Isotopologue Consistency of Semi-Empirically Computed InfraRed Line Lists and Further Improvement for Rare Isotopologues: CO_2 and SO_2 Case Studies

Xinchuan Huang (黄新川),[1a,b] David W. Schwenke,[2c] and Timothy J. Lee[3d]

[a] MS 245-6, Astrophysics Branch, Space Science and Astrobiology Division, NASA Ames Research Center, Moffett Field, CA 94035, USA

[b] SETI Institute, 189 Bernardo Avenue, Suite 200, Mountain View, CA 94043, USA

[c] MS 258-2, NAS Facility, NASA Ames Research Center, Moffett Field, CA 94035, USA

[d] MS 245-3, Planetary Science Branch, Space Science and Astrobiology Division, NASA Ames Research Center, Moffett Field, CA 94035, USA

Submit to Journal of Quantitative Spectroscopy and Radiative Transfers

[1] Corresponding Author: Xinchuan.Huang-1@nasa.gov
[2] Email: David.W.Schwenke@nasa.gov
[3] Corresponding Author: Timothy.J.Lee@nasa.gov

Abstract

The semi-empirical molecular rovibrational IR line lists, such as ExoMol, TheoReTs, and Ames, combine the experimental accuracy and theoretical power to reach better than 0.1 cm^{-1} accuracy for line positions and better than 80-90% agreement for line intensities. The quality of these existing semi-empirical IR lists allows further improvements of intensity and line positions for those unobserved minor isotopologues. This paper presents our new BTRHE (Best Theory + Reliable High-resolution Experiment) strategy implementation. For line intensity, the isotopologue consistency and the patterns of mass dependence in the Ames-296K SO_2 and CO_2 IR lists are quantitatively presented along the mass-inverse coordinates. The consistency and patterns are better than those in existing experimental data. The methodology proposed here can be used to identify inconsistencies, outliers, and mistakes in intensities, and help improve Effective Dipole Model (EDM) and molecular IR databases. We call for an experimental study on the 50006 and 60007 bands of CO_2 628. For line position predictions, a simple approach combining the variational IR line lists with Effective Hamiltonian (EH) model may refine the effective rotational constants $A_0/B_0/C_0$ and quartic centrifugal distortion constants of minor isotopologues. The prediction accuracy may be improved by *two orders* of magnitude, i.e. reaching 0-5 MHz prediction accuracy in the range of J<20-30, K_a<10-20, and 0.01-0.02 MHz accuracy for $A_0/B_0/C_0$. Several important factors have been systematically investigated and discussed, e.g. convergence, uncertainties, higher order terms, fixing EH parameters, mass coordinates, etc. A microwave (MW) line set consisting of 644,636 strong transitions for all 30 isotopologues and corresponding refined EH(Ames) parameters are reported in the supplementary material. This approach may be easily extended to rovibrational bands, hot bands, and other molecular systems.

1. Introduction

The fast growth of semi-empirically computed rovibrational IR line lists benefited from the expansion of computing power and the combination of advantages from theory and experiment. The selected, highly accurate, experimental data (usually the rovibrational levels) provides a harness on the systematic errors of first principle calculations. Accuracy in the range 0.01 ~ 0.10 cm^{-1} is achievable and has become the standard criteria for IR line lists in ExoMol,[1] TheoreTs,[2] and the Ames [3] databases. The regular procedure utilizes the reliable data of the most abundant isotopologues in the potential energy surface (PES) refinement. The refined PES is then used to generate reliable predictions for weaker-bands, higher energy bands, or the minor isotopologues. Using a high quality dipole moment surface (DMS), we can generate IR lists more complete, reliable, and consistent than empirical Effective Hamiltonian (EH) IR line lists.

The Isotopologue IR lists computed in this way are usually in high quality. This is because the rovibrational variational calculations are based on an exact quantum Hamiltonian, with the Born-Oppenheimer Diagonal Correction (BODC) [4], nonadiabatic corrections [5], and other higher order effects [6] if necessary. This built-in consistency ensures that mass-dependent isotope effects be well described from one isotopologue to another. Such prediction accuracy is the power of the "Best Theory + Reliable High-resolution Experiment" (BTRHE) strategy.[7,8] In this paper, these IR line lists are classified as BTRHE Lists, including the Ames-296K SO_2 [9–11] and CO_2 [12–15] IR lists.

However, the best prediction accuracy of minor isotopologue BTRHE lists are is limited to the level of data reproduction for the main isotopologue, or the most abundant isotopologues with the relevant high-resolution experimental data available. Can we further improve this? The answer is definitely "YES", at least partially. Since the differences between the reliable experimental data and the BTRHE IR line lists are available for a few isotopologues, the needed corrections can be quantitatively estimated for other isotopologues. In principle, it is easy to understand. Such prediction improvements can be considered the BTRHE strategy extended to a different level, i.e. use existing Δ(Expt – BTRHE) to predict new Δ corrections.

This paper explores the possibilities of such a new BTRHE extension, for both line intensity and line position. The analysis and implementations are not as simple as we had imagined in the very beginning. These investigations may provide valuable reference and insights to similar or future studies in the BTRHE field. For line intensities, we checked how the intensity, Einstein A_{21} and transition dipole moment square sum (TDM) vary along different coordinates. The Ames-2016 IR lists for 13 CO_2 isotopologues [12] and Ames-296K IR lists for 30 SO_2 isotopologues [10,11] are used in the discussion of intensities. For line positions, the MW part of the Ames-296K SO_2 IR lists are adopted in this benchmark study. They are fit to an Effective Hamiltonian (EH) model first, using the SPFIT [16,17] program. The differences between the fitted EH(Ames) parameters and experimentally determined EH parameters, i.e. Δ = EH(Ames) – EH(Expt), exhibit a clear isotope dependence. The isotope dependence is utilized to derive the Δ correction for EH(Ames) of other minor isotopologues. The refined (corrected) EH(Ames) is used in the SPCAT [16,17] program to generate the best available line positions within a limited J/K_a range.

The paper is organized as follows: after a brief introduction of the SO_2/CO_2 PES refinement, Section 3 explicitly presents the intensity consistency of the Ames IR line lists. They are well beyond the consistency level of current experimental data or techniques. This consistency can be combined with future measurement of highly accurate intensity data to improve predictions for other rare isotopologues. Section 4 reports the key points we summarized from the EH(Ames) refinement procedure. This extended BTRHE approach is compared to the purely *ab initio* method and other refinement approaches. Our prediction accuracy leads by orders of magnitude. See more details

in Section 4. Conclusions and a discussion about future studies is given in the last Section.

A direct product of this work is the SO_2 MW benchmark data, which includes the best available MW line lists for 24 SO_2 minor (rare) isotopologues at 296K. It combines Ames intensity (which we assume more reliable and consistent, especially at high J/K) with the line positions predicted by SPCAT using corrected EH(Ames) model parameters. This line set is part of the Supplementary Material. Dozens of analysis plots are also included in the Supplementary Materials (SM). Some of the figures in the SM are referred to as Fig.S*n* in this paper, *n* is the index integer.

Important Note: for the reader's convenience, 3-digit notations are used throughout to label the SO_2 and CO_2 isotopologues. The 3 digits of each notation are taken from the atomic mass of O, *X*, O isotopes rounded to one, where *X*=S for SO_2 or *X*=C for CO_2. For example, the main isotopologue of SO_2 is $^{16}O^{32}S^{16}O$, denoted **626**, a singly substituted $^{16}O^{32}S^{18}O$ is **628**; or a mixed isotopologue $^{17}O^{36}S^{17}O$ is denoted **767**, etc. The total number of SO_2 isotopologues is 30, i.e. 5 S isotopes (from ^{32}S to ^{36}S) x 6 combinations of two $^{16/17/18}O$ atoms. Similarly, the CO_2 main isotopologue $^{16}O^{12}C^{16}O$ is also denoted **626**, and $^{17}O^{32}S^{18}O$ is denoted **728**, etc. Ames-2016 IR line lists include 6 ^{12}C-isotopologues + 6 ^{13}C- isotopologues + 1 ^{14}C- isotopologue (646), for a total of 13 CO_2 isotopologues. If the two O atoms in a SO_2 or CO_2 molecule are same isotopes, we call it a "Symmetric" Isotopologue, otherwise it is an "asymmetric" isotopologue. Note all SO_2 isotopologues are still "asymmetric top" molecules, not a "symmetric top".

2. Ames-296 SO_2 IR(MW) Line Lists and Ames-2016 CO_2 IR Line Lists

The Ames IR line lists at 296K can achieve 0.01-0.02 cm^{-1} line position ***prediction*** accuracy for H_2O [4,7,18], CO_2 [12–15], SO_2 [9–11], and NH_3 [5,8,19,20] rovibrational bands and better than 80-90% agreement with those non-ab-initio intensities collected in databases such as HITRAN (*hi*gh-resolution *tran*smission molecular absorption database) [21,22] and CDMS (Cologne Database for Molecular Spectroscopy) [23–25], especially at lower J/K. Note that current HITRAN intensity data includes direct intensity measurement (e.g. NH_3 and H_2O), effective dipole moment (EDM) models derived from experiments, as well as ab initio calculations (e.g H_2O and CO_2). For SO_2, it should be noted the MW intensity in HITRAN were taken from CDMS EDM models, so they are the same.

The merit of the "Best Theory + Reliable High-resolution Experiment" (BTRHE) strategy lies in the quality of ***predictions***, e.g. from existing to unobserved, from lower quanta / energy / temperature to higher quanta / energy / temperature, and from the main isotopologue to other isotopologues. Compared to the original *ab initio* IR line lists, the accuracy has been improved by two orders of magnitude. Many predictions were verified in recent experiments, e.g. Refs.[26–28]

These semi-empirically computed IR line lists, such as the Ames-296K lists for SO_2 [10,11] and CO_2 [12], and the ExoMol [1] lists for NH_3 [29,30] and H_2CO [31], etc., can be treated as new spectra "observed" in the laboratory. We are able to carry out an Effective Hamiltonian (EH) model analysis to generate Ames IR list based EH parameters, denoted EH(Ames). For a specific EH constant reliably determined through least-squares fitting, X_{Ames} vs. X_{Expt}, the difference $\Delta = X_{Ames} - X_{Expt}$ may have small variations systematically correlated with isotopic substitutions. Therefore, the Δ for other minor isotopologues can be reliably predicted. Accordingly, the EH(Ames) parameters X_{Ames} can be corrected to higher accuracy and thus yield more accurate line positions.

3. IR/MW Intensity: Isotopologue Consistency and Prediction.

3.1 Appropriate Representation

The "isotopologue consistency" in the title of this section implicitly acknowledge certain properties (harmonic frequency, rotational constants, intensity, etc) will have systematic changes upon isotope substitutions. Different properties experience different changes. We first need to find an appropriate way to present the mass dependence of intensities or line positions, as well as their dependence on the quantum numbers or frequencies (in case of intensities).

In Ref.[12], Figs.10-15 illustrate that the isotopologue inconsistencies among existing CO_2 EDM intensities can be clearly identified with Ames-2016 CO_2 line lists. In the qualitative comparison, the relative intensity deviation $\delta\% = 50\% \times (S_{Ames}/S_{EDM} - S_{EDM}/S_{Ames})$ was plotted along m = J"(for Q/R) or -J"(for P). In this study, the isotopologue consistency test requires a more quantitative representation for Ames-296K IR intensities. The following definitions are used:

Transition dipole matrix element square $X_r = \left|\langle\varphi_f|\overrightarrow{\mu_r}|\varphi_i\rangle\right|^2$;

This work uses $\boldsymbol{TDM} = \sum_{r=x,y,z} X_r = S_{fi}/(g_i g_f)$; (1)

Line Strength $S_{fi} = g_i g_f \sum_{r=x,y,z} X_r = g_i g_f TDM$; (2)

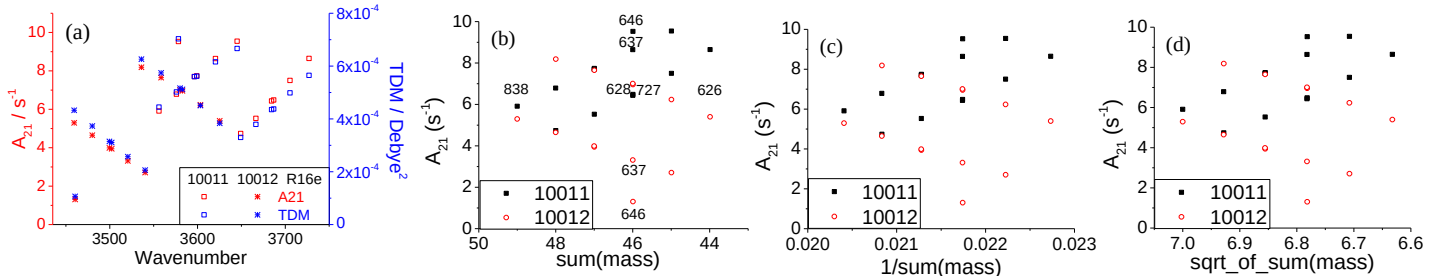
Einstein $A_{fi} = \frac{8\pi^2 \nu_{fi}^3}{3\varepsilon_0 c^3 \hbar g_f} S_{fi}$; $g_f = 2J' + 1$; $g_i = 2J" + 1$; (3)

Absolute absorption intensity $S(T) = \frac{8\pi^3}{3hc} \nu_{fi} (e^{-E"/kT} - e^{-E'/kT}) S_{fi}/Q(T)$; (4)

so $A_{fi} = 8\pi c \nu_{fi}^2 Q(T)(e^{-E"/kT} - e^{-E'/kT})^{-1} S(T)/g_f$; (5)

Oscillator strength $f_{fi} = 1.49919368 \frac{g_f}{g_i} \frac{1}{\nu_{fi}^2} A_{fi}$; (6)

The choices for the *y* axis include the TDM, absolute intensity *S*, Einstein coefficient A_{21}, line strength S_{21}, or oscillator strength f_{21}, etc. They are interconnected through formula (2)-(6), where φ' and φ'' are the rovibrational wavefunction of upper and lower levels, μ_r are the dipole components, *Q* is the partition sum, *g* is the degeneracy level = 2*J*+1, *J'*/*J"* are the rotational quantum numbers, *E*/*E*' are the energies of upper /lower levels, and ν is the transition energy in cm^{-1}. To determine the population weight, we need total *Q*, *J'* or *J"*, and *E'* or *E"*. In theoretical calculations, the line strength is directly summed over transition dipole moment squares multiplied by the degeneracy. Then it is multiplied by ν^3 to obtain Einstein A_{21}, or combined with explicit *J'*/*J"*, and the partition sum to obtain other quantities. We prefer A_{21} and TDM representations. The TDM is defined as the transition dipole moment squares summed over cartesian or principal axes. It is equal to the line strength S_{21} divided by degeneracies. The A_{21} and TDM are independent from symmetry, can be determined without *Q*(T), and so are more convenient as a consistency check. The two coordinates have qualitatively similar performance in many cases. Note all the CO_2 and SO_2 **TDMs** in this paper are **in Debye²**. All Einstein-A coefficients $\mathbf{A_{21}}$ are in $\mathbf{s^{-1}}$



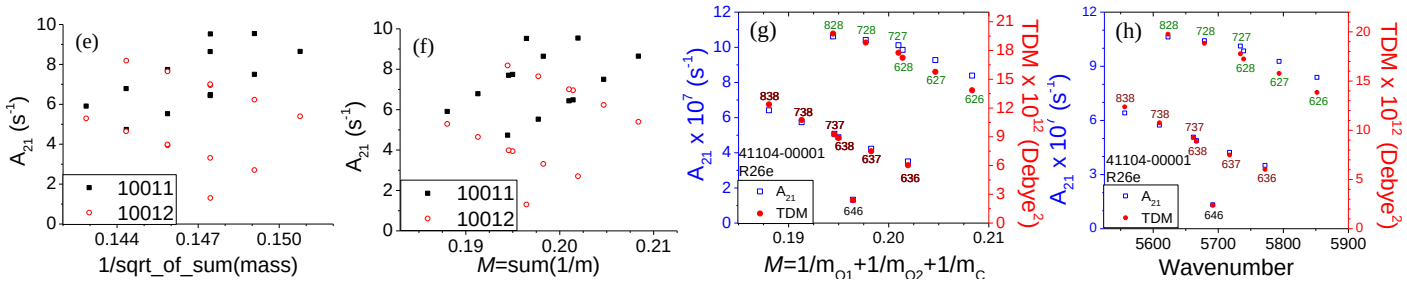


Fig.1 CO_2 Isotope effects on the R16e transition of 1001n-00001 band, n=1-2 (a-f) and the R26e of 41104-00001 band (*g*,*h*). Einstein A_{21} (left *y*) and TDM (right *y*) vs. 5 mass-related coordinates (*b-f*, *g*) or wavenumber (*a, h*). We recommend TDM and M=sum(1/m) representations.

For the *x* axis, either transition wavenumber or molecular/atomic masses may be used. Transition wavenumber is a natural choice for the *x* coordinate, as it is the ONLY experimental variable in Eqs.1-6 that can be directly determined. One of our favorite analysis is to plot TDM (or TDM_{ISO}/TDM_{626} ratios) vs. cm^{-1}. It is easy to recognize to which isotopologue a line belongs, because the line position is a solid fingerprint for isotopologues. We can also immediately identify whether a TDM value is somewhat contaminated, unreliable or plainly wrong.

A troubling question is, if the formula of a quantity contains the transition energy explicitly, such as A_{21} or the absolute intensity S_{296K} in HITRAN, can we still plot it against transition energy? Since all of the intensity related quantities have isotope / mass effects and we all agree they can be plotted against some isotope / mass coordinates, the answer to the former question should also be YES. Assuming $y=a\cdot x$, a *y* vs *x* plot will have trends or patterns different from those in an *a* vs *x* plot. Our goal is to confirm and utilize the isotopologue consistency, not to focus on the approximation formula behind such a relationship -- at least not in this study, and in any case, most of these would be molecule / band / quantum number dependent. As long as the pattern is clear and easy to guide predictions, it is worth to plot those quantities vs. transition wavenumber (energy) or wavelength. In reality, most comparisons on isotope effects can ONLY be made with whatever values and units are available in existing high-resolution IR databases. The choice in HITRAN is Einstein A_{21}, because the nuclei spin statistical weights are significantly different between symmetric and asymmetric isotopologues, and so are the intensities. More importantly, Einstein A_{21} is unique and scientists in different fields know how to convert it to other units they are specialized with.

For comparisons across several bands or transitions, transition wavenumber could be a better choice for *x*, naturally. However, absolute TDM or A_{21} values may vary by orders of magnitude from one band to another, or even from one J to another. They are not optimal choices. Instead, we recommend using the TDM_{ISO}/TDM_{626} ratio. It is always 1.0 for the CO_2 and SO_2 main isotopologues. It is also ideal for determining how the isotope effects evolve along *J*, K_a, K_c *or* ΔJ, ΔK_a, ΔK_c, etc.

For mass related coordinates, it should be noted that there does NOT exist a "standard" choice for polyatomic molecules, unlike the reduced mass coordinate established for diatomic molecules. This leaves us plenty of freedom and flexibility for systematic exploration. Possible choices include (1) total molecular mass, (2) the square root of mass, (3) the inverse of total molecular mass, (4) the inverse of the square root of total mass, (5) the sum of the inverse of atomic masses, (6) the sum of the inverse of the square root of atomic masses, or additional variants and combinations. Our analysis suggests that "(5) the sum of the inverse of atomic masses" is potentially the best, universally applicable choice, i.e.

$$M = \sum \frac{1}{m} = \frac{1}{m_{O1}} + \frac{1}{m_{C/S}} + \frac{1}{m_{O2}} \quad or \quad \frac{1}{m_{O1}} + \frac{1}{m_{O2}} + \frac{x}{m_{C/S}} \tag{7}$$

It does not necessarily mean other choices are bad or inappropriate. Our primary concern on other molecular mass coordinates is that they cannot separate equal-mass isotopologues, e.g. SO_2 666 and 648, or CO_2 628 and 727. For low resolution or qualitative analysis, this might not be an issue, but the *M* coordinate defined here is surely better. It is also our choice for EH parameter analysis and predictions for line position improvements. See more details in Sec.4. A scaling factor *x*, as shown in the last term of Eq.(7), may be used to combine all A_{21} or *TDM* to one line.

Examples for CO_2 are given in Fig.1. 13 pairs of A_{21} and TDM values of the R16e transition are extracted from the 1001*n* (*n*=1,2) and 41104-00001 bands. Not much pattern difference exists between A_{21} and TDM, or among the 5 mass coordinates, except that the *M* coordinate in Fig.1f may separate equal-mass isotopologues.

3.2 The magnitude of Isotope effects on intensity

How much can isotope substitutions change the IR/MW intensity S, A_{21} or TDM? We may think the normal magnitude is 5-10% for molecules like CO_2 and SO_2, and in extreme cases it may reach ±30-50%. Isotope substitution is thought of as a small perturbation to the Hamiltonian, wavefunctions, and intensity for these molecules. In our SO_2 and CO_2 analyses, this assumption is valid for many bands or lines. Fig.2a plots the TDM_{ISO}/TDM_{626} vs. *M* for the strongest transitions of 10 SO_2 vibrational bands rising from the ground state (GS), including GS←GS. Detail transition list is given in Sec.3.4.3. The TDM_{ISO}/TDM_{626} ratios varies from 0.6 to 1.3. We divide the ratio changes (%) by the corresponding O or S isotope mass changes, to determine δ%, the average change per unit mass, e.g. $^{17}O \rightarrow {}^{18}O$, or $^{33}S \rightarrow {}^{34}S$. The δ%(S) and δ%(O) are shown in Fig.2a. From ^{X}S to ^{X+1}S (x = 32 – 35), the average TDM change is -12% ~ +6% of TDM_{626}. From ^{y}O to ^{y+1}O (y=16 or 17), the average TDM change is -10% ~ 9% of TDM_{626}. The smallest change is on the pure rotation, GS←GS, and more detailed analysis is presented in Sec.3.4.1.

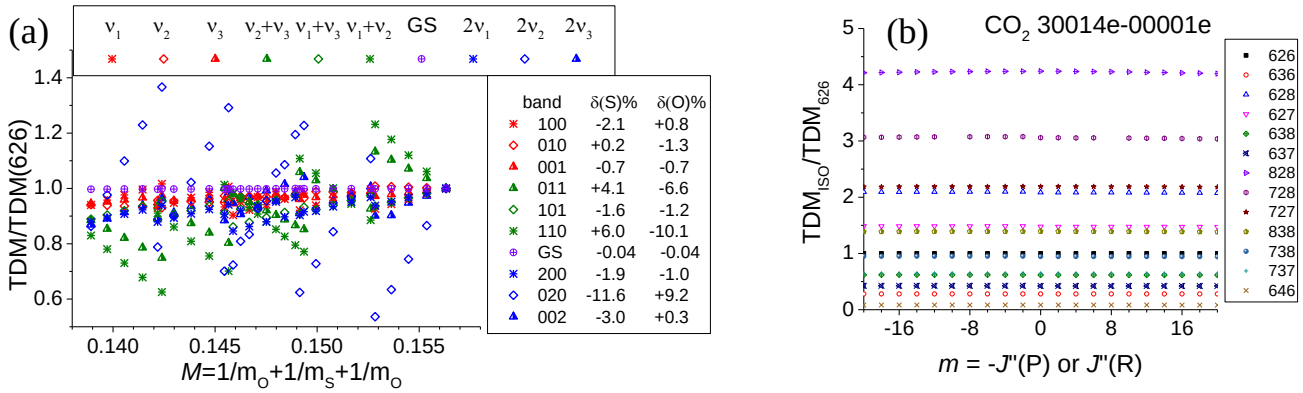


Fig.2 (a) Isotope variation of SO_2 TDM for the strongest IR lines of 10 bands at 296K. (b) the CO_2 TDM_{iso}/TDM_{626} ratios are stable for the *P* and *R* ranges of our 30014e-00001e band.

Unfortunately, not all bands or transitions follow the small perturbation assumption. The range of δ% may be surprisingly large for some regular vibrational bands. In some cases, it can be as large as ±500%, or even larger! One example in Fig.2b is the 30014e-00001e band of CO_2. From P20e to R20e, all 13 isotopologues have very stable TDM_{ISO}/TDM_{626} ratios. Examination of the isotopologues shows that heavier C isotopes significantly reduce the TDM and intensity. The TDM_{646} (for $^{14}C^{16}O_2$) is only 10% of TDM_{626}. From ^{X}O to ^{X+1}O, heavier O isotope enhance the TDM and intensity by 50-80%. The TDM_{828} (for $^{12}C^{18}O_2$) is >400% of TDM_{626}. Such "unusually large" isotope effects are actually reasonable. The 30014 band is $6\nu_2$, a high order overtone of the linear bend. For the O-C-O structure, heavier O atoms on the sides or a lighter C atom in the middle enhance the bending amplitude, while lighter O

atoms or heavier C atoms weaken the bending amplitude. Note that the agreement between UCL/HITRAN vs Ames intensities are generally very good for the bands below 8000 cm^{-1}.[32–34]

A different class of example was given in the Fig.15a of the CO_2 Ames-2016 paper.[12] The CO_2 microwave intensity for 628 and 638 are more than 3 times as strong as the intensity of 627/637/728/738. The intensity ratio is equivalent to the TDM_{ISO}/TDM_{626} ratio here. The >300% intensity ratio was also found in the UCL line lists.[12,33] This is because the effective dipole moment only exists in the vibrationally averaged structure of asymmetric isotopologues. The microwave spectra is dipole forbidden for symmetric isotopologues (including the main isotopologue 626), as there is no permanent dipole. Hence, the O isotope mass difference breaks forbidden rules and makes those lines IR active. In other words, it is not a "small perturbation" any more. The dipole moment goes from zero to non-zero. Such IR intensity ratio can also be found from those vibrational bands only active in asymmetric CO_2 isotopologues. See Fig.3 for R26e examples of the CO_2 10022, 21111, and 10001(2) bands. In these cases, A_{21} and TDM coordinates both work, *M* and wavenumber coordinates are both appropriate. Note we still prefer cm^{-1} coordinate when 2 or more bands are compared.

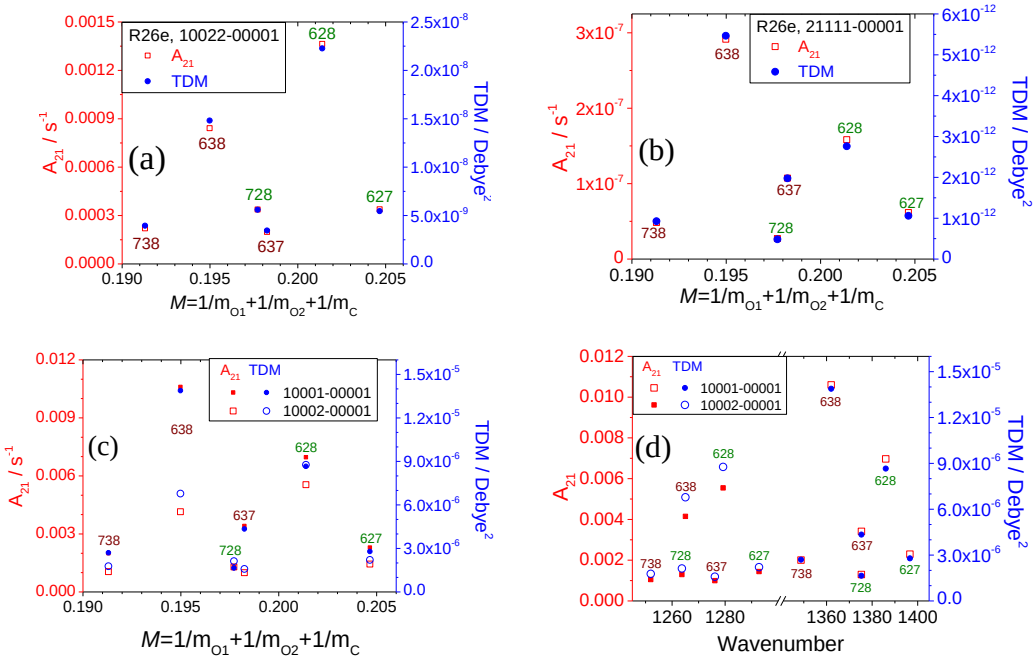


Fig.3. Examples of CO_2 rovibrational transitions where the intensity (TDM and A_{21}) are strongly affected by the mass difference of two O atoms. Panels *a,b,c* use our *M* = sum(1/*m*)coordinate, panel *d* uses cm^{-1}.

A similar ratio can be found for the MW $13_{12,1}$←$12_{12,0}$ of asymmetric SO_2 isotopologues. From 7 to 8 cm^{-1}, their intensities are in the range of 0.98-5.54E-27 cm/molecule. The 6*x*8 isotopologues have the strongest TDMs, e.g. TDM_{628}/TDM_{627}=3.622, TDM_{668}/TDM_{667}=3.628. The heavier S isotope also leads to larger TDM, e.g. TDM_{668}/TDM_{628}=1.152. Both K_C=0 and K_C>0 $J+1_{Ka,Kc+1}$←$J_{Ka,Kc}$ lines follow this pattern, .e.g. $13_{6,7}$←$12_{6,6}$ is shown in Fig.4a. Such TDM ratios are found in several MW hot bands, e.g. ν_1←ν_1, $2\nu_2$←$2\nu_2$, ν_3←ν_3, plus the fundamental bands ν_1, ν_2, ν_3, combination band $\nu_1+\nu_2$, and overtone $2\nu_1$ & $2\nu_3$.

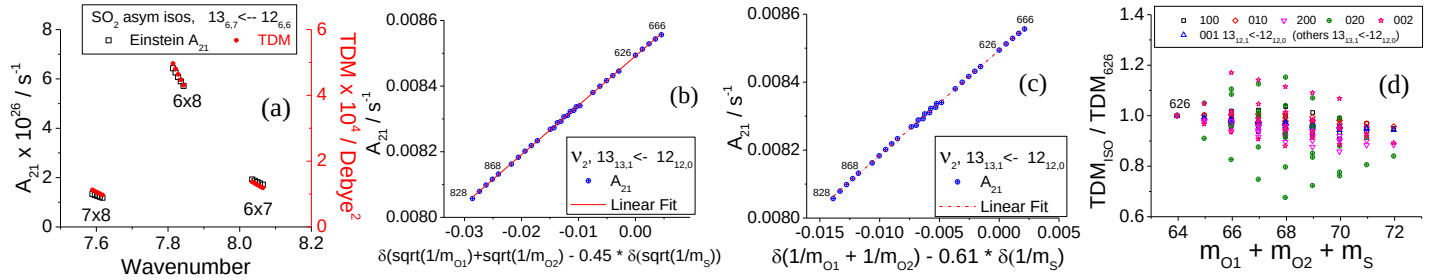


Fig.4 (a) the O isotope effects on SO_2 MW intensity of $13_{6,7} \leftarrow 12_{6,6}$ of 15 asymmetric isotopologues, the S isotope mass number x =32-36; (b-c) the A_{21} of SO_2 ν_2 transition exhibit linear relation with modified mass coordinates, m_S and m_O are atomic masses; (d) check TDM ratio changes vs. total mass of isotopologues.

For the original M=sum(1/m) coordinate, the pattern difference between A_{21} and TDM coordinates is small. The S isotope effect and O isotope effect are not colinear with each other. Fig.4b and 4c use A_{21} to show two more examples of a "combination" mass coordinate. To make all the 30 A_{21} symbols approximately linear, the *square-root* of the S atomic mass inverse is scaled by -0.45 in Fig.4b. In Fig.4c, the scaling factor is -0.61 for the S atom mass inverse (*no square-root*). Usually these scaling factors are vibrational band dependent, so not useful for multi-band comparisons. The M=sum(1/m) is universally applicable. It can easily differentiate isotopologues of the same total masses, unlike the simple total mass. In Fig.4d, there are only 9 total mass values for the 30 SO_2 isotopologues. The 6x8 and 7x7 isotopologues are not separated by this coordinate or a scaling factor on S and C atom, x =32-36 for S isotope mass.

3.3 CO_2 IR analysis

In this section, the isotopologue consistency is checked for the TDM or A_{21} is examined for the Ames-2016 IR line lists. Here, this discussion is concise, while most of the analysis figures are available in the supplementary material or upon request. "Fig.S*n*" refers to additional figures in supplementary file.

3.3.1 ν_2 and ν_3

The ν_1 band is totally symmetric, so IR inactive for symmetric isotopologues. The ν_2 state has two components, 01101e and 01101f. Their isotope effects are different. Our analysis suggests, for most ν_1=0 bands, both A_{21} and TDM follow an approximate linear relationship when plotted against wavenumber. See Fig.5 for ν_2 R16e (panel *a*), Q16e (panel *b*), and ν_3 R16e (panel *c*). Note the M coordinate plots are available in Fig.S1, i.e. the first figure in the Supplementary Material.

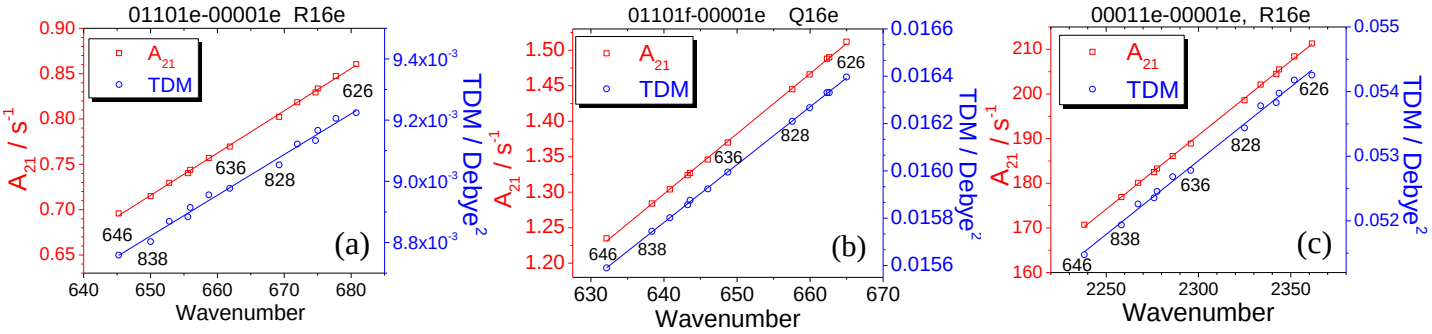
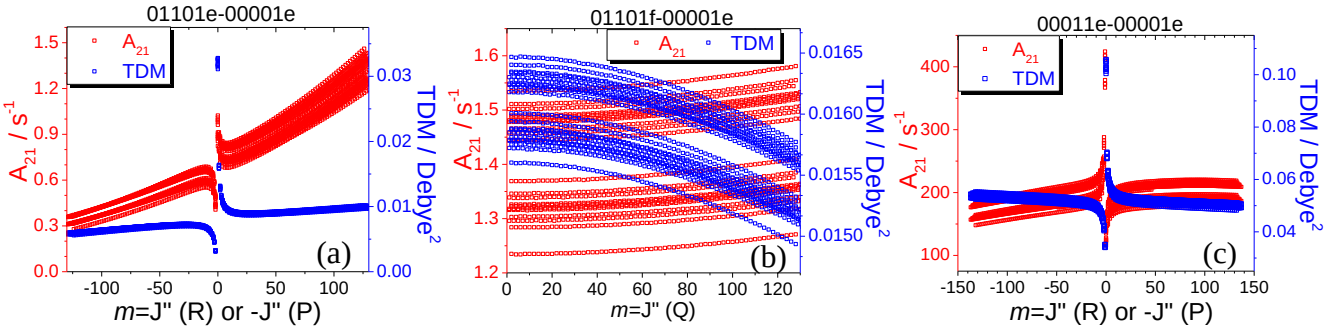


Fig.5 CO_2 Einstein A_{21} coefficients (open squares) and TDM (open circles) of (a) ν_2 R16e; (b) ν_2 Q16e; (c) ν_3 R16e. Linear approximations are also included.

Details of the linear approximation and error statistics are not our focus here. However, both of the A_{21} and TDM patterns can be utilized to make high quality predictions. The averages and σ_{RMS} of the relative fitting error (%) on A_{21} are as small as 0.2-0.3%. Is this good enough? Of course, this depends on the particular application. Currently, only a very limited number of IR intensity studies [28,35–37] actually exhibit an accuracy comparable to 0.2-0.3%. In these cases, such small deviations from the simplest linear approximation are more than satisfactory. However, the range of TDM variations in these 3 plots are just 5-6%. The 0.2-0.3% deviation is equal to 4~5% of the total isotopologue effect. This is an error large than we expect. On the other hand, the TDM patterns are clear in a) and c). Asymmetric isotopologues are above the TDM line, while the symmetric ones are below it. This is another kind of consistency we can utilize to make more accurate TDM predictions. In Fig.5b, the TDMs and A_{21}'s of Q16e are better predicted by linear approximation relations.

The isotopologue consistency of every single rovibrational transition accumulates into the isotopologue consistency for a given vibrational band. Fig.6 a-c shows the J dependence of A_{21} and TDM. Fig.6 d-f shows how the TDM_{ISO}/TDM_{626} vary along *J* in the *P* and *R* branches. In Fig.6b, there appear more than 13 lines. This is due to the small difference between even J" and odd J". In Fig.6d and 6f, those small breaks at *m*=0 for the *P* and *R* branches result from the intensity convergence issue we reported [12] for symmetric isotopologues. Because all TDM ratios are relative to TDM_{626}, the ratios of the other symmetric isotopologues are continuous. The breaks do not affect the following discussion.

In Fig.6a – 6c, the TDM changes along J are not large, except for the *m*=0 transition from *P* to *R* in 6a and 6c. We use a single color for all isotopologues. The TDM ratios of 01101f-00001e are nearly flat in Fig.6e, which are not affected by the intensity gap issue. An interesting observation is, although the TDM (and A_{21}) of *P* and *R* run into singularities at *m*=J"=0 in panel *a* and *c*, the TDM ratios are continuous (excluding the small breaks) and almost *monotonic* along the *m* coordinate in panel *d* and *f*. The TDM ratio lines of the 01101e-00001e band in Fig.6d are a good example. First, the ^{13}C substitution reduces the TDM by about 3%, the ^{14}C substitution further reduces it by 2.5%. The $^{17/18}O$ substitution lowers the TDM at J=0 by less than 1%, but it mainly changes the slope of the TDM ratio lines. The heavier the O atoms, the slope of the TDM ratio lines becomes more negative. The heavier the C atoms, the slopes are more positive. The slopes of other mixed isotopologue TDM ratio lines are a combination of C and O isotope effects.



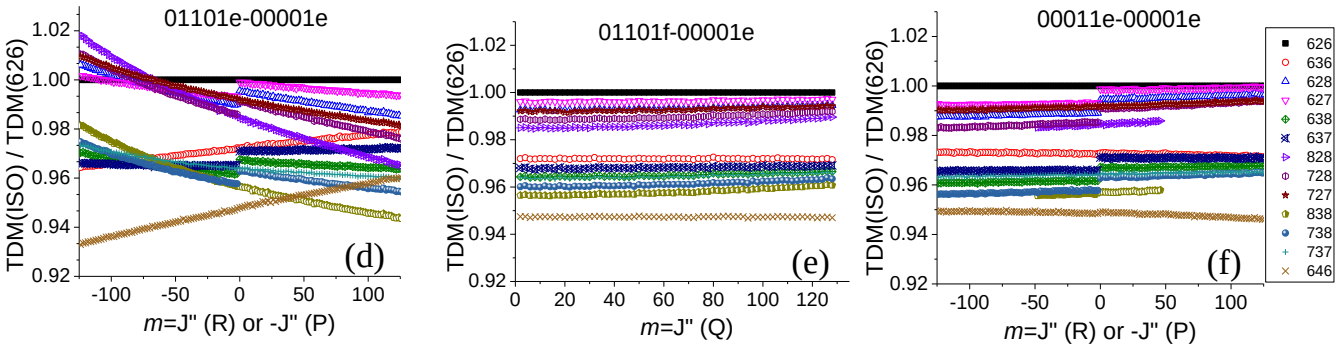


Fig.6 The *J* dependence of A_{21} & TDM (a-c) and TDM_{ISO}/TDM_{626} ratio (d-e) of 13 isotopologues, for CO_2 01101e-00001e (a,d), 01101f-00001e (b,e), and 00011e-00001e (c,f) band components.

Fig.S1 CO_2 01101e-00001e analysis [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

Fig.S2 CO_2 01101f-00001e analysis [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

Fig.S3 CO_2 00011e-00001e analysis [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

3.3.2 Polyad analysis: 1001n, 2001n, 3001n

Unlike those ν_1=0 CO_2 bands, the A_{21} and TDM of bands involved in resonance polyads do not follow a linear relationship when plotted against the cm^{-1} coordinate. However, their patterns are easily recognized. In Fig.7, the R16 transition of the 1001*n*, 2001*n*, 3001*n* bands clearly demonstrates a pattern, or a new consistency. *n* is the index number for a band in the polyad. When plotted vs. cm^{-1}, the highest energy band (*n*=1) and the lowest energy band (*n*=ν_1+1) are roughly symmetric with respect to the center of energy range. Their C and O isotope effects for A_{21} and TDM have opposite signs. For example, the TDM_{646}/TDM_{626} ratio of the R16e transition in the 20011 and 20013 bands is 2.34 vs. 0.137, respectively. It is significantly reversed. The corresponding TDM_{828}/TDM_{626} ratios are 0.363 (20011) and 2.79 (20013), similarly reversed. These changes can be explained by the nature of the bands. The $\nu_1$0011 band is ν_3+symmetric stretch. The vibration and intensity of the symmetric stretching part may be hampered by heavier O atoms. The $\nu_1$001(ν_1+1) band is the ν_3+bending overtone. The vibration and intensity of bending overtones are significantly reduced by heavier C atoms. Then the 20012 band is a transition between 20011 and 20013, but its intensity is higher than both. In Fig.7c, there are two transition bands between the 30011 and 30014 bands. Similar TDM/A_{21} patterns can be found in other higher $\nu_1$001*n* (*n*=1,…,ν_1+1) vibrational polyads.

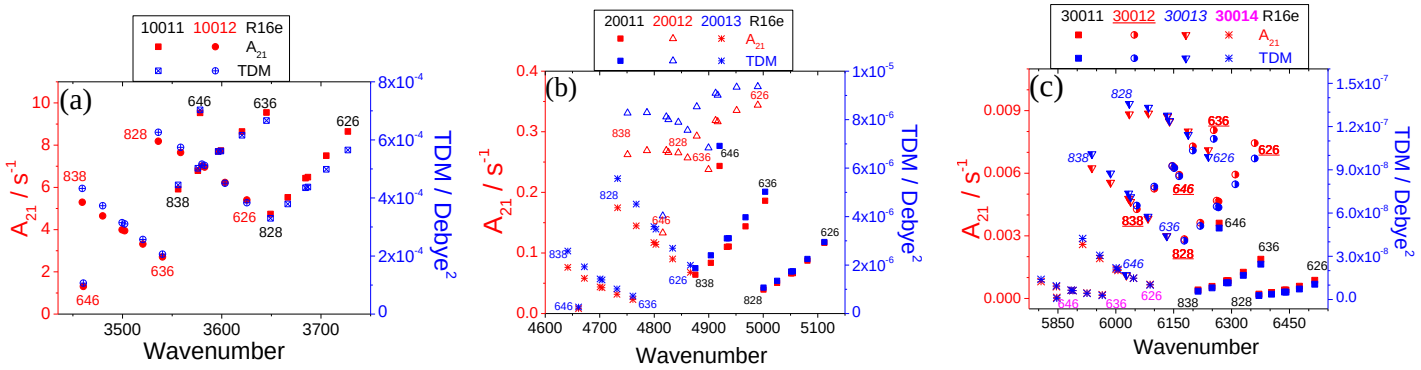


Fig.7 Einstein A_{21} and TDM of R16e vs. cm^{-1} in 3 vibrational polyads: (a) 1001n, n=1,2; (b) 2001n, n=1-3; (c) 3001n, n=1-4.

