Uncertainties Associated with Theoretically Calculated \( \text{N}_2 \)-Broadened Half-Widths of \( \text{H}_2\text{O} \) Lines

Q. Ma\textsuperscript{a)}

NASA/Goddard Institute for Space Studies and Department of Applied Physics and Applied Mathematics, Columbia University, 2880 Broadway, New York, NY 10025

R. H. Tipping

Department of Physics and Astronomy, University of Alabama, Tuscaloosa, AL 35487

R. R. Gamache

Department of Environmental, Earth and Atmospheric Science, University of Mass. Lowell, Lowell, MA 01854.

Abstract

With different choices of the cut-offs used in the theoretical calculations, we have carried out extensive numerical calculations of the \( \text{N}_2 \)-broadened Lorentzian half-widths of the \( \text{H}_2\text{O} \) lines using the modified Robert-Bonamy formalism. Based on these results, we are able to thoroughly check the convergence. We find that with low-order cut-offs commonly used in the literature, one is able to obtain converged values only for lines with large half-widths. Conversely, for lines with small half-widths much higher cut-offs are necessary to guarantee the convergence. We also analyze uncertainties associated with calculated half-widths, and these are correlated as above. In general, the smaller the half-widths are, the poorer the convergence and the larger the uncertainty associated with them. One can divide all \( \text{H}_2\text{O} \) lines into three categories: large, intermediate and small according to their half-width values. One can use this division to judge whether the calculated half-widths are converged or not, based on the cut-offs used, and also to estimate how large their uncertainties are. We conclude that with the current Robert-Bonamy formalism for lines in category 1, one can achieve the accuracy requirement set by HITRAN while for lines in the category 3, it is absolutely impossible to meet this goal.

1. Introduction

The modeling of the atmosphere from satellite-based, balloon-based, and Earth-based instruments requires an accurate spectroscopic database such as HITRAN [1]. This widely-used database has spectral line data for the most important molecules in bands from the microwave to the ultraviolet spectral regions. Some of the data for the water vapor lines are experimental, but because of the large number of lines and the ambient temperature range in different parts of the atmosphere, many have to be calculated theoretically. The accuracy of these calculated values depends on many factors such as the line-shape theory used, the sophistication and accuracy of the intermolecular potential model, the need to obtain converged results, etc. For instance, for accurate atmospheric...
applications involving the very important water vapor molecule, it is desirable to know the half-widths and their temperature dependence better than a 3% uncertainty for strong lines and 10% for weaker lines [2??]. Unfortunately, of all spectroscopic parameters in HITRAN for H\textsubscript{2}O, the pressure-broadened self and air half-widths and their temperature dependences contain the largest uncertainties. Recently, in order to improve this situation by combining all available data both experimental measurements and theoretical calculations, the H\textsubscript{2}O database in HITRAN has been updated [2]. In developing this updated version in 2006, in cases where there are no half-width values to the desired accuracy available in the vast experimental literature, theoretically calculated values have been adopted. The latter have been derived based on the Complex Robert-Bonamy (CRB) line-shape theory and potential models consisting of a short-range site-site interaction and long-range interactions by Gamache and his collaborators [3-6]. This new database represents our cutting-edge knowledge of these parameters. Since then, attempts to refine the theoretical calculations continue [7].

The Robert-Bonamy (RB) formalism was developed more than three decades ago [8]. In comparison with previous formalisms, the main features of this theory are the non-perturbative treatment of the S scattering matrix though the use of the linked cluster theorem, and a convenient description of classical trajectories for large impact parameters as well as for the closest approach. However, the RB formalism contains several basic assumptions whose applicability was not thoroughly verified. Since then, the formalism has been widely used in calculating the Lorentzian half-widths and shifts for spectral lines for many molecular systems. Some progress has been made recently to improve the original version, but most of these efforts have been related to the extension to more complicated potentials, vibrational dependences, more accurate trajectories, etc. None has focused on making further improvements to its core part. As a result, the calculations using the RB formalism have become routine. In cases where the calculated values did not match the experimental data well, the strategy commonly used for improvements was to tune parameters, rather than to scrutinize the RB formalism itself. Given the fact that the current RB formalism remains almost the same as 30 years ago, one can ask if this theory can meet the cutting age requirement from HITRAN? Furthermore, in order to make the calculations tractable with potentials including short-range site-site models, one usually introduces two cut-offs to limit the number of terms considered, although until now no thorough convergence checks have been made. Thus, it has becomes necessary to scrutinize current theoretical calculations to address this problem.

