, P.B. Carroll, M.C. McCarthy, Characterization of the simplest hydroperoxide ester, hydroperoxymethyl formate, a precursor of atmospheric aerosols, PCCP 21 (33) (2019) 18065–18070.

[18] W. Gordy, R.L. Cook, Microwave Molecular Spectra, Wiley, 1984.

[19] B.S. Ray, Über die eigenwerte des asymmetrischen kreisels, Zeitschrift für Physik 78 (1–2) (1932) 74–91.

[20] B. Gough, GNU scientific library reference manual, Network Theory Ltd. (2009).

[21] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, et al., Numerical recipes, Vol. 3, Cambridge University Press Cambridge, 1989.

[22] P.B. Carroll, B.J. Drouin, S.L.W. Weaver, The submillimeter spectrum of glycolaldehyde, Astrophys J 723 (1) (2010) 845.