The A_{21} and TDM patterns can also be found vs. the mass coordinate, *M*=sum(1/m). See Fig.8. A_{21} symbols are red, and TDM symbols are blue. Component bands use different symbols, just like those in Fig.7. The largest *M* at right end of each panel is 626, and the left end is 838. Fig.8 seems like those patterns in Fig.7 are squeezed together. Although it makes it difficult to compare the isotopologue consistency patterns across bands, it does provide better

opportunity to see how A_{21} and TDM of a specific isotopologue vary by band. Comparing Fig.7 and Fig.8, we conclude again that the energy coordinate (cm^{-1}) is at least comparable to the best mass coordinate we can come up with, *M*. The difference between the A_{21} and TDM coordinates are not significant enough to change the patterns (isotopologue consistency) observed for single transition. This is consistent with what we found in other bands.

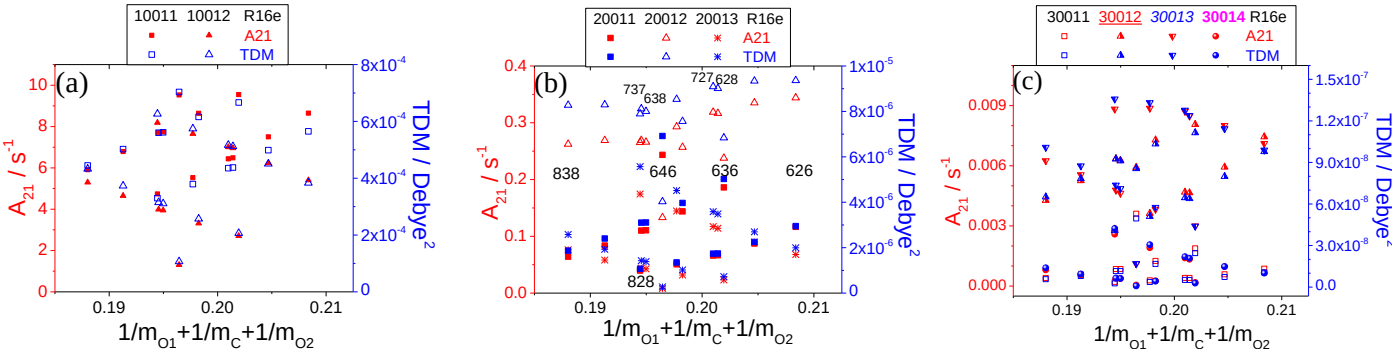


Fig.8 Einstein A_{21} and TDM of R16e transition vs. *M*=sum(1/*m*) in 3 vibrational polyads: (a) 1001n, n=1,2; (b) 2001n, n=1-3; (c) 3001n, n=1-4.

For J < 50, the TDM_{ISO}/TDM_{626} ratios are stable for all 4 components of the 3001*n* polyad - see Fig.9. The absolute TDM values of transitions in the 30012, 30013 bands are higher than those of the 30011 and 30014 bands, but the range for most of the TDM ratios are 40%-140%, much smaller than the range of TDM ratios for the 30011 and 30014 bands, i.e. 10%-480%. Readers are able to easily deduce the roles of C and O isotope effects as they vary from one band to next.

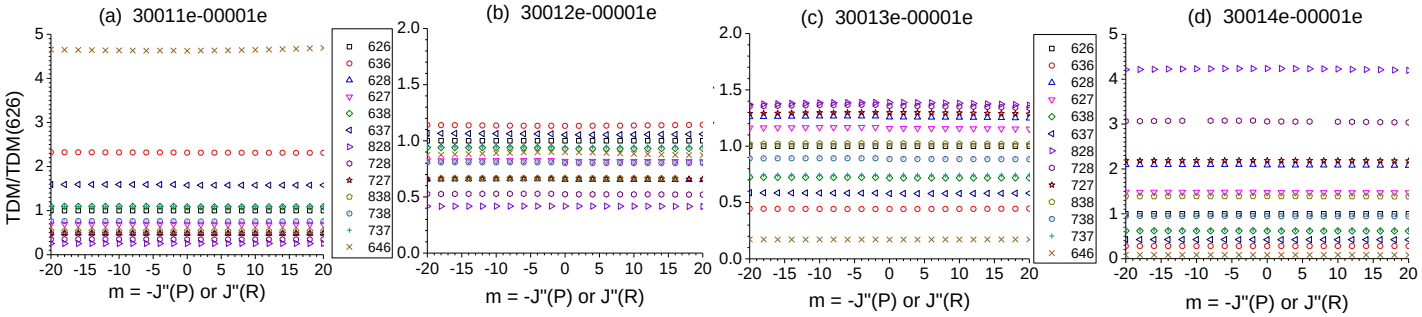


Fig.9. TDM_{ISO}/TDM_{626} ratios of 13 CO_2 isotopologues, for the *P* and *R* branches of 4 bands in 3001*z* polyad, z=1-4.

Everything we show from Fig.7 to Fig.9 is part of the isotopologue consistency contained in the variationally computed IR line lists. Full-range J dependence of A_{21}, TDM, and TDM_{ISO}/TDM_{626} ratios are also reported in Fig.S4-S6. We recommend that interested readers examine the TDM ratio variations at J~100 and beyond. It would be interesting to compare those with highly accurate experimental intensities,[28,35,37,38] which means they can be determined with 1% or less uncertainty.

Fig.S4 1001x-00001 analysis [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

Fig.S5 2001x-00001 analysis [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

Fig.S6 3001x-00001 analysis [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

3.3.3 Hot bands 0nn11-0nn01 (n=0-3)

In addition to the vibrational fundamentals and resonance polyads above, it is worthwhile to examine how the A_{21} and TDM vary for a series of hot bands lying in the same cm^{-1} range. The $\nu_3+n\nu_2$ (n=0-3) series is reported in Fig.S7. It includes the ν_3 fundamental, 00011-00001. Just like ν_3, when plotted vs. cm^{-1}, both the TDM and A_{21} of the hot

bands can be linearly approximated. For A_{21} linear approximations, σ_{RMS} of relative fitting residual = 0.16-0.17%, with maximum |δ%| = 0.25-0.32%. For TDM linear approximations, σ_{RMS} = 0.12-0.14%, maximum |δ%| = 0.15-0.18%.

Please remember, the TDM residual difference between asymmetric and symmetric isotopologues is real and systematic. It is part of the intensity "consistency" or "pattern" which we consider useful in future calibration. In other words, intensity approximation is not our target, getting small approximation residuals is not our target either. Approximations are just visualizations for *part* of the consistency. On the other hand, even with fitting residuals, linear approximations in Fig.S7 still give more consistent (or even more accurate) intensities than existing EDM models or experimental measurements.

The J dependence of ν_3 TDM ratios is shown in Fig.6f and Fig.S3. For the three ν_2 related hot bands, there are 4 types of transitions, *e/f* ←→ *e/f*. From Fig.S8. We can see how they differ from each other. Please ignore the small *P*-*R* breaks at J=0. Note that the Ames-2016 line lists did not complete the 1-1 vibrational quantum number conversion from variational CI bases to conventional polyad labels. Consequently, some isotopologues have incomplete data but this does not impact the overall comparison.

Fig.S7 The Einstein-A coefficient A_{21} and TDM (sum of transition dipole element squares) of 13 CO_2 isotopologues, for the R16e transition in 0nn11-0nn01 ($n\nu_2 + \nu_3 \leftarrow n\nu_2$, n=0-3) bands. [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

Fig.S8 The TDM_{iso}/TDM(626) ratios of 4 e/f←→e/f branches of 01111-01101 and 02211-02201 bands. [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

3.3.4 CO_2 analysis Summary

One of the most important lessons we learned in this CO_2 and SO_2 Isotopologue consistency study is that, all related calculations and analysis procedures need to be done with the highest consistency. The analysis provides a very sensitive measure to check the convergence and consistency of rovibrational IR line list computations. In our previous studies, we were not concerned with the consistency of CO_2 calculations. They did agree pretty well with many recent high quality experiments.[26–28,37,38] The J=0 gap between the *P* and *R* branches is fixable, and the ~0.6% intensity difference is smaller than the uncertainty of most experimental data or EDM models. At the beginning of this work, we were even satisfied with ~0.3% linear approximation error on A_{21}. Now our understanding is different. For example, Fig.5b shows seemingly satisfactory A_{21} linear approximation for Q16e. We are not sure if *every* data point along the line is free of deviations at the 0.01-0.10% level, which may directly affect the fitting accuracy. The TDM ratios look very consistent in Fig.6 and Fig.9, but it may be just because the scale is too large. Base on the data oscillations in Fig.S8, we quantitatively estimate the noise level for the 0nn11-0nn01 bands is 0.05-0.09%, close to 0.1%. Two possible explanations: (1) the calculations are >99.5% converged, but maybe not >99.95% converged; (2) the intensity data we computed has kept only 4 significant figs, which may introduce small uncertainty 0.004-0.04%. In the next round of CO_2 calculations, the noise level should be reduced to 0.001% or smaller. This is NOT an easy job, as evidenced by our SO_2 MW spectra analysis- see the Section 4.

We also suggest that our colleagues perform isotopologue consistency tests on their IR line lists: to establish reliable spectra pattern; to identify noise in calculation; to help identify unreliable experimental data; and to help improve EDM models, etc. Note that we are promoting the "consistency", not claiming "accuracy". Based on recent experimental work and understanding of our calculations, we may accept intensity deviation as large as 20-50% for those very weak bands or some bands higher than 10,000-13,000 cm^{-1}. However, the isotopologue consistency cannot be qualitatively wrong, if the isotope effects on the computed intensities are >5-10%.

In real applications, the computed CO_2 intensity quantities can be combined with a few experimentally determined

values to give more accurate predictions for rare isotopologues. In other words, our calculations can be further "calibrated" using experimental data. From this perspective, this essentially is an extension of the BTRHE strategy. As we have demonstrated in Ref.[12], the isotopologue consistency of the computed Ames IR intensity is above the current consistency level of most experimental measurements and EDM models. The best EDM isotopologue consistency is 3~5%, with exceptions up to 10-50% or even larger.

The 3rd application of these isotopologue consistency tests, is to help the CO_2 vibrational polyad quantum number assignments for the Ames-2016 line lists and energy levels. The wavefunctions of clustered energy levels can be distinguished by related intensities. Together with J→J+1 search and SPFIT fitting, most difficult assignments can be accomplished. The only exceptions are those transitions strongly perturbed by resonances. This will be included in future studies. More systematic investigations and physical/mathematical studies will be more appropriate for next stages. Interested readers can do their own analysis using either the Ames-2016 lists [3,12] or the UCL lists [32,33].

Recently, Karlovets et al [39,40] reported CO_2 effective dipole moment parameters for the $\Delta P=8/9$ series. Along with their previous work published in 2011 [41] and 2015 [42], this confirms that the independently fitted EDM parameters for various isotopologue also exhibit a pattern, or trend. The approximation they adopted in the latest CDSD-296 is that the EDM parameters are (approximately) independent of isotope substitution for even ΔP series, or proportional to the mass difference of the two O atoms for odd ΔP series. They reported this simple approximation can predict the strongest bands of rare isotopologues with errors less than 10%. Therefore, we can safely estimate those predicted EDM parameters may have inconsistency around 3~5%, including noise or uncertainty. Our isotopologue consistency discussed in this work is at least 2 or even 3 orders of magnitude higher, i.e. 0.005-0.05%. However, to make a more direct comparison, Ames-2016 intensities should be re-generated with highest possible consistency, then fit with the same EDM models. The fitted Ames sets of EDM parameters should still have the isotopologue consistency we discussed in this section. This would be an interesting topic for a mutually beneficial collaboration between experiment and theory.

3.3.5 Call for New Experiments on 50006/60007 bands.

The proposal in this section is based on our strong confidence in the isotopologue intensity consistency of the Ames-2016 IR line lists, plus the existing experimental data of CO_2 626 and 628.

In the Ames-2016 paper[12], we pointed out that the 50006 and 60007 states have the lowest energy in all the band origins having δ(Ames-CDSD) > 0.05 cm^{-1}. The data in HITRAN/CDSD was extrapolated from EH models, not experimentally measured. This is partially because the $\nu_1$0000n-00001 bands do not exist for symmetric isotopologues, n=1,2,…ν_1+1. It is important to determine the real 50006 and 60007 band origins. Both the CDSD (Tomsk) [21,43] and Ames/UCL groups [12,15,32–34] need to determine here the discrepancy comes from and how to fix it. It is a prediction quality test for EH models and theoretical calculations.

At 296K, the strongest 626 band involving the 50006 levels is 50006←01101, with line intensity 1E-30~1E-29 cm/molecule, so it is hard for detection and analysis. Fortunately, for 628, 50006-00001 is the 2nd strongest band in the 6135 – 6155 cm^{-1} range with intensity up to 1.5E-26 cm/molecule. It is >1000 times stronger than the 626 lines. Fig.10a shows the intensity of 628 50006-00001 lines and their distance to neighboring lines of 30013-00001 band. There are very little line mixings. We estimate at least 20 R lines can be reliably, accurately measured in this range. In Fig.10b example, the R8e transition is in the middle of two strong 30013-00001 transitions, R11e and R12e. All the rest of nearby lines are weaker by at least one order of magnitude. No experimental difficulties are expected.

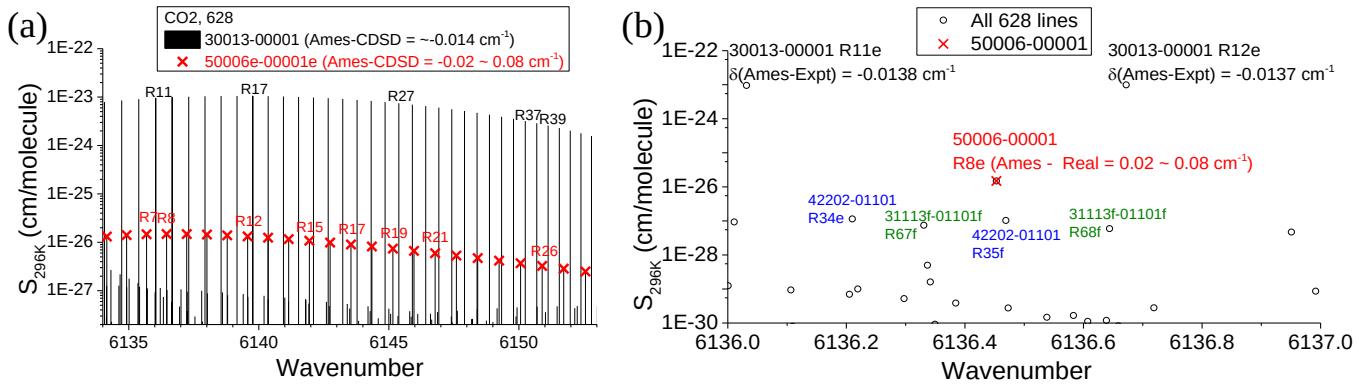


Fig.10 Easy-to-find 50006-00001 lines of CO_2 628 isotopologue. (a) the R branch region as the 2nd strongest band; (b) R8e example.

For the 60007-00001 band, the 628 isotopologue is still the best opportunity. At 296K, most low J lines are weaker than 2E-30 cm/molecule, too difficult to measure due to many stronger bands nearby. However, the *P*66e and *R*64e (and neighboring lines) intensities are significantly enhanced by resonance with 40014-00001. The R64e intensity in the Ames IR line list was 2.3E-28 cm/molecule. It is more than 100 times stronger than most J<55 lines. More conveniently, the R64e becomes the 2nd strongest line in the 7370.7 – 7371.1 cm^{-1} range – see Fig.11. The neighboring 40014 lines have been studied. The *P*66e transition has a similar intensity, but it is not a good choice. Fig.11b shows a possible way we suggest to find the R64e transition.

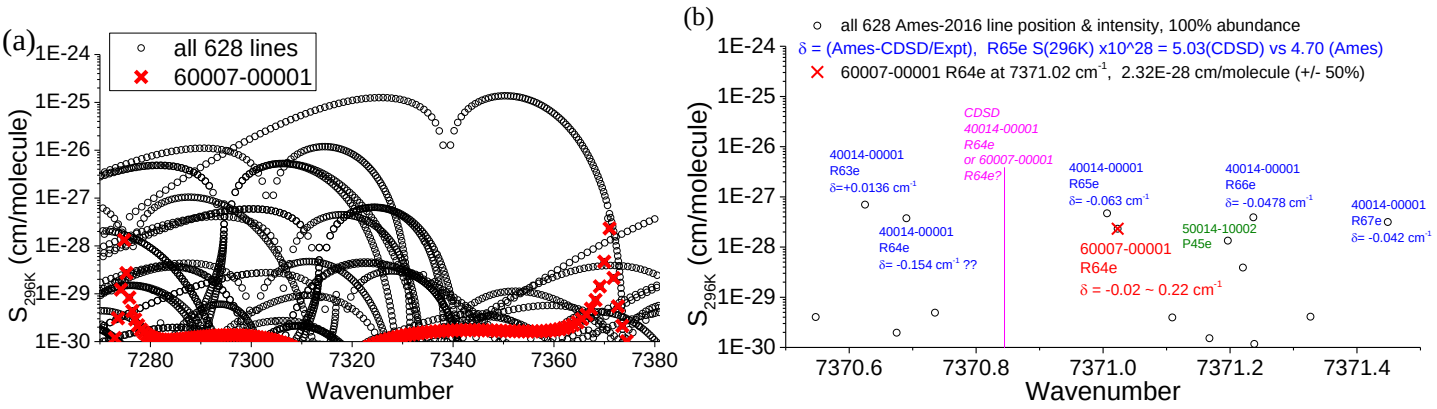


Fig.11 Hard-to-find 60007-00001 lines of CO_2 628 isotopologue. (a) the P & R branches (in red crosses); (b) the strongest line in 628 60007-00001 band, R64e, and other 40014-00001 lines nearby.

We hope experimentalists may carry out CO_2 628 measurement for the 50006-00001 and 60007-00001 bands. These would aid in future improvements of the CO_2 PES, and also provide justification for claims about the extrapolation accuracies. Karlovets et al [40] only reported the 60006/5/4 bands of the 628 isotopologue, but not the 60007 band.

3.4. SO_2 IR/MW Intensity analysis: Isotopologue Consistency

After performing some tests, we found the previously published Ames-296K SO_2 IR lists did not have enough precision for the isotopologue consistency study reported here, especially the cold MW spectra of 30 SO_2 isotopologues. It was necessary to re-generate all lines with intensity precision doubled. The transitions used in this section have their absolute intensity (100% abundance), sum of square of dipole matrix element in Debye^2 (TDM), population weight of lower level, and $(1 - e^{-(E''-E')/kT})$ explicitly printed with 8 significant figures.

It should be noted that, before the regeneration, we have re-computed all J=0-75 levels for all 30 isotopologues in the most consistent way. This is the requirement for the EH(Ames) analysis and refinements discussed in Section 4. In Section 4, the new energy levels and MW line positions are fit to Effective Hamiltonian (EH) models to get EH(Ames) parameters, then the rotational constants and quartic centrifugal distortion constants of EH(Ames) are semi-empirically refined with the available EH(Ames)-EH(Expt) differences. The refined EH(Ames) set is fed to the SPCAT program [16,17] which generates all the EH-EDM based MW transitions, including line position, intensity, and TDM data that we use for the comparisons in this section.

3.4.1 TDM ratio disagreement for ʳR lines, J'=K_a'

In Fig.2a, the pure rotational MW spectra has the smallest TDM_{ISO}/TDM_{626} changes. The strongest MW line at 296K is $13_{13,1}\leftarrow 12_{12,0}$. We extract the TDM_{ISO}/TDM_{626} ratios for all $J+1_{J+1,1}\leftarrow J_{J,0}$ lines, J=0-45. First we plot the J=0,5,10…45 TDM ratios along the *M*=sum(1/m) coordinate. See Fig.12a. The pattern is very clear: heavier isotopes have a smaller TDM; higher *J* and K_a further reduce the TDM. The right end of the *M* coordinate is the main isotopologue 626, or $^{16}O^{32}S^{16}O$. The left end is the heaviest 868, or $^{18}O^{36}S^{18}O$. The total amount of TDM ratio reduction at J=45 is only 2%. At J=13, the reduction is only about 0.3%. One will obtain an almost linear relation if the A_{21} values are plotted against cm^{-1}, but that is misleading. As a consistency check, it is more meaningful to compare the TDM ratios at all J from 0 to 45. See Fig.12b, it is crystal clear that the TDM ratio reductions are proportional to the molecular mass increases, i.e. the δ_{mass} with respect to 626. Isotopologues with approximately the same mass have very similar TDM ratio changes. The relation is stable and consistent for all data. The only range having different isotopic patterns is J<4.

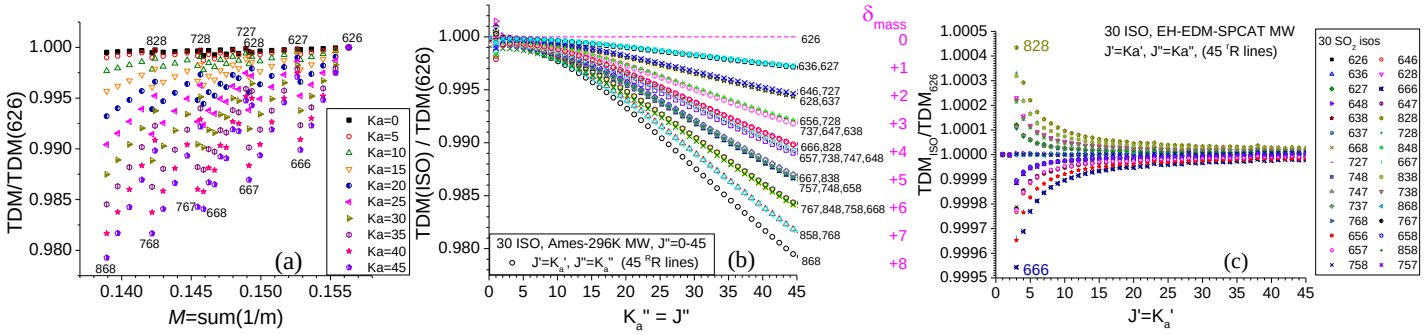


Fig.12. The TDM_{ISO}/TDM_{626} ratios of SO_2 MW transitions, $J+1_{J+1,1}\leftarrow J_{J,0}$, J=0-45. The systematic isotope and J-dependence of Ames-296K intensities in *a* and *b* are in sharp contrast with EH-EDM-SPCAT model predictions in *c*.

These TDM ratio reductions at high J are the higher order effects discovered by the theoretical calculations. In contrast, the EDM based intensities predicted by the SPCAT program do not have such reductions. The SPCAT runs with just a single effective dipole term, 1.63308 Debye. That is the exact same value adopted in CDMS [23–25] models.[44] In Fig.12c, the range of TDM ratio differences monotonically decreases from 1±0.0004 to 1±0.00003. The noise level for those data in Fig.12c is estimated as 3~6E-6. It is mainly numerical. For a given *J*, 828 has the largest TDM while 666 has the smallest TDM. In other words, the S and O isotope effects are opposite. This is different from what we saw in Fig.12a, where the heavier S or O isotopes always reduce the TDM and intensity.

For microwave spectra, our understanding is that the state-of-the-art technology can achieve intensity accuracy to about 1%. This is not necessarily good enough to justify or calibrate the theoretically computed isotopologue consistencies. If the experimental uncertainty can be reduced to 0.3-0.5%, measurement data for strong lines (J=8-20) can be used in a qualitative comparison. If the uncertainty is 0.1% or less, J=10-30 measurements can help

calibrate the pattern in Fig.12b and improve intensity predictions for all rare isotopologues. In principle, data from 4-8 isotopologues should be more than adequate for calibration: 626, 828, 666, 868, plus 628/727/668/767.

3.4.2 $J,K_a,\Delta K_a$ dependence at J"=32: Ames vs. EH-EDM

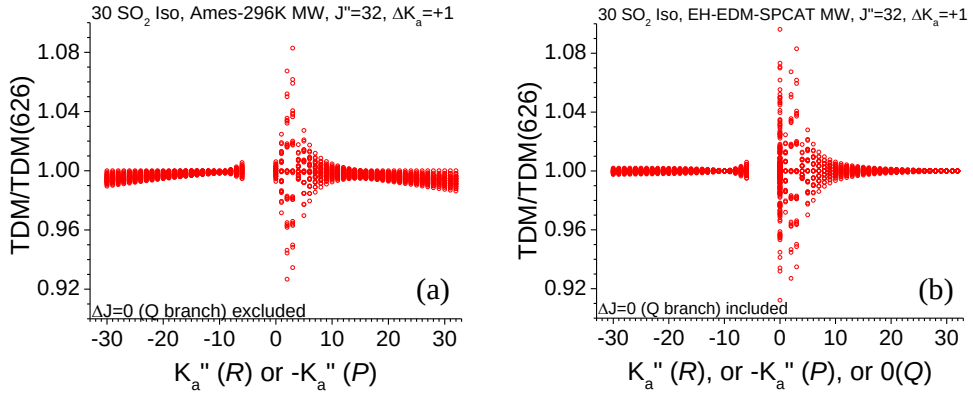


Fig.13 The K_a dependence of TDM ratios at J"=32, ΔK_a=+1. (a) P and R branches of Ames intensity data; (b) P, Q, and R branches of EH-EDM-SPCAT intensity data.

After checking the J dependence, we look into the K_a dependence of TDM ratios at J"=32. Fig.13 a) and b) shows two sets of TDM ratios in the same scale: Ames (a) vs. EH-EDM (b) for the ΔK_a=+1, ΔJ=±1 lines. The agreement between them is much better than what was found in the $J+1_{J+1,1}\leftarrow J_{J,0}$ series, especially when K_a < 10 (*P*) or K_a < 13 (*R*). For higher K_a transitions, the range of the EH-EDM TDM ratio monotonically decreases to 0.01% (*R* branch, K_a=32) or remains stable at ±0.2% (*P* branch). This is different from the trends of Ames TDM ratios in panel *a*. For both the *P* and *R* branches, the range of Ames TDM ratios slowly, but consistently, increases to 1.2-1.5%. The right-end K_a=32 data are consistent with those in Fig.12.

Now we expand the analysis from ΔK_a=+1 to ΔK_a=-1,1,3,5,7,9. Those ΔK_a>3 lines are much weaker. The magnitude of corresponding TDMs drops exponentially from 1E-3 at ΔK_a=3 to 1E-7 at ΔK_a=7 and 1E-10 at ΔK_a=9, as summarized in Fig.14a. The TDM ratio series are plotted along the *M*=sum(1/m) × ΔJ coordinate, where $\Delta J = J' - J''$, the rotational quantum number change. Both the range and pattern of the TDM_{ISO}/TDM_{626} ratios show good agreements between Ames (Fig.14a) and EH-EDM (Fig.14b).

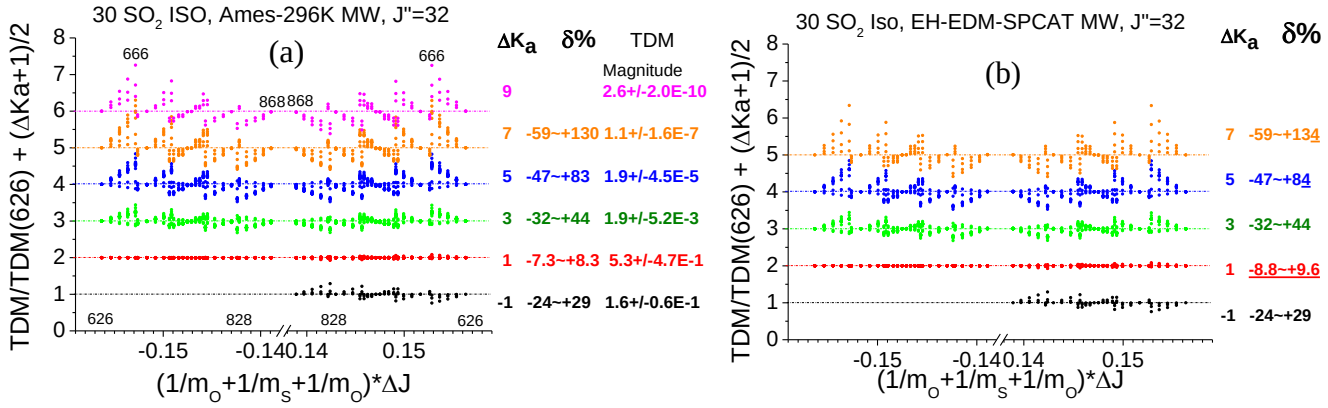


Fig.14 the TDM ratios along *M* coordinate for ΔK_a=-1,1,3,5,7, and 9, with J"=32. (a) Ames-296K IR lists data; (b) EH-EDM-SPCAT model data using single EDM parameter.

Fig.S9 K_a" dependence of TDM ratios: a) J"=32; b) J"=12 [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

Fig.S9 investigates the K_a" and J" dependences of TDM ratios, using EH-EDM data at J"=12 and J"=32. Our observations include: (1) at a specific J, the isotope effects on the *R* branch is larger than those on the *P* branch; (2) the largest TDM ratio changes are found near the lowest K_a; (3) the K_a for maximum TDM changes may increase along with J" increasing, e.g. from K_a"=1 at J"=12 to K_a"=3 at J"=32; (4) the magnitude of the isotope effects on the TDM ratios also rises when J" increases from 12 to 32.

3.4.3. Strongest lines: in hot band pure rotations and Vibrational bands

The intensity order of the SO_2 626 vibrational bands with the lowest quanta is:

$$\nu_3 > \nu_2 > \nu_1 > \nu_1+\nu_3 > \nu_2+\nu_3 > 2\nu_1 > 2\nu_3 > \nu_1+\nu_2 > 2\nu_2$$

The strongest transition in these bands are the ΔJ=ΔK=1 lines with J=K_C, K_a=0/1: $21_{1,21}\leftarrow 20_{0,20}$ (ν_1 and ν_2), $19_{0,19}\leftarrow 18_{0,18}$ (ν_3 and $\nu_1+\nu_3$), $18_{1,18}\leftarrow 19_{1,19}$ ($\nu_2+\nu_3$), $22_{0,22}\leftarrow 21_{1,21}$ ($2\nu_1$), $23_{1,23}\leftarrow 24_{0,24}$ ($2\nu_2$), $24_{0,24}\leftarrow 23_{1,23}$ ($2\nu_3$), and $28_{0,28}\leftarrow 27_{1,27}$ ($\nu_1+\nu_2$). Their TDM ratios have been given in Fig.2a.

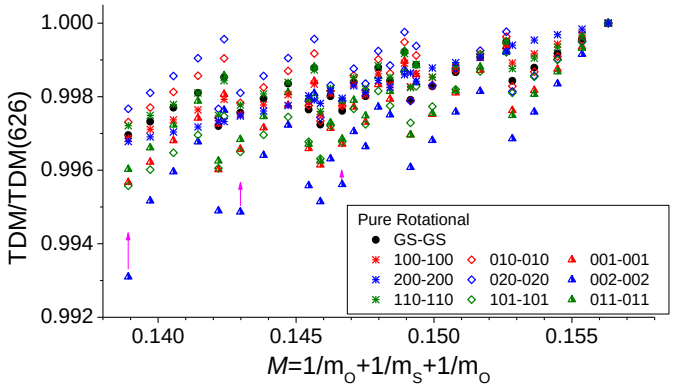


Fig.15. TDM_{ISO}/TDM_{626} of the strongest ʳR lines in 10 pure rotational bands. (compare to Fig.2a)

For the $\nu_1\nu_2\nu_3 \leftarrow \nu_1\nu_2\nu_3$ pure rotations, the strongest line at 296K is $13_{13,1}\leftarrow 12_{12,0}$ (ν_3=even) or $13_{13,0}\leftarrow 12_{12,1}$ (ν_3=odd). Their TDM ratios are shown in Fig.15. Here it explicitly confirms that both S and O heavier isotopes reduce the TDM of the strongest ΔJ=ΔK=1 transition of *all* pure-rotational bands. The TDM reductions from 626 to 828 are approximately 0.3% - 0.6%. However, there exist convergence defects for the $2\nu_3\leftarrow 2\nu_3$ data (blue triangles) of the 868, 758, and 658 isotopologues. The location of expected TDM ratios is shown by cyan arrows. The deviations are as small as 0.02-0.10%. Such minor defects are difficult to detect in regular spectroscopic analysis. This further confirms the TDM ratio as a very sensitive tool in IR list analysis. Those outlier data cannot be used for approximations.

Fig.S10 The J dependence of the TDM ratios of strong ΔJ=ΔK=1 transitions in 9 bands: $12_{12,0}$(Δν=0) or $12_{0,12}$(Δν>0). Note the $2\nu_2$ need more investigation / confirmation for higher J effects, where the TDM ratios of 6x6 isotopologues (x=3-6) spans from 1E-4 to 10. The local minima of TDM ratios are found at J=48,44,37,32 for x=3,4,5,6, respectively. [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

4. EH Approach for Line Position Improvement: SO_2 MW spectra

The pattern recognition and future calibration of CO_2 and SO_2 properties relating to intensity are beyond the isotopologue consistency level of current experimental data or Effective Dipole Models (EDM). Yet it is a different story for the realm of line position. Theoretical calculations try very hard to follow (not catch up with) the accuracy of experiments. For example, the best accuracy for polyatomic rovibrational line positions is 0.01-0.05 cm^{-1}.[1–3] In

this study, we demonstrate the possibility to improve the line position accuracy of Ames-296K MW lists by *two* more orders of magnitude, i.e. from a few MHz to 0.01-0.02 MHz accuracy for $A_0/B_0/C_0$ and from hundreds of MHz to 1-10 MHz for J<30 and K_a<15. The accuracy is expected to be *independent* of isotope mass substitutions. In other words, predictions of all un-observed minor isotopologues have uniform accuracy. Note the mass-dependent nonadiabatic corrections are important, but from our experience [5,6] it is still insufficient to achieve the goal of accuracy improvement. This is also true for other higher order correction terms. The only feasible way is to combine the experimental and theoretical efforts again, i.e. further implementation of BTRHE strategy.

Effective Hamiltonian (EH or H_{eff}) modeling is the standard approach for rovibrational IR analysis and predictions in laboratory experiments. Measured spectra lines are fit to appropriately constructed EH models. Least-squares fits usually generate a few dozens of EH(Expt) parameters, denoted X_{Expt}. It includes effective rotational constants $A_0/B_0/C_0$, quartic/sextic centrifugal distortion constants D/H, and even higher order terms L/S, etc. The fitted EH(Expt) parameters are used to derive the complete set of rovibrational transitions. The EH approach has the highest accuracy for most line positions within the quantum number range of reliable measurements. The best EH(Expt) models for SO_2 MW spectra are available in CDMS [44] and recent experimental studies.[45–49] Therefore, they are quoted as Expt(CDMS/Expt).

The SO_2 Ames-296K line lists have reliable and consistent intensities, but their MW line positions could be off by a few hundreds of MHz. See Fig.16a for 646 as an example. To improve the line position accuracy, the Ames-296K MW line lists are fed into the EH model based least-squares fitting program, SPFIT [16,44,50,51], as "new" spectra recorded in the "computer lab". The Ames list based EH(Ames) parameters, denoted X_{Ames}, are compared to EH(Expt) parameters, X_{Expt}. The isotope dependent, systematic variations of $\Delta = X_{Ames} - X_{Expt}$ are utilized to implement a BTRHE refinement on EH(Ames) and SPCAT predictions for other minor isotopologues. The EH model we chose for this benchmark study of SO_2 MW spectra, is the 26-parameter EH model that CDMS adopted for 626/636/646/628.[44] Interested users can check its details online.[44]

The structure of this section is as follows. In Sec.4.1, we first discuss a bad BTRHE strategy and its seemingly good improvement. In Sec.4.2, several sub-sections are devoted to the important technical details necessary for a more appropriate, reliable and accurate implementation. Many are the caveats or lessons we have learned during this study. As the 1st paper investigating a BTRHE strategy to such unprecedented isotopologue-mass-independent prediction accuracy, it is reasonable to include more important factors. Although numeric results may vary from one research group to another, from one molecule to another, etc. most discussion in Sec.4.2 is a useful memo(guide) for colleagues who are interested in the general topic of fitting variationally computed IR line lists to EH models, and / or EH constant predictions. To avoid confusion, the role of a particular factor in the big picture is explained at the beginning of each sub-section. Note Sec.4.2.8 compares the prediction accuracy of our approach to that of a formula adopted by the Ulenikov group.[ref] But to save space, some of the details have been moved to the Supplementary material.

With all critical factors appropriately considered and included, Sec.4.3 presents the most appropriate BTRHE implementation for SO_2 EH(Ames) constant refinement, plus the best available $A_0/B_0/C_0/D_K/D_{JK}$ constants we predict for 30 SO_2 isotopologues. Then Sec.4.4 carefully discusses the difference between our strategy and those from other researchers, emphasizing the uniqueness of this work, and why our implementation has a higher accuracy standard from beginning to the end.

4.1 Two-orders of magnitude improvement from Over-Simplified (BAD) model.

This section uses a seemingly successful, but bad BTRHE implementation to illustrate the possibility of *two orders* of magnitude improvement based on the mass-dependent $\Delta = X_{Ames} - X_{Expt}$. It introduces the basic idea that the *lower* order EH(Ames) terms can be successfully refined, then combined with the original *higher* order EH(Ames) terms to give more accurate SPCAT predictions. It ends up with corrections and leads us to more critical factors to be discussed in Sec.4.2.1-4.2.7, because the reliability and accuracy of EH(Ames) and EH(Expt) are the pre-requisite of any BTRHE implementation.

For rotational constants, the difference $\Delta = X_{Ames} - X_{Expt}$ seems correlated linearly with isotopic substitutions. See Fig.16 b) and c) for the S and O isotope effects on Δ(A/B/C), plotted against isotope mass. Can we use the linear relationship to predict the Δ of unobserved isotopologues, e.g. 656, then use the predicted Δ to correct X_{Ames}(656)?

The short answer is YES. A primitive test was attempted by refining 646 EH(Ames) using the Δ of 626 and 636. In other words, the goal was to predict Δ(646) from Δ(626) and Δ(636), to use the Δ(646) to improve the 646 EH(Ames), and then to reduce the line position deviations in Fig.16 a). Accordingly, the GS←GS MW spectra of 626, 636 and 646 were first extracted from published Ames-296K IR lists [10], fit to EH(Ames) models, and compared to EH(CDMS).[44] All transitions with 296K intensity > 1E-23 cm/molecules are included in SPFIT fits, since stronger lines are more likely to be first measured. A symmetric (or asymmetric) SO_2 isotopologue has ~4000 (or ~7000) transitions. The uncertainty of Ames line positions was set to 1E-5 cm^{-1}. The rejection threshold is set to 5 times of the uncertainty, i.e. 5E-5 cm^{-1} or 1.5 MHz. Lines with a fitting residual larger than this threshold are rejected.

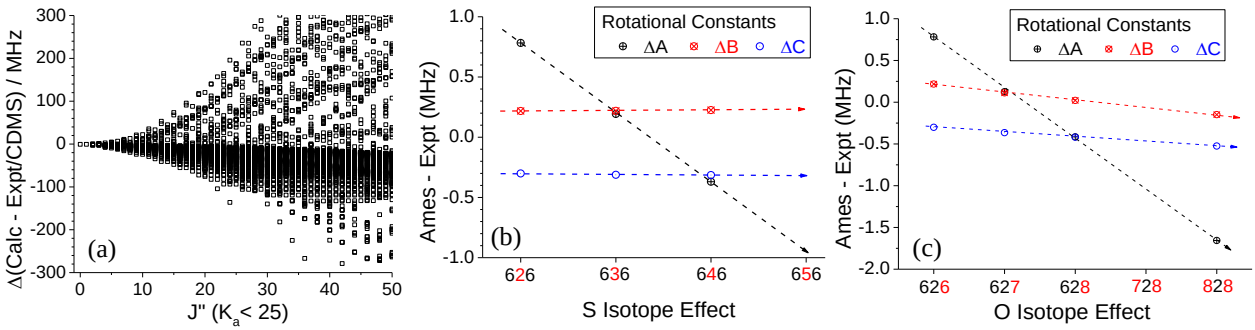


Fig.16 (a) the line position errors of SO_2 646 MW spectra, can reach ~300 MHz at J=30, K_a~25; (b) and (c) show the rotational constants deviations between EH(Expt) and EH(Ames) for SO_2 isotopologues, $\Delta = X_{Ames}-X_{Expt}$: (b) the S isotope effects 626, 636, and 646; c) the O isotope effects 626, 627, 628, and 828.

Based on Fig.16b and 16c, our over-simplified model assume all S isotope have integer masses and the 626→636 perturbation is very close to the 636→646 perturbation. So the formula for the primitive EH(Ames) refinement test is the simplest :

$$X_{Predicted}(646) = X_{Ames}(646) - \Delta(646),$$

where $\Delta(646) = 2\Delta(636) - \Delta(626)$, and $\Delta = X_{Ames} - X_{Expt}$, and $X_{Expt}=X_{CDMS}$ in this test [8]

The correction formula, Eq.8, is applied to all 26 EH(646) parameters, using the available 626 and 636 EH(Ames) and EH(CDMS) parameters. Both the original EH(Ames) and the "Predicted" (or "refined", "corrected") EH(Ames) parameters are compared to the CDMS parameters. The relative differences (in %) with respect to EH(CDMS) are shown in Fig.17a. Another rough prediction is also included as open squares in Fig.17a. It is purely based on EH(Expt): $X(646)=2X_{CDMS}(636)-X_{CDMS}(626)$. This linear approximation is too simple to discuss numerical details.

Fig.17 does show the accuracy improvement. The seemingly satisfactory predictions were reported in recent

scientific conferences.[52,53] From Fig.17a, we see the prediction formula, Eq.8, works well for ***lower order*** terms, which include the rotational constants and quartic centrifugal distortion constants. The *B* and *C* deviations are reduced by nearly 2 orders of magnitude, but it does not improve the accuracy of ***higher order*** terms, i.e. sextic and above. Instead, it is getting worse in some cases. However, this is easy to understand, because some higher order terms of 636 and 646 EH(CDMS) were taken from 626 EH(CDMS).[44] Consequently, the assumptions behind Eq.8 are invalid and the over-simplified model fails. Therefore, it is more appropriate to keep the original EH(Ames) higher order terms. The "Predicted" lower order terms represents a significant improvement over the original EH(Ames) lower order terms.

Now we have three sets of *lower* order terms for 646: original EH(Ames), "Predicted" (i.e."refined"), and EH(CDMS/Expt). And two sets of *higher* order terms: EH(Ames) and EH(CDMS/Expt). Combining the "Predicted" lower order terms with either EH(CDMS/Expt) or EH(Ames) higher order terms, we get two new EH parameter sets. The SPCAT program [16,17] read in the two new EH sets, the original EH(Ames) set, and the EH(CDMS/Expt) set, to generate corresponding line positions and intensities with the EDM term 1.63308 *D*. Taking the EH(CDMS/Expt) based line positions as "exact", we compute the line position deviations of the other three sets and plot in Fig.17b and 17c. Note the three EH sets are explicitly labeled as "lower order terms" + "higher order terms".