In Gamache’s research works, both the potential models used by them and their final results are available [7]. Thus, it is possible for others to make independent verifications. In the present study, we repeat and extend their calculations beyond their choices for cut-offs. Based on these calculations, we find that one can divide the H\textsubscript{2}O lines into three categories according to the values of the half-widths. It turns out that this division coincides with a division according to uncertainties associated with the calculated values. The smaller the half-width is, the larger the uncertainty. For lines with large half-width values, by using a realistic potential model one is able to use the RB formalism and meet the accuracy requirement set by HITRAN, but for lines with small half-widths, the RB formalism definitely does not meet this requirement.

2. General formalism of the line half-width
2.1. Brief outline of the RB formalism and convergence challenges

Within the modified RB formalism, the expressions of the half-widths is given by [9]
\[
\gamma = \frac{n_e}{2\pi c} \int_0^{\infty} v f(v) dv \int_0^{\infty} 2\pi b db [1 - \cos(S_i + \Im S_j) e^{-\Re S_j}]
\]
\[
\approx \frac{n_e v}{2\pi c} \int_0^{\infty} 2\pi b db (1 - e^{-\Re S_j}) \int_0^{\infty} 2\pi \left( \frac{v'}{v} \right)^2 r_c dr_c (1 - e^{-\Re S_j}),
\]
where \(S_1\) and \(S_2\) are the first and second orders of the perturbation expansion of the Liouville operator \(\hat{S} = S_1 \cdot S_2^*\), where \(S_1\) and \(S_2\) are scattering matrices in Hilbert space. In the above expression, \(f(v)\) is the Maxwell-Boltzmann distribution function, \(\nu\) is the average velocity, \(r_c\) is the closest distance of approach along trajectories, \(r_{c,\min}\) is the minimum of \(r_c\) corresponding to strictly head-on collisions, and the “apparent” velocity \(v'_c\) is defined as [8]
\[
v'_c = v \left[1 + \frac{8e}{mv^2} \left[5\left(\frac{\sigma}{r_c}\right)^2 - 2\left(\frac{\sigma}{r_c}\right)^6\right]\right]^{1/2}.
\]
Usually, \(S_2\) is represented by three terms labeled as \(S_{2,\text{outer},1}\), \(S_{2,\text{outer},2}\), and \(S_{2,\text{middle}}\), respectively. In the present study, we follow the same custom, and as an example we show how to calculate \(S_{2,\text{outer},1}\) in detail here. The original expression for \(S_{2,\text{outer},1}\) is given by
\[
S_{2,\text{outer},1} = \frac{1}{\hbar^2 (2j_1+1)} \sum_{l_{z1}} \rho_{l_{z1}} \int_{-\infty}^{\infty} dt \int_{-\infty}^{t} dt' \sum_{(m)} j_{l_1} \tau_{l_1} m_{l_1} m_{l_2} \langle V(\bar{R}(t)) \bar{V}(\bar{R}(t')) | j_{l_1} \tau_{l_1} m_{l_1} m_{l_2} >
\]
where the quantum numbers \(j_{l_1}\) and \(\tau_{l_1}\) specify the energy levels of the initial states of the \(\text{H}_2\text{O}\) line of interest and the Hilbert interaction operator \(\hat{V}(\bar{R}(t))\) is defined by
\[
\hat{V}(\bar{R}(t)) = e^{i(H_{e}+H_{b})t/\hbar} V(\bar{R}(t)) e^{-i(H_{e}+H_{b})t/\hbar}.
\]
Usually, in order to evaluate the potential matrix elements in Eq. (3), one prefers to express \(V(\bar{R}(t))\) in terms of the standard spherical expansions
\[
V(\bar{R}(t)) = \sum_{L_1 L_2 L; K, R} u(L_1 L_2 L; K, R(t)) \times \sum_{m_1 m_2 m} C(L_1 L_2 L, m_1 m_2 m) D^{L_1}_{m_1 k_1} (\Omega_1) D^{L_2}_{m_2 k_2} (\Omega_2) Y^*_{l m}(\Omega(t)).
\]
In the present study, we adopt potential models consisting of a long-range dipole-quadrupole and quadrupole-quadrupole interactions \(V_{dq}(\Omega_1, \Omega_2, \bar{R}) + V_{qq}(\Omega_1, \Omega_2, \bar{R})\) and a short-range interaction \(V_{\text{atom-atom}}(\Omega_1, \Omega_2, \bar{R})\) modeled by the site-site model. For \(V_{dq}\) and \(V_{qq}\), their spherical expansions are well known and numbers of their expansion terms are very limited. For \(V_{dq}\), the expansion contains only one term (i.e., \(L_1 = 1, K_1 = 0, L_2 = 2, \text{and} L = 3\) which varies with \(R\) as \(R^3\)). Meanwhile, for \(V_{qq}\) it has three terms (i.e., \(L_1 = 1, K_1 = 0, \pm 2, L_2 = 2, \text{and} L = 3\) which vary as \(R^5\).

With respect to \(V_{\text{atom-atom}}\), the site-site model expression is given by
where \( \sigma_{ij} \) and \( \epsilon_{ij} \) are parameters and \( r_{ij} \) are distances between the i-th atom of the absorber molecule a (i.e., H\(_2\)O) and the j-th atom of the bath molecule b (i.e., N\(_2\)). In order to evaluate its matrix elements, one needs to rewrite \( V_{\text{atom-atom}} \) in terms of the standard spherical expansions

\[
V_{\text{atom-atom}}(\Omega_1, \Omega_2, \vec{R}) = \sum_{i \neq a} \sum_{j \neq b} 4\epsilon_{ij} \left( \frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right),
\]

(6)

where \( \sigma_{ij} \) and \( \epsilon_{ij} \) are parameters and \( r_{ij} \) are distances between the i-th atom of the absorber molecule a (i.e., H\(_2\)O) and the j-th atom of the bath molecule b (i.e., N\(_2\)). In order to evaluate its matrix elements, one needs to rewrite \( V_{\text{atom-atom}} \) in terms of the standard spherical expansions

\[
V_{\text{atom-atom}}(\Omega_1, \Omega_2, \vec{R}(t)) = \sum_{L_1} \sum_{L_2} \sum_{n_{ij}} \sum_{wq} \frac{U(L_1 K_1 L_2 L_n_{ij}, wq)}{R^{L_1+L_2+q+2w}}
\]

\[
\times \sum_{m_1 m_2 m} C(l_1 l_2 m_1 m_2 m) D_{m_1 K_1}^{l_1}(\Omega_1) D_{m_2 0}^{l_2}(\Omega_2) Y_{L_m}^*(t),
\]

(7)

where \( n_{ij} \) runs over all pairs of atoms in Eq. (6), \( q = 6 \) or 12, \( w \) is an integer index from 0 to infinity, and the definitions for \( U(L_1 K_1 L_2 L_n_{ij}, wq) \) can be found in the literature. It turns out that except that \( L_1 + L_2 + L \) and \( L_2 \) must be even, and the well-known requirements from the Clebsch-Gordan coefficients and the rotational Wigner matrices, no other restrictions on them are enforced. Meanwhile, although choices of \( q \) and \( n_{ij} \) are limited, the number of combinations of \( q \) and \( n_{ij} \) with all possible values of \( w \) are infinite. Therefore, one must introduce cut-offs.

There are two kinds of cut-offs associated with two summation indices in Eq. (7). The first is a cut-off to limit the set of irreducible tensor rank \( L_1 \) plus a subsidiary \( K_1 \), and ranks \( L_2 \) and \( L \) needed in the expansions. For example, one can choose 2 as upper limits for both of \( L_1 \) and \( L_2 \). By enforcing these restrictions for \( L_1 \) and \( L_2 \), selections for two remaining \( L \) and \( K_1 \) are also limited. The second kind of cut-off is to set an upper limit for the index \( w \). Thus, the possible choices of the combinations of \( q \), \( n_{ij} \), and \( w \) become finite. In the literature, if one chooses 8 as the maximum of 2w, the cut-off is said to be 8-th order. Usually, a justification of the usage of the selected cut-offs can be made by an estimation of how small the magnitudes of correction terms in the spherical expansion of \( V_{\text{atom-atom}} \) would be, in comparison with the leading terms. Thus, it appears that the expansion problem can be easily solved.

Unfortunately, by taking these steps the trouble does not go away because in evaluating \( S_{2,\text{outer},i} \) one needs to consider the products of two potential operators. For later convenience, we add subscript \( V_{\text{tot}} \) to emphasize that it is the total potential (i.e., \( V_{d \bar{q}} + V_{q \bar{q}} + V_{\text{atom-atom}} \)). Indeed, considerations of the products of two potentials makes a big impact on deciding if the cut-offs are adequate. The obvious one is that the total number of summation terms for the products of two \( V_{\text{tot}} \) is the square of those for an individual \( V_{\text{tot}} \). We know that the number of terms of \( V_{\text{tot}} \) is mainly determined by the number of terms of \( V_{\text{atom-atom}} \). Thus, if the number of terms for \( V_{\text{atom-atom}} \) is large, the total number for the product of two \( V_{\text{tot}} \) is much larger.