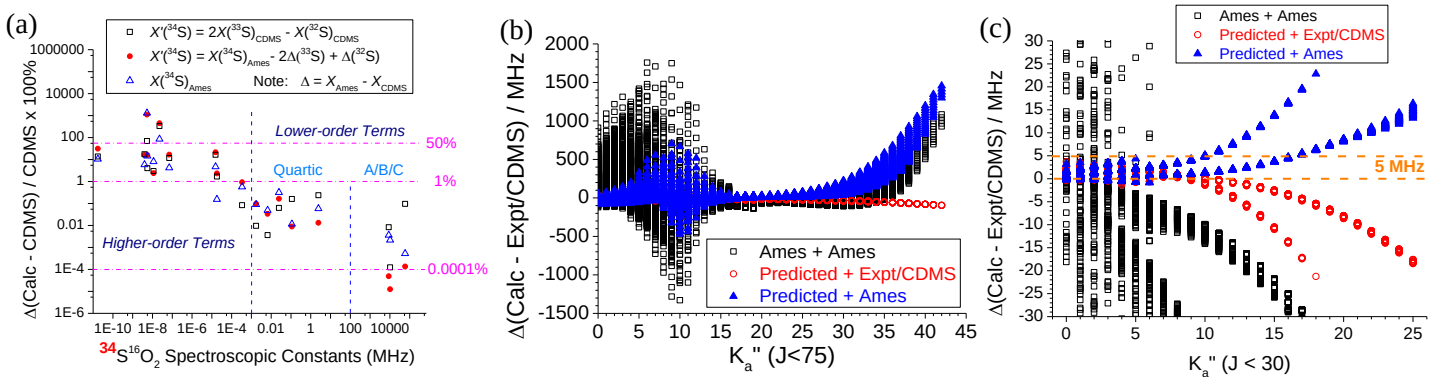


Fig.17. (a) the relative differences (%) of original X_{Ames}, and X_{Ames} predicted from two formula, with respect to the 646 X_{CDMS} parameters; (b) and (c) compare the $^{34}S^{16}O_2$ line positions predicted by three sets of EH parameters vs. original EH(CDMS/Expt) line positions: (b) J up to 75, K_a > 40; (c) reliable range, J" = 0 – 30 and K_a" = 0 – 25.

Fig.17b shows the overall comparison with J up to 70+ and K_a>40. Significant deviations are found in the lower and higher K_a region of the original EH(Ames) and EH(Predicted+Ames) parameter sets. For low K_a<10, the black open squares of the Ames lists rise to as large as 0.05 cm^{-1}. At high J, the errors for the *A/B/C* constants (lower order) become very significant. More than 80% of these deviations disappear when the lower order terms are replaced by the "Predicted" values, i.e. the blue triangles for EH(Predicted+Ames). At high K_a > 35, both "Ames+Ames" and "Predicted+Ames" show fast-growing deviations. These deviations probably result from the over-extrapolation of the EH(CDMS) model, mainly the higher order terms. Similar over-extrapolation errors were reported for the 626 EH(CDMS) model in Fig.9d of Ref.[9], where the old EH(CDMS) 626 model [44] was extrapolated from K_a<=23 to K_a=35-45. In this 646 test, the higher order terms are fit with K_a" up to 31 in EH(Ames) vs. up to 20 in EH(CDMS).[44] So a similar magnitude of over-extrapolation errors can be expected at high K_a for the EH(CDMS) parameters. "Predicted + Expt/CDMS" line positions noted by red open circles have much smaller deviations at high K_a, because it uses the exactly same EH(CDMS/Expt) high order terms. Therefore, we can reduce the comparison to K_a<25, see Fig.17c. In this range, the EH(CDMS/Expt) extrapolations should be still accurate. The original EH(Ames), i.e. "Ames+Ames", can be ignored due to too large errors.

With K_a<25, the line position deviations of "Predicted + Ames" increase more rapidly than those of "Predicted + CDMS/Expt" when J goes beyond 70 (not shown). It might be traced to certain higher order J terms. The J range is 0-52 in 646 EH(Ames) line set, while 0-55 in the 646 EH(CDMS) line set. The J term differences might only become noticeable at higher J extrapolations. Such differences between the two EH sets are not our primary concern, not before we get accurate lower K_a/J transitions first. To avoid the high J effects, we further reduce the J range to 0-30.

Fig.17c shows the final accuracy comparison in the range of J<30 and K_a"<25. The "Predicted+Ames" deviations are 0-5 MHz for most transitions with K_a<20. Deviations of most K_a<10 transitions are within 2-3 MHz. Compared to those deviations in Fig.16a, the accuracy is improved by nearly 2 orders of magnitude.

The results of this simple EH(646) refinement suggest the *possibility* of 1-2 orders of magnitude improvement on the MW line position accuracy of Ames-296K IR lists for the minor isotopologues. The discussion about large deviations in the high K_a / J region should be also applicable to other isotopologues or molecules. In Fig.17c, the "Predicted + Ames" data is only worse than the "Predicted + Expt/CDMS" data by a few MHz at most. We believe such accuracy may offer definitive reference data and facilitate astronomical data analysis and assignments, from individual line search to collective pattern match. Theoretically, other SO_2 isotopologue MW line lists can be improved in the same way, including those SO_2 pure-rotational bands in Sec. 3.3.3 and Fig.S7. Vibrational bands can also be improved if a few high quality EH(Expt) models are available.

The remaining issue is: is it possible to guarantee such accuracy improvement? If the answer is "Yes", how to do that? If no, why? And how bad may the predictions become?

Unfortunately, the short answer is "NOT YET". The analysis in Fig.16b and 16c are bad approximations, being not strictly correct. In addition, the 646 test and comparison in Fig.17 did not consider many important factors. After reporting the seemingly satisfactory 646 test in conferences [52,53], we realized how primitive it was. Firstly, the isotope effect should not correlate simply with mass. The M =sum(1/m) coordinate we used in the intensity / TDM analysis is a much better choice. Secondly, the uncertainty of EH(Ames) constants fitted at 1E-5 cm^{-1} level is too large. In addition, it is far from a real apples-to-apples comparison. For example, 8 high order terms in 646 EH(CDMS) were fixed at their 626 EH(CDMS) values.[44] The higher order terms fixed at 626 EH(CDMS) values vary from one isotopologue to another, while the 636 and 646 EH(CDMS) models [44] were taken as "exact" references in the test.

4.2 Critical Factors

Prediction errors for low J/K_a MW line positions mainly come from the errors on the three rotational constants and quartic centrifugal distortion constants, especially D_K/D_{JK}. Deficiencies of other EH terms generally show up at higher J/K_a. The accuracy of the refinement is pre-determined by the reliability and consistency of both EH(Ames) and EH(Expt/CDMS) constants. Ideally, all EH constants in the "Prediction" step should be determined with the highest possible precision, and 100% consistency for all isotopologues. Unfortunately that is rather difficult, if not totally impossible. For now, the isotopologue prediction accuracy on most molecules cannot always be as good as shown in Fig.17c, due to various factors to be briefly discussed in this section. These critical factors impact EH(Ames) and EH(CDMS/Expt), so they strongly affect our refinement, especially the isotopologue consistency of final improvements. Readers will find some additional discussion in the Supplementary material, and do their own evaluations (if interested). Note the role and magnitude of some factors vary by molecules, bands, the quality of available data, and even the variational programs generating the semi-empirical IR line lists, but the basic ideas

behind the "Ames lists + EH SPFIT → mass-dependent correction → mass-independent accuracy" should probably remain the same.

4.2.1. Consistency of EH(Expt) parameters

Isotopologue inconsistency among EH(Expt) parameters, if not identified and minimized, will pass to the final results. In principle, all SPFIT/SPCAT analyses should use the same EH model, because different EH models may have different formula form or include different parameters. The weight for a specific EH parameter should be same for all isotopologues. No parameters shall be fixed at any value, neither the main isotopologue value nor some "predicted" value. If all isotopologue EH fits have a pre-defined group of parameters fixed, it may introduce too much weights or unnecessary noises. The EH(CDMS) models are the best set we can find, but they are still affected by these inconsistencies. For consistency, we exclude ^{33}S and ^{17}O hyperfine splitting terms, assuming they are not strongly coupled with other major terms.

The transition line sets should be equivalent, or as similar as possible, for different isotopologue EH fits. The SPFIT program prefers that not only transition quantum numbers, but also the experimental uncertainty associated with each individual line needs to be consistent among all isotopologues. This is extremely difficult or essentially unrealistic. It is even more impossible to have theoretical IR lists carrying the same or similar data uncertainty as the experimental data. In this aspect, experimental line position represents the lab reality. But the reality is, many rovibrational EH(Expt) models of minor isotopologues were fit with data far less than that of main isotopologue. This is also a source of inconsistency among EH(Expt) models and parameters.

Take the 646 EH(Ames) as example, we find the effects of line set difference are -0.006 MHz, 0.004 MHz, and 0.001 MHz for the $A_0/B_0/C_0$. These are computed as the differences between Section 4.1 values fit from ~4000 transitions, and new SPFIT results using a transition set of 282 lines, which is equivalent to the corresponding 646 EH(CDMS) line set. It is highly recommended to converge EH(Expt) with as many as possible reliable measured transitions.

4.2.2 Convergence of EH(Ames) constants: A_0 and D_k

In the *A*-reduced Watson Hamiltonian, A_0 and D_k are explicitly correlated. This calls for extra effort to converge them. Otherwise the predicted A_0 and D_k may have large deviation. Accordingly, the line list prediction for other isotopologues will lose accuracy and consistency.

To achieve 0.01 MHz prediction accuracy on the $A_0/B_0/C_0$ of EH(666) or EH(646), linear approximation requires better than 0.005 MHz accuracy on the $A_0/B_0/C_0$ of EH(646) or EH(636). In our tests, the B_0/C_0 in EH(Ames) are usually converged better than 0.005 MHz. Nevertheless, the SPFIT in Sec.4.1 was carried out at the 1E-5 cm^{-1} uncertainty level. It cannot guarantee A_0 converged better than 0.01 MHz, at least not for the ~4000 transition set of the EH(646) fit in Fig.16b and Fig.17a. We found two *stable* local minima of A_0/D_k at 58990.81394 MHz / 2.440615 MHz, and 58990.88397 MHz / 2.441753 MHz. Such "oscillation" is too large for reliable predictions. The 0.07MHz discrepancy on A_0 may lead to ~60 MHz uncertainty at J=30, and 0.00144 MHz on D_k will bring ~70 MHz uncertainty at K_a=15. In addition, there is a reaction-path like A_0/D_k series connecting the two minima. Two EH(646) fit examples on the "path" are given in Supplementary Material (SM): A_0/D_k = 58990.8267 / 2.44082 MHz, and 58990.8622 / 2.44147 MHz, respectively. They have similar fitting quality. Interested readers can try the fit by themselves. Note the corresponding changes on B_0 / C_0 are -0.0039 MHz and -0.0024 MHz, i.e. they are still good for the 0.01 MHz goal.

The solution to this issue is to reduce the uncertainty level from 1E-5 cm^{-1} to 1E-6 cm^{-1}, or even 1E-7 cm^{-1}. At 1E-6 cm^{-1} level, the A_0 "oscillation" is reduced to ~0.006 MHz, and even smaller at 1E-7 cm^{-1} level. It should be good

enough for the following analysis and refinement. All following analysis and refinements use 1E-6 cm^{-1} level results, unless specified.

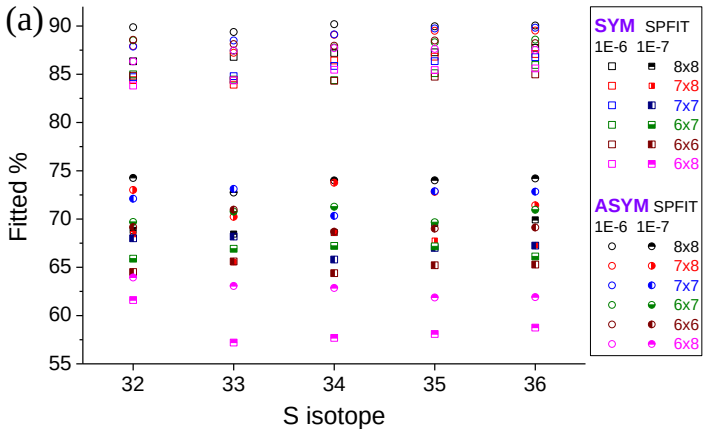
This convergence defect at 1E-5 cm^{-1} level raises doubt on the A_0 / D_K accuracy of experimental IR analysis at higher energies. Most IR lines reported for polyatomic molecules were not published with 1E-6 cm^{-1} (or smaller) error bar. We have serious concerns on those experimental A_0 associated with error bar as large as 1E-3 ~ 1E-4 cm^{-1} -- unless those experimental line positions did carry precision higher than 1E-5 cm^{-1}. However, such multi-minima issue did not appear in our SO_2 tests using just hundreds of transitions. It may only become significant for considerable number of transitions and/or relatively large uncertainty. On the other hand, most transitions in the original 646 EH(CDMS) fit carry an uncertainty as small as 0.05-0.20 MHz, or 1.7-7E-6 cm^{-1}. It is in the safe range we find in the tests. Therefore we believe the fitted EH(CDMS) parameters have been reliably converged.

4.2.3 New EH(Ames) Line Set and Fitting Accuracy

To guarantee the data purity for our BTRHE implementation, it is necessary to understand the isotope dependence of EH model fitting quality. This sub-section explains how the most isotopologue-consistent SO_2 MW line sets are generated and used to figure out how the fitting accuracy and percentage vary systematically with isotope substitution and uncertainty levels. These trends may result from the K_C split discrepancy at high J/K_a, but they provide an effective check for the sanity and consistency of both Ames line lists and the EH(Ames) fits.

In contrast to experimental data, theoretical MW/IR lists may have complete isotopologue data with consistent (and sometimes higher) precision. They are good for systematic convergence check on the fitted EH constants with respect to uncertainty levels. Nonetheless the SO_2 MW line sets used in previous EH(Ames) test only carry 1E-5 cm^{-1} precision, i.e. only 5 digits after the decimal point. To ensure the rotational constants converged to better than 2 kHz, or 0.002 MHz, all Ames-296 MW lists were re-computed with 1E-8 cm^{-1} precision. The re-computation process has the highest consistency we can reach.

Firstly, J=0-75 rovibrational energy levels of all 30 isotopologues have been re-computed, using an uniform set of VTET[54] input parameters for Asymmetric isotopologues. Then all the $J \leqslant 60$, $K_a \leqslant 35$ transitions with S(296K) > 1E-32 cm/molecule of isotopologue 668 are selected to be the "standard" line sets. Equivalent line sets are extracted from other 29 isotopologue GS←GS lists. To study the difference between symmetric (Sym) and asymmetric (Asym) isotopologues, two line sets are prepared for each isotopologue. The Sym set only contains those transitions allowed in the main isotopologue 626. The Asym set contains everything allowed for asymmetric isotopologues. Each Sym set has 13,744 lines, and each Asym set has 40,311 lines, in the range of 0 – 370~410 cm^{-1}.



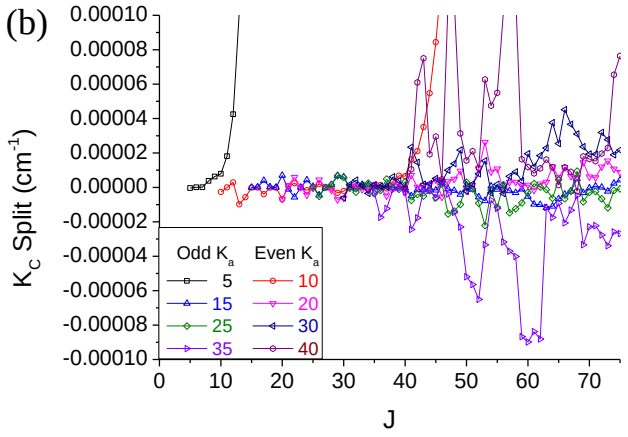


Fig.18. (a) Percentage of fitted lines in 30 EH(Ames) analysis, using Sym or Asym datasets, and two different uncertainty levels from 1E-6 to 1E-7 cm^{-1}. Sorted by line set, uncertainty level, and S & O isotopes. (b) The K_c splits of Ames SO_2 668 ground state levels. Split $\Delta = E(J, K_a, K_c=J-K_a) - E(J, K_a, J+1-K_a)$.

There are 4 levels of uncertainty: 1E-4 cm^{-1}, 1E-5 cm^{-1}, 1E-6 cm^{-1} and 1E-7 cm^{-1}, and the SPFIT [16,17,50] error bar is set to uncertainty x 5. Due to the nature of over-determined linear equations, we cannot 100% guarantee the "best" EH parameters have been determined for all the 1E-6 and 1E-7 fits. The patterns are crystal clear. The fitted % at 1E-6 cm^{-1} are 86-90% (Asym) or 83-87% (Sym). They are reduced to 62-74% (Asym) or 57-70% (Sym) at 1E-7 cm^{-1}. At each level, an Asym set always has higher % fitted than the related Sym set. Interestingly, for symmetric isotopologues, heavier O isotopes are correlated with smaller frequency range, reduced fitting residuals and higher fitted %, i.e. 8x8 > 7x7 > 6x6; for asymmetric isotopologues, less O atom (relative) mass differences means higher fitted %, i.e. 7x8 > 6x7 > 6x8. As a result, the 6x8 isotopologue group consistently has the lowest fitted %. See Fig.18a and Fig.S11 for fitting % details. Note the fitting % count the transitions with fitting deviations < 5 × uncertainty.

Fig.S11 Percentage of fitted lines in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm^{-1}. (a) overview, by fitting accuracy level; (b) detailed comparison, by fitting accuracy, S and O isotopes. [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

The fitted % reduction at 1E-7 cm^{-1} is probably related to the difference between Ames rovibrational energies and EH models. We compute the K_C split, $\Delta = E(J_{Ka,J-Ka}) - E(J_{Ka,J+1-Ka})$, for all 668 ground state rotational levels ($J \leq 70$). For high J levels, the Δ at odd/even K_a oscillates between negative and positive regions. See Fig.S12. The Δ magnitude is up to 1E-5~1E-4 cm^{-1} in range of K_a >30 and J=40-70. See Fig.18b and Fig.S12. Such Δ does NOT exist in the EH/SPCAT results we acquired in this work. For example, at J=60, all Δ of $K_a \geq 15$ levels in EH/SPCAT outputs are less than 1E-7 cm^{-1}. Two possible explanations deserve future investigation for this *two* orders of magnitude discrepancy: either Ames calculation at high K_a contains numerical impurities, or the EH model is incomplete for K_a higher order effects. The Δ oscillations make many Ames frequencies deviated from EH model predictions by more than 5E-7 ~ 5E-6 cm^{-1}, but mainly on high J/K_a. This suggests the corresponding EH(Ames) constants determined at 1E-6 and 1E-7 cm^{-1} level be more reliable for low J/K_a, especially 1E-7 cm^{-1}.

Fig.S12 The K_c splits of Ames SO_2 668 ground state levels. Split $\Delta = E(J, K_a, K_c=J-K_a) - E(J, K_a, K_c=J+1-K_a)$. Left side is overview; Right panel shows more details. [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

4.2.4 D_K and D_{JK}: Convergence and Prediction

Because the D_K and D_{JK} are the two most important quartic constants in EH(Ames) refinement, this sub-section presents the quantitative estimation of their convergence, uncertainty, and if they have nearly linear mass dependence.

The D_K convergence is carefully checked in Fig.19a and Fig.S13. The D_K values are mass-dependent. From 626 to 868, it drops from 2.59 MHz to 2.04 MHz. There are 8 D_K values for each isotopologue: two datasets (Asym vs. Sym) x 4 accuracy levels (1E-4~1E-7 cm^{-1}). The spread of D_K values (with respect to the mean averaged from 8 D_K values) shrinks from ±20Hz to ±15Hz. This means the relative spreads are approximately same for all isotopologues. An interesting observation is that D_K (Asym) are slightly higher than D_K(Sym), while their differences decrease from ~10 Hz (1E-4 fit) to 1-2 Hz (1E-6 fit), or even smaller (1E-7 fit). This is an indication of convergence. The range of $\delta = D_K(1E-7) - D_K(1E-6)$ is -15 Hz ~ -5 Hz, while the D_K (1E-4) and D_K (1E-5) are not recommended.

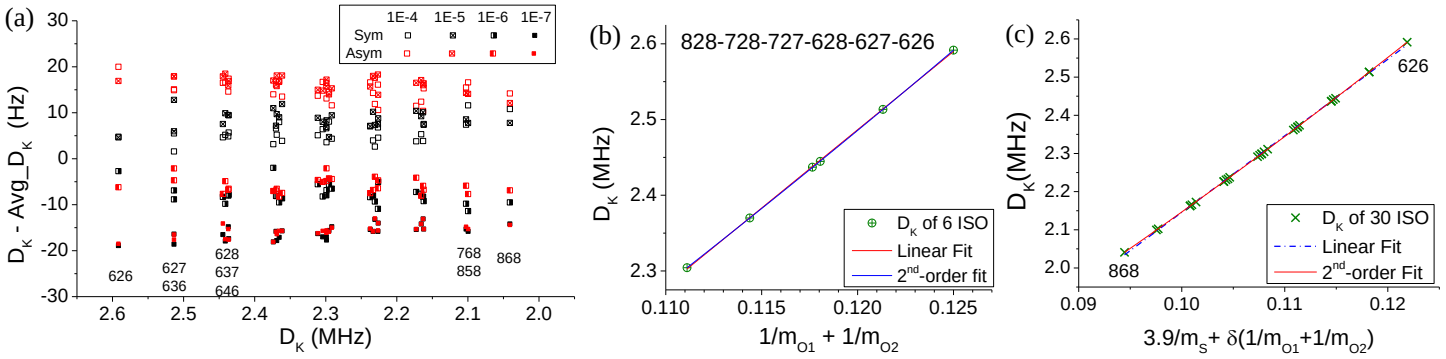


Fig.19 (a) D_K convergence in the EH(Ames) of 30 isotopologues, using Sym or Asym datasets, and four different uncertainty/accuracy level from 1E-4 cm^{-1} to 1E-7 cm^{-1}, with respect to the D_K value averaged from 8 SPFIT fits. (b) The linear and 2nd-order fits of EH(Ames) D_K values (determined at 1E-6 cm^{-1} level) along the mass inverse coordinate, *M*; (c) a coordinate integrating the S & O isotope effects. See text for details and why we can do better than b) and c).

Fig.S13. D_K analysis in the EH(Ames) of 30 isotopologues, using Sym or Asym datasets, and four different uncertainty/accuracy level from 1E-4 cm^{-1} to 1E-7 cm^{-1}. (a) overview, by fitting accuracy level, data are with respect to the value averaged from 8 D_K; (b) the D_K convergence with respect to fitting accuracy: δ(Asym-Sym) are black squares, and $\delta(D_K(n+1) - D_K(n))$ are triangles, where *n* is the accuracy/uncertainty level, e.g. 5-4 means D_K(1E-5) – D_K(1E-4). [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

In Fig.19b and 19c, the D_K values are plotted against the *M* coordinates. Just like in Section 3, linear or 2nd-order approximations can work well. In Fig.19b, both D_K(628) and D_K(727) follow the approximation lines, and can be separated from each other. This strongly supports the validity of *M* coordinate. In Fig.19c, the scaling factor between $1/m_S$ and $(1/m_{O1}+1/m_{O2})$ is 3.9. It is handpicked for best overall linearity. The linear fit has relative deviations from -0.11% to +0.33%, with mean ± σ = 4E-4 ± 0.11%. The 2nd order fit has relative deviations in -0.038 ~ +0.024%, with mean ± σ = 0.0 ± 0.016%. The prediction accuracy on many isotopologues is already better than the performance of the formula we are going to discuss in Sec.4.2.8. From our experience in Sec.4.2.2, the D_K convergence should achieve 0.1 kHz (or better) to ensure A_0 accuracy. For SO_2, 0.1 kHz equals to 0.006% of D_K. By analyzing D_K(Ames)-D_K(Expt), we can further improve the D_K predictions. See Sec.4.3. Nevertheless more factors affecting the data need to be discussed first.

Fig.S14 D_{JK} in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm^{-1}. (a) different D_{JK} values, by fitting accuracy level and data type, with respect to the value averaged from 8 D_{JK}'s; (b) the D_{JK} convergence with respect to fitting accuracy: δ(Asym-Sym) are black squares, and $\delta(D_{JK}(n+1) - D_{JK}(n))$ are triangles; (c) the D_{JK} overview of 30 SO_2 isotopologues. [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

Please note that not all EH constants can find such linear relations easily. The 2nd most important EH constant at quartic level is D_{JK}. Convergence for D_{JK} has reached Δ(Asym-Sym) < 1 Hz and δ(1E-7 – 1E-6) = 2~3 Hz. But the D_{JK} vs. *M* cannot find any linear relations, not even roughly. See Fig.S14. However, what really matter are the patterns of the EH(Expt) - EH(Ames) differences, not the pattern of EH(Expt) or EH(Ames) itself.

4.2.5 Uncertainty Check for EH(Ames) parameters

In addition to D_K and D_{JK}, this sub-section carries out similar quantitative check for other quartic terms and all those *higher* order EH(Ames) constants. The original purposes of the check are to determine their volatility and noise, to confirm their normal range and identify (if any) outliers, and to justify the choice of refining lower order terms. The uncertainties of *higher* order terms are large enough to prevent meaningful formula-based predictions. These are the limitation of our "EH/SPFIT + Ames" approach. We believe these findings may also provide valuable reference to EH(CDMS/Expt) analysis, because the *complete* sets of MW lines and related benchmark EH analysis are not a

feasible option for most polyatomic molecules.

Please note the difference of two kinds of "uncertainty". One is the SPFIT value range for a EH constant, the other is the numerical error bar for the solution of least-squares problem. We call them "constant uncertainty" and "fitting uncertainty", respectively. But only the relative uncertainties concern us. The relative "fitting uncertainty" are extracted from the SPFIT results at 1E-6 cm^{-1} or 1E-7 cm^{-1} level, then divided by the mean value of corresponding constants. The relative "constant uncertainty" is estimated by δ|EH(1E-7)-EH(1E-6)| differences, which are divided by the corresponding constants to get the relative δ%. Both uncertainties are averaged across 30 isotopologues. Fig.20a plots absolute mean values of constant uncertainty (right axis, blue circles with crosses), relative constant uncertainty (left axis, half filled red squares), and relative fitting uncertainty (left axis, open triangles). The EH parameter values on *x* axis are also averaged from 30 isotopologues. This is appropriate in the log scale.

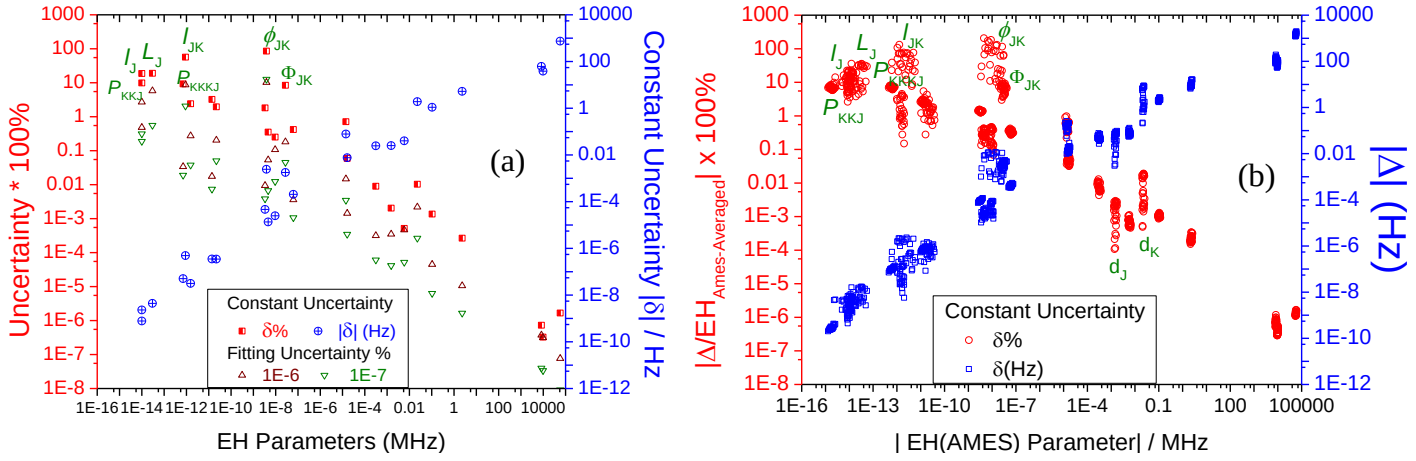


Fig.20 (a) The fitting uncertainty and constant uncertainty (convergence) of 26 EH(Ames) parameters, all averaged from 30 SO_2 isotopologues. Left axis: the relative fitting uncertainty % from SPFIT (triangles), and relative constant uncertainty δ%|EH(1E-6) – EH(1E-7)| (red squares, half filled); Right axis: absolute constant uncertainty δ|EH(1E-6) – EH(1E-7)| in Hz; (b) constant uncertainty of 26 EH(Ames) parameters in the 1E-6 cm^{-1} and 1E-7 cm^{-1} SPFITs of 30 isotopologues. Right axis: absolute uncertainty in Hz (blue squares); Left: relative constant uncertainty % (red circles).

Compare to Fig.20a, Fig.20b adds back the isotopologue variance of all quantities. The EH(Ames) parameters are available in supplementary file or upon request. Interested readers can apply unique color & symbol to every EH constants and/or every isotopologue, for further explorations.

In Fig.20, our main observations include: (1) the constant uncertainty is the major factor, 2-4 orders of magnitude larger than the fitting uncertainty; (2) rough linear relation can be established between constant uncertainty (in Hz) and parameter values; (3) in general, relative constant uncertainties are 0.0001% ~ 0.01% for quartic constants, 0.1~10% for sextic terms, and 1-100% for higher order terms, so it supports our choice not to refine them; (4) not all the "higher order terms" have similar uncertainty or relative variations, the largest δ% are: ~10% for P_{KKJ}, P_{KKKJ}, and Φ_{JK}; ~20% for L_J and l_J, and 50-100% for ϕ_{JK} and l_{JK}. (5) in all "lower order terms" to refine, the isotopologue variances of d_J and d_K constant uncertainty cross more than one order of magnitude, significantly wider than others. Fortunately, even the largest δ% are still less than 0.05% (d_K) or 0.005% (d_J). (6) Even with 40,311 transitions (J/K_a up to 60/35), not all EH parameters can be precisely determined.

4.2.6 EH(CDMS/Expt) Higher-order Term Effects & Inconsistency

Since we have justified the choice to keep the original EH(Ames) *higher* order terms, the small deviations in the EH(Ames) *higher* order terms will determine the upper limit of the final line position accuracy after refinements. This

sub-section presents the EH(646) limit as example. Because the EH(Ames) models are considered highly consistent among isotopologues, it also serves as a further check on the isotopologue consistency of EH(CDMS/Expt) *higher* order terms.

This check can serve 2 purposes. First one is to determine the upper limit of EH(Ames) refinement accuracy. Second one is to check the isotopologue consistency of published EH(CDMS/Expt) models. Simply replacing the EH(CDMS/Expt) higher order terms with the EH(Ames) counterpart, we re-run SPCAT to generate the line positions and compare with pure EH(CDMS/Expt) based predictions. The line position differences are shown in Fig.21.

In Fig.21a are the results on 646, where the high-order term effects (red open squares) are -3 MHz – 5MHz for most K_a'<15, J<60 lines. This is the best prediction we can expect. Although certain combinations of lower + higher order terms may lead to even better accuracy, but it is not systematically achievable. In Fig.21b are the results collected on 7 EH models: 5 from CDMS[44], 2 from Ulenikov et als [45,49]. The Ames-296 MW line list based EH(Ames) are self-consistent. If the 7 EH(CDMS/Expt) models are also consistent, we should be able to see some pattern in Fig.21b. Unfortunately, it is hard to find any trend along 626→636→646, or 626→627→628 →828. This is understandable. Because different higher order terms in EH(CDMS) were fixed at 626 values. The 828 EH model difference is related to the Ulenikov prediction formula [49] to be discussed in Sec.4.2.8. The 636 EH(Expt) model of Blake and Lafferty [47,48] is not included here due to consistency issues found in their published constants.[55] For 626, we all agree the Ulenikov EH model [45] is a upgrade from (better than) CDMS model, while its higher-order effects seem slightly larger than those of CDMS model. A L_{JJK} discrepancy is recently found [56] between the two EH model fits, but an inclusive investigation is beyond the scope of this study. In short, 100% apple-to-apple comparisons are still unavailable. For example, it requires to include the δ(Fixed – Relaxed) corrections.

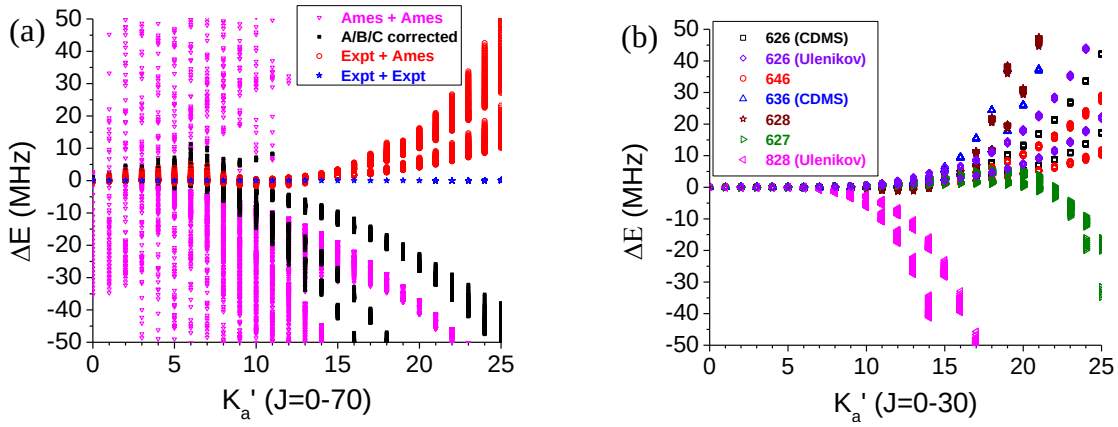


Fig.21. Line position differences caused by A/B/C, quartic level terms, and higher order terms. (a) SPCAT line position predictions using 4 sets of EH(646) parameters, with respect to pure EH(CDMS/Expt) model based line position predictions. EH sets are labeled as "lower order terms" + "higher order terms". "Ames" is original EH(Ames), "A/B/C corrected" includes expt A/B/C + Ames quartic +higher order terms. See text for explanations. (b). the EH(Expt + Ames) model based SPCAT line positions for 6 isotopologues, with respect to the pure EH(CDMS/Expt) model based SPCAT line position predictions.

4.2.7 Fixed vs. Relaxed.

Another important factor contributing to EH constant uncertainty is the regular practice fixing some minor isotopologue EH constants at corresponding main isotopologue values. The magnitude of the corresponding impacts on EH constants needs to be quantitatively determined so that we can decide if extra correction is necessary before the EH constant refinements.

The impacts are easy to compute, and necessary for refinement. Original EH(Ames) constants are denoted "relaxed". Then we fix certain 627, 628, 636, 646, and 828 higher order terms at 626 EH(Ames) values, re-run the SPFIT at 1E-5, 1E-6 and 1E-7 cm^{-1} levels. The terms fixed are the same ones fixed in the EH(CDMS) analysis. Note 828 also follows the 628 EH(CDMS) "fixings". The new EH(Ames) parameter values are denoted "Fixed". Differences Δ are defined as X(Fixed) – X(Relaxed), or the "Fixed-Relaxed" effect. Note the X terms are not fixed, but their values are different when other terms are fixed vs. relaxed.

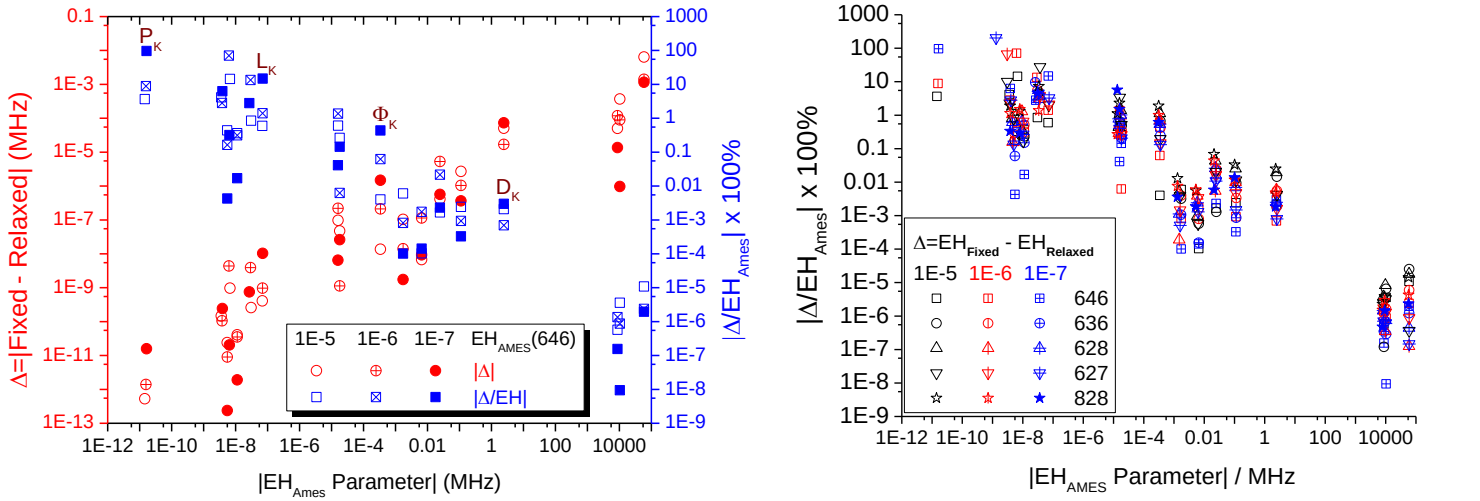


Fig.22 (a-b): Absolute (Left y axis, circles) and relative (Right y axis, squares) differences of the EH parameters acquired between fully relaxed fit (all vary freely) and the partially fixed fit (some parameters are fixed at 626 values): (a) 646 only; (b). the relative differences (%) of 646, 636, 628, 627, and 828, at 3 fitting accuracy levels. [See Supp.Figures.and.Sec4.Discussions.updated.pdf]

Fig.22a shows the Δ and Δ% of 646 EH(Ames) constants. In general, smaller constants have larger "Fixed" effects. The "Fixed" effects are smaller at 1E-6 and 1E-7 cm^{-1} level. At 1E-7 cm^{-1} level, the effects on $A_0/B_0/C_0$ are only 10^{-5}-10^{-3} MHz, similar to the constant uncertainty in Fig.20a. But it is hard to predict at which uncertainty level the effects on those not-fixed higher order terms are the largest, or the smallest.

Fig.22b expands the Δ% data to 5 isotopologues at 3 fitting accuracy levels. The general trend remains valid for most X >1E-9 MHz. In other words, higher fitting accuracy leads to smaller Δ% "Fixed" effects. On the other hand, the Δ% magnitude of different isotopologues could differ by orders of magnitude. In 5 isotopologues, the Δ% (646) effects are the smallest. Possible explanations may include that X(646) is closer to X(626), or the terms to be fixed are selected more wisely for 646; etc.

The X(828) terms between 1E-4 MHz and 5 MHz have "Fixed" effects Δ% larger than other 4 isotopologues. Not sure if this suggests the optimal choice of terms to fix be isotopologue dependent. The 828 fits in Fig.22b panels fixed the same EH term set that was fixed in the 628 EH(CDMS) fits. Please note they are NOT the same kind of 828 EH models in Fig.21b, where Ulenikov et al [57] used a "prediction" scheme to estimate higher order terms. The scheme is discussed and compared in next sub-section.

4.2.8 EH Parameter Prediction Accuracy Using Ulenikov Formula

The investigation in 4.2.7 does not apply to interesting cases when some EH constants were fixed at theoretically predicted values, instead of main isotopologue values. This sub-section tries to answer the question by running systematic checks on the performance (or accuracy) of such a prediction scheme. The goal is to prove our BTRHE implementation is still more accurate and can help improve its performance.

The EH(828) analysis in this section follows the procedure given in the section 3.2 of Ulenikov et al[49], which utilizes the A_{828}/A_{626} and C_{828}/C_{626} ratios to estimate other 828 EH constants. The isotope substitution theory [58–60] predicts the centrifugal distortion constants as:

$$X_{828}^{ij} = X_{626}^{ij} \cdot \alpha_A^i \cdot \alpha_C^j \,,\ \alpha_A = A_{828}/A_{626},\ \alpha_C = C_{828}/C_{626},$$

For a specific EH constant, if its total power is m, the power of its J_Z^2 part is i, and $j = m/2 - i$. The formula applies to symmetric C_{2v} type isotopologues. In Ref.[49], α_A and α_c were determined from established 626 *A/B/C* values and the 828 *A/B/C* constants fit from J=0-4 transitions. Then all sextic and higher order terms of the 828 EH(Expt) fit were fixed at the values predicted with this formula, except H_K and H_{KJ}.

Compare to EH(Expt) models, the EH(Ames) model set has much higher isotopologue consistency. The 30 isotopologue IR lists were computed with a consistent VTET [54] parameter set and fitted to a uniform EH model with accuracy control and satisfactory constant uncertainties (Fig.20). The formula performance check was carried out with the EH(Ames) model pair of *any* two symmetric isotopologues. Relative deviations % of predicted values are computed with respect to original EH(Ames) term values.

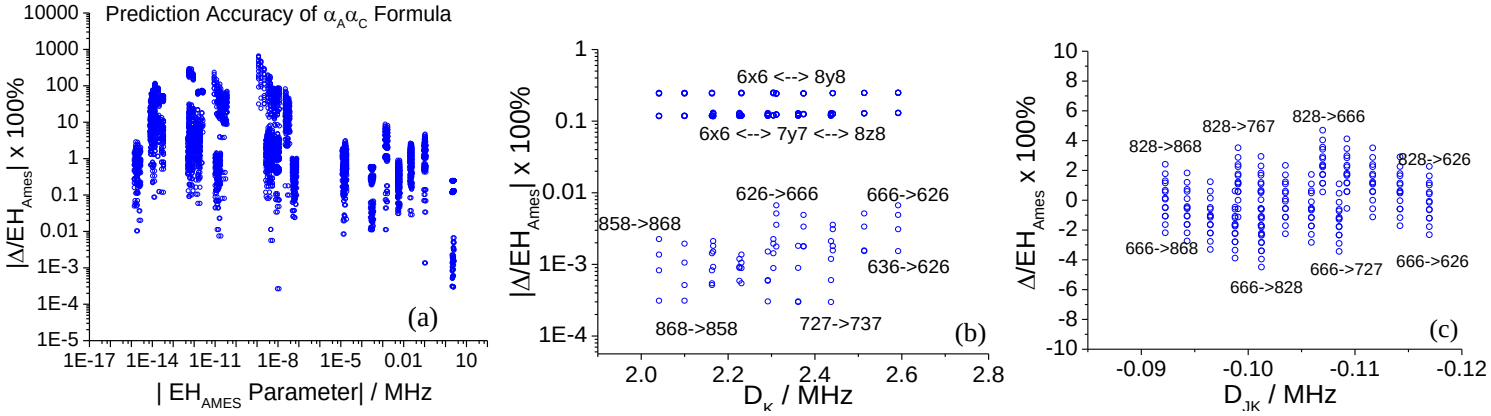


Fig.23 Accuracy check for the prediction formula Ulenikov et al [49] adopted in their EH(828) model. The 30 SO_2 isotopologue EH(Ames) parameter sets acquired at 1E-6 cm^{-1} level are taken as standard reference for relative prediction deviations Δ%. (a) overview for 30 isotopologues; (b) the D_K prediction accuracy, *x/y/z* =2-6 for $^{32}S - {}^{36}S$ isotope, x≠y, y≠z; (c) the D_{JK} prediction accuracy.