The second impact is more subtle than the first and it results only from the second kind of cut-off. Due to different symmetries, there are no couplings between two potential terms whose irreducible tensor ranks are not identical. As a result, it is certain that weak terms ignored in evaluating \( V_{\text{atom-atom}} \) by the first cut-off can only make small contributions to \( S_{2,\text{outer},i} \). Then, a criterion established by checking the convergence of \( V_{\text{atom-atom}} \) for this cut-off can be, more or less, applicable for \( S_{2,\text{outer},i} \). In contrast, with respect to the second kind of cut-off, there are couplings between terms ignored by the
cut-off in evaluating $V_{\text{atom-atom}}$ and terms considered with the same tensor ranks in $V_{\text{tot}}$. We would like to emphasize here that the latter consists not only of the terms remaining in the spherical expansion of $V_{\text{atom-atom}}$, but also components labeled by the same tensor ranks in the expansions of $V_{dq}$ and $V_{qq}$. As a result, to ignore weak terms in $V_{\text{atom-atom}}$ using the second kind of cut-off could cause significant errors because these could make significant contributions to $S_{2,\text{outer},i}$ through couplings with other strong remaining components. Thus, one can conclude that it becomes more difficult to establish a reliable convergence criterion because the smallness of the terms in $V_{\text{atom-atom}}$ ignored by the second kind of cut-off itself is not adequate to justify the convergence of $S_{2,\text{outer},i}$.

Above, we have outlined the usual method used in calculating $S_{2,\text{outer},i}$. A drawback of this method is that one could encounter convergence troubles when site-site potential models are considered. In these cases, to adopt higher cut-offs in the spherical expansion and to consider all contributions including couplings between terms with the same categories of tensor ranks whose origins could be one of any type (e.g., the long-range electrostatic, the long-range induction, the short-range site-site interactions, etc.) requires very much tedious work and numerical evaluations of a large number of resonance functions so that one may not be able to obtain converged results.

In summary, the convergence over these two cut-offs could become a formidable obstacle in practical calculations of Re($S_2$) for molecular pairs except for large separation distances where the electrostatic interaction is overwhelmingly dominant. Careful readers may notice that at this stage we have avoided the use of the phase “convergence of calculated half-widths.” Although half-widths are determined by calculating Re($S_2$), but the convergence of Re($S_2$) and the convergence of half-widths are not always the same. We will return to this subject later. In conclusion, most previous theoretical calculations have been carried out using low-order cut-offs and their accuracy could be limited by convergence problems; thus a better way to deal with this problem is needed.

2.2. Coordinate representation and new formalism

The possibility of convergence errors forces one to wonder if this method is the best way to proceed. Based on our experience in dealing with complicated potential models in treating far-wing line shapes and other problems [10,11], we know that the coordinate representation used in those studies has advantages in dealing with complicated potentials.

With the standard method, the basis set in Hilbert space is constructed from $| j, \tau, m_1 > \otimes | i, m_2 >$, the product of the internal states H$_2$O and N$_2$. On the other hand, instead of this choice, one can select the orientations of the pair of molecules as the basis set in Hilbert space; i.e., $| \delta(\Omega_1 - \Omega_{1\alpha}) > \otimes | \delta(\Omega_2 - \Omega_{2\alpha}) >$, where $\Omega_{1\alpha}$ and $\Omega_{2\alpha}$ represent orientations of H$_2$O and N$_2$ specified by $\alpha$, respectively. By introducing the coordinate representation, the potential becomes a diagonal operator and the matrix elements become multi-dimensional integrations [10,11]. In the standard method, the functions required to be evaluated are the resonance functions and their total number depends on how many combination choices of summation indices $L_1$, $K_1$, $L_2$, $L$, and $q$, $n_{(ij)}$, $w$ are used in Eq. (7) with the restrictions enforced by the two kinds of cut-offs. As these two cut-offs increase, the number of resonance functions increases very quickly and to evaluate all these resonance functions become formidable.
On the other hand, with the new method, the functions required to be evaluated are the correlation functions defined later and their total number is determined only by how many choices of the tensor ranks $L_1$, $K_1$, $L_2$, and $L$ restricted by the first kind of cut-off. This means that no matter how high the second kind of cut-off is, the number of correlation functions remains unchanged. In addition, because $V_{atom-atom}$ becomes ordinary functions, one can choose any order cut-off one wants. In fact, the second kind of cut-off affects only how accurately one needs to evaluate the coefficients $u(L_1 L_2; K_1; R(t))$ in Eq. (5). No matter how high this cut-off goes in calculating these coefficients, there are only small differences determined by the needed computational resources. As a result, we can use both cut-offs which are sufficiently high to guarantee convergence in practical calculations.

We have developed the new formalism to calculate half-widths for a simpler system consisting of two linear molecules (i.e., the $N_2 - N_2$ pair) [12]. In the present study, we make an extension to the $H_2O - N_2$ pair. In this case, $S_{2,outer,i}$ can be expressed as

$$S_{2,outer,i} = \frac{1}{\hbar^2 (2j_i + 1) \sum l_i} \int dt \int dt' \sum \int d\Omega_2 \int d\Omega_1 e^{i(\theta_{j_i} + \theta_{l_i}) (t-t')} \times <j_i \tau_i m_i l_i | \alpha > V_{\alpha} (\vec{R}(t)) < \alpha | j_i' \tau_i' m_i' l_i' >
$$

$$< j_i' \tau_i' m_i' l_i' | \beta > V_{\beta} (\vec{R}(t')) < \beta | j_i \tau_i m_i l_i | >,$$

where $|\alpha>$ is a shorthand notation for the basis set of the molecular pair in the coordinate representation and the subscript $\alpha$ of $V_{\alpha}$ represents the potential evaluated at a specified orientation labeled by $\Omega_{1a}$ and $\Omega_{2a}$. The inner products $<j_i; \tau_i; m_i; l_i | \alpha>$ represent a transformation between two basis sets of these two representations and are nothing but products of the complex conjugates of the H$_2$O wave functions at the orientation $\Omega_{1a}$ and the complex conjugates of the N$_2$ wave functions at the orientation $\Omega_{2a}$.

In order to simplify Eq. (8), we introduce the correlation functions which is defined as a convolution integration

$$F_{t_i; K_1; l_2} (t) = \int dt' G_{t_i; K_1; l_2} (t' + t/2, t' - t/2),$$

where $G_{l_i; K_1; l_2} (t, t')$ is given by

$$G_{l_i; K_1; l_2} (t, t') = \frac{-1}{4\pi(2l_i + 1)^2 (2l_2 + 1)^2 (-1)^{K_1 + K_1'} \sum L (-1)^{(L_1 + L_2 + L)} (2L + 1)}$$

$$\times u(L_1 L_2 L; K_1; R(t)) u(L_1 L_2 L; K_1'; R(t')) P_L (\cos \Theta_{tt'}).$$

In Eq. (10), the factor $(-1)^{(L_1 + L_2 + L)} = 1$ because the summation index $L$ must satisfy $L_1 + L_2 + L = even$ and $\Theta_{tt'}$ are angles between two vectors $\vec{R}(t)$ and $\vec{R}(t')$. The set used to label the correlation functions consist of one tensor rank $L_1$ with two subsidiary indices $K_1$, $K_1'$ related to H$_2$O and another tensor rank $L_2$ for N$_2$. Because N$_2$ is a diatomic molecule, $L_2$ must be even. Thus, the number of sets is determined by the upper limits of $L_1$ and $L_2$. If one chooses the II R representation to develop the H$_2$O wave functions where two H atoms are symmetrically located in the molecular-fixed frame, values of $K_1$
and $K_1'$ must also be even. Then, it is easy to show that if one sets 2 as the maximums for both $L_1$ and $L_2$, the number of sets is 20. Meanwhile, if one increases the maximum for $L_1$ from 2 to 3 or 4, the number of sets increases from 20 to 38 or 88, respectively. We list all these assignments in the Appendix.

As shown by Eqs. (9) and (10), the correlation functions $F_{L_1K_1',K_1''L_2'}(t)$ are convolution integrations over $t'$ from -∞ to ∞ whose integrands are proportional to products of $u(L_1L_2L;K_1;\hat{R}(t'+t/2))$ and $u(L_1L_2L;K_1';\hat{R}(t'-t/2))$. The latter are the products of the radial expansion components of potentials $V_\alpha(\hat{R}(t'+t/2))$ and $V_\beta(\hat{R}(t'-t/2))$ appearing in Eq. (8). Besides, the integrands of the convolutions also depend on the angles between the two distance vectors $\hat{R}(t'+t/2)$ and $\hat{R}(t'-t/2)$ which are arguments of $V_\alpha$ and $V_\beta$, respectively. In order to find the physical meaning of the correlation functions, let’s consider how the convolution integration over $t'$ given by Eq. (9) is carried out. One can imagine that as the integration variable $t'$ varies from -∞ to +∞, both $\hat{R}(t'+t/2)$ and $\hat{R}(t'-t/2)$ move along a same trajectory. One distance vector follows its predecessor’s movement and always keeps its delay time exactly all the way. The delay time is nothing but the time displacement $t$, the argument of the correlations. We note that the products of $V_\alpha(\hat{R}(t'+t/2))$ and $V_\beta(\hat{R}(t'-t/2))$ represent their overlaps. It is obvious that as the delay time $t$ increases, the separations between $V_\alpha$ and $V_\beta$ increase and their overlap decreases. Therefore, the correlation functions represent how the total overlaps between corresponding expansion components of $V(\hat{R}(t'+t/2))$ and $V(\hat{R}(t'-t/2))$ accumulated over a whole trajectory change with the time displacement $t$.