Fig.23a collected all one-to-one prediction deviation δ% of 23 EH terms, i.e. without *A/B/C*. Except for D_K, quartic term predictions have |δ|% in 0.01 ~ 10%. The δ% for sextic and higher order terms are noticeably higher, |δ| = 0.1 ~ 1000%. This is normal. However, not all higher order terms are fitted with adequate line set. If we need to know what percentage of those 10-1000% deviations were resulted from constant uncertainty, it requires systematic check for each EH parameter. If the constant uncertainty is insignificant, corresponding |δ%| is expected to have clear patterns. The patterns probably relate to the effective mass changes and vary by specific EH constants. Fig.23b and 23c present the D_K and D_{JK} pattern checks.

The |δ%| in Fig.23a spread in a wide range of 2-4 orders of magnitude, but hard to see detail variations from one isotopologue to another. In Fig.23b, $|\delta(D_K)|$ are plotted with linear scale on horizontal axis. Obviously, there are two groups of $|\delta(D_K)\%|$: the 0.1-0.3% group for O isotope effects, and 0.0003-0.007% group for S isotope effects. It simply indicates the isotopologue consistency of the prediction formula breaks up on O isotopes, although |δ%| are still systematic. In both groups, smaller isotope mass differences are associated with smaller |δ%|. For example, 7y7←8z8 is more accurate then 6x6←8z8, and 868←858 prediction is more accurate than 666←626 by more than one order of magnitude. The O isotope effects are dominant, so in log scale the S isotope effects are negligible in first group

and only discernible in second group.

We compare the first group of prediction error, $|\delta(D_K)\%|$ = 0.1-0.3%, to the approximations in Fig.19b. The simple linear fit on the 626-627-628-727-728-828 D_K series has relative σ_{RMS} = 0.054% and $|\delta\%|_{max}$=0.07%, with respect to $1/m_{O1}+1/m_{O2}$. The 2nd-order fit has σ_{RMS} = 0.01% and $|\delta\%|_{max}$=0.018%. Prediction accuracy of our approximations leads by one order of magnitude. It benefits from the choice of *M* coordinate, and the isotopologue consistency of EH(Ames) models. Please note these are the approximations without knowing EH(Expt), i.e. *before* any refinements. If we include the 0.00~0.05% Ames-Expt difference, simple linear approximation still works better than the formula on D_K(8x8) ←→D_K(6x6), *x*=2-6. Prediction accuracy of simple approximations is the evidence for isotopologue consistency of EH(Ames).

For D_K(6*x*6)←→D_K(6*y*6) predictions (*x,y* = 2-6, *x*≠*y*), the formula works really well. The $|\delta\%|$ range in Fig.23b are 1e-6~1e-4, or 0.0001-0.01%. It is significantly smaller than the 0.02-0.04% deviations we get from linear fitting of corresponding EH(Ames) D_K values with respect to $1/m_S$. A 2nd order fit can reduce $\delta\%$ to 1e-8 ~ 1e-7, which are within the constant uncertainties. So the formula's accuracy is between the 1st and 2nd order fits. Up to now, we have not found a simple 1st-order fit doing better than the formula on S isotope substitutions.

D_{JK} is the 2nd most important quartic order constants in EH models, for low J/K. In Fig.23a, the $|\delta\%|$ of D_{JK} are in the range of 0.002~10%, but no details. By changing both axis to linear scale, and using original $\delta\%$ instead of $|\delta\%|$, Fig.23c clearly demonstrates how the $\delta\%$ systematically vary along with S / O isotope substitutions. The prediction errors of that formula are evenly distributed within ±5%. It is interesting to note the $\delta\%$ distribution is highly symmetric, while their S and O isotope effects are opposite to each other. The smaller S or O isotope mass differences between two isotopologues, the smaller the $\delta\%$ magnitude will be. But in general, most δ and $\delta\%$ are too large for D_{JK}.

Compare to CDMS/Expt D_{JK} values reported around 0.11 MHz, the original EH(Ames) D_{JK} fitted at 1E-7 cm^{-1} accuracy level are only off by 0.001-0.002% (626), -0.001%(636), -0.0003% (646), -0.001% (627), -0.013% (628) and -0.012% (828). They are 2-4 orders smaller than the errors of prediction formula. It has also reached the point that our deviations are comparable to or even smaller than the fitting uncertainty of EH(Expt) models.[44] For example, D_{JK}(646) of EH(CDMS) is 0.11165788(261), the fitting uncertainty = 2.3e-5, or 0.0023%, larger than our deviation. For the main isotopologue 626, CDMS fitting uncertainty is 3.8e-6, or 0.0004%, smaller than our deviation. But the D_{JK} reported in 2013 Ulenikov et al [45] is out of the uncertainty range of CDMS value: D_{JK}(Ulenikov)/D_{JK}(CDMS) -1 = 1.32e-5, or 0.0013%, similar to our deviation. The 2013 EH(Expt) model was as an upgrade to EH(CDMS). Therefore, it is not worth to look for $\delta(D_{JK})$ patterns out of uncertainties and noises. One kind of "noise" is the "Fixed" effect in Fig.23b, with relative magnitude 1e-6 ~ 1e-4 at 1E-7 cm^{-1} level. Another kind is the constant uncertainty discussed in Sec.4.2.5 and Fig.20a,b. The D_{JK} convergence is estimated 1~2 Hz at 1E-7 cm^{-1} fitting level. It equals to 1-2e-5 relative uncertainty. These factors all contribute to the differences between EH(Ames) and EH(CDMS/Expt) quantities. Another extreme case was found in our most recent studies on 626 [56] and 636 [61]. If a higher order EH term was fixed at a wrongly quoted value, it might cause collateral damages on all other correlated EH terms or even noticeable changes on the primary rotational constants.

In short, the D_{JK} case is different from the D_K case. In addition to built-in self-consistency, the accuracy of SO_2 EH(Ames) D_{JK} is comparable to EH(CDMS/Expt) D_{JK}. Such agreement is more like accidental. On the other hand, the error pattern in Fig.23c provides a solid base to further calibrate those Ulenikov formula based D_{JK} predictions. Given a few reliable experimental D_{JK}, its $\delta\%$ errors maybe reduced by 1-2 orders of magnitude.

Interested readers may run additional checks to find out if other higher order terms have noticeable systematic $|\delta|$

patterns or noises. Just remember to include the magnitude and the spread of EH(Ames) constant uncertainties in Fig.20 into consideration, plus the "Fixed vs Relaxed" effects, and the "Prediction Formula" effects. The "Fixed vs. Relaxed" effect refers to the EH constant changes caused by fixing some higher order terms at their main isotopologue values. The "Prediction Formula" effect refers to the differences between the original EH(Ames) constants fitted from uniform weights and new EH(Ames) constants fitted with sextic and higher order terms fixed at values given by Ulenikov formula.

Back to the formula, the performance of the $\alpha_A^i\alpha_C^J$ scaling factors is not bad, in some cases is excellent. But it pre-requires the knowledge of *A/C*. The scaling factor $\alpha_A^i\alpha_C^J$ will magnify any *A/C* deviations exponentially and add on to existing formula deviations. Our D_K and D_{JK} tests are done with "exact" *A/C*. Even with the "exact" *A/C*, the predicted D_{JK} may contain a few percent errors. But it is a useful check for the noise (or consistency) level of *some* higher order EH constants. The difficult part is to guarantee the accuracy of higher order EH(Expt) terms used as "standard base". Since all "base" values must come from high-quality least squares fit, most discussions from Fig.20 to Fig.23 are applicable. Note the SPFIT fitting residuals in our 1E-7 cm^{-1} EH(Ames) fits are less than 5E-7 cm^{-1} or 15 kHz, with σ_{RMS} ~ 10 kHz. But the uncertainties associated with certain EH(CDMS) parameters might be larger than scientists may have expected. For example, the 626 EH(CDMS) model we adopted in this study [44] is *not* the one Ulenikov quoted [62] to compare with their EH(Expt) analysis [45]. The ϕ_{JK} discrepancy is -4.3e-9 MHz (from a ground state only fit) vs. -7.0e-9 MHz (from a combined fit, see Table 3 of Ref.[63]). Our 626 SPFIT fits support the former one, -4.3e-9 MHz. The higher order term values of that ground-state-only fit were also adopted in the EH(CDMS) of other minor isotopologues.[44]

Last but not the least important, the small deviations in higher order terms (caused by fixing some EH terms at values predicted by the formula of Ulenikov group) may accumulate to cause detectable changes in lower order terms, including rotational constants. From 1E-5 cm^{-1} to 1E-7 cm^{-1} fitting level, the "Prediction Formula" effects on 828 rotational constants are 0.005-0.011 MHz (A), 0.001-0.002 MHz (B) and 0.003-0.004 MHz (C).

4.3 Refined Linear & Quadratic Predictions (Corrected)

The reason for the detailed analysis in 4.2.1.-4.2.8 is that they are all connected to the final refinement (prediction) we report in this section: we need to know the internal inconsistency among EH(CDMS/Expt) models (4.2.1); to ensure the highest consistency for Ames line sets and EH(Ames) fits (4.2.3); to ensure the convergence of A_0 and D_K is good enough for refinement (4.2.2); to correctly estimate the convergence and uncertainty of D_K / D_{JK} (4.2.4) and higher order terms in EH(Ames) (4.2.5); to have reasonable expectation for prediction accuracy (4.2.6); to ensure we do understand the "Fixed vs. Relaxed" impacts (4.2.7) and "Prediction Formula" impacts (4.2.8). Now it is time to move forward.

In Section 4.1, we have demonstrated the possibility to reach 0.01-0.02 MHz accuracy on rotational constants, and 0-5 MHz deviations for $K_a/J<10$-15 transitions, and higher order term effects are mainly significant in higher K_a/J region. But the old approximation neglected many factors, e.g. the convergence, mass coordinate and "Fixed vs Relaxed" uncertainties in the least-squares fitting, etc. "Prediction Formula" effects should also be taken into account for 828. So before the lower order EH(Ames) parameter refinement, both the "Fixed vs Relaxed" and "Prediction Formula" effects have been removed from EH(CDMS/Expt) values.

Here we give a quick update on the ΔA/B/C, D_K and $\delta(D_K)$ analysis, using the EH(Ames) terms from SPFIT least-squares fits at 1E-7 cm^{-1} accuracy level.

For rotational constants, Fig.24a plots the linear approximations of O isotope effects on the Δ(Ames – Expt/CDMS) vs. M=sum(1/m_O). It uses original, published, experimental or CDMS data of 626, 627, 628 and 828. '_f' means some higher order terms of minor isotopologues were "fixed" at their 626 values. '_p' means most higher order terms of 828 EH(Expt) were fixed at values predicted by Ulenikov formula. Ranges of the linear fit residuals are: -0.011 ~ 0.018 MHz (A), -0.007 ~ 0.005 MHz (B), -0.003 ~ 0.002 MHz (C). The "Fixed-Relaxed" effects are on the order of 0.0001-0.0005 MHz so can be safely neglected. The "Prediction effects" on 828 rotational constants are about one order of magnitude larger: 0.005-0.011 MHz (A), 0.001-0.002 MHz (B), and 0.003-0.004 MHz (C). Interested readers may incorporate them as adjustments on the Expt/CDSD constant values and re-do the linear fits, but fitting quality will remain similar.

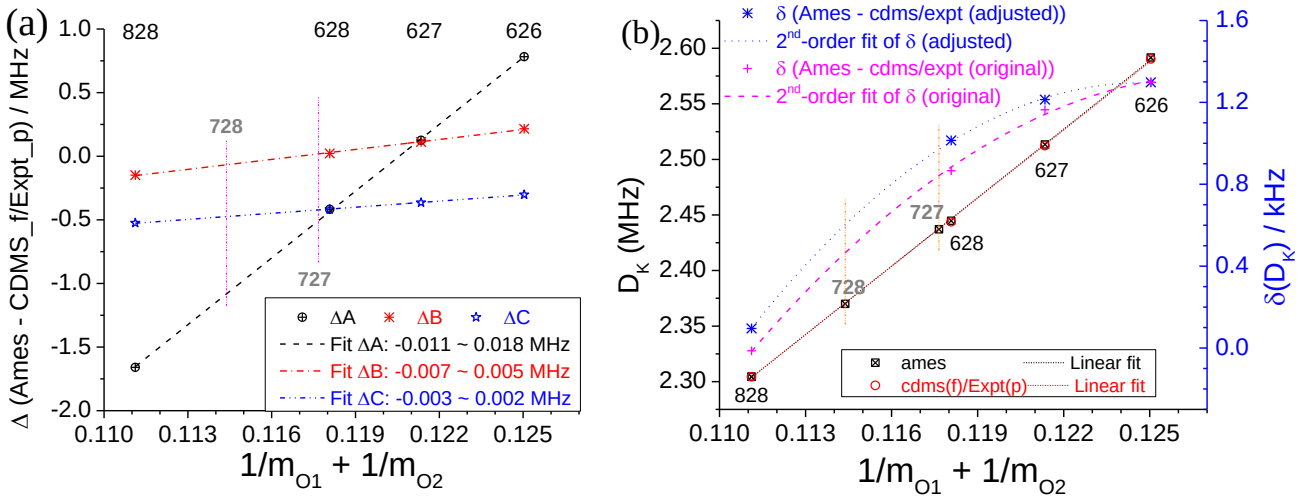


Fig.24. The corrected version of O isotope effects: 626, 627, 628 and 828. (a) the absolute values and the linear fits of ΔA/ΔB/ΔC; (b) the linear and 2nd order fits of D_K values (left), and the 2nd order fits of the δ(D_K) = D_K(Ames) – D_K(CDMS/Expt) differences with and without "fixed" and "prediction" adjustments. Use EH(Ames) constants fit at 1E-7 cm^{-1} accuracy level.

In Fig.24b, the original EH(Ames) D_K values are plotted for 626, 627, 628, 727, 728 and 828, and the EH(CDMS/Expt) D_K values are plotted for 626, 627, 628 and 828. In D_K scale, their symbols and linear approximations overlap with each other (Left Axis, squares and circles). Note the CDMS constants for 627 and 628 have "Fixed-Relaxed" shifts 0.00002 – 0.00014 MHz, and Expt D_K (828) has the "Prediction-effect" shifts 0.0001-0.0002 MHz. Such 0.1-0.2 kHz differences can be only distinguished by plotting the δ(D_K) with and without these shifts. See the blue crosses (dotted curve) and cyan crosses (dashed curve) for the Right axis in Fig.24b, which represent the adjusted and original D_K. Now we can see the O isotope effect on δ(D_K) is *not* linear, but quadratic. With the adjustments made based on the "Fixed vs Relaxed" and "Prediction effect" shifts, the quadratic fit (blue dotted curve) residuals are less than those of original δ(D_K) fit (cyan dashed curve) at 627 and 628. It suggests that higher order polynomials be more appropriate for *some* sextic and higher order EH constants, instead of linear or 2nd-oder approximations. This could be another problem of Fig.17a: inappropriate (too simple) approximation was used for those higher order terms. If one wants to determine the best approximation or prediction, he/she needs to get first reliable, accurate higher order EH constants for minor isotopologues with small enough uncertainty or noises. Various uncertainties and effects do have discernible impacts on the isotopologue consistency and prediction accuracy of EH(Ames) constant refinements. Need more studies to estimate the applicable range of Ames – Expt adjustments.

Finally, we can re-state the claim with strong confidence: (1) EH(Ames) A/B/C constants can be refined to 0.01-0.02 MHz (A) or < 0.01 MHz (B and C); (2) with quartic constants accurately computed or refined, line position predictions may reach δ deviations as small as 0-5 MHz for low K_a/J transitions; (3) such accuracy enables our predictions to be

helpful in astronomical observations, laboratory identification, and rovibrational IR analysis or simulations.

Up to now, the most uncertain factor in the whole procedure is the A_0. Good news is that the B and C predictions have been accurate and stable enough. They can always facilitate future spectra analysis. In Table 1, we give the summary of available $A_0/B_0/C_0/D_K/D_{JK}$ constants for 30 isotopologues, either from EH(CDMS/Expt) models or the corrected EH(Ames) values. Partition function values are given for J=0-75 and E'<8000 cm^{-1} above the zero-point. Partition sums at higher temperature will be reported later. In our recent test, the SO_2 666 line positions predicted by SPCAT program using the refined EH(Ames) parameters nicely match the 29 transitions in old microwave data table [64], with $|\delta|<2$MHz and mean $\pm \sigma(\delta) = -0.06 \pm 0.69$ MHz.

Table 1. Effective Rotational Constants $A_0/B_0/C_0$, and D_J/D_{JK} of $^{32-36}S$ $^{16-18}O$ isotopologues (in MHz). Partition sums (J=0-75) at 296K and estimated abundances are included for reference.

Iso.	Abundance estimate	Partition Sum[&] (296K)	A_0	B_0	C_0	D_K	D_{JK}
626[*]	0.94469	6333.93	60778.5500	10318.0734	8799.7032	2.590253	-0.116956
646[#]	0.04269	6459.66	58991.1824	10318.5098	8761.3024	2.440180	-0.111658
636[#]	0.00756	6398.13	59856.4785	10318.2981	8780.1393	2.512337	-0.114217
628[#]	0.00194	13674.02	59101.1801	9724.6449	8331.5620	2.443872	-0.108416
627[#]	3.59857E-4	13220.11	59883.7281	10008.2142	8555.1366	2.512272	-0.112540
666	1.99029E-4	6575.13	57399.6285	10318.9029	8725.3873	2.309894	-0.106985
648	8.77313E-5	13947.81	57314.7696	9724.3248	8294.6166	2.298276	-0.103355
647	1.62624E-5	13441.30	58096.6792	10008.4380	8517.5949	2.364584	-0.107372
638	1.55421E-5	13813.31	58179.5953	9724.4861	8312.7475	2.368265	-0.105800
828[**]	3.98943E-6	7400.23	57384.9709	9170.5065	7889.5737	2.304325	-0.101258
637	2.88098E-6	13313.75	58961.8190	10008.3310	8536.0146	2.435586	-0.109868
728	7.39505E-7	14228.99	58185.8433	9435.5193	8100.7771	2.369558	-0.104752
668	4.09004E-7	14205.36	55724.1172	9723.9961	8260.0052	2.171963	-0.098889
848	1.80287E-7	7533.83	55598.0432	9170.9104	7854.9110	2.163273	-0.096479
727	1.37079E-7	246851.40	58978.0583	9709.2957	8317.8314	2.436257	-0.108533
667	7.5815E-8	13685.57	56505.4229	10008.6287	8482.4529	2.236416	-0.102813
748	3.3419E-8	14517.69	56398.9779	9435.7636	8065.1232	2.226383	-0.099846
838	3.1939E-8	7515.96	56463.1261	9170.7143	7871.9297	2.231067	-0.098787
747	6.195E-9	251744.67	57190.9218	9709.7152	8281.3749	2.290973	-0.103508
738	5.92E-9	14375.91	57264.0291	9435.6464	8082.6246	2.295202	-0.102215
737	1.097E-9	249340.16	58056.1059	9709.5116	8299.2664	2.360813	-0.105935
868	8.41E-10	7677.30	54006.8831	9171.2747	7822.3799	2.041003	-0.092264
768	1.56E-10	14789.58	54807.8837	9435.9747	8031.6904	2.102224	-0.095518
767	2.9E-11	256409.58	55599.5746	9710.0933	8247.2177	2.164941	-0.099077
656	Trace	6517.97	58171.4193	10318.7117	8743.0138	2.372679	-0.109246
658	<1.d-13	14078.82	56495.4646	9724.1610	8276.9988	2.232822	-0.101050
657	<1.d-14	13559.03	57277.0676	10008.5373	8499.7037	2.298175	-0.105019
858	<1.d-15	7607.16	54778.4827	9171.0975	7838.3592	2.099904	-0.094303
758	<1.d-16	14655.93	55579.4501	9435.8732	8048.1091	2.162040	-0.097612
757	<1.d-17	254151.76	56371.2651	9709.9094	8263.9888	2.225666	-0.101221

[&] nuclei spin degeneracies are excluded from $^{33/35}S$ (×4) and ^{17}O (×6 for partially substituted, or ×15,21 for fully substituted) related isotopologues.

[*] Ulenikov et al (2013) JQSRT, 130, 220–232. See Ref.[45].

[#] CDMS values are taken from online data pages, see Ref.[44]

[**] Ulenikov *et al.* (2015), JQSRT, 166, 13–22. See Ref.[49]

4.4 How are we different from other prediction approaches?

The idea of BTRHE strategy is old. The applications of BTRHE like strategy are abound. The combination of variationally computed IR line lists with Effective Hamiltonian (EH) models is not new, either. First-principle calculations of EH constants through the 2nd-order Vibrational Perturbation Theory (VPT2) have existed for decades. Using experimentally determined EH constants to improve *ab initio* EH constants has also been a regular exercise. What make this study unique?

Answers lie in the accuracy level at our starting point, the accuracy level set in our goal, and the systematic investigation we have carried out.

For more than 20 years, our group at NASA Ames has been active in the field of molecular VPT2 analysis using high quality *ab initio* Quartic Force Field (QFF).[65–68] As far as we know, the ***best*** $A_0/B_0/C_0$ agreements scientists reported between the computed and experimental values are δ% = 0.01-0.10%. For example, the best *c*-$C_3H_3^+$ predictions we published in 2011 [69] was verified by 2014 experiment [70] with δ(B)=9 MHz and δ(C)=4 MHz, or δ% = 0.025%. It matches the statistical analysis of Puzzarini, Stanton and Gauss given in 2008 [71] and 2010 review [72]. By including basis set extrapolation, core correlation, triples and quadruples excitation correction, vibrational correction, and electronic contribution etc., their best mean B_0 deviations are 0.04-0.06%, see Tables 6-7 of their 2010 review [72]. Core-correlation effect plays the primary role in the accuracy improvements. This means better than 10 MHz for constants 10 GHz (or 0.3 cm^{-1}), and better than 100 MHz for constants ~ 100 GHz (or 3 cm^{-1}). The "surprisingly small" δs reported in their Table 8 for HSOH isotopologues [72] are less than 70 MHz for *A* (~200 GHz) and a few MHz for *B* and *C* (13-15 GHz). Such state-of-the-art accuracy surely can "guide experiment and in particular help to restrict the searches to narrow frequency ranges".

For $^{32}S^{16}O_2$, such accuracy equals to 5-60 MHz (A_0), 1-10 MHz (B_0 and C_0). Look at Fig.16 and Fig.24a, the A/B/C accuracy of *original* EH(Ames) fitted from Ames-296K MW line lists is -2~1 MHz for A (~60 GHz), -0.5 ~ 0.5 MHz for B and C (~10 GHz). In other words, all δ% are within ±0.003-0.005% for all 30 isotopologues. This is *already* higher than the "state-of-the-art" *ab initio* accuracy by *one order of magnitude*. It is the achievement of the BTRHE strategy refining high quality *ab initio* PES using $^{32}S^{16}O_2$ experimentally determined rovibrational energy levels. But it is just our *starting point*.

In this work, the goal of our A/B/C (and other constants) refinement through 2nd-order implementation of BTRHE strategy is, to further improve the prediction accuracy by *two more* orders of magnitude. For example, δ% was reduced from ±0.003-0.005% (Fig.16) to ±0.00002-0.00007% (Fig.24a). More importantly, the accuracy should be consistent throughout all 30 isotopologues. We do not know *any* previous work ever pushing the accuracy & consistency to this level. The reason is simple: such kind of work requires quantum exact rovibrational calculations (within Born-Oppenheimer approximations or plus nonadiabatic correction and higher order corrections) on semi-empirically refined PES, and enough EH(Expt) work on a few isotopologues. Without a highly accurate PES, or just going with regular VPT2 analysis, the starting deviations will be *at least* one order of magnitude larger. Without quantum exact calculations, the defects would introduce noises (i.e. inconsistencies) to IR line lists and EH(Ames) parameters. Without enough EH(Expt) studies, the systematic improvements for both S and O isotope substitutions will not be possible. In addition, without careful investigations on various coordinates, convergence, uncertainties, and minor effects, etc., we cannot properly estimate the limitation, accuracy and uncertainties of improvements.

The deviations of best VPT2 predictions are a few percent for quartic centrifugal distortion constants. For example, the δ% of $^{33}SO_2$ quartic terms reported by Gauss and Puzzarini [73] are -12% (D_K), 6% (D_{JK}), -2% (D_J), -2% (d_J), and -15% (d_K), with respect to EH(CDMS). The corresponding δ% of *original* 636 EH(Ames) terms are: 0.046% (D_K), 0.001% (D_{JK}),

-0.043% (D_J), 0.012% (d_J), and -0.096% (d_K). See Figs.19b and 19c. Again, our starting accuracy level is 2-3 orders higher. Partially this is because the vibrational average correction of quartic centrifugal distortion constants is not available in most VPT2 programs yet. Then, after refinement, the D_K quadratic fit in Fig.24b has reached 0.01-0.02 kHz accuracy for all SO_2 isotopologues, or δ% ~ 0.0004%. Again, this is the *two-orders* of magnitude improvement. The EH(Ames) D_{JK} accuracy is so close to EH(CDMS/Expt), no need to do extra refinements.

Different accuracy levels have decided we need different approaches to improve the predictions.

When the δ% are *not* very small, a common approach is to scale the calculated EH(iso) constants by a factor (constant) determined from the experimental and computational information of main isotopologue:

$$X^{ISO}_{scaled} = X^{ISO}_{Calc} \times (X^{Main}_{Expt} / X^{Main}_{Calc})$$ where "Main"/"ISO" refers to the main/minor isotopologue.

For example, Cazzoli, Puzzarini and Gauss [74,75] used this formula to predict $H_2{}^{33\text{-}36}S$ spectroscopic constants, with $H_2{}^{32}S$ scaling factors: 0.998275164 (A), 1.002117637 (B), 0.999583165 (C), 1.07633826 (D_K), and 1.05794646 (D_{JK}). For $H_2{}^{33\text{-}34}S$, they reported δ% = 0.0001% accuracy for A/B/C, and 0.001-0.01% for quartic constants. After scaling, corresponding δ(A/B/C) are reduced from 536/571/59 MHz to 0.43 / 0.008 / 0.13 MHz for $H_2{}^{33}S$, or from 536/571/59 MHz to 0.87 / 0.02 / 0.27 MHz for $H_2{}^{34}S$. Compare to $H_2{}^{32}S$, the relative changes on A/B/C constants are only 0.35% / 0.00027% / 0.161% for $H_2{}^{34}S$, or 0.64% / 0.00051% / 0.305% for $H_2{}^{36}S$. The δ(A/B/C) of $H_2{}^{36}S$ are 1.54 / 0.005 / 0.44 MHz. The scaling algorithm treats δ% as a constant, this works perfectly when the δ% isotopic changes are really negligible. The δ(B) is such a case, because from $H_2{}^{32}S$ to $H_2{}^{36}S$, B_{Calc} and B_{Expt} only change by 1 MHz, or 0.0004%. After-scaling δs also reached the range of fitting uncertainty, ~0.01 MHz.

But if δ% is not a constant among isotopologues, after-scaling δ and δ% will have noticeable mass-dependence. This is what happened to δ(A) and δ(C) in Ref.[75]. *After* scaling, δ still follows simple mass relations like those we showed in Fig.16b. Along $^{33}S - {}^{34}S - {}^{36}S$, δ(A) = 0.43 – 0.87 – 1.54 MHz, and δ(C) = 0.13 – 0.27 – 0.44 MHz. The δ magnitude is comparable to the δ(A/B/C) of original SO_2 EH(Ames) in Fig.16. The relative δ% is between our starting accuracy and post-refinement accuracy. The goal of our BTRHE implementation is set to minimize these residuals by approximating and removing the mass-dependent part. This is a major difference: we track the isotope-related δ and δ% changes from/to a higher accuracy level, but the scaling algorithm treats δ% as a constant. Furthermore, in the case of quartic centrifugal distortion constant D_K, Fig.24b has proved that a 2nd-order fit is more appropriate than linear fit for δ(D_K). This treatment is two orders more accurate, since the scaling algorithm is just a 0th-order approximation.

More convincing examples are the δ(A) in Fig.16b, 16c, and the δ(B) in Fig.16c. Upon S or O isotope substitutions, they change from positive to negative. Scaling factor does not work for the δD_K either, see Fig.24b. Under such circumstances, the over-simplified scaling algorithm would introduce unnecessary errors to δ, instead of any improvements. This is easy to understand, but not many theoretical EH models can have such accuracy.

For $HD^{34}S$ and $D_2{}^{34}S$, the scaling factors need be determined from new parent species, $HD^{32}S$ and $D_2{}^{32}S$, instead of $H_2{}^{32}S$. This is because deuterations significant change the constants (and symmetry). Note the $HD^{34}S$ EH(Expt) quoted in Ref.[75] were unreliable, should use CDMS data.[76] In contrast, our simple linear approximations in Fig.16c and Fig.24a cover both symmetric and asymmetric isotopologues, from 626, 627, 628 to 727, 728 and 828, with respect to a single *M* coordinate. We have tried the scaling algorithm on the SO_2 *B* constants. It is found that, to ensure consistent ~0.01 MHz accuracy, we need 6 scaling factors for the 6x6, 6x7, 6x8, 7x7, 7x8 and 8x8 isotopologue series, where *x*=2-6 for $^{32\text{-}36}S$. Otherwise δ(B) will be as large as 0.1-0.3 MHz. Even with appropriate

scaling factors, δ(B) will be still mass-dependent. In contrast, our O isotope approximations in Fig.24a is essentially equivalent to finding the connections among 6 scaling factors. In other words, we treat both S and O effects together, while the scaling factor algorithm could only focus on one of them. Usually it is the one with smaller effects, for better prediction accuracy.

The scaling factor method for EH(iso) improvement is somehow parallel to the 1st-order BTRHE approach behind semi-empirical IR line lists. Both algorithms use experimental data (of main isotopologue, in many cases) to improve the accuracy of isotopologue rovibrational data. Both are very successful, both find *A* harder to improve than *B*/*C*, but both leave mass-dependent δ and δ%. The approach we propose in this work focuses on how to minimize δ and δ% by finding and approximating the pattern of the isotope/mass-dependence in residuals. Therefore, it is a 2nd-order implementation of BTRHE strategy. It is not really an appropriate apple-to-apple comparison if this study is compared to the scaling method.

Theoretically, a more advanced method is to predict minor isotopologue EH constants from EH(Expt) constants of main isotopologue. One example is the formula adopted by Ulenikov et al [49] in their SO_2 828 EH(Expt) analysis. We have examined its requirement and performance in Sec.4.2.8, and concluded that the deviations on D_K and D_{JK} are large. Our strategy can further improve their formula based predictions.

5. Conclusion and Future work

We report systematic and quantitative exploration for the isotopologue consistency of Ames-216 CO_2 IR lists and Ames-296K SO_2 IR lists. For intensity, the goals of our consistency analysis are: (a) to demonstrate the consistency level of IR line lists generated by BTRHE strategy and alike; (b) to lay down a solid base for future calibration using a few measured intensity data to improve the rare isotopologue intensity predictions. We found both Einstein A_{21} and TDM (sum of transition dipole element squares) are good indicators of isotopic effects. Both energy (cm^{-1}) and *M*=sum(1/m) coordinates are reasonable choices of horizontal axis for a specific transition. *M* is the best, or most universally applicable choice in all mass-related coordinates. The isotopologue consistency allows us to make intensity predictions with δ% less than ±0.3% (linear) or ±0.02% (quadratic). However, it must be repeated that intensity approximations are not our goal. They are just useful visualizations for *part* of the pattern (consistency). A derived coordinate, $TDM_{ISO}/TDM_{626(main)}$ can help us track the isotope effect variations along J, K, energy, or mass, etc. Relative TDM changes from one isotopologue to another may be as large as several hundred percent.

For CO_2, the linear bending and symmetric stretch vibrations have opposite isotope effects. As a result, the TDM and A_{21} patterns in a vibrational polyad have predictable transition from the first band to the last band. Combined with non-polyad A_{21} or TDM patterns, we quantitatively conclude the isotopologue consistency among Ames-2016 CO_2 lists exceeds the consistency level of most experimental intensity data and EDM models. In the coming era of high accurate IR intensity, the consistency may also help scientists determine other rare isotopologue intensity with similar accuracy, but without real measurements in lab. Based on the intensity consistency, we hope a few 628 transitions can be measured for 50006-00001 and 60007-00001 bands. It will help elucidate the discrepancy between CDSD and Ames band origin (G_v) extrapolations.

For SO_2, we focus on the comparison between EDM intensities and variationally computed Ames-296K intensities. In several tests, the TDM_{ISO}/TDM_{626} ratios have generally good agreement between the two sets of intensity. On the strongest MW lines, $J+1_{J+1,0} \leftarrow J_{J,0}$, heavier isotopes lead to noticeable intensity reductions. The reduction is

approximately proportional to the increment of total molecular mass. At $J=K_a=45$, the reductions are up to 2%, while EH(Ames)-EDM based TDM_{ISO}/TDM_{626} ratios are converged to 1. See Fig.12b,c. Compare to EH(Ames)-EDM, the TDM ratios of Ames-296K IR list have more higher order effects in high K_a range, e.g. K_a>10-20. Current SO_2 MW EDM models use single effective dipole term. Ames MW(IR) intensities (including A_{21} and TDM) can give valuable guidance on future EDM upgrades.

For line positions, we demonstrate the power of a 2nd-order implementation of BTRHE strategy. The semi-empirically computed IR line lists are fit to the same EH model adopted in experimental analysis. The fitted EH(Ames) constants may be refined with available EH(CDMS/Expt). The mass/isotope dependence of δ is minimized. The δ of rotational constants are reduced from ~1 MHz to 0.01-0.02 MHz, or even less. Refined *lower* order EH terms, including A/B/C and quartic centrifugal distortion constants, can provide δ = 0 – 5 MHz for J<30, K<10-20 transitions. The whole procedure is easy to follow. But several factors play important roles and need be carefully addressed.

(1) all the MW/IR list data should be generated in the most consistent way. They should carry line position precision 1e-7 cm^{-1} (or better). EH constants, A_{21} and TDM patterns are highly sensitive to small inconsistencies or errors in experimental data, SPFIT analysis, or theoretical calculations. We recommend 100% equivalent line set and uniform EH model/weight for all isotopologues.

(2) Correlation among EH constants may lead to unwanted uncertainties. In the SO_2 case involving thousands of transitions, only the SPFIT analysis at 1E-6 cm^{-1} ~ 1E-7 cm^{-1} level can converge *A* constant better than 0.005 ~ 0.010 MHz, and D_K better than ~0.1 kHz. Both are required for our target accuracy.

(3) K_C-split check suggests that part of Ames line list data may contain numerical noise up to 1E-7~1E-6 cm^{-1}. SPFIT analyses at 1E-7 cm^{-1} level are expected more reliable in lower K_a/J region. "Constant uncertainty" (convergence defects) is much more significant than "fitting uncertainty" (least-squares solutions). In our SO_2 tests, it is 0.0001% ~ 0.1% for quartic terms, and 1~100% for higher order terms.

(4) *M*=sum(1/*m*) coordinate is good for isotope effects on *A/B/C* and D_K, while not all quartic constants have recognizable pattern for its δ (Calc – Expt) vs. *M*. *m* is the atom mass.

(5) Both "Fixed" effects and "Prediction Formula" effects are real and may cause changes on EH constants. They should be identified and removed before final refinement and predictions.

(6) Second or higher order fits may be more appropriate for the δ of quartic and *higher* order terms. See D_K example in Fig.24b. Not all *higher* order terms are determined with adequate data. But most line position errors associated with EH(Ames) higher order terms are 0-3 MHz for $K_a \le$ 10-15 and J≤30.

We have carefully discussed the differences between this 2nd-order BTRHE implementation and the regular scaling factor algorithm that many colleagues have used. The highly consistent EH(Ames) parameter sets are also utilized to check the accuracy level of the $\alpha_A\alpha_C$ dependent formula that Ulenikov et al adopted in their SO_2 828 study. In short, we present a higher order refinement procedure starting with higher accuracy and ending with the best prediction accuracy but without isotope/mass dependency.

Such unprecedented, consistent accuracy has gone beyond the point to "guide experiment and in particular help to restrict the searches to narrow frequency ranges". It can provide reference EH constants (lower order terms) closer to *real* spectroscopic accuracy. The semi-empirically refined, highly accurate global PES and the quantum exact rovibrational computations are the pre-requisites for the isotopologue consistency among IR list data and the two orders of magnitude improvement. The 2nd-order BTRHE implementation can be easily extended to other pure-rotational hot bands, or lowest rovibrational bands. We have been working on the SO_2 vibrational fundamentals, first overtones, and combination bands, i.e. vibrational quanta ≤ 2. It will be reported later in due course. We expect

more mutual beneficial interaction and collaboration between experimentalists and theoreticians.

Based on the consistent Ames-296K IR line lists, SPFIT/SPCAT EH analysis, linear or quadratic approximation procedures discussed in this study, we have generated a microwave dataset including 644,636 GS←GS pure rotational transitions for all 30 SO_2 isotopologues, with reliable, consistent Ames-296K line list intensity (100% abundance) > 1E-32 cm/molecule, and Einstein A_{21} coefficients. They are generated by SPCAT using the "corrected" or "refined" lower order EH(Ames) terms, plus the original higher order EH(Ames) terms. Here is the procedure:

(1) The Asym line lists reported in 4.2.3 are fit by SPFIT with the CDMS EH model (4.2.1), at 1E-7 cm^{-1} uncertainty level. The A_0 and D_K convergence are guaranteed in all 30 sets of EH(Ames) constants. (see 4.2.2). The best available EH(CDMS/Expt) models for 626, 636, 646, 627, 628 and 828 (see Fig.21b in 4.2.6) are used to compute the $\Delta = X_{Ames} - X_{CDMS/Expt}$.

(2) Use linear approximation for $\Delta(A/B/C)$ and quadratic approximation for Δ of quartic terms along the M=sum(1/m) coordinate. For example:

Δ(6x6) - Δ(626) = k * (M(6x6) – M(626)), k is the slope of fit on Δ(626), Δ(636) and Δ(646).

Couplings between S and O isotope effects on $\Delta(A/B/C)$ are assumed negligible. We do not use Ulenikov's formula since we have proven its D_K accuracy is not as good as our BTRHE implementation, see 4.2.8 and 4.3.

(3) The $\Delta(A/B/C)$ of both 656 and 666 are predicted from the linear extrapolation along $\Delta(A/B/C)$ of 626→636→646. Then the predicted B/C of 666 are taken as "accurate" and fixed in a SPFIT on the experimental MW line set [ref]. In this way, we successfully determine the experimental A and $\Delta(A)$ for 666. The new $\Delta(A)$ and the linearly extrapolated $\Delta(A)$ agree to better than 0.0015 MHz. This confirms the S isotope effects have been accurately modeled. The $\Delta(A/B/C)$ of 727 and 728 are interpolated between those of 628 and 828, using the O isotope effects averaged from 627→628→828. The averaged O effect (per M unit) is defined as

$$y = \{ [\Delta(627)-\Delta(626)]/[M(627) - M(626)] + 2*[\Delta(628)-\Delta(626)]/[M(628) - M(626)] + 4[\Delta(828)-\Delta(626)]/[M(828) - M(626)] \} / 7.$$

Then the Δ(727) and Δ(728) are computed as:

$$\Delta(727) - \Delta(626) = \{ [\Delta(627)-\Delta(626)+y*(M(727) - M(627))] + [\Delta(828)-\Delta(626)-y*(M(828) - M(727))] + [y*(M(727) - M(626))] \}/3$$

$$\Delta(728) - \Delta(626) = \{ [\Delta(628)-\Delta(626)+y*(M(728) - M(628))] + [\Delta(828)-\Delta(626)-y*(M(828) - M(728))] + [y*(M(728) - M(626))] \}/3$$

(4) The quadratic approximation takes the M (626) and Δ(626) as zero point. For example,

$$\Delta(6x6) - \Delta(626) = a\delta_M^2 + b\delta_M , \qquad \delta_M = M(6x6) - M(626)$$

The coefficients a and b for S isotope effects can be numerically determined from Δ(636) - Δ(626) and Δ(646) - Δ(626). For O isotope effects, two sets of a and b coefficients are computed from 626-627-628 and 626-628-828, respectively. Their weighted sum (2:1) are taken as final a and b. Every EH quartic term has two sets of a,b, one set for S isotope effects, the other set for O isotope effects.

(5) The final set of EH(Ames) model include "refined" or "corrected" *lower* order terms + original EH(Ames) *higher* order terms. These refined EH(Ames) parameters can be found in Supplementary file. The SPCAT program reads the parameters and generates the MW line lists for 30 isotopologues. All EDM intensity is replaced by the corresponding Ames-296K line list intensities. The superior consistency has been demonstrated in the first part of this paper.

As presented in sections 4.1 and 4.2.6, the general prediction accuracy is expected to be better than 5 MHz for most line positions in the range of J<20 and K_a<10-15. It is also possible to reach σ_{RMS}<1-2 MHz for J<10-15 *and* K_a<10. The MW line positions of 626, 636, 646, 627, 628 and 828 are replaced by original EH(CDMS/Expt) model based line positions. Please note that we choose not to include the "Fixed – Relaxed" correction or the "Prediction Formula" correction in the line list generation procedure (see section 4.2.7 and 4.2.8). This choice seems conflicting with the central idea of this paper. As shown in Fig.24b, the D_K predicted for 727 and 728 may be off by 0.00013 – 0.00015 MHz, this will lead to line position deviation 1.3-1.5 MHz at K_a=10. They are noticeable, but not very large. It is interesting to observe, without these two corrections, how good / bad our predictions will compare to future experimental data.

Interested readers can download this line set, plus the refined EH(Ames) model parameters, from supplementary files, or from Ames Molecular Spectroscopic Data for Astrophysical and Atmospheric Studies, http://huang.seti.org.

Acknowledgement

DWS, TJL, and XH gratefully acknowledge the financial support from the NASA 16-PDART16_2-0080 grant, 13-PATM12-0012 grant, and early stage support from NASA 12-APRA1-0107 grant and Venus Express Supporting Investigator Program. Huang acknowledges the NASA/SETI Co-operative Agreement NNX15AF45A and NNX17AL02G. We sincerely thank helpful discussions with Dr. Holger Müller (Cologne), Dr. Michael McCarthy (Harvard CfA), Dr. Shanshan Yu (JPL), Dr. Keeyoon Sung (JPL), and Dr. Iouli Gordon (Harvard CfA).

References:

[1] Tennyson J, Yurchenko SN, Al-Refaie AF, Barton EJ, Chubb KL, Coles PA, et al. *The ExoMol database: Molecular line lists for exoplanet and other hot atmospheres.* J Mol Spectrosc 2016;327:73–94. doi:10.1016/j.jms.2016.05.002.