Then, in terms of the correlation functions one is able to rewrite Eq. (3) as

$$S_{2,outer,i} = \sum_{L_1K_1K_1'K_1''L_2} \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' G_{L_1K_1K_1'K_1''L_2}(t,t') W^{(a)}_{L_1K_1K_1'}(t-t') W^{(b)}_{L_2}(t-t')$$

$$= \sum_{L_1K_1K_1'K_1''L_2} \int_{0}^{\infty} dt W^{(a)}_{L_1K_1K_1'}(t) W^{(b)}_{L_2}(t) F_{L_1K_1K_1'K_1''L_2}(t).$$

In deriving the above expression, two functions which are independent of the potential are defined by

$$W^{(a)}_{L_1K_1K_1'}(t) = \sum_{j,l} (2i' + 1) D(j_i,\tau_i,j_i'\tau_i';L_1K_1) D(j_i,\tau_i,j_i'\tau_i';L_1K_1') e^{i\theta_{j_i,j_i'}t},$$

where $D(i\tau;j';L K)$ is defined by

$$D(j_i,\tau_i,j_i'\tau_i';L K) = \sum_{k} (-1)^{k} U_{k_{\tau_i}}^{j_i} U_{k_{\tau_i'}}^{j_i'} C(j_i,j_i';L_1L_2,kK - kK)$$

and

$$W^{(b)}_{L_2}(t) = \sum_{i_2} (2i_2 + 1)(2i_2' + 1) \rho_{i_2} C^2(i_2,i_2'L_2,000) e^{i\theta_{i_2,i_2'}t},$$

respectively. Finally, by introducing the Fourier transform of the correlation functions $H_{L_1K_1K_1'K_1''L_2}(\omega)$ defined by
\[ H_{L_i K_i, L_j K_j} (\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} F_{L_i K_i, L_j K_j} (\omega) dt, \]  

one is able to obtain an expression for the real part of \( S_{2,\text{outer},i} \) as

\[
\text{Re}(S_{2,\text{outer},i}) = \sqrt{\frac{\pi}{2}} \sum_{L_i K_i L_j K_j} \left\{ \sum_{i_2 j_2} (2i_2 + 1)(2j_2 + 1) \rho_{i_2} C^2 (i_2 j_2 L_2, 00) \right\} 
\times \left\{ \sum_{j_1' j_1} (2j_1' + 1)D(j_1', j_1'; L_1 K_1)D(j_1, j_1; L_1 K_1) \right\} H_{L_i K_i, L_j K_j} (\omega, j_1', j_1; \theta_i + \omega_{i_2 j_2}).
\]

We note that all these \( F_{L_i K_i, L_j K_j} (t) \), \( H_{L_i K_i, L_j K_j} (\omega) \), and \( \text{Re}(S_2) \) are associated with a specified trajectory. As a result, all of them depend on \( r_c \). For simplicity, we have omitted the argument \( r_c \) in their expressions.

### 2.2. Trajectories

Within the RB formalism, the trajectory model and the potential model are two main factors in governing calculated results. We would like briefly outline some trajectory models first. When two molecules collide, both their isotropic and anisotropic interactions play roles in determining a trajectory along which their translational coordinates vary with the time. However, to consider effects from the anisotropic interactions is a formidable problem, and has not been treated except for work by Green [13], based on close-coupling scattering theory applicable only for simple systems usually consisting of one atom and one molecule. Therefore, we would like to emphasize that there is a basic assumption that the trajectory is governed by the isotropic potential only. Of course, this is not true and this limitation is a big drawback suffered by most theoretical width and shift calculations.

Since Robert and Bonamy developed the RB formalism, there are two trajectory models [8] that are commonly used in calculations. We label them as the “parabolic” and the “modified parabolic” trajectories here. The first one assumes the trajectory labeled by \( r_c \) is a straight path, but the modulus \( R(t) \) is given by

\[ R(t) = \sqrt{r_c^2 + v'_c t^2}, \]

where \( v'_c \) is defined by Eq. (2) and we have set \( t = 0 \) when the translational motion reaches the closest distance \( r_c \) along the trajectory. This implies the translation motion moves with an “apparent” velocity. The second one is a modified version in which the straight line is replaced by a curve defined by

\[
\sin \theta(t) = \frac{v_c t}{\sqrt{v'_c^2 + v_c^2 t^2}} \quad \text{and} \quad \cos \theta(t) = \sqrt{1 - \sin^2 \theta(t)}
\]

where \( v_c \) is the velocity at \( R = r_c \) and can be calculated by

\[ v_c = v \left\{ 1 - \frac{8\sigma}{m v^2} \left[ \frac{\sigma}{r_c} \right]^{12} - \left( \frac{\sigma}{r_c} \right)^6 \right\}^{1/2}. \]

However, it seems that one cannot make this replacement everywhere because \( \sin \theta(t) \) could be larger than 1 as \( t \) increases for some trajectories for which \( v_c > v'_c \). This implies that one has to manipulate it further, more or less, arbitrarily. Different researchers use different manipulation techniques and usually they do not provide detailed instructions.
about their method. Therefore, in order to avoid arbitrarily effects resulting from adopting the modified parabolic trajectory model, we only consider the original “parabolic” model in the present study.