[2] Rey M, Nikitin A V., Babikov YL, Tyuterev VG. *TheoReTS – An information system for theoretical spectra based on variational predictions from molecular potential energy and dipole moment surfaces*. J Mol Spectrosc 2016;327:138–58. doi:10.1016/j.jms.2016.04.006.

[3] Huang X, Schwenke DW, Lee TJ. *Ames Molecular Spectroscopic Data For Astrophysical And Atmospheric Studies*. See http://huang.seti.org.

[4] Schwenke DW. Beyond the Potential Energy Surface: *Ab initio* Corrections to the Born−Oppenheimer Approximation for H_2O. J Phys Chem A 2001;105:2352–60. doi:10.1021/jp0032513.

[5] Huang X, Schwenke DW, Lee TJ. *Rovibrational spectra of ammonia. I. Unprecedented accuracy of a potential energy surface used with nonadiabatic corrections.* J Chem Phys 2011;134:044320. doi:10.1063/1.3541351.

[6] Polyansky OL, Csaszar AG, Shirin S V., Zobov NF, Barletta P, Tennyson J, et al. *High-Accuracy ab Initio Rotation-Vibration Transitions for Water*. Science 2003;299:539–42. doi:10.1126/science.1079558.

[7] Partridge H, Schwenke D. *The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data.* J Chem Phys 1997;106:4618–39. doi:10.1063/1.473987.

[8] Huang X, Schwenke DW, Lee TJ. *An accurate global potential energy surface, dipole moment surface, and rovibrational frequencies for NH_3.* J Chem Phys 2008;129:214304. doi:10.1063/1.3025885.

[9] Huang X, Schwenke DW, Lee TJ. *Highly accurate potential energy surface, dipole moment surface, rovibrational energy levels, and infrared line list for $^{32}S^{16}O_2$ up to 8000 cm^{-1}*. J Chem Phys 2014;140:114311. doi:10.1063/1.4868327.

[10] Huang X, Schwenke DW, Lee TJ. *Empirical infrared line lists for five SO_2 isotopologues: $^{32/33/34/36}S^{16}O_2$ and $^{32}S^{18}O_2$*. J Mol Spectrosc 2015;311:19–24. doi:10.1016/j.jms.2015.01.010.

[11] Huang X, Schwenke DW, Lee TJ. *Ames $^{32}S^{16}O^{18}O$ line list for high-resolution experimental IR analysis*. J Mol Spectrosc 2016;330:101–11. doi:10.1016/j.jms.2016.08.013.

[12] Huang (黄新川) X, Schwenke DW, Freedman RS, Lee TJ. *Ames-2016 line lists for 13 isotopologues of CO_2: Updates, consistency, and remaining issues*. J Quant Spectrosc Radiat Transf 2017;203:224–41. doi:10.1016/j.jqsrt.2017.04.026.

[13] Huang X, Freedman RS, Tashkun SA, Schwenke DW, Lee TJ. *Semi-empirical $^{12}C^{16}O_2$ IR line lists for simulations up to 1500K and 20,000 cm^{-1}.* J Quant Spectrosc Radiat Transf 2013;130:134–46. doi:10.1016/j.jqsrt.2013.05.018.

[14] Huang (黄新川) X, Gamache RR, Freedman RS, Schwenke DW, Lee TJ. *Reliable infrared line lists for 13 CO_2 isotopologues up to E'=18,000 cm^{-1} and 1500 K, with line shape parameters*. J Quant Spectrosc Radiat Transf 2014;147:134–44. doi:10.1016/j.jqsrt.2014.05.015.

[15] Huang X, Schwenke DW, Tashkun SA, Lee TJ. *An isotopic-independent highly accurate potential energy surface for CO_2 isotopologues and an initial $^{12}C^{16}O_2$ infrared line list*. J Chem Phys 2012;136:124311. doi:10.1063/1.3697540.

[16] Pickett HM. *The fitting and prediction of vibration-rotation spectra with spin interactions*. J Mol Spectrosc 1991;148:371–7. doi:10.1016/0022-2852(91)90393-O.

[17] Novick SE. *A beginner's guide to Pickett's SPCAT/SPFIT*. J Mol Spectrosc 2016;329:1–7. doi:10.1016/j.jms.2016.08.015.

[18] Schwenke DW. *First principles prediction of isotopic shifts in H_2O*. J Chem Phys 2003;118:6898–904. doi:10.1063/1.1561053.

[19] Huang X, Schwenke DW, Lee TJ. *Rovibrational spectra of ammonia. II. Detailed analysis, comparison, and prediction of spectroscopic assignments for $^{14}NH_3$, $^{15}NH_3$, and $^{14}ND_3$*. J Chem Phys 2011;134:044321. doi:10.1063/1.3541352.

[20] Sung K, Brown LR, Huang X, Schwenke DW, Lee TJ, Coy SL, et al. *Extended line positions, intensities, empirical lower state energies and quantum assignments of NH_3 from 6300 to 7000 cm^{-1}*. J Quant Spectrosc Radiat Transf 2012;113:1066–83. doi:10.1016/j.jqsrt.2012.02.037.

[21] Gordon IE, Rothman LS, Hill C, Kochanov RV, Tan Y, Bernath PF, et al. *The HITRAN2016 molecular spectroscopic database*. J Quant Spectrosc Radiat Transf 2017;203:3–69. doi:10.1016/j.jqsrt.2017.06.038.

[22] Hill C, Gordon IE, Kochanov R V., Barrett L, Wilzewski JS, Rothman LS. *HITRANonline: An online interface and the flexible representation of spectroscopic data in the HITRAN database*. J Quant Spectrosc Radiat Transf 2016;177. doi:10.1016/j.jqsrt.2015.12.012.

[23] Müller HSP, Thorwirth S, Roth DA, Winnewisser G. *The Cologne Database for Molecular Spectroscopy, CDMS*. Astron Astrophys 2001;370:L49–52. doi:10.1051/0004-6361:20010367.

[24] Müller HSP, Schlöder F, Stutzki J, Winnewisser G. *The Cologne Database for Molecular Spectroscopy, CDMS: a useful tool for astronomers and spectroscopists*. J Mol Struct 2005;742:215–27. doi:10.1016/j.molstruc.2005.01.027.

[25] Endres CP, Schlemmer S, Schilke P, Stutzki J, Müller HSP. *The Cologne Database for Molecular Spectroscopy, CDMS, in the Virtual Atomic and Molecular Data Centre, VAMDC*. J Mol Spectrosc 2016;327:95–104. doi:10.1016/j.jms.2016.03.005. See more at https://cdms.astro.uni-koeln.de.

[26] Petrova TM, Solodov AM, Solodov AA, Lyulin OM, Tashkun SA, Perevalov VI. *Measurements of $^{12}C^{16}O_2$ line parameters in the 8790–8860, 9340–9650 and 11,430–11,505 cm^{-1} wavenumber regions by means of Fourier transform spectroscopy.* J Quant Spectrosc Radiat Transf 2013;124:21–7. doi:10.1016/j.jqsrt.2013.03.017.

[27] Karlovets EV, Kassi S, Tashkun SA, Perevalov VI, Campargue A. *High sensitivity Cavity Ring Down spectroscopy of carbon dioxide in the 1.19–1.26μm region.* J Quant Spectrosc Radiat Transf 2014;144:137–53. doi:10.1016/j.jqsrt.2014.04.001.

[28] Yi H, Liu Q, Gameson L, Fleisher AJ, Hodges JT. *High-accuracy $^{12}C^{16}O_2$ line intensities in the 2 μm wavelength region measured by frequency-stabilized cavity ring-down spectroscopy*. J Quant Spectrosc Radiat Transf 2018;206:367–77. doi:10.1016/j.jqsrt.2017.12.008.

[29] Yurchenko SN. *A theoretical room-temperature line list for $^{15}NH_3$.* J Quant Spectrosc Radiat Transf 2015;152:28–36. doi:10.1016/j.jqsrt.2014.10.023.

[30] Yurchenko SN, Barber RJ, Tennyson J. *A variationally computed line list for hot NH_3*. Mon Not R Astron Soc 2011;413:1828–34. doi:10.1111/j.1365-2966.2011.18261.x.

[31] Al-Refaie AF, Yachmenev A, Tennyson J, Yurchenko SN. *ExoMol line lists – VIII. A variationally computed line list for hot formaldehyde*. Mon Not R Astron Soc 2015;448:1704–14. doi:10.1093/mnras/stv091.

[32] Zak EJ, Tennyson J, Polyansky OL, Lodi L, Zobov NF, Tashkun SA, et al. *Room temperature line lists for CO_2 symmetric isotopologues with ab initio computed intensities.* J Quant Spectrosc Radiat Transf 2017;189. doi:10.1016/j.jqsrt.2016.11.022.

[33] Zak EJ, Tennyson J, Polyansky OL, Lodi L, Zobov NF, Tashkun SA, et al. *Room temperature linelists for CO_2 asymmetric isotopologues with ab initio computed intensities*. J Quant Spectrosc Radiat Transf 2017;203. doi:10.1016/j.jqsrt.2017.01.037.

[34] Zak E, Tennyson J, Polyansky OL, Lodi L, Zobov NF, Tashkun SA, et al. *A room temperature CO_2 line list with ab initio computed intensities*. J Quant Spectrosc Radiat Transf 2015;177:1–12. doi:10.1016/j.jqsrt.2015.12.022.

[35] Polyansky OL, Bielska K, Ghysels M, Lodi L, Zobov NF, Hodges JT, et al. *High-Accuracy CO_2 Line Intensities Determined from Theory and Experiment.* Phys Rev Lett 2015;114:243001. doi:10.1103/PhysRevLett.114.243001.

[36] Benner DC, Devi VM, Sung K, Brown LR, Miller CE, Payne VH, et al. *Line parameters including temperature dependences of air- and self-broadened line shapes of $^{12}C^{16}O_2$: 2.06-μm region*. J Mol Spectrosc 2016;326:21–47. doi:10.1016/j.jms.2016.02.012.

[37] Kang P, Wang J, Liu G-LL, Sun YRR, Zhou Z-YY, Liu A-WW, et al. *Line intensities of the 30011e – 00001e band of $^{12}C^{16}O_2$by laser-locked cavity ring-down spectroscopy.* J Quant Spectrosc Radiat Transf 2018;207:1–7.

doi:10.1016/j.jqsrt.2017.12.013.

[38] Devi VM, Benner DC, Sung K, Brown LR, Crawford TJ, Miller CE, et al. *Line parameters including temperature dependences of self- and air-broadened line shapes of $^{12}C^{16}O_2$: 1.6-μm region*. J Quant Spectrosc Radiat Transf 2016;177:117–44. doi:10.1016/j.jqsrt.2015.12.020.

[39] Karlovets EV, Campargue A, Kassi S, Tashkun SA, Perevalov VI. *Analysis and theoretical modeling of 18 O enriched carbon dioxide spectrum by CRDS near 1.35 μm: (II) 16O13C18O, 16O13C17O,12C18O2, 17O12C18O, 12C17O2, 13C18O2 and 17O13C18O*. J Quant Spectrosc Radiat Transf 2017;191:75–87. doi:10.1016/j.jqsrt.2017.01.038.

[40] Karlovets EV, Čermák P, Mondelain D, Kassi S, Campargue A, Tashkun SA, et al. *Analysis and theoretical modeling of the ^{18}O enriched carbon dioxide spectrum by CRDS near 1.74 μm.* J Quant Spectrosc Radiat Transf 2018;217:73–85. doi:10.1016/j.jqsrt.2018.05.017.

[41] Karlovets EV, Perevalov VI. *Calculation of the carbon dioxide effective dipole moment parameters of the qJ and q2J types for rare isotopologues.* Atmos Ocean Opt 2011;24:101–6.

[42] Karlovets E V., Perevalov VI. *The influence of isotopic substitution on the effective dipole moment parameters of CO_2 molecule*. Opt Spectrosc 2015;119:16–21. doi:10.1134/S0030400X15070139.

[43] Tashkun SA, Perevalov VI, Gamache RR, Lamouroux J. *CDSD-296, high resolution carbon dioxide spectroscopic databank: Version for atmospheric applications.* J Quant Spectrosc Radiat Transf 2015;152:45–73. doi:10.1016/j.jqsrt.2014.10.017.

[44] Müller HSP. Online EH(CDMS) model data at https://cdms.astro.uni-koeln.de/classic/predictions/daten/SO2/: 34SO2/s34.par for 646; SO2/alt.1/so2.par for 626; 33SO2/s33.par for 636; SO18O/o18.par for 628; SO17O/17.par for 627.

[45] Ulenikov ON, Onopenko GA, Gromova OV, Bekhtereva ES, Horneman V-M. *Re-analysis of the (100), (001), and (020) rotational structure of SO_2 on the basis of high resolution FTIR spectra*. J Quant Spectrosc Radiat Transf 2013;130:220–32. doi:10.1016/j.jqsrt.2013.04.011.

[46] Gueye F, Manceron L, Perrin A, Tchana FKK, Demaison J. *First far-infrared high-resolution analysis of the ν_2 band of sulphur dioxide $^{32}S^{16}O^{18}O$ and $^{32}S^{18}O_2$*. Mol Phys 2016;114:2769–76. doi:10.1080/00268976.2016.1154619.

[47] Flaud J-M, Blake TA, Lafferty WJ. *First high-resolution analysis of the ν_1, ν_3 and $\nu_1+\nu_3$ bands of sulphur dioxide $^{33}S^{16}O_2$*. Mol Phys 2017;115:447–53. doi:10.1080/00268976.2016.1269966.

[48] Blake TA, Flaud J-M, Lafferty WJ. *First analysis of the rotationally-resolved ν_2 and $2\nu_2-\nu_2$ bands of sulfur dioxide, $^{33}S^{16}O_2$*. J Mol Spectrosc 2017;333:19–22. doi:10.1016/j.jms.2016.12.011.

[49] Ulenikov ON, Bekhtereva ES, Krivchikova YV, Morzhikova YB, Buttersack T, Sydow C, et al. *High resolution analysis of $^{32}S^{18}O_2$ spectra: The ν_1 and ν_3 interacting bands*. J Quant Spectrosc Radiat Transf 2015;166:13–22. doi:10.1016/j.jqsrt.2015.07.004.

[50] Introduction, program and documents can be found at https://cdms.astro.uni-koeln.de/classic/pickett

[51] Examples can be found at https://cdms.astro.uni-koeln.de/classic/predictions/pickett/beispiele/

[52] Workshp on Astrophysical Opacities, Aug.1-4, 2017, Kalamazoo, MI, USA

[53] 254th ACS National Meeting, in Molecules in Space: Linking the Interstellar Medium to (Exo)-Planets, August

2017, Washington D.C., USA

[54] Schwenke DW. *Variational calculations of rovibrational energy levels and transition intensities for tetratomic molecules*. J Phys Chem 1996;100:2867–84. doi:10.1021/jp9525447.

[55] Blake TA, Flaud J-M. private communications 2017.

[56] Huang X. unpublished work and private communications 2018.

[57] Ulenikov ON, Bekhtereva ES, Krivchikova YV, Zamotaeva VA, Buttersack T, Sydow C, et al. *Study of the high resolution spectrum of $^{32}S^{16}O^{18}O$: The ν_1 and ν_3 bands*. J Quant Spectrosc Radiat Transf 2016;168:29–39. doi:10.1016/j.jqsrt.2015.08.010.

[58] Makushkin YS, Ulenikov ON. *Isotopic relationships for polyatomic molecules.* Opt Spectrosc 1975;39:629–36.

[59] Bykov AD, Makushkin YS, Ulenikov ON. *On isotope effects in polyatomic molecules. Some comments on the method*. J Mol Spectrosc 1981. doi:10.1016/0022-2852(81)90217-4.

[60] Bykov AD, Makushkin YS, Ulenikov ON. *On the displacements of centers of vibration-rotation bands under isotope substitution in polyatomic molecules*. J Mol Spectrosc 1982. doi:10.1016/0022-2852(82)90273-9.

[61] Huang X. unpublished work and private communications 2017.

[62] the other SO_2 626 EH model: https://cdms.astro.uni-koeln.de/classic/predictions/daten/SO2/SO2/SO2.par.

[63] Müller HSP, Brünken S. *Accurate rotational spectroscopy of sulfur dioxide, SO_2, in its ground vibrational and first excited bending states, v_2=0, 1, up to 2 THz.* J Mol Spectrosc 2005;232:213–22. doi: 10.1016/j.jms.2005.04.010.

[64] Lovas FJ. *Microwave spectral tables II. Triatomic molecules*. J Phys Chem Ref Data 1978;7:1445–750. doi:10.1063/1.555588.

[65] Lee TJ, Scuseria GE. *The vibrational frequencies of ozone*. J Chem Phys 1990. doi:10.1063/1.459548.

[66] Lee TJ, Martin JML, Taylor PR. *An accurate ab initio quartic force field and vibrational frequencies for CH_4 and isotopomers*. J Chem Phys 1995;102:254–61. doi:10.1063/1.469398.

[67] Huang X, Lee TJ. *A procedure for computing accurate ab initio quartic force fields: Application to HO_2^+ and H_2O*. J Chem Phys 2008;129:044312. doi:10.1063/1.2957488.

[68] Huang X, Valeev EF, Lee TJ. *Comparison of one-particle basis set extrapolation to explicitly correlated methods for the calculation of accurate quartic force fields, vibrational frequencies, and spectroscopic constants: Application to H2O, N2H+ , NO2+ and C2H2.* J Chem Phys 2010;133:244108. doi:10.1063/1.3506341.

[69] Huang X, Taylor PR, Lee TJ. *Highly Accurate Quartic Force Fields, Vibrational Frequencies, and Spectroscopic Constants for Cyclic and Linear $C_3H_3^+$*. J Phys Chem A 2011;115:5005–16. doi:10.1021/jp2019704.

[70] Zhao D, Doney KD, Linnartz H. *LABORATORY GAS-PHASE DETECTION OF THE CYCLOPROPENYL CATION (c-$C_3H_3^+$).* Astrophys J Lett 2014;791:1–4. doi:10.1088/2041-8205/791/2/L28.

[71] Puzzarini C, Heckert M, Gauss J. *The accuracy of rotational constants predicted by high-level quantum-chemical calculations. I. molecules containing first-row atoms*. J Chem Phys 2008;128:194108. doi:10.1063/1.2912941.

[72] Puzzarini C, Stanton JF, Gauss J. *Quantum-chemical calculation of spectroscopic parameters for rotational spectroscopy.* Int Rev Phys Chem 2010;29:273–367. doi:10.1080/01442351003643401.

[73] Helgaker T, Gauss J, Cazzoli G, Puzzarini C. *^{33}S hyperfine interactions in H_2S and SO_2 and revision of the sulfur nuclear magnetic shielding scale*. J Chem Phys 2013;139:244308. doi:10.1063/1.4849177.

[74] Puzzarini C, Cazzoli G, Gauss J. *The rotational spectra of $HD^{17}O$ and $D_2{}^{17}O$: Experiment and quantum-chemical calculations.* J Chem Phys 2012;137:154311. doi:10.1063/1.4758316.

[75] Cazzoli G, Puzzarini C, Gauss J. *Rare isotopic species of hydrogen sulfide: the rotational spectrum of $H_2{}^{36}S$.* Astron Astrophys 2014;566:A52. doi:10.1051/0004-6361/201323298.

[76] Müller HSP. $HD^{34}S$, https://cdms.astro.uni-koeln.de/cgi-bin/cdmsinfo?file=e037503.cat.

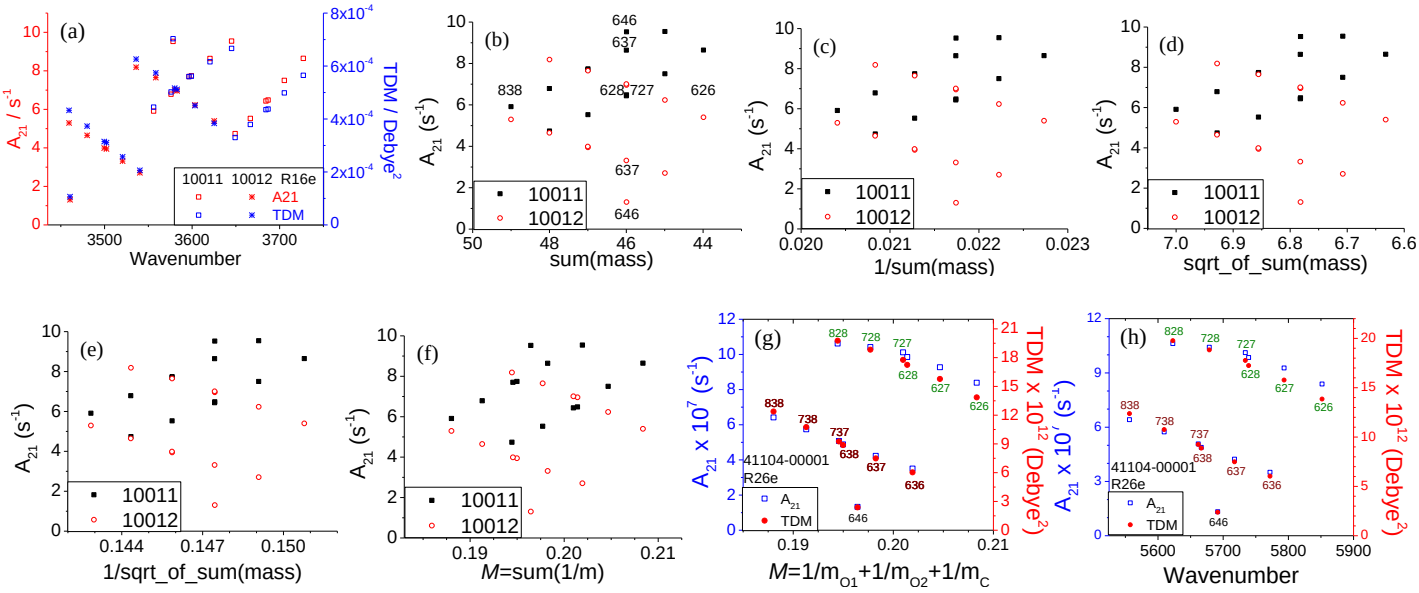


Fig.1 CO_2 Isotope effects on the R16e transition of 1001n-00001 band, *n*=1-2 (a-f) and the R26e of 41104-00001 band (*g*,*h*). Einstein A_{21} (left *y*) and TDM (right *y*) vs. 5 mass-related coordinates (*b-f*, *g*) or wavenumber (*a, h*). We recommend TDM and *M*=sum(1/*m*) representations.

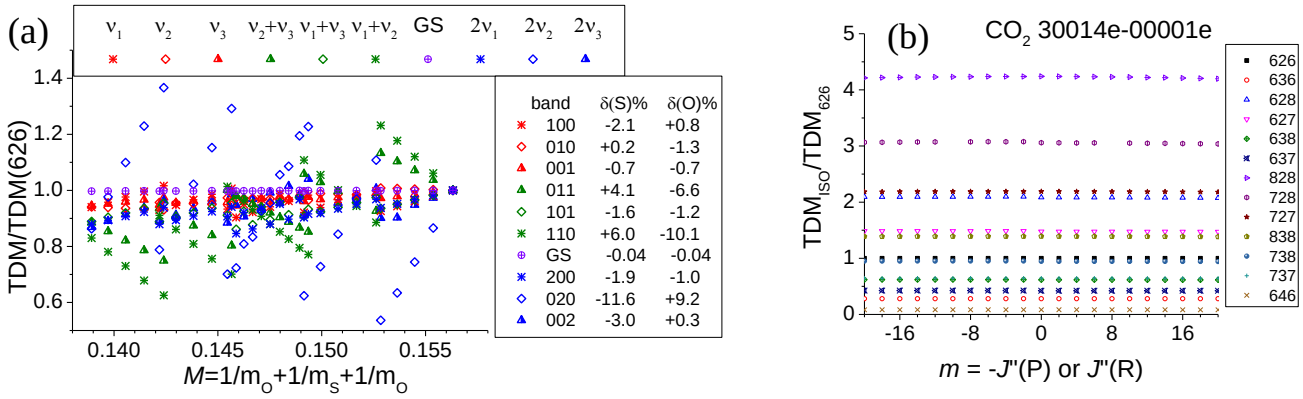


Fig.2 (a) Isotope variation of SO_2 TDM for the strongest IR lines of 10 bands at 296K. (b) the CO_2 TDM_{iso}/TDM_{626} ratios are stable for the *P* and *R* ranges of our 30014e-00001e band.

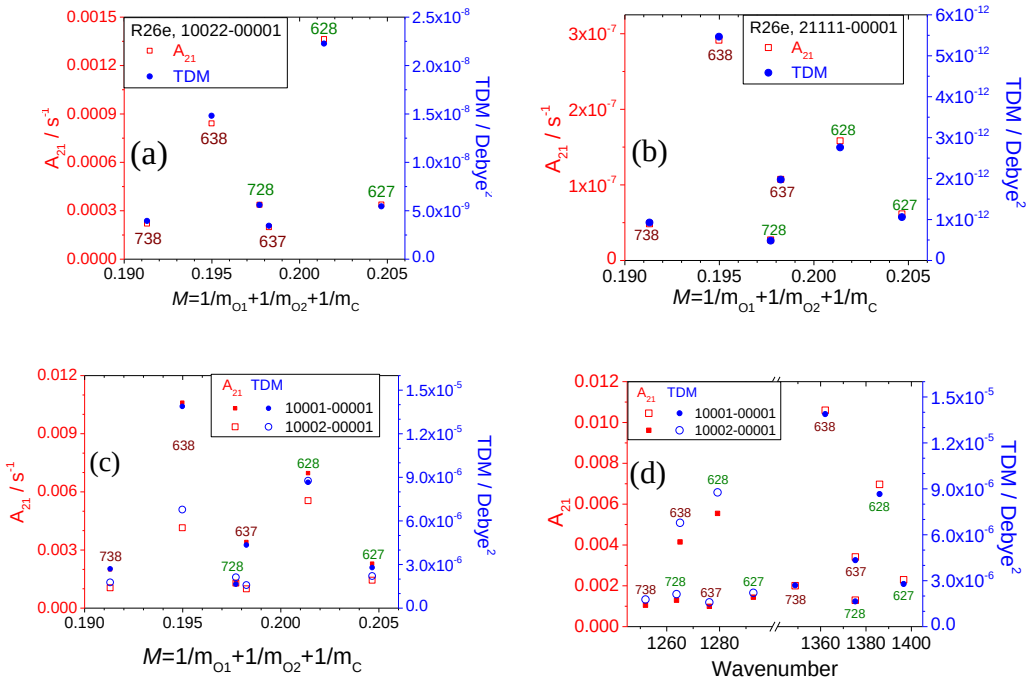


Fig.3. Examples of CO_2 rovibrational transitions where the intensity (TDM and A_{21}) are strongly affected by the mass difference of two O atoms. Panels *a,b,c* use our *M* = sum(1/*m*)coordinate, panel *d* uses cm^{-1}.

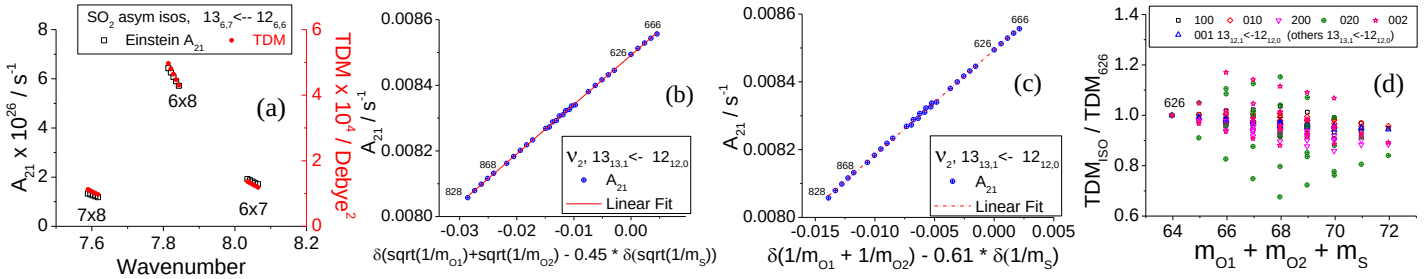


Fig.4 (a) the O isotope effects on SO_2 MW intensity of $13_{6,7}\leftarrow12_{6,6}$ of 15 asymmetric isotopologues, the S isotope mass number *x* =32-36; (b-c) the A_{21} of SO_2 ν_2 transition exhibit linear relation with modified mass coordinates, m_S and m_O are atomic masses; (d) check TDM ratio changes vs. total mass of isotopologues.

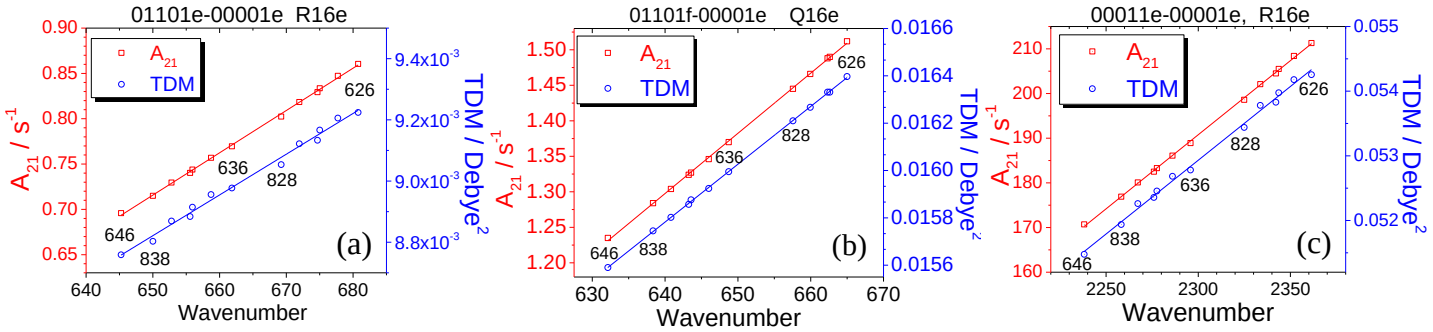


Fig.5 CO_2 Einstein A_{21} coefficients (open squares) and TDM (open circles) of (a) ν_2 R16e; (b) ν_2 Q16e; (c) ν_3 R16e. Linear approximations are also included.

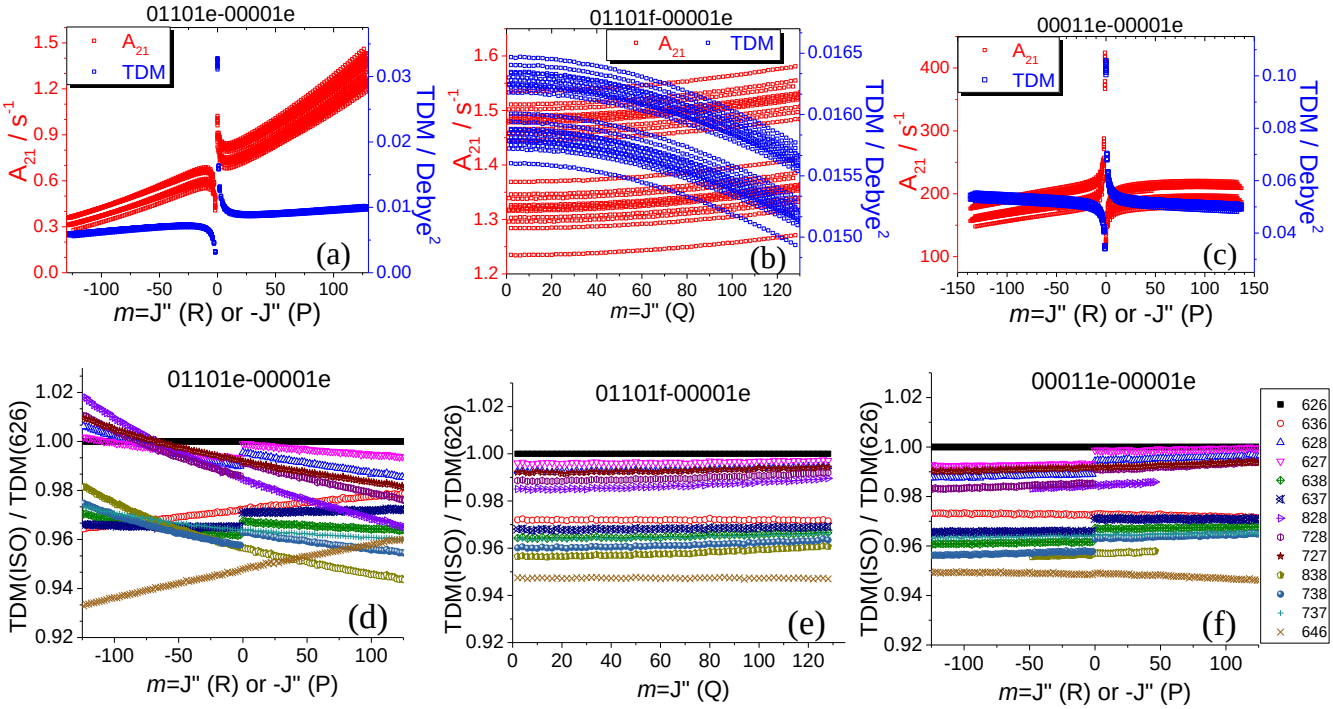


Fig.6 The *J* dependence of A_{21} & TDM (a-c) and TDM_{ISO}/TDM_{626} ratio (d-e) of 13 isotopologues, for CO_2 01101e-00001e (a,d), 01101f-00001e (b,e), and 00011e-00001e (c,f) band components.

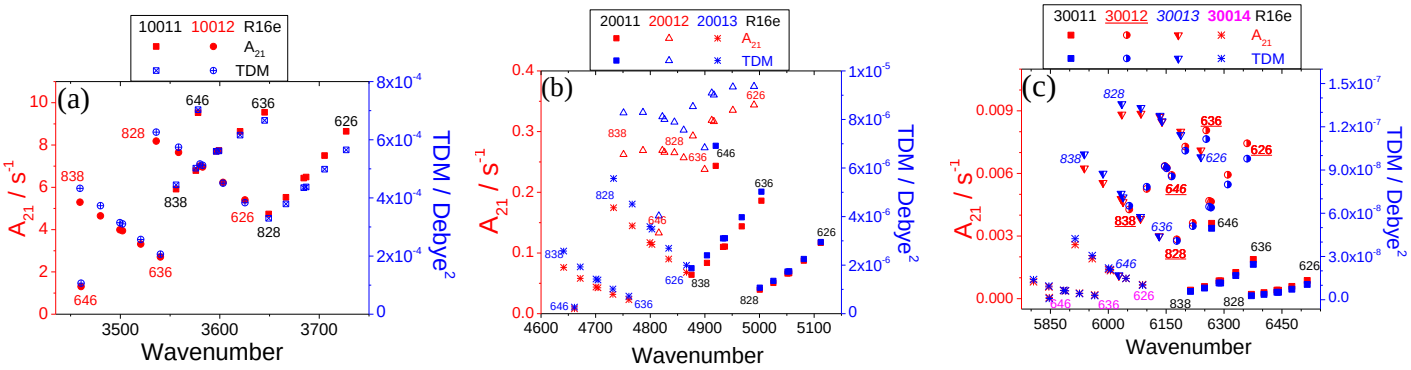


Fig.7 Einstein A_{21} and TDM of R16e vs. cm^{-1} in 3 vibrational polyads: (a) 1001n, n=1,2; (b) 2001n, n=1-3; (c) 3001n, n=1-4.

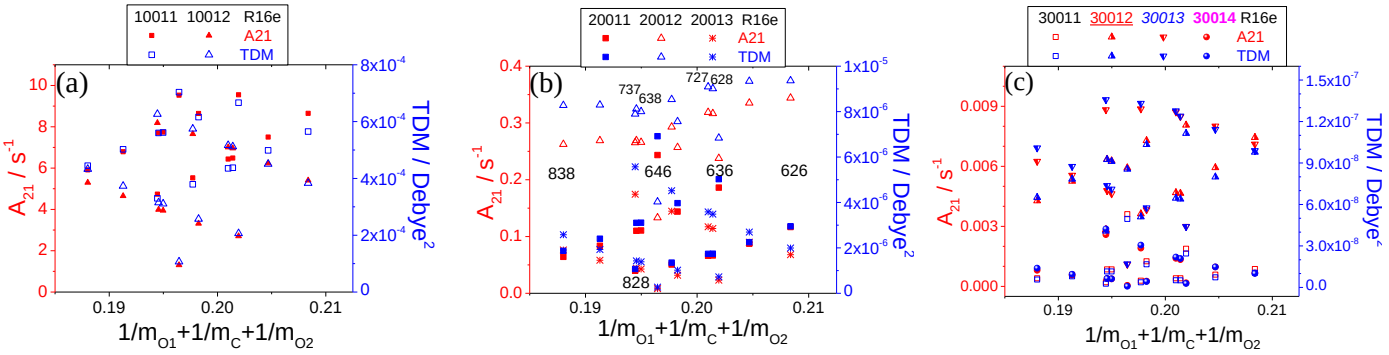


Fig.8 Einstein A_{21} and TDM of R16e transition vs. *M*=sum(1/*m*) in 3 vibrational polyads: (a) 1001n, n=1,2; (b) 2001n, n=1-3; (c) 3001n, n=1-4.

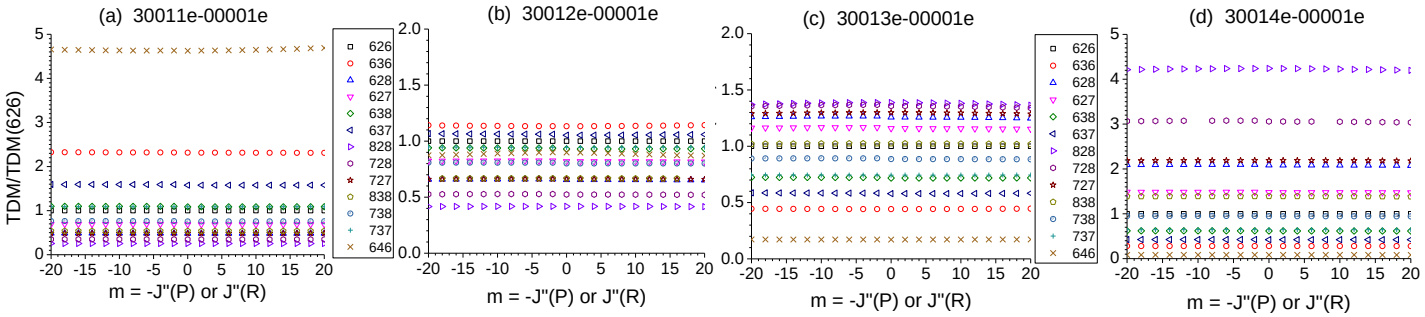


Fig.9. TDM_{ISO}/TDM_{626} ratios of 13 CO_2 isotopologues, for the *P* and *R* branches of 4 bands in 3001*z* polyad, z=1-4.

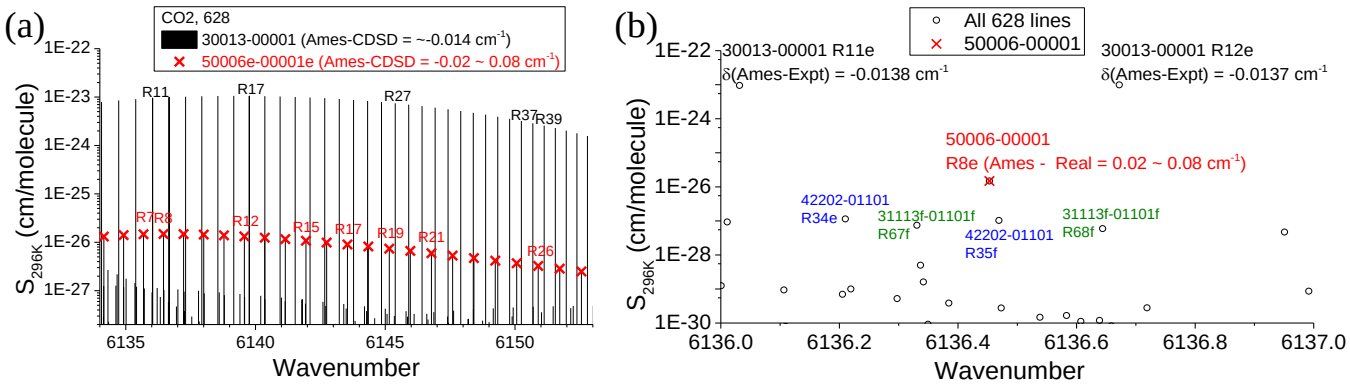


Fig.10 Easy-to-find 50006-00001 lines of CO_2 628 isotopologue. (a) the R branch region as the 2nd strongest band; (b) R8e example.

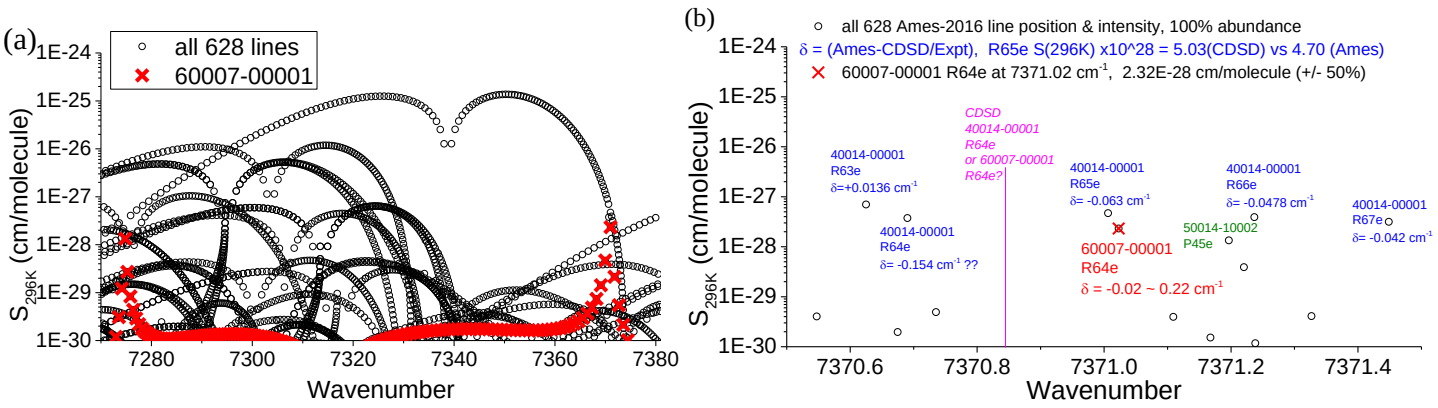


Fig.11 Hard-to-find 60007-00001 lines of CO_2 628 isotopologue. (a) the P & R branches (in red crosses); (b) the strongest line in 628 60007-00001 band, R64e, and other 40014-00001 lines nearby.

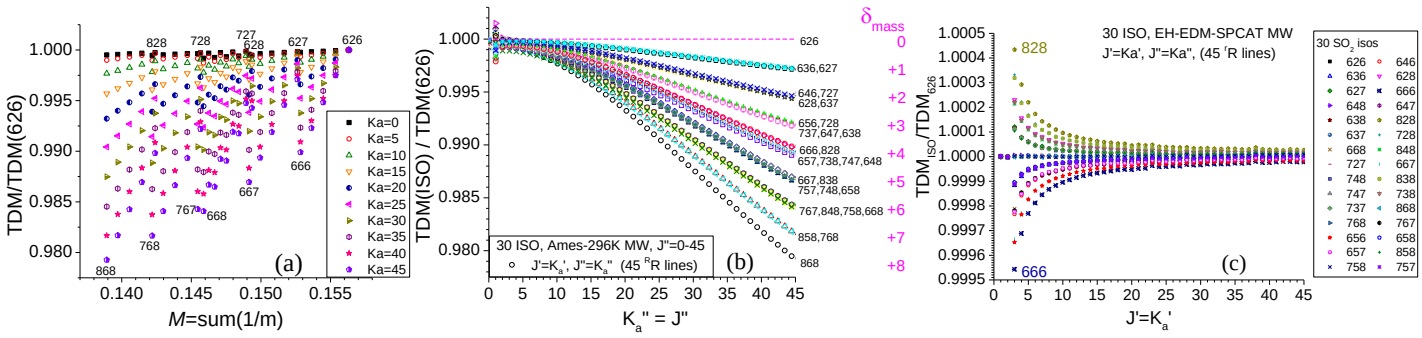


Fig.12. The TDM_{ISO}/TDM_{626} ratios of SO_2 MW transitions, $J+1_{J+1,1} \leftarrow J_{J,0}$, J=0-45. The systematic isotope and J-dependence of Ames-296K intensities in *a* and *b* are in sharp contrast with EH-EDM-SPCAT model predictions in *c*.

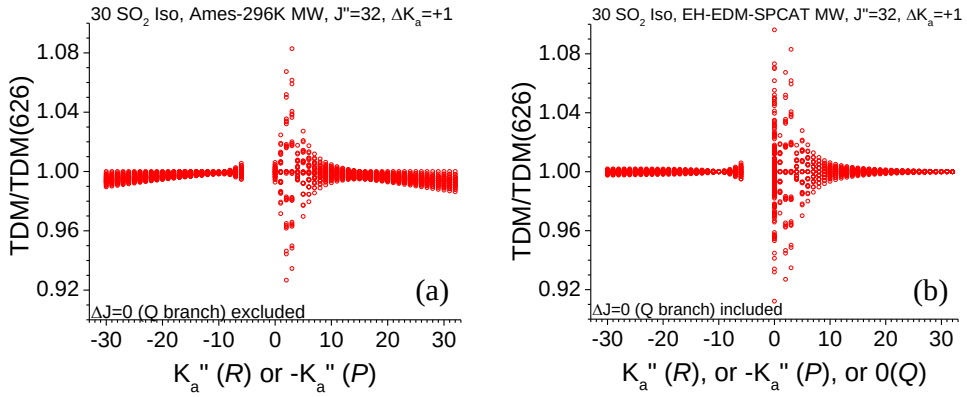


Fig.13 The K_a dependence of TDM ratios at J"=32, ΔK_a=+1. (a) P and R branches of Ames intensity data; (b) P, Q, and R branches of EH-EDM-SPCAT intensity data.

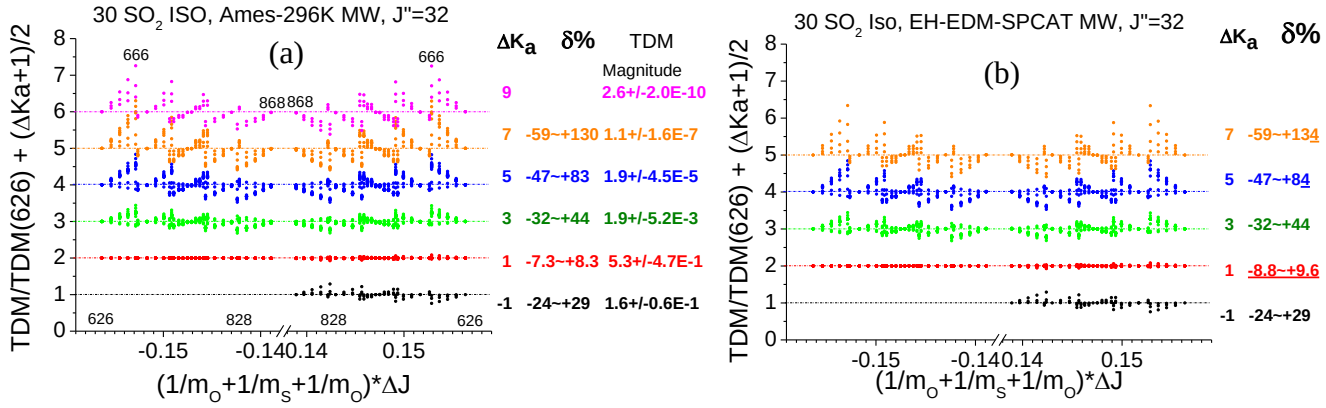


Fig.14 the TDM ratios along *M* coordinate for ΔK_a=-1,1,3,5,7, and 9, with J”=32. (a) Ames-296K IR lists data; (b) EH-EDM-SPCAT model data using single EDM parameter.

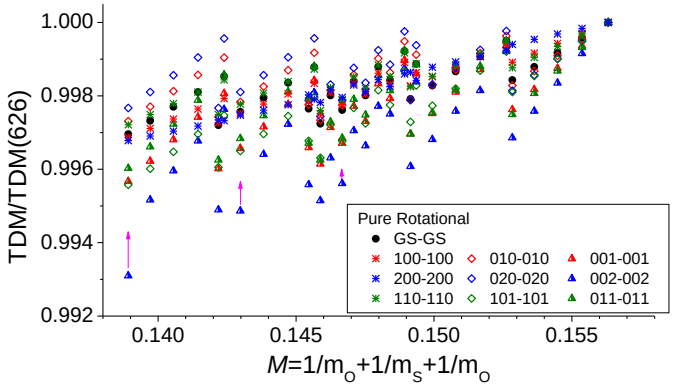


Fig.15. TDM_{ISO}/TDM_{626} of the strongest ʳR lines in 10 pure rotational bands. (compare to Fig.2a)

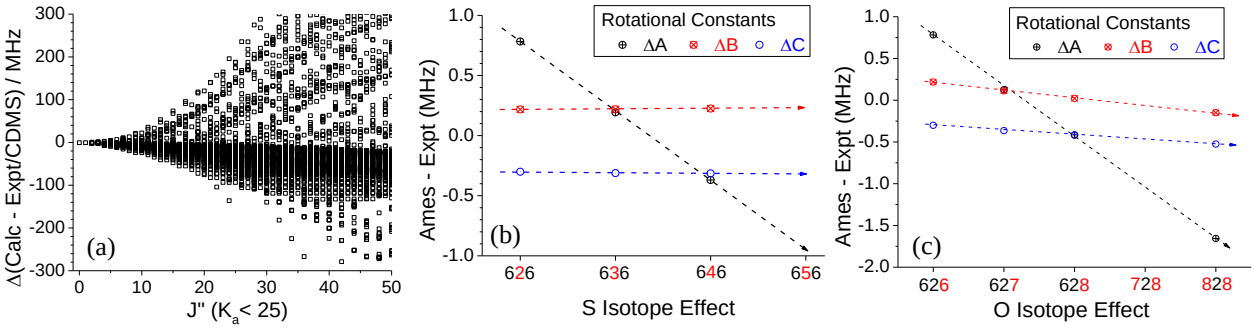


Fig.16 (a) the line position errors of SO_2 646 MW spectra, can reach ~300 MHz at J=30, K_a~25; (b) and (c) show the rotational constants deviations between EH(Expt) and EH(Ames) for SO_2 isotopologues, $\Delta = X_{Ames}-X_{Expt}$: (b) the S isotope effects 626, 636, and 646; c) the O isotope effects 626, 627, 628, and 828.

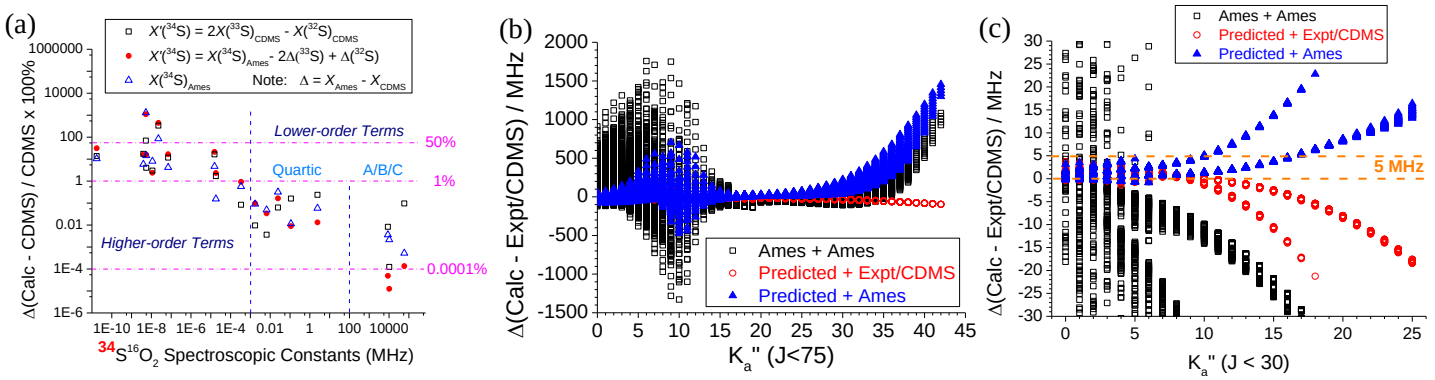


Fig.17. (a) the relative differences (%) of original X_{Ames}, and X_{Ames} predicted from two formula, with respect to the 646 X_{CDMS} parameters; (b) and (c) compare the $^{34}S^{16}O_2$ line positions predicted by three sets of EH parameters vs. original EH(CDMS/Expt) line positions: (b) J up to 75, $K_a > 40$; (c) reliable range, J" = 0 – 30 and K_a" = 0 – 25.

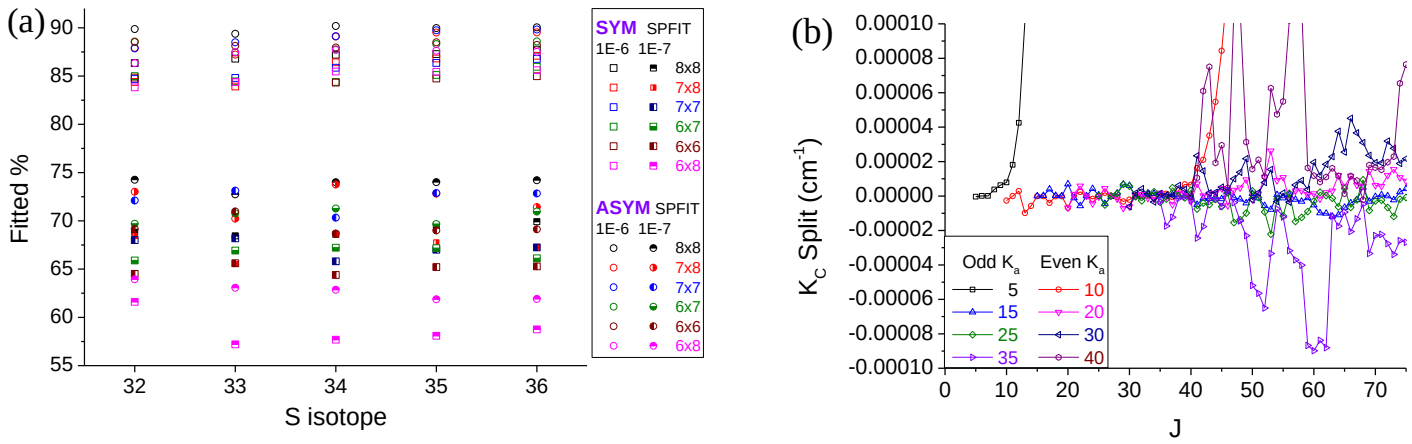


Fig.18. (a) Percentage of fitted lines in 30 EH(Ames) analysis, using Sym or Asym datasets, and two different uncertainty levels from 1E-6 to 1E-7 cm^{-1}. Sorted by line set, uncertainty level, and S & O isotopes. (b) The K_C splits of Ames SO_2 668 ground state levels. Split Δ = E(J, K_a, K_c=J-K_a) – E(J, K_a, J+1-K_a).

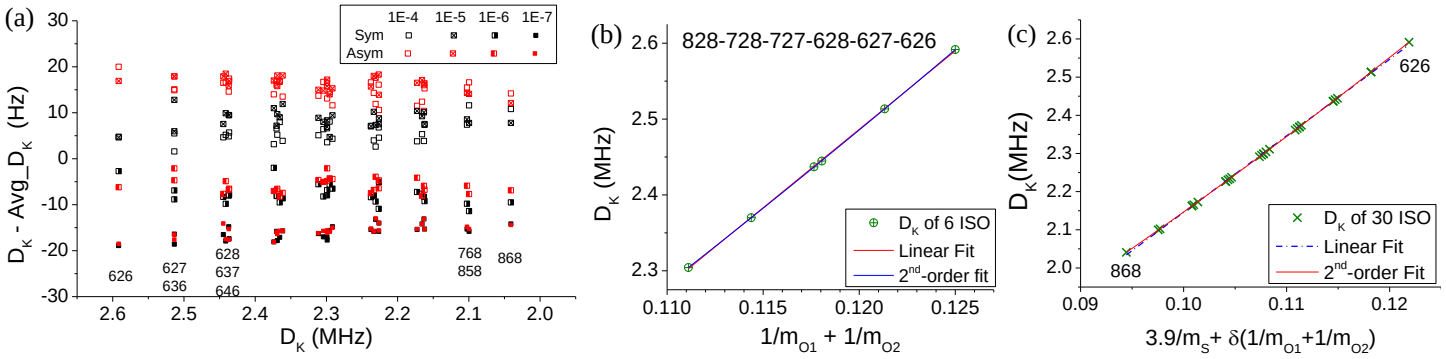


Fig.19 (a) D_K convergence in the EH(Ames) of 30 isotopologues, using Sym or Asym datasets, and four different uncertainty/accuracy level from 1E-4 cm^{-1} to 1E-7 cm^{-1}, with respect to the D_K value averaged from 8 SPFIT fits. (b) The linear and 2nd-order fits of EH(Ames) D_K values (determined at 1E-6 cm^{-1} level) along the mass inverse coordinate, *M*; (c) a coordinate integrating the S & O isotope effects. See text for details and why we can do better than b) and c).

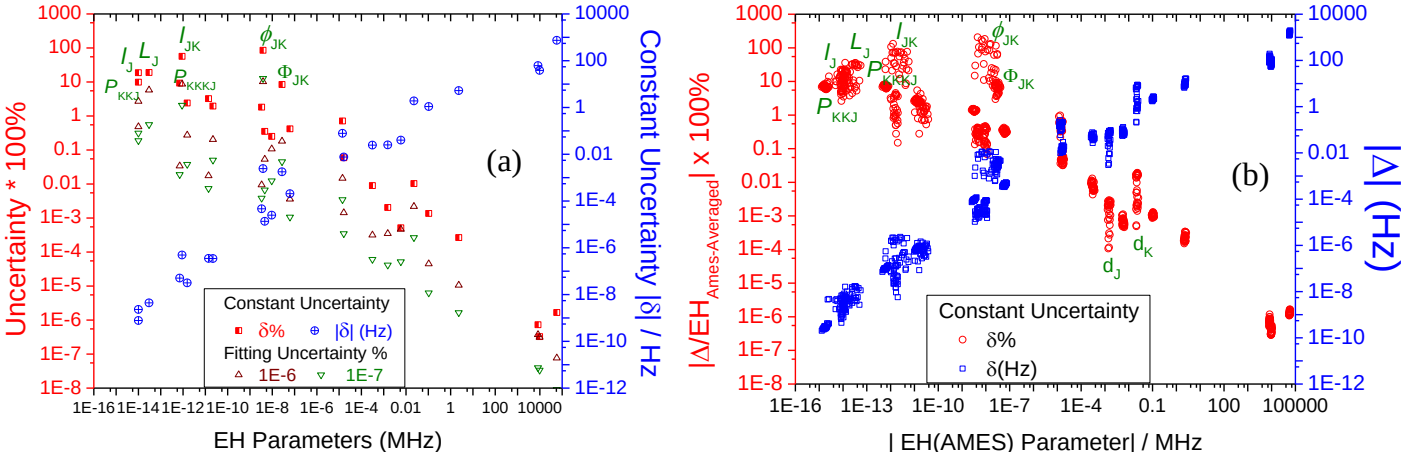


Fig.20 (a) The fitting uncertainty and constant uncertainty (convergence) of 26 EH(Ames) parameters, all averaged from 30 SO_2 isotopologues. Left axis: the relative fitting uncertainty % from SPFIT (triangles), and relative constant uncertainty $\delta\%|EH(1E\text{-}6) - EH(1E\text{-}7)|$ (red squares, half filled); Right axis: absolute constant uncertainty $\delta|EH(1E\text{-}6) - EH(1E\text{-}7)|$ in Hz; (b) constant uncertainty of 26 EH(Ames) parameters in the 1E-6 cm^{-1} and 1E-7 cm^{-1} SPFITs of 30 isotopologues. Right axis: absolute uncertainty in Hz (blue squares); Left: relative constant uncertainty % (red circles).

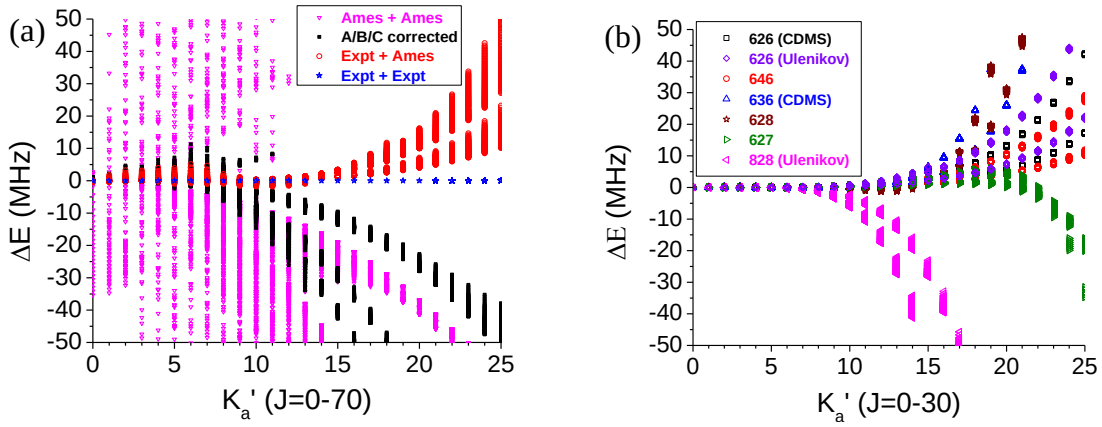


Fig.21. Line position differences caused by A/B/C, quartic level terms, and higher order terms. (a) SPCAT line position predictions using 4 sets of EH(646) parameters, with respect to pure EH(CDMS/Expt) model based line position predictions. EH sets are labeled as "lower order terms" + "higher order terms". "Ames" is original EH(Ames), "A/B/C corrected" includes expt A/B/C + Ames quartic +higher order terms. See text for explanations. (b). the EH(Expt + Ames) model based SPCAT line positions for 6 isotopologues, with respect to the pure EH(CDMS/Expt) model based SPCAT line position predictions.

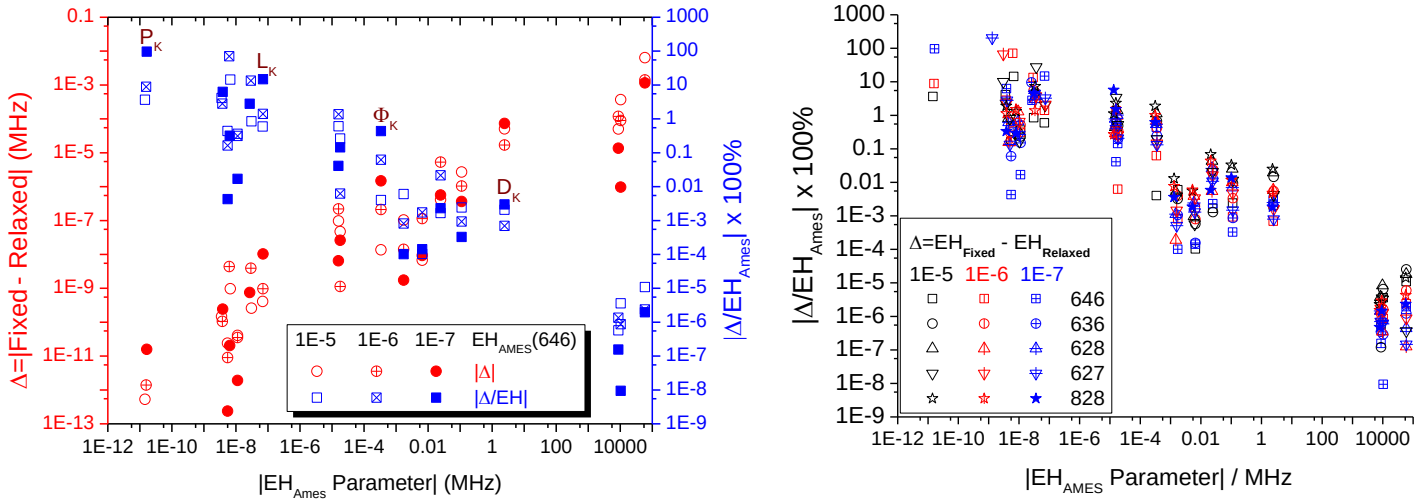


Fig.22 (a-b): Absolute (Left y axis, circles) and relative (Right y axis, squares) differences of the EH parameters acquired between fully relaxed fit (all vary freely) and the partially fixed fit (some parameters are fixed at 626 values): (a) 646 only; (b). the relative differences (%) of 646, 636, 628, 627, and 828, at 3 fitting accuracy levels.

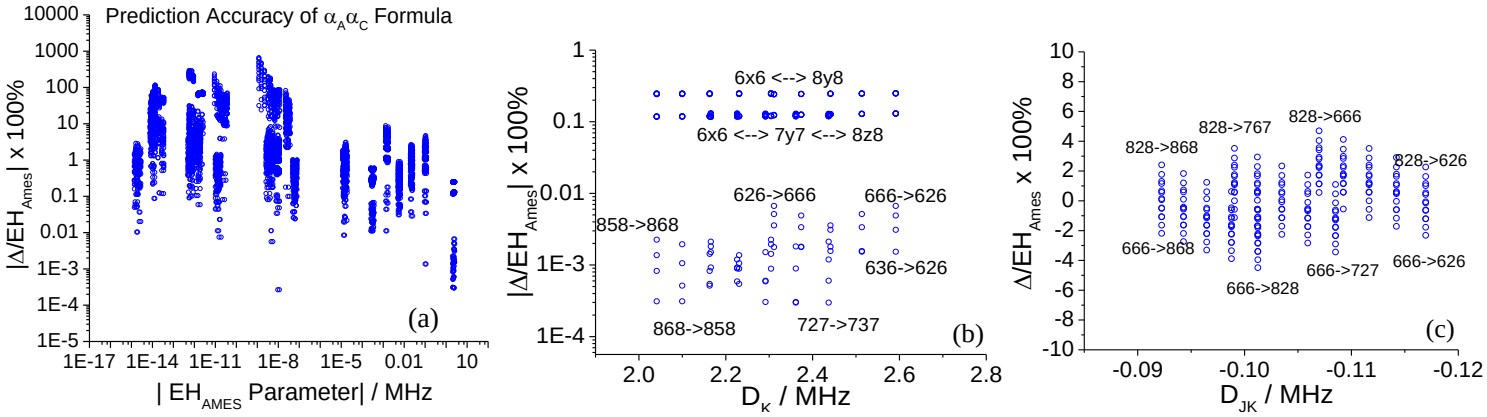


Fig.23 Accuracy check for the prediction formula Ulenikov et al [49] adopted in their EH(828) model. The 30 SO_2 isotopologue EH(Ames) parameter sets acquired at 1E-6 cm^{-1} level are taken as standard reference for relative prediction deviations Δ%. (a) overview for 30 isotopologues; (b) the D_K prediction accuracy, *x/y/z* =2-6 for $^{32}S - {}^{36}S$ isotope, x≠y, y≠z; (c) the D_{JK} prediction accuracy.

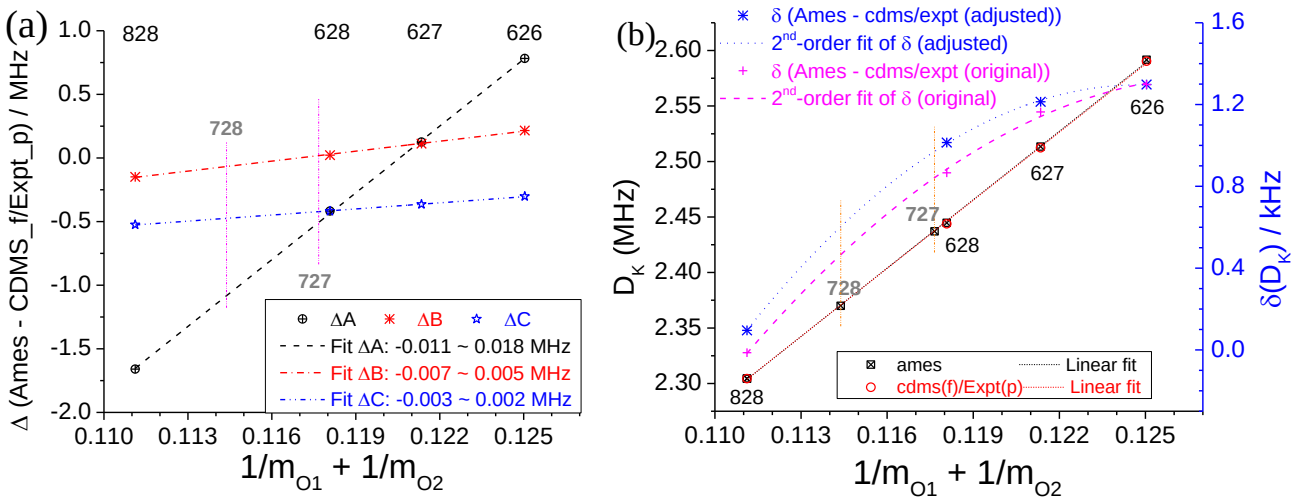


Fig.24. The corrected version of O isotope effects: 626, 627, 628 and 828. (a) the absolute values and the linear fits of ΔA/ΔB/ΔC; (b) the linear and 2nd order fits of D_K values (left), and the 2nd order fits of the $\delta(D_K) = D_K(\text{Ames}) - D_K(\text{CDMS/Expt})$ differences with and without "fixed" and "prediction" adjustments. Use EH(Ames) constants fit at 1E-7 cm^{-1} accuracy level.

Supplementary Figures for Isotopologue Consistency Paper

Fig.S1 CO_2 01101e-00001e analysis

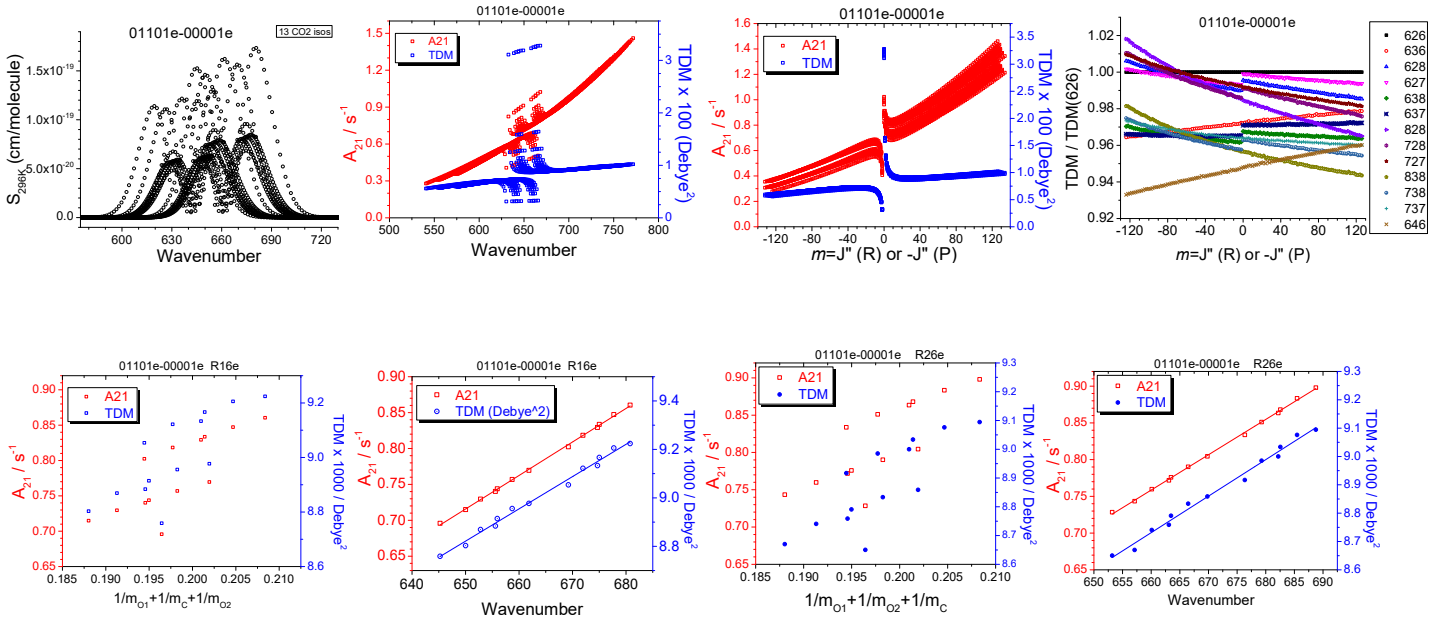


Fig.S2 CO_2 01101f-00001e analysis

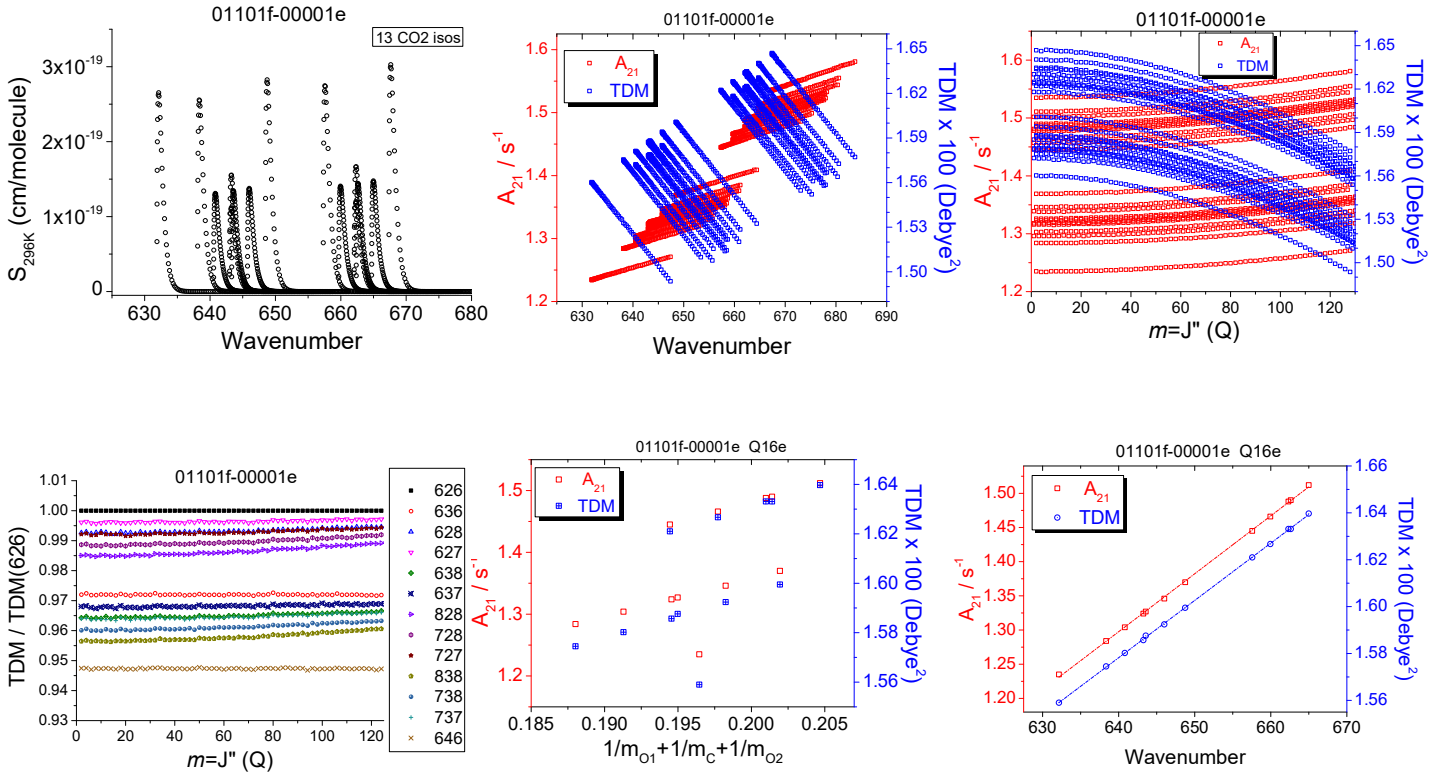


Fig.S3 CO_2 00011e-00001e analysis

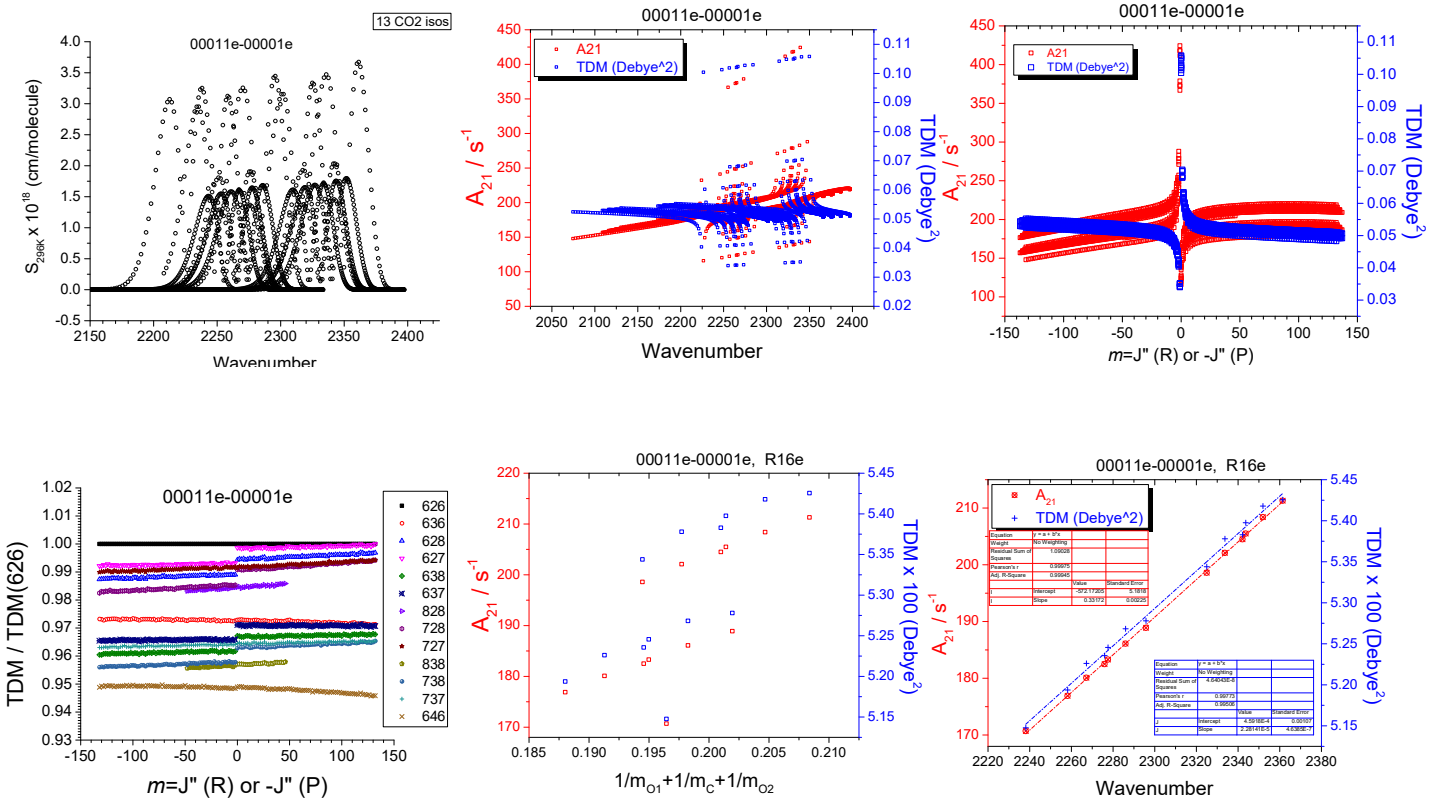
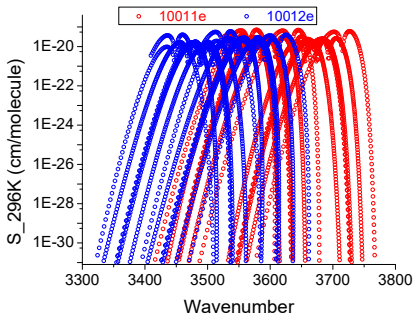
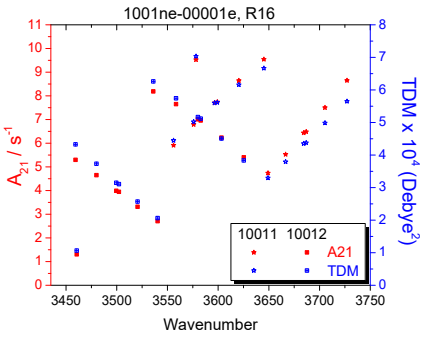
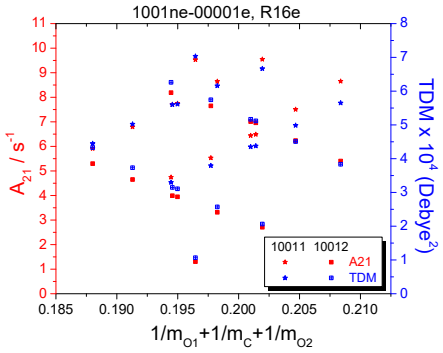
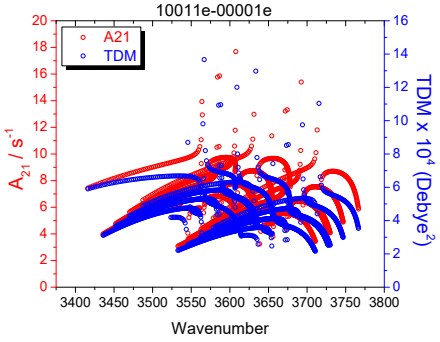


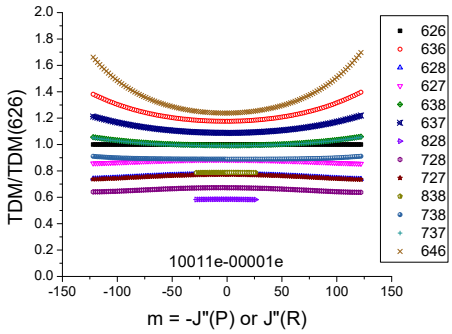
Fig.S4 CO_2 1001x-00001 analysis

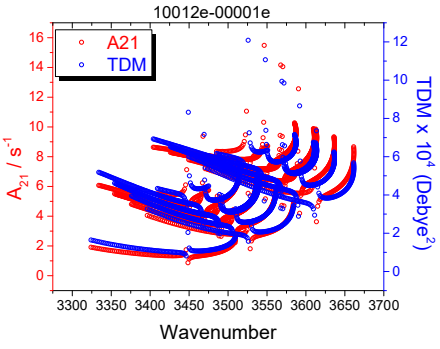












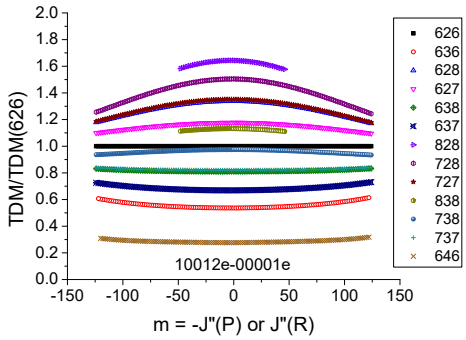


Fig.S5 CO_2 2001x-00001 analysis

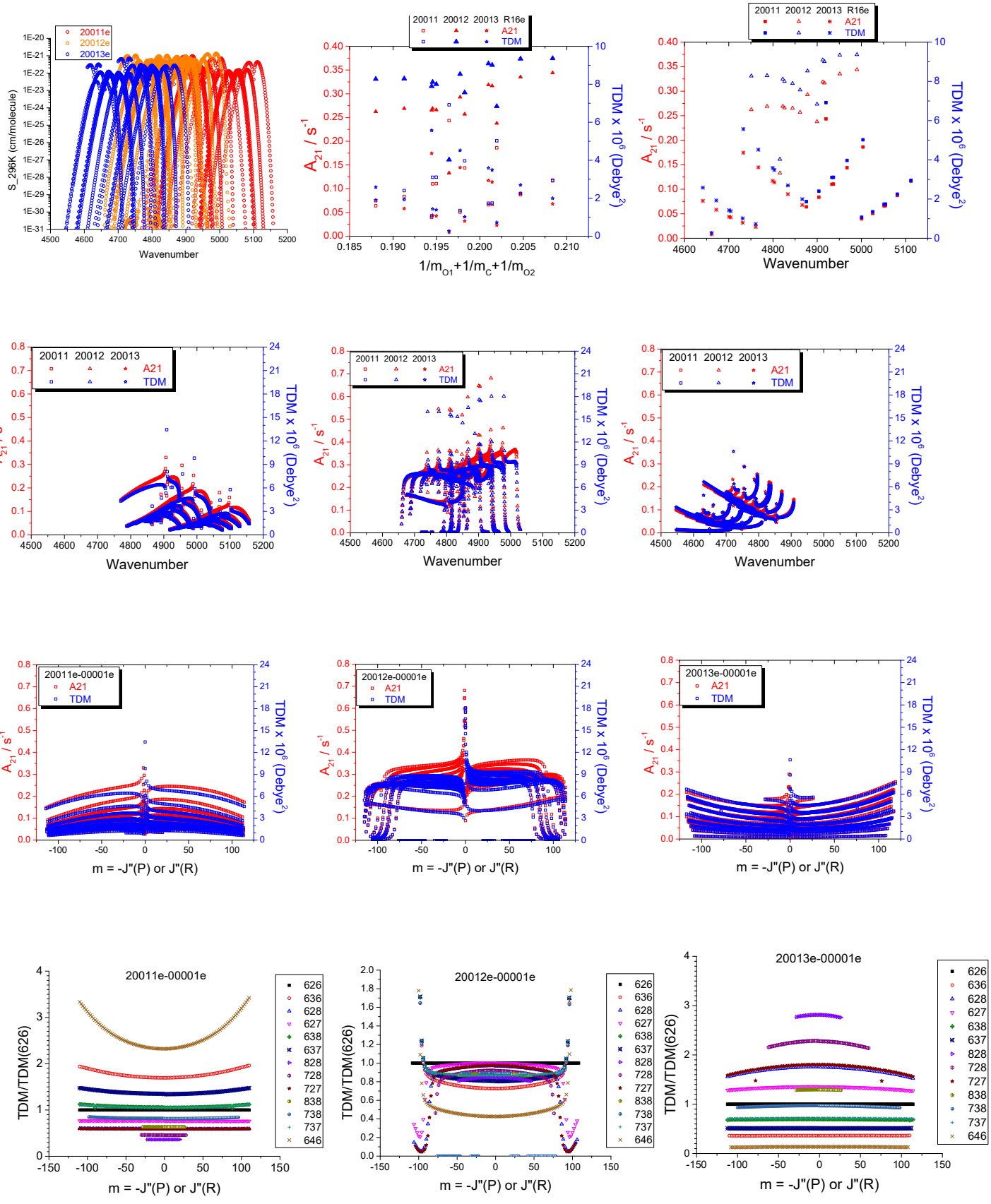


Fig.S6 CO_2 3001x-00001 analysis

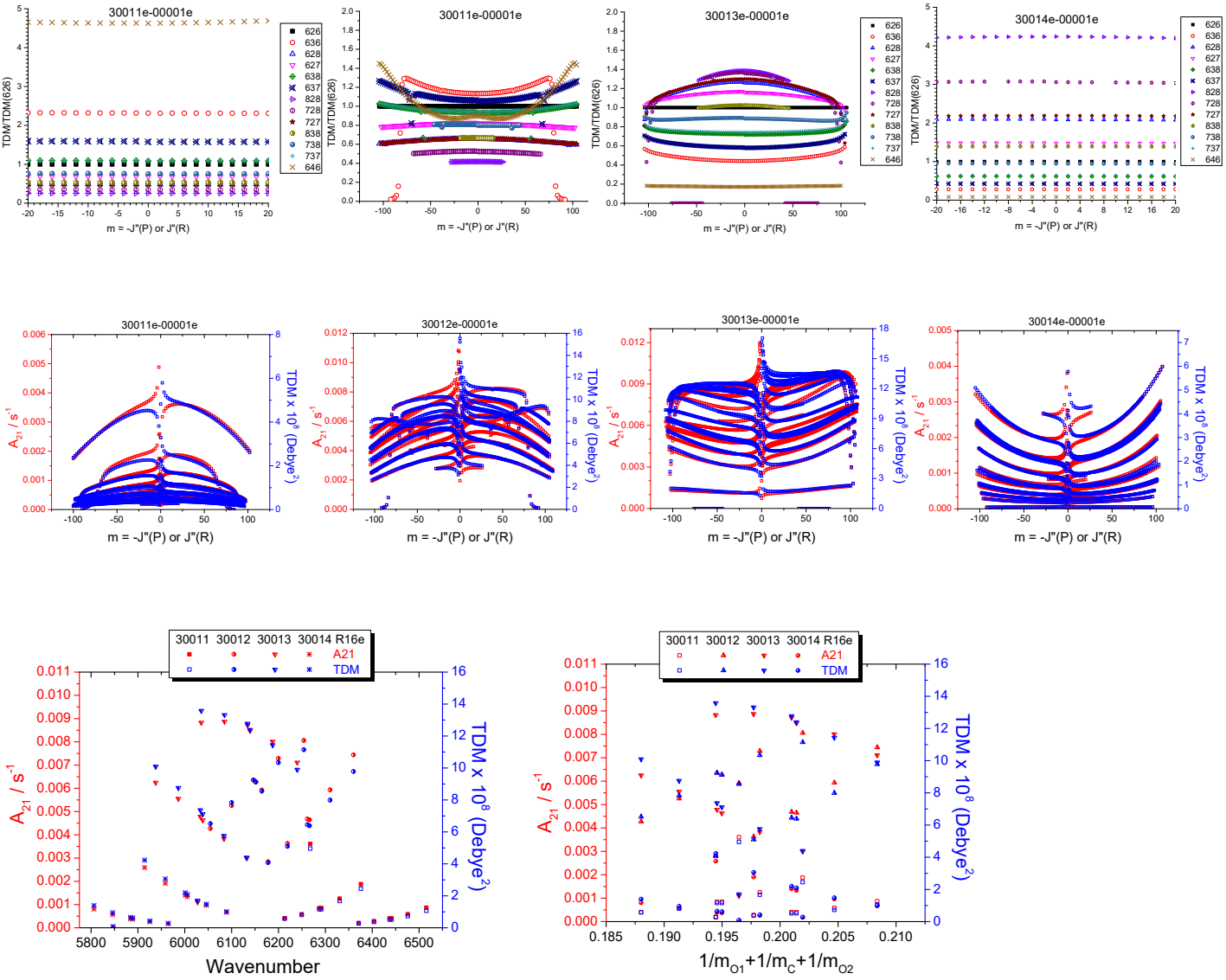


Fig.S7 The Einstein-A coefficient A_{21} and TDM (sum of transition dipole element squares) of 13 CO_2 isotopologues, for the R16e/f transitions in 0nn11-0nn01 ($n\nu_2 + \nu_3 \leftarrow n\nu_2$, n=0-3) bands.