More recently, there have been attempts by Buldyreva et al. to consider the “exact” trajectory [14,15]. Their work is based on Bykov’s method presented 18 years ago, but no computation of line widths was performed at that time [16]. A basic idea of Bykov is that by changing the integration variable from the time \( t \) to the distance \( r \) in calculating the resonance functions, the trajectory dependence of the resonance functions can be explicitly taken into account. It turns out that by introducing the coordinate representation, the “exact” trajectory model can be more straightforwardly incorporated into the RB formalism. As shown by Eq. (10), when one calculates 

\[ G_{L_1K_1,K'_1L_2}(t,t') \]

(or \( G_{L_1K_1,K'_1L_2}(t,t') \)), one needs to know how \( \tilde{R}(t) \) varies with the time \( t \) along the trajectory. Because \( \tilde{R}(t) \) no longer is an operator, to take into account its trajectory dependence is more natural here. Interested readers can find our detailed explanation on the “exact” trajectory model in our paper [12].

It is well known that different collision trajectories have different the closest distances \( r_c \) and the strictly head-on collision has the smallest \( r_{c\text{-min}} \). After selecting the trajectory model and the potential, one can easily find values of \( r_{c\text{-min}} \) for temperatures of interest. For example, with the “parabolic” trajectory model \( r_{c\text{-min}} \) can be derived from the expression directly

\[
r_{c\text{-min}} = \sigma \frac{2^{1/6}}{\left(1 + \sqrt{1 + \frac{mv^2}{2\epsilon}}\right)^{1/6}}. \tag{19}
\]

Meanwhile, for the “exact” trajectory governed by \( V_{\text{iso}}(R) \), the value of \( r_{c\text{-min}} \) can be derived numerically from the equation

\[
\frac{2V_{\text{iso}}(r_{c\text{-min}})}{mv^2} - 1 = 0. \tag{20}
\]

With the potential model used in the present study, a value of \( r_{c\text{-min}} \) at \( T = 296 \) K derived from the “parabolic” trajectory model and with the 14-th or the 20-th order cut-off is 3.43564 Å. If one adopts the 8-th order cut-off, \( r_{c\text{-min}} \) becomes 3.42070 Å. Meanwhile, the value of \( r_{c\text{-min}} \) is 3.52242 Å for the “exact” trajectory model and with the 20-th order cutoff. The value of \( r_{c\text{-min}} \) is very important in calculating half-widths because it serves as the lower integration limit of Eq. (1). In addition, when one investigates profiles of potential models, one can use \( r_{c\text{-min}} \) as the lower boundary of the distance \( R \). Of course, \( r_c \) and \( R \) are two different concepts. The former is used to label or to represent the whole collision trajectory and the latter is the argument of potential and it could vary from \( r_c \) to infinity along the trajectory. However, for our present study purpose, there is an important link between them: for a specified \( r_c \), all things of interest to us along this trajectory are independent of the behavior of the potential \( V(R) \) for \( R < r_c \). As a result, one only needs to investigate profiles of the potential in the range of \( R > r_c \). In addition, it turns out that in calculating \( N_2 \) broadened half-widths of the \( H_2O \) lines, the potential ranges of most interest are those \( R \) staring from the closest distance \( r_c \) of collisions to a little bit beyond. Therefore, we will pay extra attention to profiles of the potential in small \( R \) ranges.
As an example, we present several typical “parabolic” and “exact” trajectories for H$_2$O – N$_2$ at T = 296 K in Fig. 1. For the “exact” model three nearly head-on trajectories associated with $r_e$ = 3.55, 3.60, and 3.80 Å, respectively, and three glancing ones with $r_e$ = 4.3, 5.0, and 6.0 Å are shown. Because the repulsive force is dominant for the former, the first three curves bend away from the scattering center (at the origin of the plane). In contrast, due to the dominant attractive force the last three curves bend towards the scattering center. Besides, there is a curve with $r_e$ = 4.0 Å in between these two groups. The corresponding seven “parabolic” trajectories are also represented. They are horizontal straight lines with $r_e$ values as their ordinates. In addition, by adding small solid circles on the curve and small empty circles on the straight lines, we explicitly show how fast the relative motion of the pair moves along the trajectories. A travel time from one circle to its next is 0.05 picoseconds. To show this feature is important because one has to rely on both this and the trajectory to know how $\ddot{R}(t)$ varies with time. Careful readers can find that the relative motion moves with higher speed along the “parabolic” path than the “exact” trajectory which bends away from the center. The higher “apparent” velocity $v'$ applicable in the straight path is required to compensate for the bending effect in the “exact” trajectories such that the increase of the modulus R(t) can, more or less, keep a similar pace with the latter. Based on the same reason, it moves with lower speed along the “parabolic” path than the “exact” trajectory which bends towards the center.