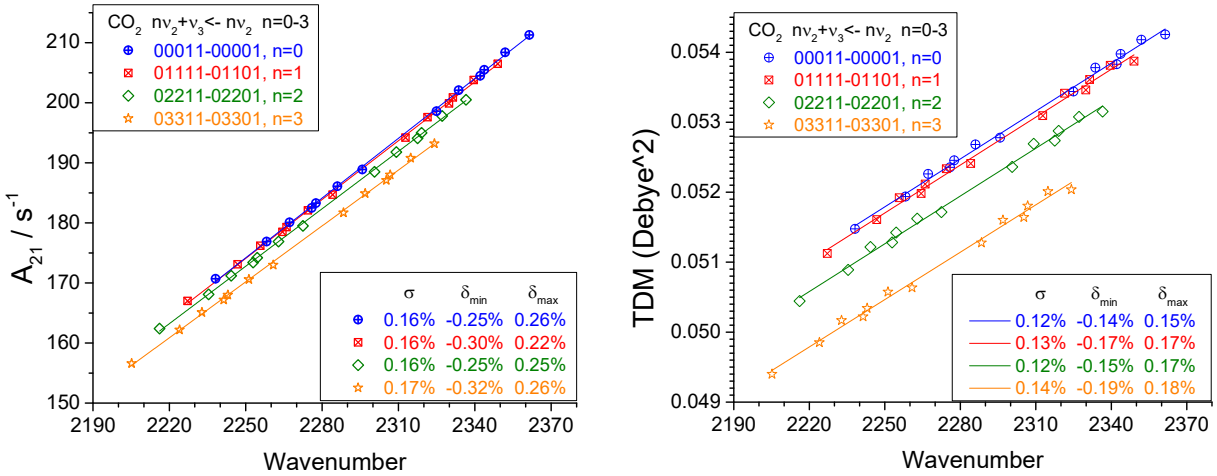


Fig.S8 The TDM_{ISO}/TDM(626) ratios of 4 e/f←→e/f branches of CO_2 01111-01101 and 02211-02201 bands.

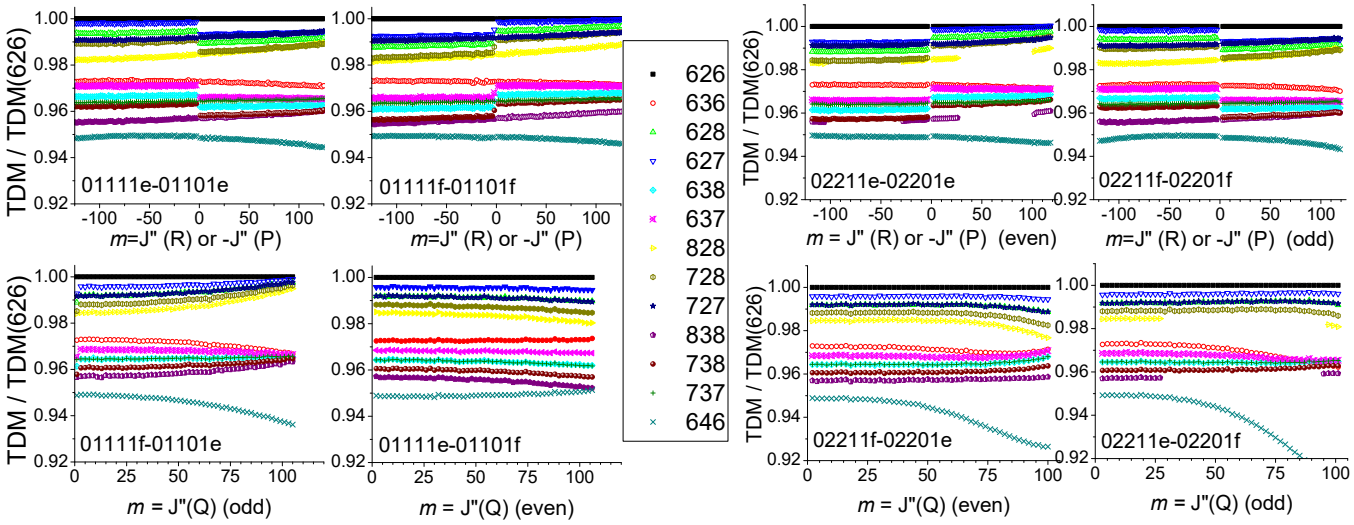
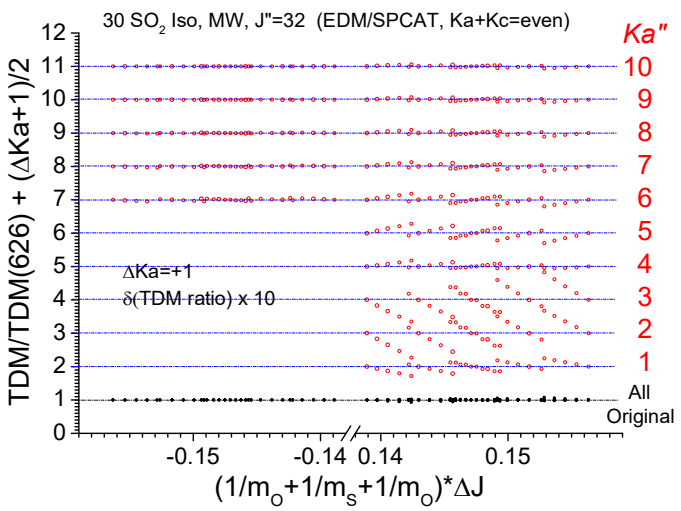


Fig.S9 Ka" dependence of SO_2 TDM ratios: a) J"=32; b) J"=12



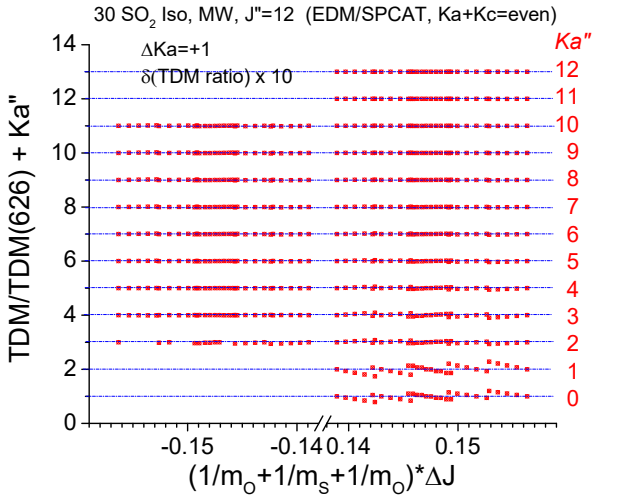
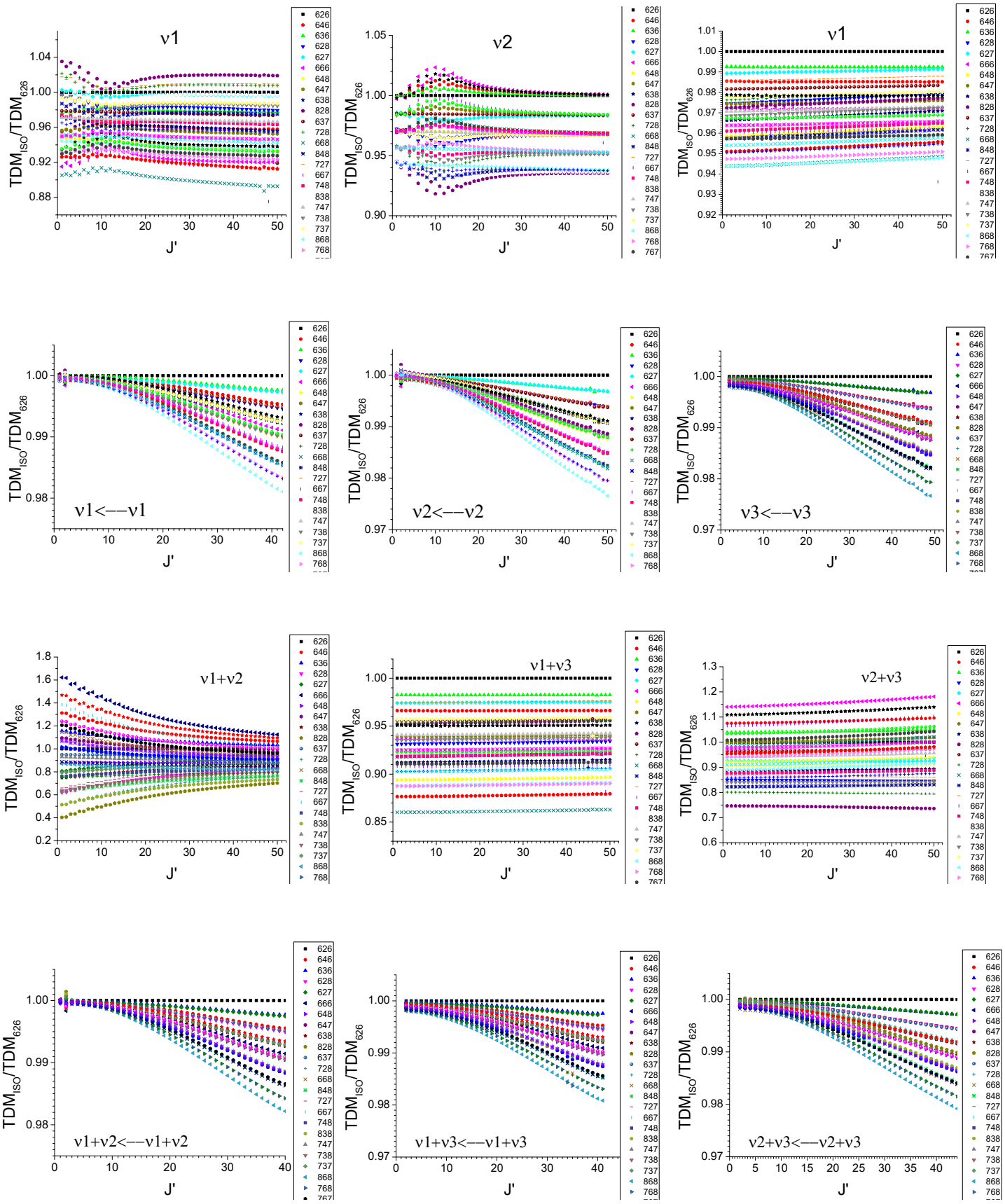


Fig.S10 The J dependence of the TDM ratios of ^{r}R transition in 9 bands: J"$_{Ka",Kc"}$ = $12_{12,0}$($\Delta\nu$=0), or $12_{0,12}$($\Delta\nu$>0). Note the 2ν2 need more investigation / confirmation for higher J effects, where the TDM ratios of 6x6 isotopologues (x=3-6) spans from 1E-4 to 10. The local minima of TDM ratios are found at J=48,44,37,32 for x=3,4,5,6, respectively.



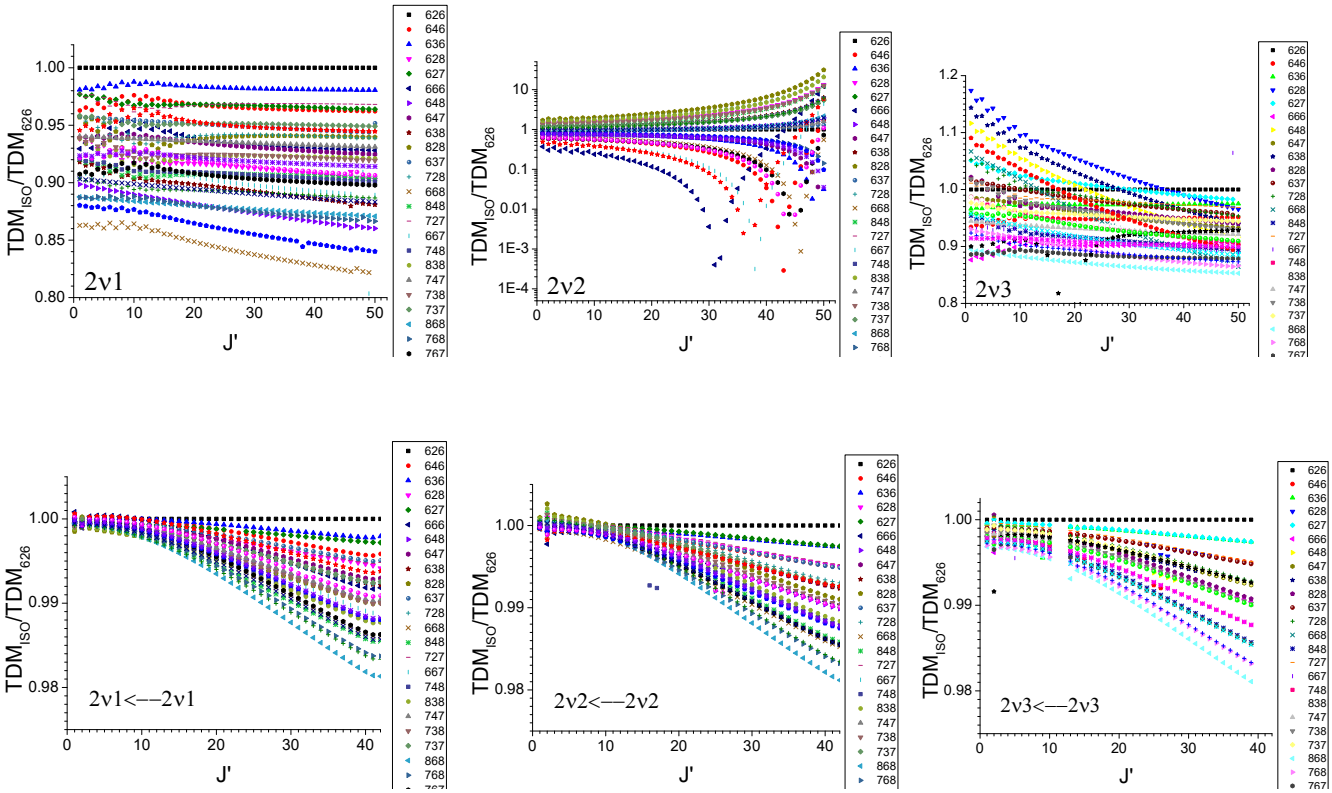

TDM_ISO/TDM_626
J'
2ν1
2ν2
2ν3
2ν1<—2ν1
2ν2<—2ν2
2ν3<—2ν3
626
646
636
628
627
666
648
647
638
828
637
728
668
848
727
667
748
838
747
738
737
868
768
767

Fig.S11 Percentage of fitted lines in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm^{-1}. (a) overview, by fitting accuracy level; (b) detailed comparison, by fitting accuracy, S and O isotopes.

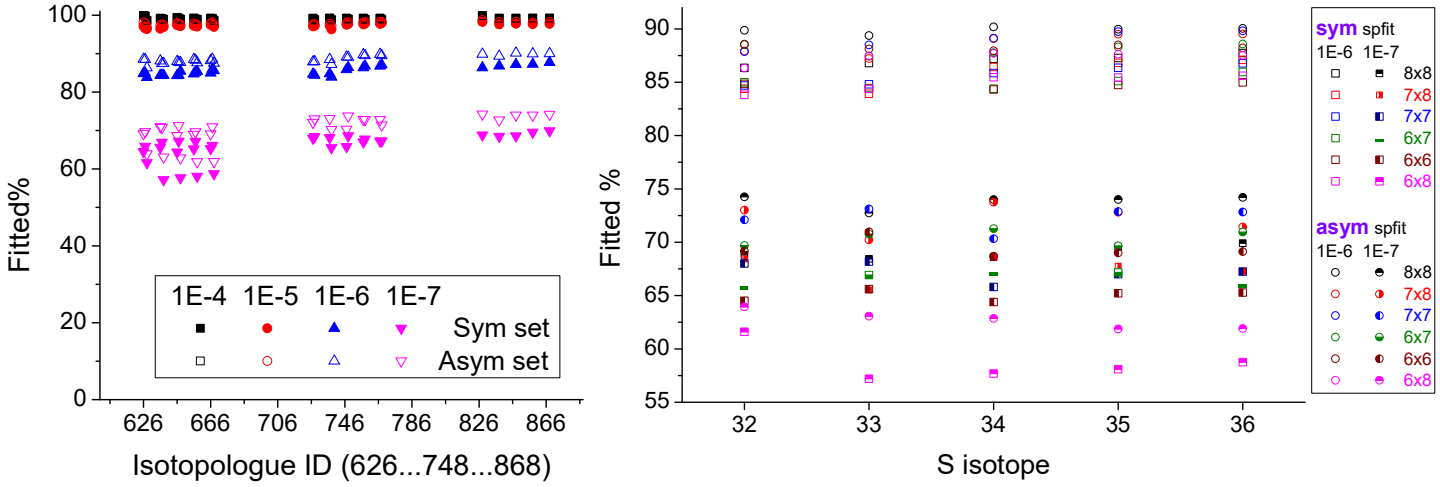


Fig.S12 The K_C splits of Ames SO_2 668 ground state levels. Split Δ = E(J, K_a, K_c=J-K_a) – E(J, K_a, J+1-K_a). Left side is overview; Right panel shows more details.

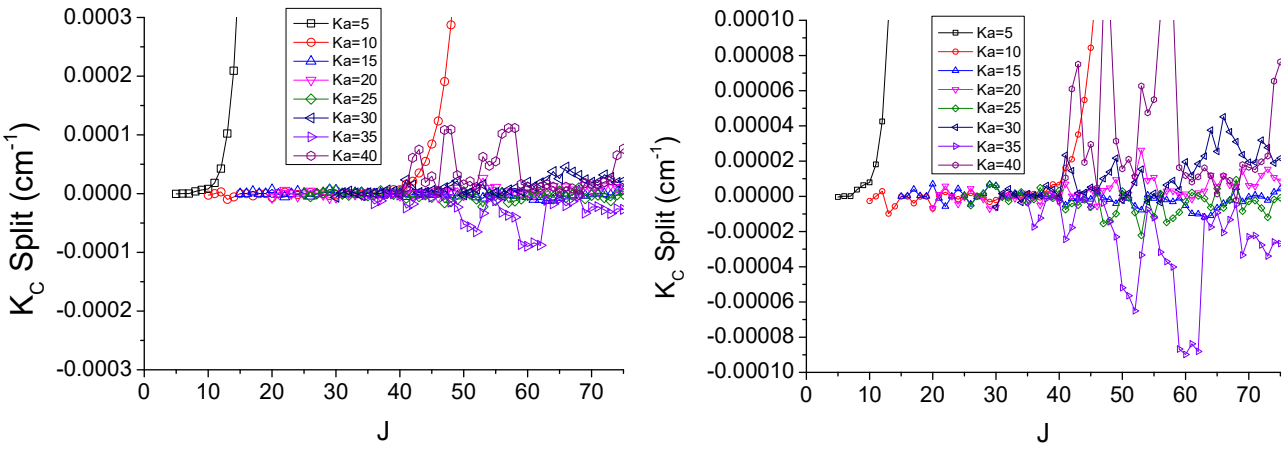


Fig.S13. D_K analysis in the EH(Ames) SPFIT of 30 isotopologues, using Sym or Asym datasets, and four different uncertainty/accuracy level from 1E-4 cm^{-1} to 1E-7 cm^{-1}. (a) overview, by fitting accuracy level, data are with respect to the value averaged from 8 D_K; (b) the D_K convergence with respect to fitting accuracy: δ(Asym-Sym) are black squares, and $\delta(D_{K,}(n+1) - D_K(n))$ are triangles, where *n* is the accuracy/uncertainty level, e.g. 5-4 means D_K(1E-5) – D_K(1E-4).

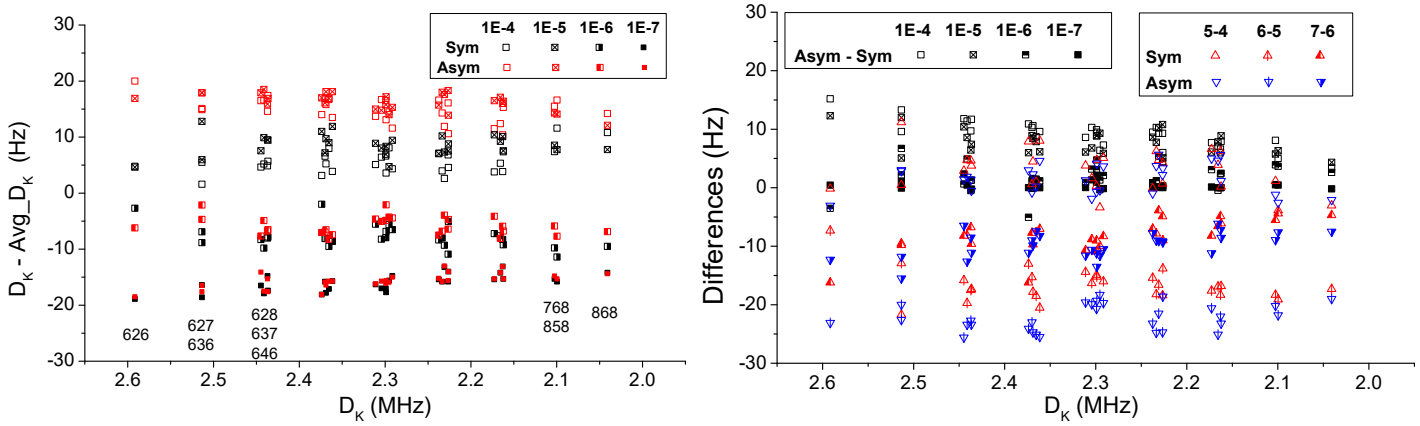


Fig.S14 D_{JK} in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm^{-1}. (a) different D_{JK} values, by fitting accuracy level and data type, with respect to the value averaged from 8 D_{JK}'s; (b) the D_{JK} convergence with respect to fitting accuracy: δ(Asym-Sym) are black squares, and $\delta(D_{JK,}(n+1) - D_{JK}(n))$ are triangles; (c) the D_{JK} overview of 30 SO_2 isotopologues.

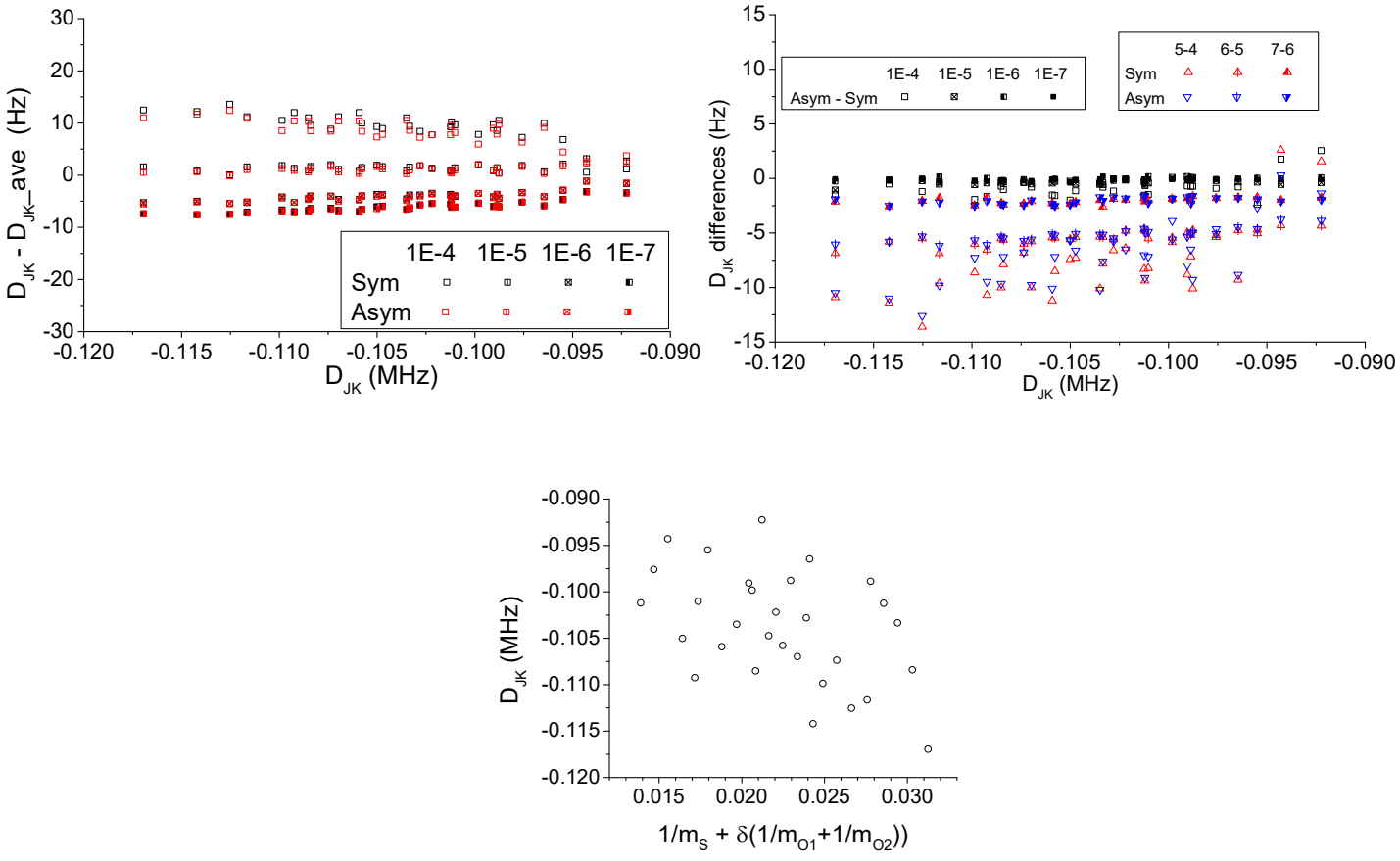
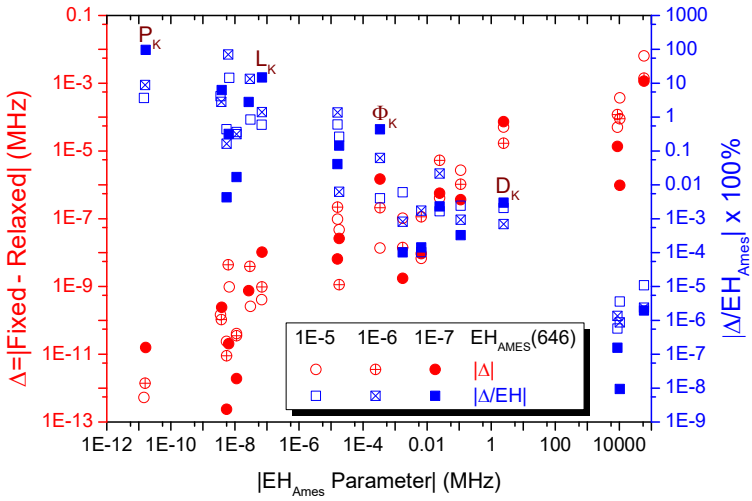
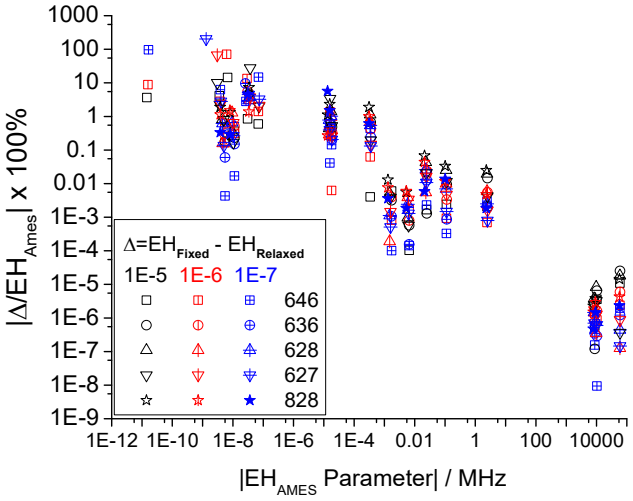


Fig.S15 (a-b): Absolute (Left y axis, circles) and relative (Right y axis, squares) differences of the EH parameters acquired between fully relaxed fit (all vary freely) and the partially fixed fit (some parameters are fixed at 626 values): (a) 646 only; (b). the relative differences (%) of 646, 636, 628, 627, and 828, at 3 fitting accuracy levels.





Note: The Fig.S15 has been moved back to manuscript → See Fig.22.

Discussions in Ver.0, Section 4, for EH(Ames) and Line Positions

1. Isotopologue-consistency of EH model fits

To ensure the Δ(Ames-Expt/CDMS) in Figs.16-17 are highly self-consistent, a uniform EH model has to be used for *all* the 30 isotopologues. We know the fitting residuals depend on the completeness of the EH model terms included in the fit. Adding more terms to describe higher order effects or inter-molecular couplings and perturbations usually can help reduce the fitting residuals. Experimentalists prefer the EH fitting residuals as small as, or even smaller than the data uncertainty. This is easy to understand, but sometimes it may lead to isotopologue-specific terms, which means EH model differences and introduces inconsistency in the Δ calculations. The 646 EH model from CDMS website is selected because the EH model was also used for 626, 628, 627, 636, and recent 828 experimental vibrational IR band studies.[1] Extra terms (if any) accounting for the ^{33}S and ^{17}O related hyperfine splittings are excluded, which we assume not strongly coupled with remaining terms.

Given a uniform EH model on an isotopologue, its parameters can carry different least-squares weights. But the weight for any specific parameter should be a constant across all 30 isotopologues. However, the CDMS EH models of minor isotopologues had various higher order terms fixed at CDMS EH(626) values, i.e. their corresponding weights were set to essentially zero. It effectively breaks the consistency of EH(CDMS/Expt) data between 626 and the rest. If we ignore 626 EH parameters and make predictions with the Δs of minor isotopologues, the Δ consistency is still not very good, because different higher order terms were fixed in SPFIT. This is part of the bottleneck damaging our prediction accuracy for general applications.

Secondly, the datasets used in SPFIT should be as consistent as possible, too. Ideally, they should have the same set of transition quantum numbers, so can be noted as "equivalent" datasets. This is trivial for EH(Ames) analysis because the all microwave transitions > 1E-36 cm/molecule at 296K have been computed. But it is highly unlikely for experimental analysis, if not totally unrealistic. Less abundant isotopologues naturally have less number of transitions observed in lab, and scientists always need the most reliable spectroscopic constants fit from the maximum number of reliable, identified transitions.

Even with "equivalent" transition datasets, in the least-squares fitting step, the experimental uncertainty associated with every transition is almost impossible to maintain 100% consistency for all isotopologues. Accordingly, the error bar set for rejection cannot be synchronized, either. For EH(Expt) fits, it is unwise to manually change the error bar (which in SPFIT correlates with uncertainty), no matter larger or smaller. But the equivalent dataset extracted from Ames-296K IR lists have highly consistent uncertainties. In addition, we can also modify the uncertainty to check the convergence of fitted EH constants, see details in Sec.4.2.5.

In short, the best isotopologue-consistency should come from uniform model, identical fitting weights and error bar, and equivalent datasets (including uncertainties) for both EH(Ames) and EH(Expt) analysis on all involved isotopologues. Since the published or online EH parameters were usually fit from different models and datasets, one may need to extract the largest common transition dataset from the original experimental data, run least-squares fits independently but consistently, and use the results for Δ predictions. But less transitions could hurt the quality of EH(Expt) constants.

Now we can tell that the Fig.17 a) is partially wrong in the "higher-order terms" region, because 8 (12) higher order terms in the 646 (636) EH(CDMS) models were fixed at EH(CDMS) 626 values, they cannot work with any prediction formula. The values of the rest 6 sextic centrifugal distortion constants will be affected, too. Another choice is to re-fit the Ames datasets by fixing the parameters at EH(Ames) 626 values, too. But the results would still have broken consistency from fitting weight discrepancies. Third choice is to re-fit the experimental data of minor isotopologues with the original (fully relaxed) parameter set, but the transition sets do not contain enough information to determine those higher order terms reliably.

OK, it is clear both the dataset and EH model must be consistent for all isotopologues. Existing EH(CDMS/Expt) models for SO_2 isotopologues are not quite consistent. How about EH(Ames) then? Are we sure the Ames IR list based EH

parameters fitted from SPFIT are truly reliable, accurate and consistent?

2. Reliability of EH(Ames) constants: A_0/DK

In Fig.16 b), the S isotope effect from ^{32}S to ^{36}S suggests that the $\Delta(^{34}S)$ in the middle is critical, if one wants to predict $\Delta(^{35}S)$ and $\Delta(^{36}S)$ with ~0.01 MHz accuracy. A 0.01 MHz error on ^{34}S may become 0.02 cm^{-1} or larger at ^{36}S. So less than 0.005 MHz deviations are preferred. In our tests, the B_0/C_0 in EH(Ames) usually are well converged, but the largest uncertainty lies in the A_0 constant of EH(Ames).

In the *A*-reduced Watson Hamiltonian, A_0 and D_K are clearly correlated. The consequence is that, the SPFIT at the 1E-5 cm^{-1} uncertainty level cannot guarantee A_0 convergence better than 0.01 MHz, at least impossible for the ~3000 transitions included in our EH(646) fit. The EH(Ames) reported in Fig.16 b) has the A_0/D_K at 58990.81394 MHz / 2.440615 MHz. But there does exist another ***stable*** local minima where A_0/D_K are 58990.88397 MHz / 2.441753 MHz. The corresponding changes on B_0 and C_0 are -0.0039 MHz and -0.0024 MHz, respectively. Obviously, such +0.07MHz / +0.00114 MHz differences are too large to ensure the prediction accuracy of EH(Predicted). For example, +0.07MHz on A_0 may cause deviations ~60MHz at J=30, and +0.001 MHz on D_K can bring ~50 MHz errors at K_a=15.

If we fix the D_K somewhere between the two minima, the fitted A_0 also comes between the two A_0 values at minima. You can imagine a "solution" curve of all possible A_0/D_K combinations that give similar least-squares fits, e.g. same no. of failed transitions and total rms. We did locate two more local minima along the "solution" curve, with A_0/D_K fit at 58990.8267 / 2.44082 MHz and 58990.8622 / 2.44147 MHz, respectively. These two EH(Ames) 646 fits have similar fitting quality, with 298 and 278 transitions rejected, respectively. The two fits are included in supplementary material. Interested readers can try SPFIT on your own.

The issue we describe above is all on EH(Ames) side. The solution is easy: no convergence at 5E-5 cm^{-1}? Try 5E-6 or 5E-7 cm^{-1}! At 5E-6 cm^{-1}, the A_0 comes out in the middle +0.07MHz A_0 variance is reduced to ~0.006 MHz. This seems good enough for linear approximations in Fig.16 b). The variance is further reduced at 5E-7 cm^{-1} level.

Apparently, this issue and fix are not new to experimental spectroscopists. They have accumulated rich experiences in numerous EH model fits where some parameters cannot be accurately determined. In most cases, it is attributed to the dataset deficit, i.e. too few transitions or quantum numbers not high/low enough, or the EH model deficit, e.g. to add additional coupling terms or higher order terms. Note such multi-minima scenario did not occur when we reproduced the CDMS fit (using the CDMS dataset) or run SPFIT with the equivalent Ames set (CDMS quantum numbers + Ames-296K line positions). The ~3000 strongest MW transitions at 296K included in the EH(Ames) model fit are more than 10 times of EH(CDMS) fit. It is reasonable to suspect such multi-minima issue may only become significant for considerable number of transitions and/or relatively large uncertainty. For example, most transitions in CDMS EH(646) fit carry 0.05-0.20 MHz uncertainty, or 1.7-7E-6 cm^{-1}.

It raises a question on the rovibrational band EH analysis. In those experimental rovibrational IR work fitting thousands of unperturbed transitions carrying 1E-5~1E-4 cm^{-1} accuracy, does the A_0 still have uncertainty as large as 0.05 MHz (or larger)? Many molecular rovibrational IR studies choose to report the observed line positions and fitting errors in unit of 1E-3 cm^{-1} or at most 1E-4 cm^{-1}, which are limited by the sig figs of the frequency of the "standard reference" transition. Now we are telling people that even the fitting deviation σ_{RMS} as small as 5E-5 cm^{-1} can ***not*** 100% guarantee the A_0 convergence better than 0.05 MHz. No need to mention EH fits with errors ~1E-3 cm^{-1}. In short, we seriously doubt the uncertainty range of published A_0 constants for many molecules, and bands. It could be larger than we had thought.

On the other hand, we have observed the B_0/C_0 are fairly stable. With less than 0.005 MHz variations, they probably can be safely adopted in the linear approximation formula (7) in Section 4.1.

In Fig.16 a) analysis, ~3000 transitions were fit with equally weighted EH parameters, while there were only 282 transitions in CDMS fit. How the rotational constants become if we use the equivalent Ames set of 282 lines? At 1E-5 cm^{-1} level, the A/B/C differences are -0.006, +0.004, and +0.001 MHz, respectively. But this test has limited value, because the results may rely on the specific J/K_a data range, uncertainty level, as well as the molecular type and vibrational bands.

After checking the reliability and convergence of A_0/D_K constants in EH(Ames) analysis, next we check the EH(Ames) consistency across all isotopologues, then discuss more about convergence and uncertainty.

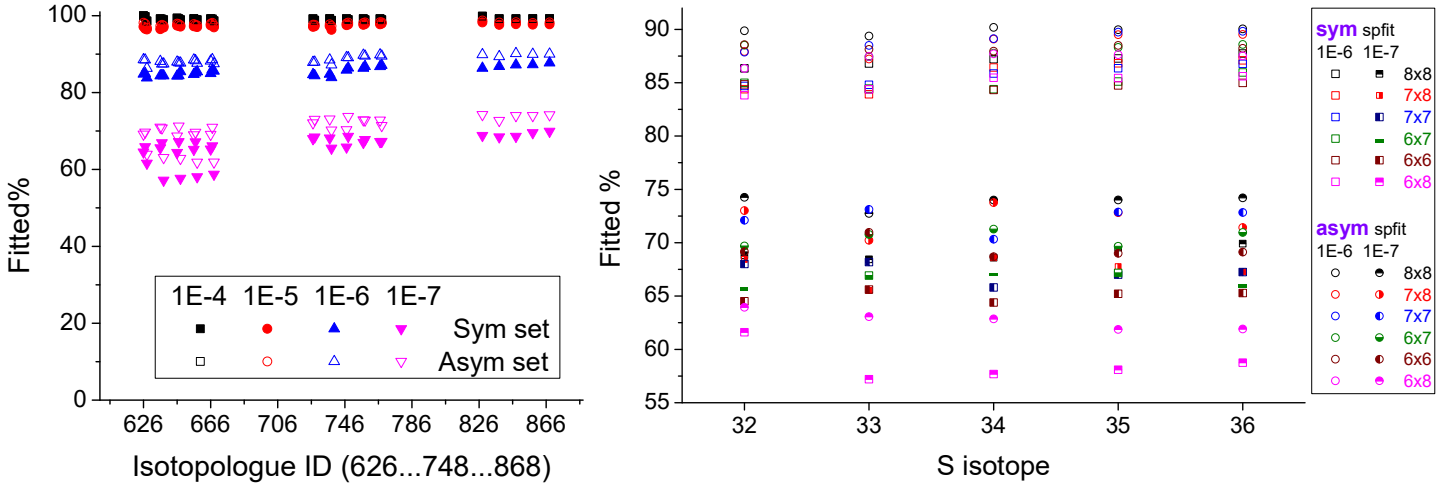
3. Isotopologue Consistency Benchmark: EH(Ames) Dataset and Fitting Accuracy.

This section explores how isotopologue-consistent the EH(Ames) analysis may become, if we trust the first principle calculations on the empirically refined PES. To ensure the rotational constants converged to better than 0.002 MHz, we need larger dataset. To check the convergence with respect to the uncertainty level and SPFIT error bars, we need more significant figures for line position. The Ames MW lists used in Sec.4.2.2 only carry 1E-5 cm^{-1} resolution, now all the Ames-296K MW line sets are regenerated with 1E-8 cm^{-1} (or 3E-4 MHz) resolution.

Asymmetric and 7x7 isotopologues have more IR active transitions than symmetric isotoplogues, due to nuclei spin statistical weights. To maintain the best consistency, we created two sets of transitions for each isotopologue. One set is the asymmetric (Asym) isotopologue style, the other set is the symmetric (Sym) isotopologue style. For Asym and 7x7 isotopologues, it is straightforward to create a Sym dataset by filtering out those Sym-forbidden transitions from the Asym dataset. For other Sym isotopologues, those Asym-type transitions can only be created from the rovibrational energy levels we computed, with zero intensities. During this step, we found and fixed small inconsistency on a few isotopologue line lists, because their old IR lists were computed using slightly different parameters. At the end, the J=0-75 rovibrational energy levels of all 30 isotopologues have been re-computed, using an uniform set of VTET input parameters for Asym isotopologues. Sample input file and the rovibrational levels are included in supplementary files.

At the beginning, we chose those transitions with $J \leq 60, K_a \leq 35$, and S(intensity) > 1E-32 cm/molecule for each isotopologue. But the transition sets constructed in this way are not strictly consistent, because the intensity of a specific transition will vary by isotopologue. See Sec.3 for examples. As a result, the 30 datasets for SPFIT will have slightly different size and quantum numbers. To solve this problem, we randomly picked the 668 isotopologue Asym and Sym datasets as the "standard" transition sets for all SPFIT least-squares fitting analysis. We go through other 29 Ames-296K lists, and extract all equivalent transitions that form our final SPFIT sets. In this way, we avoid dataset contaminations on the isotopologue consistency.

Fig.S11 Percentage of fitted lines in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm^{-1}. (a) overview, by fitting accuracy level; (b) detailed comparison, by fitting accuracy, S and O isotopes.



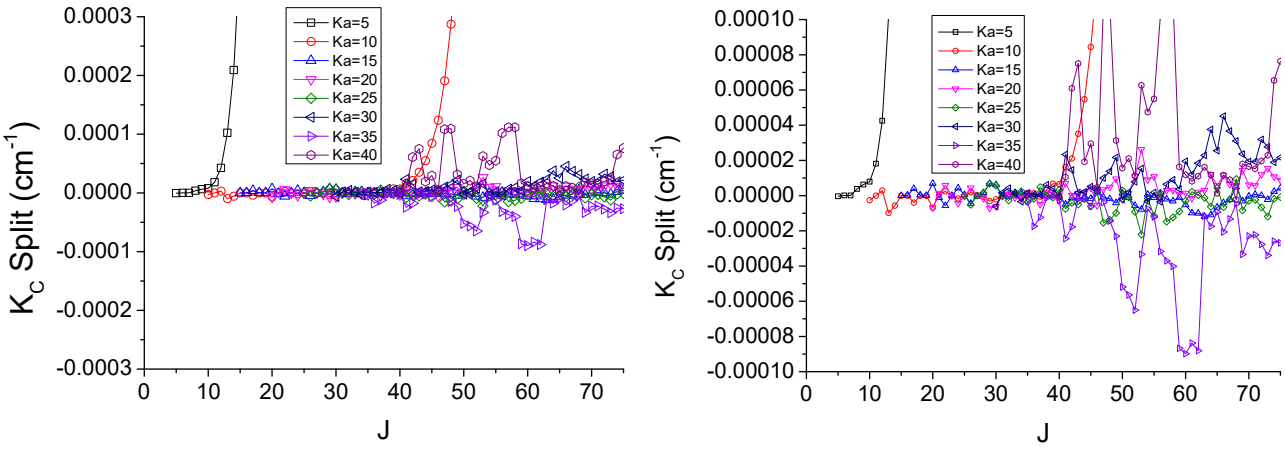
Each Sym set has 13,744 lines, and each Asym set has 40,311 lines, in the range of 0 – 370~410 cm^{-1}. There are 4 levels of uncertainty: 1E-4 cm^{-1}, 1E-5 cm^{-1}, 1E-6 cm^{-1} and 1E-7 cm^{-1}, and the SPFIT error bar is set at 5 times of uncertainty. Usually more data or smaller uncertainty requires more SPFIT cycles to find the "global minimum" in the phase space of EH parameters. Fig.S11 summarizes the percentage of lines successfully reproduced within the error bar. In general, heavier isotopes bring more lines fitted within the error bar, and the percentage of failed transitions rises on smaller uncertainty . For 1E-5 cm^{-1} and 1E-4 cm^{-1} cases, 96-100% lines can be reproduced, but some fitted EH parameters might not be tightly converged (Sec.4.2.2 and 4.2.5). 83-90% lines can be reproduced at 1E-6 cm^{-1} level. Even for the most difficult 1E-7 cm^{-1}

level Sym fits, we still have 57-70% lines that can be successfully reproduced. It should be noted that the Asym sets have fitted percentages higher than those of the corresponding Sym sets. Does this suggest the EH(Expt) of asymmetric isotopologues are slightly more reliable than symmetric ones? Not necessarily, but it is an interesting topic to explore in future.

Due to the nature of over-determined linear equations, we cannot 100% guarantee the "best" EH parameters have been determined for all the 1E-6 and 1E-7 fits. However, we can confidently claim one more layer of consistency: the consistency of SPFIT fitted percentages. In Fig.S11b, detailed fitted % are presented by S isotope (horizontal axis), dataset Asym (square) vs Sym (circle), accuracy level 1E-6 (open) vs. 1E-7 (half filled), and O isotope combinations (6 colors). In all isotopologue groups, the 6x8 isotopologues consistently have the lowest fitted %. Looking at the figure, two general trends are mixed together: for symmetric isotopologues, heavier O isotopes means higher fitted %, i.e. 8x8 > 7x7 > 6x6; for asymmetric isotopologues, less O atom (relative) mass differences means higher fitted %, i.e. 7x8 > 6x7 > 6x8.

Which factors affect the fitted % at higher accuracy level, i.e. 1E-6 and 1E-7 cm^{-1}? We have identified part of them from the K_c split analysis, Ames vs. EH model. Use the 668 ground state rotational levels, we compute all K_C splits at selected K_a and J up to 70, $\Delta = E(K_c=J-K_a) - E(K_C=J+1-K_a)$. Results are plotted along *J* in Fig.S12. K_a=5 and K_a=10 splits shoot up at J~12 and J~45, respectively. For K_a=15 and up, the magnitude of K_C split oscillation in J=40-70 region keeps getting larger. The pattern is clear: the K_C splits of odd K_a (15,25,35,…) go negative, while the K_c splits of even K_a (20,30,40) go positive. See Fig.S12b. As far as we know, such K_C splits of 1E-5~1E-4 cm^{-1} in high K_a region do not exist in the EH models we adopt in this work. For example, at J=60, all the K_C splits of $K_a \geq 15$ levels predicted by EH models are less than 1E-7 cm^{-1}, i.e. two orders of magnitude less than Ames K_C splits. Currently we are inclined to believe the higher K_a calculations may contain some numerical impurities that we should try to fix in future. Some EH(Ames) oscillations at higher J/Ka are larger than the rejection threshold, i.e. 5E-6~5E-7 cm^{-1}. This means the EH(Ames) models acquired at 1E-6~1E-7 cm^{-1} accuracy level are fitted with less transitions at higher J/Ka region. It also suggests better reliability for the fitted EH(Ames) constants, including both lower order terms (A/B/C + quartic) and higher order terms.

Fig.S12 The K_C splits of Ames SO_2 668 ground state levels. Split $\Delta = E(J, K_a, K_c=J-K_a) - E(J, K_a, J+1-K_a)$. Left side is overview; Right panel shows more details..



Next we check the convergence and quality of EH(Ames) constant, D_K.

4 Isotopologue Consistency Benchmark: D_K and Improved Prediction.

Except the rotational constants, the quartic centrifugal distortion constant D_K has the largest impact on prediction accuracy of rotational energy levels and line positions. The D_K values are plotted in Fig.13a. There are 8 different D_K for each isotopologue: two datasets (Asym vs. Sym) x 4 accuracy levels (1E-4~1E-7 cm^{-1}). The averaged D_K is taken as the zero line, all 8 values are adjusted with respect to it. Note the y axis unit is Hz, not MHz or kHz.

In Fig.S13a, from the 626 at left end to the 868 at the right end, the magnitude of D_K consistently decreases from 2.59 to 2.05 MHz, the spread of D_K values shrinks from ±20Hz to ±15Hz, while the ratio of spread/magnitude approximately remains the same. The spread is the difference between D_K(1E-4) and D_K(1E-7). It is evident that D_K monotonically drops

from 1E-4 cm^{-1} level to 1E-7 cm^{-1} level, which is probably the result of less higher J/Ka levels or transitions. An interesting observation is that Asym D_K are slightly higher than the Sym D_K, while their differences drop from ~10 Hz (1E-4 fit) to 1-2 Hz (1E-6 fit), then negligible (1E-7 fit). This is an indication of convergence: Asym vs Sym dataset difference can be minimized by raising the fitting accuracy level. The trend of $\delta[D_K(\text{Asym})-D_K(\text{Sym})]$ differences are clearly demonstrated by black squares in Fig.S13b.

Fig.S13. D_K analysis in the EH(Ames) SPFIT of 30 isotopologues, using Sym or Asym datasets, and four different uncertainty/accuracy level from 1E-4 cm^{-1} to 1E-7 cm^{-1}. (a) overview, by fitting accuracy level, data are with respect to the value averaged from 8 D_K; (b) the D_K convergence with respect to fitting accuracy: δ(Asym-Sym) are black squares, and $\delta(D_{K,}(n+1) - D_k(n))$ are triangles, where *n* is the accuracy/uncertainty level, e.g. 5-4 means D_K(1E-5) – D_K(1E-4).

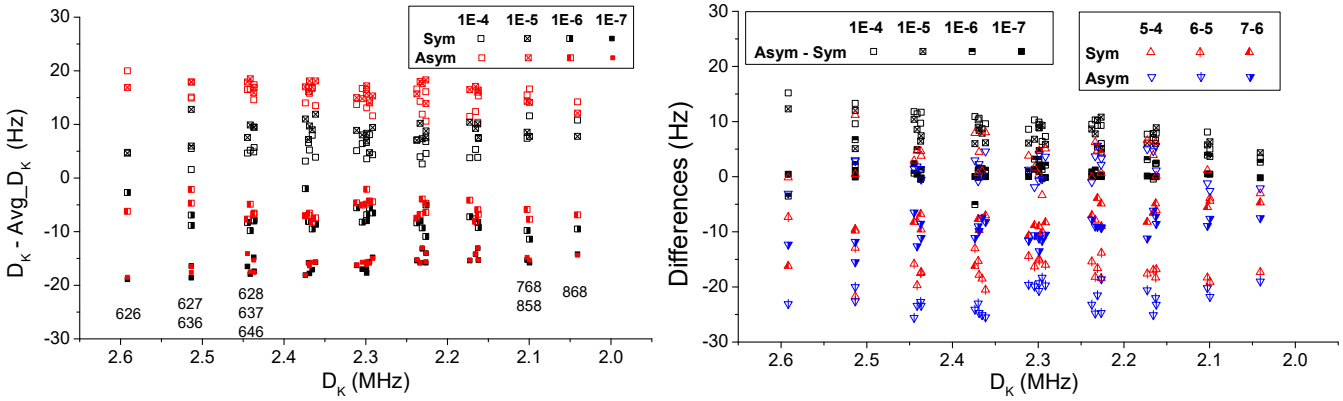


Fig.S13b also reports the D_K convergence with respect to fitting accuracy, $\delta(D_{K,}(n+1) - D_k(n))$, where *n* is the accuracy/uncertainty level, e.g. δ(6-5) = D_K(1E-6) – D_K(1E-5). Interestingly, the δ(5-4) are small, ±5Hz; the δ(6-5) are the largest, -25~-15Hz; then the δ(7-6) are reduced to -15~-5 Hz. This may indicate that the 1E-5 results are not well converged, while the 1E-6 and 1E-7 are more reliable.

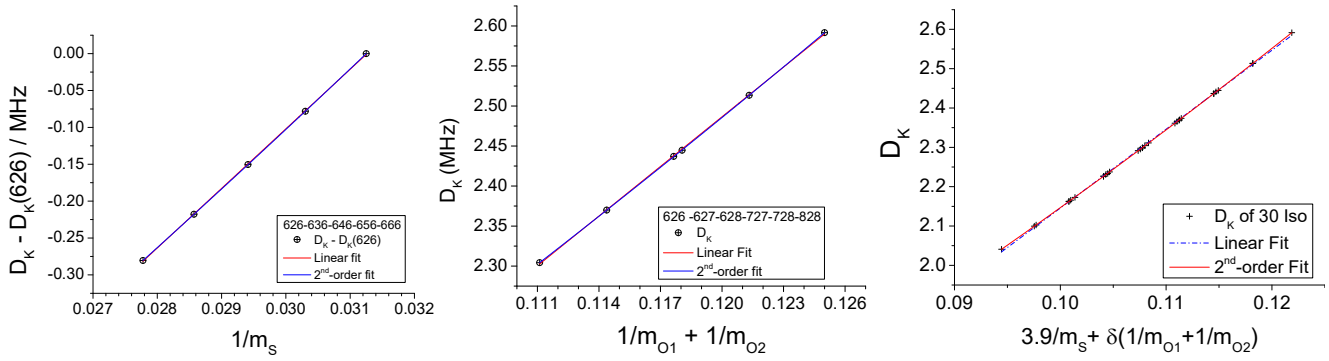


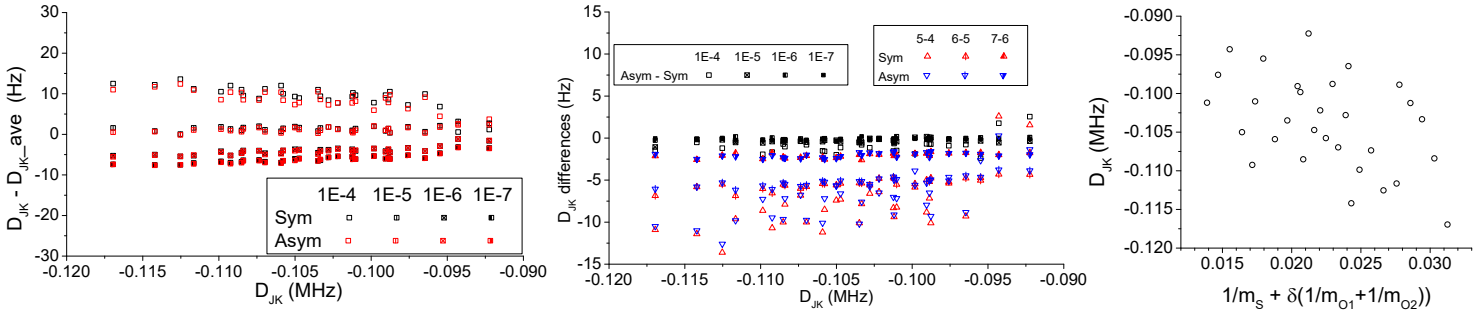
Fig.SS1. The linear and 2nd-order fits of EH(Ames) D_K values (determined at 1E-6 cm^{-1} level) along the S and O isotopes. (a). S isotope; (b) O isotope; (c) S & O isotope effects combined. See text for averaged fitting residuals and σ_{rms}.

Now it is time to re-check the validity of prediction in Fig.16. It turns out the simple approximation we used there was inadequate to distinguish 6x8 and 7x7 isotopologues. The x axis better uses the inverse of mass, instead of the mass itself. As shown in Fig.SS1, it works well. The S isotope effects and O isotope effects are in panel (a) and (b), respectively. The D_K values in panel (a) are with respect to the D_K(626), while panel (b) uses original D_K. The linear fits (red) and 2nd-order fits (blue) have fitting residuals similar to what we reported in Section 4.1. However, now we can separate 628 and 727 predictions, and both are reliable. Note Fig.SS1 and Fig.19 have same b and c panels.

Fig.SS1c tries to combine the S and O isotope effects. The scaling factor between $1/m_S$ and $(1/m_{O1}+1/m_{O2})$, 3.9, is handpicked so that it can reproduce the linearity reasonably well. The linear fit has relative deviations from -0.11% to +0.33%, with mean ± σ = 4E-4 ± 0.11%. The 2nd order fit has relative deviations in -0.038 ~ +0.024%, with mean ± σ = 0 ± 0.016%. It represents the isotopologue consistency of EH(Ames) D_K parameters, which can be utilized to predict or verify the experimental D_K of rare isotopologues. It is a better way to use the mass inverses.

On the other hand, not all EH constants can find such linear relations easily. The 2nd most important parameter at quartic level is D_{JK}. Fig.S14 reports the distribution of $\delta(D_{JK}-D_{JK}\_avg)$ in panel (a), the δ(Asym-Sym) and $\delta(D_{JK}(n+1)-D_{JK}(n))$ in panel (b). Similarly, along the horizonal axis, larger D_{JK} magnitude goes with larger D_{JK} spread on vertical direction, see (a). Along the vertical axis in (a), higher fitting accuracy leads to more negative D_{JK}, and smaller δ(Asym-Sym) differences. Note the D_{JK} δ(Asym-Sym)differences are one-order-of-magnitude smaller than D_K. In Fig.S14b, the $\delta(D_{JK}(n+1)-D_{JK}(n))$ monotonically rises from -12~-10 Hz to ~-2 Hz. Even with such tight convergence, we still failed to find a D_{JK}—mass inverse relation to align all D_{JK} on a single line, not even very approximately, see Fig.S14c. However, please remember what really matters is the pattern of the differences between EH(Expt) and EH(Ames), not either EH(Expt) or EH(Ames) only.

Fig.S14 D_{JK} in 30 EH(Ames) analysis, using Sym or Asym datasets, and four different uncertainty/error bar thresholds from 1E-4 to 1E-7 cm^{-1}. (a) different D_{JK} values, by fitting accuracy level and data type, with respect to the value averaged from 8 D_{JK}'s; (b) the D_{JK} convergence with respect to fitting accuracy: δ(Asym-Sym) are black squares, and $\delta(D_{JK,}(n+1) - D_{JK}(n))$ are triangles; (c) the D_{JK} overview of 30 SO_2 isotopologues.



5. Isotopologue Consistency Benchmark: Uncertainty of all 26 EH constants.

Follow the D_K and D_{JK} analysis, we check the uncertainty (or convergence) of other EH constants. In fact, there are two different "uncertainty" to discuss. The first kind is the width of data range for a specific EH constant from 8 SPFIT values, e.g. Fig.SS1a and Fig.S14a. The second kind of uncertainty refers to the numbers in the parentheses at the end of each constant, which is the uncertainty estimated for the specific over-determined least-squares problem. We call them the "constant uncertainty" and the "fitting uncertainty", respectively.

For each isotopologue, the "fitting uncertainty" are extracted from the SPFIT results at 1E-6 cm^{-1} or 1E-7 cm^{-1} level, then divided by the averaged value of corresponding constants. This is relative fitting uncertainty in %. The "constant uncertainty" is estimated by $\delta|$EH(1E-7)-EH(1E-6)| differences, which are divided by the corresponding constants to get the relative δ%. For each EH constant, the relative constant uncertainty and relative fitting uncertainties are averaged across 30 isotopologues and shown as the red squares in Fig.20a, see left Y axis (%). The EH parameter values on the *x* axis are also averaged from 30 isotopologues. This is appropriate in the log scale. Obviously, the constant uncertainty (squares) is the major factor, which are 2-4 orders of magnitude larger than the fitting uncertainties.

The constant uncertainty is also averaged and reported as blue circles with crosses, see the right axis in unit of Hz. Obviously one may try some linear approximation to get the relation between constant uncertainty (in Hz) and parameter values. For example,

$Log_{10}(\delta$ in Hz) = Log_{10}(EH parameter in MHz) x 5/7 - 11.

The trend of relative constant uncertainty % can be approximated, too, but might not be quite useful. In general, we have 0.01% ~ 1E-4% for quartic constants, 0.1~10% for sextic terms, and 1-100% for higher order terms. For many constants below 1E-7 MHz, their relative constant uncertainties (red squares, half filled) are much smaller than the "prediction" errors in Fig.17a. Note we already know those "predictions" in Fig.17a were unreliable because some constants were fixed at 626 values. In Fig.20a, the largest δ% goes with: ~10% for P_{KKJ}, P_{KKKJ}, and Φ_{JK}, ~20% for L_J and l_J, and 50-100% for ϕ_{JK} and l_{JK}. Further investigations at higher J and K_a may help clarify if we have adequate information to determine these terms, e.g. the ϕ_{JK} around 1E-8~1E-7 MHz. Some of those might be correlated with other EH parameters.

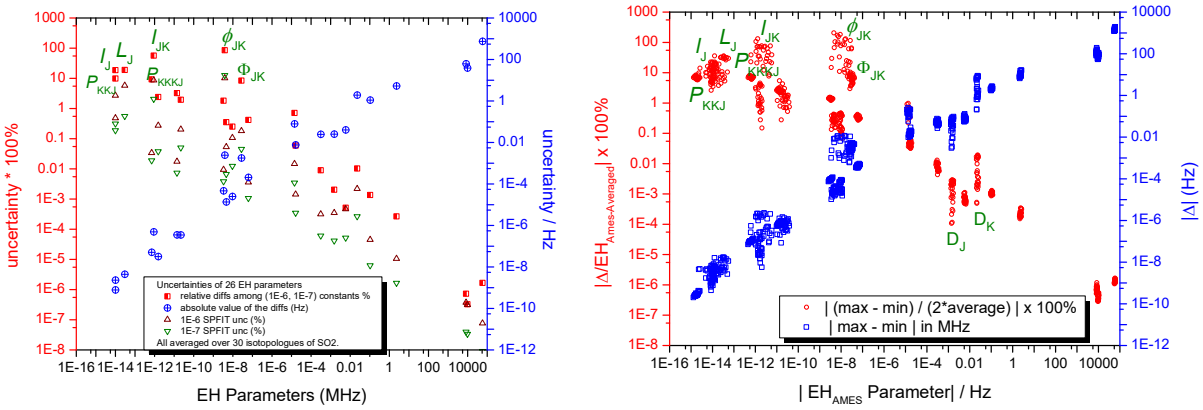


Fig.20 (a) The fitting uncertainty and constant uncertainty (convergence) of 26 EH(Ames) parameters, averaged from 30 SO_2 isotopologues. Left axis: the relative fitting uncertainty % from SPFIT (triangles), and $\delta\%|EH(1E\text{-}6) - EH(1E\text{-}7)|$ (red squares, half filled); Right axis: $\delta|EH(1E\text{-}6) - EH(1E\text{-}7)|$ in Hz; (b) constant uncertainty of 26 EH(Ames) parameters in the 1E-6 cm^{-1} and 1E-7 cm^{-1} SPFITs of 30 isotopologues. Right axis: absolute uncertainty in Hz (blue squares); Left: relative constant uncertainty % (red circles).

From another perspective, the averaged δ and $\delta\%$ values in Fig.20a can be taken as an indicator for the compatibility between Ames datasets and the EH models, i.e. less compatibility leads to larger uncertainty. To have a more precise understanding about the consistency vs. uncertainty, Fig.20b reports the δ and $\delta\%$ for all 26 EH constants of all 30 isotopologues. The spread of EH constants in panel (b) is affected by two factors: the isotopologue difference, and the EH model or data inadequacy. In other words, they are molecule specific and model/data specific, but some general trends may be still recognizable. For example, in the "lower order terms" which include rotational constants and quartic level constants, the $\delta\%$ variation of D_J and D_K is as wide as more than 1 order of magnitude, and significantly wider than those of rest. Fortunately, even the largest $\delta\%$ are less than 0.05% (D_K) or 0.005% (D_J), since our transition datasets are more than adequate for D_J or D_K determination. The EH(Ames) parameters are included in supplementary file and also available upon request. Interested readers can apply unique color & symbol to every EH constants and/or every isotopologue, for further explorations.

This uncertainty check sheds some lights on future EH analysis. For example, not all the "higher order terms" have similar reliability or relative variations. Even with ~40,000 transitions (J/K_a up to 60/35), not all EH parameters can be precisely determined. For experimental data based EH model analysis using a few thousands or even just hundreds of transitions, the constant uncertainty may be larger than reported.

6. Isotopologue Consistency Benchmark: Higher order terms.

In CDMS analysis,[2] all 626 EH constants were fit with non-zero weights, then part of the fitted EH higher order terms were used for other minor isotopologues, including 636, 646, 627 and 628 (not for 828). Therefore, the higher order terms of other minor isotopologues are inconsistent from one to another. Since we have studied the isotopologue consistency of EH(Ames) constant uncertainties, we can compare the higher order terms, EH(Ames) vs. EH(CDMS), checking if there exists any pattern or trends.

To do that, the EH(Ames) lower order terms need to be replaced by EH(CDMS) values. We re-run the "prediction" procedure for 646, replace the A/B/C constants and quartic level constants, feed different EH sets to SPCAT and predict line positions, compare the predictions to the pure EH(CDMS/Expt) based reference predictions, then plot the line position differences in Fig.21a. The two terms before and after "+" refer to the source of "lower order" and "higher order" terms, respectively, while "A/B/C corrected" means the A/B/C constants in EH(Ames) models have been replaced with EH(CDMS/Expt) values. The EH(Ames) parameters set being used here was fit from Asym datasets with 1E-7 cm^{-1} accuracy level, which we believe have been converged to better than 0.003 MHz, and more appropriate for the low J/K_a line positions.

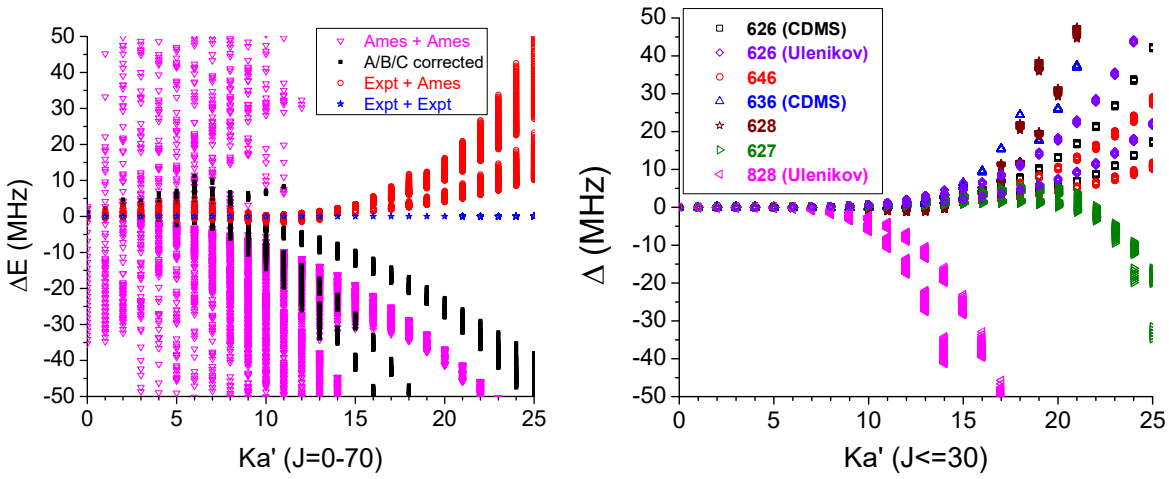


Fig.21. Line position differences caused by A/B/C, quartic level terms, and higher order terms. (a) SPCAT line position predictions using 4 sets of EH(646) parameters, with respect to pure EH(CDMS/Expt) model based line position predictions. See text for explanations. (b). the EH(Expt + Ames) model based line positions for 6 isotopologues, with respect to the pure EH(CDMS/Expt) model based SPCAT line position predictions.

As expected, the A/B/C constant is the primary source of deviations, and quartic centrifugal distortion constants are the 2nd most important factor with significant enough contributions. The "Expt + Ames" set based line positions do carry 0-5 MHz accuracy for $K_a' \leq 15$ transitions. Deviations may rise up to 5-15 MHz at $K_a'=20$. Note in Fig.21a the J goes up to 70, so we can see higher-J related deviations at both $K_a \sim 5$ and $K_a=20$-25. If J is limited to ≤ 30, the deviations at $K_a'=15$ would be only 1-2 MHz, see Fig.21b.

Same procedure is repeated for minor isotopologues for which EH(CDMS/Expt) analysis are available. Results are all presented in Fig.21b, with $J \leq 30$ limit. Either 626 → 636 → 646 or 626 → 627 → 628, no obvious trend was found. If we multiply the 628 deviations by -1, the Δ would have apparent trend going from positive to negative, along with the O isotope substitution 626→627→628→828. Here we assume the EH(CDMS) is OK for 626, but we can notice the differences between the EH(CDMS) model and the improved EH(Expt) model from Ulenikov group. Please note those differences are beyond the scope of this study, as we just noticed a L_{JJK} related discrepancy between the CDMS EH model [2] and Ulenikov EH model [1] for 626.[3]

Short conclusion: We have confirmed again the possibility to have 0-10 MHz prediction accuracy for isotopologue MW line positions, but current EH(CDMS/Expt) models do not have noticeable trend or consistency along the isotope substitutions. A more appropriate "apple-to-apple" comparison needs the EH(Ames) analysis to be done in the same way as in EH(Expt/CDMS) for minor isotopologues.

7. Isotopologue Consistency Benchmark: Fixed vs. Relaxed.

By fixing part of "higher order" terms at 626 values, we re-compute EH(Ames) constants for 636, 646, 627 and 628, denoted as "Fixed", and compare the values of rest "not fixed", i.e. relaxed terms with the original fully "Relaxed" EH(Ames) constants. The differences Δ are defined as X(Fixed) – X(Relaxed). For each isotopologue, we computed three sets of Δ at 1E-5, 1E-6 and 1E-7 levels. Then the Δ are divided by original "Relaxed" values to get Δ/EH ratios, or the relative Δ.

Fig.S15a has results for 646 EH constants. In general, smaller constants have larger Δ differences, larger Δ/EH ratios, which indicate less reliability for the fitted EH parameters. Compare to 1E-5 cm^{-1} results, the 1E-6 cm^{-1} and 1E-7 cm^{-1} parameters usually give smaller Δ/EH ratios. For A/B/C constants, there is no doubt the 1E-7 cm^{-1} level fits have the least impact of fixing part of higher order terms, i.e. Δ% in 10^{-8}~10^{-10}. It is similar or slightly less than the constant uncertainty shown in Fig.20. For other terms, the order of Δ or Δ% is hard to predict along with higher accuracy from 1E-5 cm^{-1} to 1E-7 cm^{-1}. For example, the P_K at the left end of Fig.S15a have "Fixed vs. Relaxed" values (x 1E+11) as: 1.50 vs. 1.45 (1E-5), 1.71 vs. 1.57 (1E-6), and 3.23 vs 1.65 (1E-7). The Δ% of Φ_K and L_K also follow such reversed order, i.e. rising along 1E-5 → 1E-6 →

1E-7. Compared to the constant uncertainty we discussed in Sec.4.2.5, the "Fixed" effects of these three EH terms (P_K, Φ_K and L_K) are notably larger, even by orders of magnitude. Another example is D_K. From 1E-5 cm^{-1} to 1E-7 cm^{-1}, the D_K(Fixed) increases from 2.44124 MHz to 2.44132 MHz, while the original D_K(Relaxed) decreases from 2.44129 MHz to 2.44126 MHz. Their trends are just opposite.

Fig.S15 (a-b): Absolute (Left y axis, circles) and relative (Right y axis, squares) differences of the EH parameters acquired between fully relaxed fit (all vary freely) and the partially fixed fit (some parameters are fixed at 626 values): (a) 646 only; (b). the relative differences (%) of 646, 636, 628, 627, and 828, at 3 fitting accuracy levels.

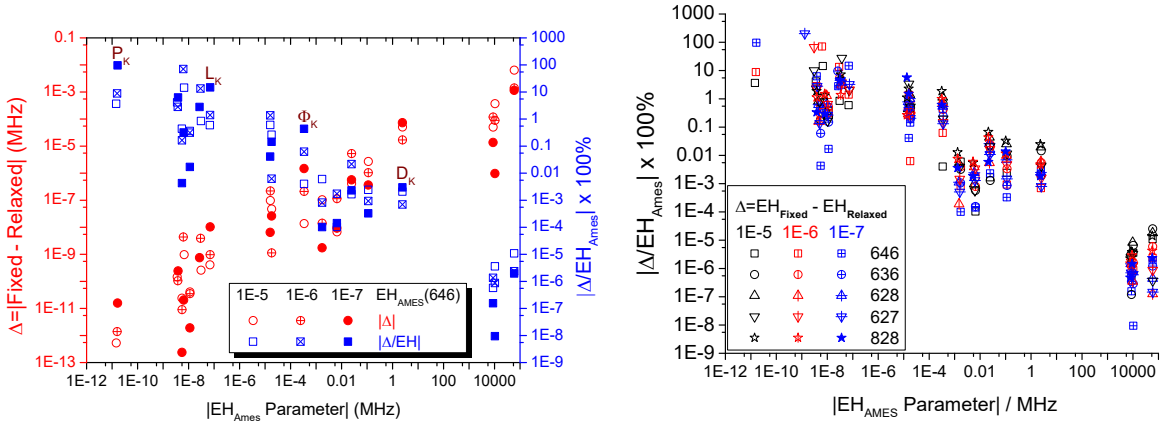


Fig.S15b shows the relative Δ% of 5 minor isotopologues at 3 fitting accuracy levels. In general, the effects on A/B/C constants are insignificant. The Δ% of 1E-7 cm^{-1} level fits are usually the smallest, but with exceptions. Using higher J/K_a dataset with more systematic SPFIT investigations may help find a path to optimize the EH model or determine certain parameters with higher confidence. It uses different filling pattern to separate the three fitting accuracy levels. Note the 828 fits in b panel were generated by fixing the same EH term set that was fixed in EH(628) fits. They are NOT the 828 fits we discuss in next section, which use a "prediction" scheme to estimate higher order terms.

8. Isotopologue Consistency Benchmark: Higher order term prediction.

The EH(828) analysis in this section follows the procedure given in the section 3.2 of Ulenikov et al [1], which utilizes the 828 vs. 626 ratios of A and C constants to estimate other 828 EH constants. The isotope substitution theory [4,5,6] predicts the centrifugal distortion constants as:

$$X_{828}^{ij} = X_{626}^{ij} * \alpha_A^i * \alpha_C^j \,,\ \alpha_A = A_{828}/A_{626},\ \alpha_C = C_{828}/C_{626},$$

For a specific EH constant, if its total power is *m*, the power of its J_Z^2 part is *i*, and *j* = *m*/2 - *i*.

In Ref.1, this rotational-constant dependent prediction formula is applied using the established 626 A/B/C, and the 828 A/B/C constants fit from J=0-4 transitions (to get α_A and α_c). In their EH(828) least-squares fit, all the higher order terms are fixed at the predicted values, except H_K and H_{KJ}.

Since part of our goal is to report our capability to achieve high prediction accuracy on isotopologues, it is definitely necessary to run parallel check and determine which prediction is more accurate. Compare to experimental EH models, our biggest advantage is we have 30 isotopologue IR lists which are computed with a consistent VTET [7] parameter set and fitted to a uniform EH model with accuracy control and satisfactory constant uncertainties (Fig.20). The Ames isotopologue consistency is far better than experimental EH models which were constructed independently from each other and with various limitations. We can take EH(Ames) model fits of any two symmetric isotopologues (the formula used in Ref.1 is for symmetric C_{2v} type), predict one set from the other set, compare to the original EH(Ames) constants that we take as reliable reference, calculate relative errors, and collected in Fig.22a. This plot may combine with Fig.20 to get more interesting results, while Fig.20 includes both symmetric and asymmetric isotopologues.

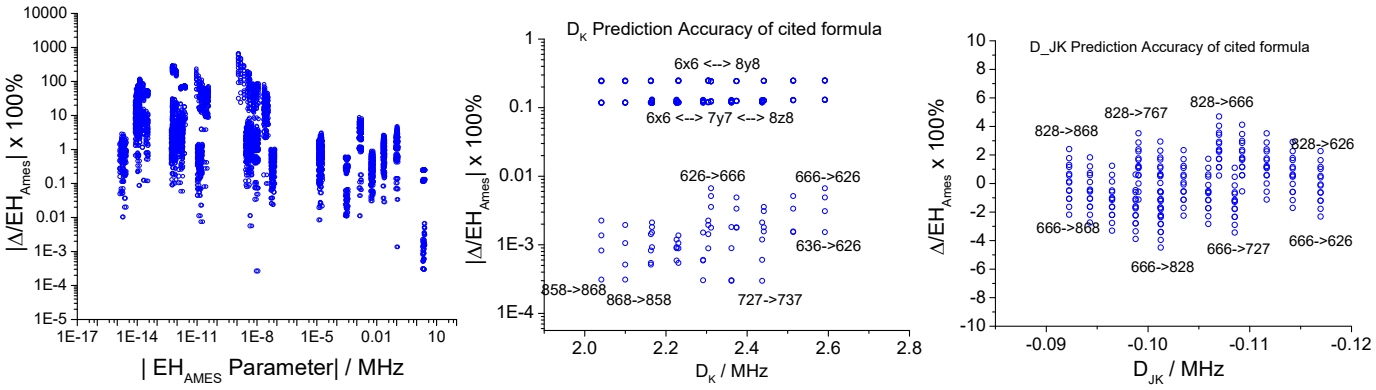


Fig.22 Accuracy benchmark for the prediction formula Ref.1 adopted in their EH(828) model. The 30 SO_2 isotopologue EH(Ames) parameter sets acquired at 1E-6 cm^{-1} level are used to compute the relative prediction deviations Δ%. (a) overview for 30 isotopologues; (b) the D_K prediction accuracy; (c) the D_{JK} prediction accuracy.

Look at the performance of prediction formula in Fig.22a. Except for D_K, quartic term predictions have relative deviations $|\delta|$% in 0.1 ~ 10%. For sextic and higher order terms, the deviations are noticeably higher, $|\delta|$ = 0.1 ~ 100%. Apparently, they all spread in a wide range of 2-3 orders of magnitude. But the log scales is not an ideal choice if we want to find out how the deviations vary from one isotopologue to another. So in Fig.22b, $|\delta(D_K)|$ terms are plotted with linear scale on horizontal axis. Obviously, there are two groups of $|\delta(D_K)|$: the 0.1-0.3% group for O isotope substitutions, and 0.0003-0.007% group for S isotope substitutions. In second group, heavier O isotopes are associated with smaller $|\delta|$, e.g. 868. This probably relates to the effective mass changes between two isotopologues, which vary by specific EH constants.

In panel (a), we can only tell the $|\delta|$ of D_{JK} in the range of 0.002~10%. By changing the vertical axis to linear scale, and using original δ instead of $|\delta|$, the Fig.22c clearly demonstrates how the δ correlates with S and O isotope substitutions. The isotope effects on the prediction deviations are linear again in the range of 0-5%, similar to the patterns we observed in Section 3 or Fig.SS1 (and Fig.19b,c). With such EH(Ames) based deviation patterns, it is possible to further calibrate the D_{JK} predictions, if D_{JK} can be accurately determined for a few isotopologues. But many other higher order terms do not have such systematic $|\delta|$ patterns.

Compare Fig.20 and Fig.22, it is clear our EH(Ames) fits do provide far better results for quartic level centrifugal distortion constants. For higher order terms, both the average magnitude and the spread of EH(Ames) constant uncertainties in Fig.20 are usually smaller than those of $|\delta|$ here. In principle, the performance of the α_A/α_C based formula is not bad at all, with mass-difference dependent accuracy. However, the formula requires knowledge of A/C, and our test is done with "accurate" rotational constants. If one use roughly estimated A/B/C, the scaling factor $\alpha_A^i\alpha_C^J$ will inherit and magnify the A/C deviations, which will add up to the method-dependent deviations we report in Fig.22. Even with the "exact" A/C constants, the quartic level constants from EH(Ames) are still 2-4 orders of magnitude more accurate than those predicted by the A/C dependent formula.

In principle, we are inclined to believe the formula predictions can serve as useful tool to check the noise level of *some* higher order EH constants. But in reality, there is a bottleneck which obviously requires much more thorough investigation, i.e. the quality of higher order EH(Expt) terms which scientists need to take as standard reference. They must come from reliable least squares fitting of another isotopologue, to which our findings in Fig.20 and related discussions are surely applicable. The SPFIT fitting residuals of our 1E-7 cm^{-1} EH(Ames) fits are less than 5E-7 cm^{-1} or 15 kHz, and σ_{RMS} ~ 10 kHz. We believe this is close to the accuracy level of some experimental MW studies. The EH(expt) constant uncertainty will add to the method deviation patterns, if it does not ruin the latter first. In our most recent studies on 626 [3] and 636 [8], if one higher order term was wrong but used in EH analysis of any isotopologue, it would cause collateral damages on other correlated parameters or even noticeably alter the value of rotational constants.

[1] Ulenikov ON, Bekhtereva ES, Krivchikova YV, et al. *High resolution analysis of 32S18O2 spectra: The $v1$ and $v3$ interacting bands*. J Quant Spectrosc Radiat Transf. 2015;166:13-22. doi:10.1016/j.jqsrt.2015.07.004.

[2] Online EH(CDMS) model data https://www.astro.uni-koeln.de/site/vorhersagen/daten/SO2/ : 34SO2/s34.par for 646; SO2/alt.1/so2.par for 626; 33SO2/s33.par for 636; SO18O/o18.par for 628; SO17O/17.int for 627.

[3] Huang X, unpublished work and private communications (2018)

[4] Bykov AD, Makushkin YuS, Ulenikov ON. On isotope effects in polyatomic molecules: some comments on the method. J. Mol. Spectrosc. 1981;85:462–79.

[5] Makushkin YuS, Ulenikov ON. Isotopic relationships for polyatomic molecules. Opt. spektrom. 1975;39:629–36.

[6] Bykov AD, Makushkin YuS, Ulenikov ON. On the displacements of centers of vibration-rotation bands under isotope substitution in polyatomic molecules. J. Mol. Spectrosc. 1982;93:46–54

[7] Schwenke DW, Variational calculations of rovibrational energy levels and transition intensities for tetratomic molecules. J. Phys. Chem. 1996; 100:2867-2884.

[8] Huang X, unpublished work and private communications (2017)