![Figure 1. Calculated trajectories for the molecular pair of H$_2$O - N$_2$ at T = 296 K with $r_e$ = 3.55, 3.6, 3.8, 4.0, 4.3, 5.0, and 6.0 Å. They are represented by curves with different colors in order from bottom to top. The “exact” trajectories are plotted by the solid curves and the “parabolic” trajectories are given by the dotted straight lines. Small circles are added on the trajectories such that a travel time between circles is 0.05 picoseconds. The scattering center is at the origin of the XY plane.](image-url)
2.4. Profiles of the short-range and long-range parts of potentials

In the present study, we adopt the most updated potential model by Gamache et al. [7] which consists of the short-range interaction $V_{\text{atom-atom}}(\Omega_1, \Omega_2, \vec{R})$ and the long-range dipole-quadrupole and quadrupole-quadrupole interactions $V_{dq}(\Omega_1, \Omega_2, \vec{R}) + V_{qq}(\Omega_1, \Omega_2, \vec{R})$. Generally speaking, our knowledge of the long-range interactions are better than the short-range ones. In fact, most recent attempts to refine the potential models are mainly made for the latter.

We present two simple figures to depict profiles of this potential. First of all, by choosing $\vec{R}$ to lie along the $z$ axis of the space-fixed frame, one can replace the vector $\vec{R}$ by its magnitude $R$. Because these interactions also depend on the orientations of $\text{H}_2\text{O}$ and $\text{N}_2$, it is impossible to depict them well with a two dimensional plot. However, by plotting their maxima and minima as functions of $R$ we can grasp the ranges of their magnitudes. With the Monte Carlo method to choose enough random selections of the orientations of the $\text{H}_2\text{O}$ and $\text{N}_2$ pair, one can easily find the maxima and minima of $V_{\text{atom-atom}}(\Omega_1, \Omega_2, R)$ and $V_{dq}(\Omega_1, \Omega_2, R) + V_{qq}(\Omega_1, \Omega_2, R)$. Meanwhile, by averaging the former’s values over all these random selections of the orientation, one can obtain a corresponding isotropic component. We note that in general, the isotropic part of $V_{\text{atom-atom}}(\Omega_1, \Omega_2, R)$ does not match the LJ isotropic model exactly. However, when one calculates half-widths with the “parabolic” trajectory model, one usually assumes that the isotropic interaction can be represented by the LJ model. In this case, one needs to find the LJ parameters $\sigma$ and $\epsilon/k_B$ from a fitting procedure. In the present study, we have found that $\sigma = 3.83$ Å and $\epsilon/k_B = 53.4$ K is an optimized choice to fit the isotropic part of $V_{\text{atom-atom}}(\Omega_1, \Omega_2, R)$. We note that these values differ slightly from those (see Table 3) derived by Gamache et al. The reason is their values are obtained from adopting the 8-th order cut-off and 20 correlations while ours are from higher cut-offs and are already converged. On the other hand, if one uses the “exact” trajectory model in calculations, it is better to use the isotropic potential derived directly without using the secondhand LJ model. With respect to $V_{dq}(\Omega_1, \Omega_2, R)$ and $V_{qq}(\Omega_1, \Omega_2, R)$, it is well known that they are purely anisotropic.
Figure 2. Maxima and minima of $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ and $V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R)$ as functions of $R$. They are plotted by two solid black lines and two dot-dashed red lines, respectively. An isotropic part of $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ is also plotted by a dotted green curve. $V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R)$ do not contain an isotropic part.

We present these results in Fig. 2. As shown by the figure, $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ could have very large repulsive forces at short distances. Beyond $R = 4.3$ Å, the maximum and minimum of $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ are well within the range of the maximum and the minimum of $V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R)$. This is consist with a general trend that $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ is the major interaction in the short distance region and $V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R)$ play a dominant role in the long distance region. Meanwhile, one can find that in Fig. 1 the distance at which the isotropic potential changes signs and also its maximum depth match values of $\sigma$ and $\varepsilon/k_B$ mentioned above, respectively. However, to plot Fig. 2 alone does not provide enough information to judge between $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ and $V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R)$ which is a major component in a range of $r_{c_{\text{min}}} < R < 4.3$ Å. Although the maxima of $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ are much larger than those of $V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R)$, this does not necessarily mean the former are more likely to dominate the latter there. The reason is these maxima only represent extreme events happening in special orientations. In practice, the importance of an event depends not only on the event itself, but also on how often it happens. Thus, we present two subsidiary figures to provide more in-depth analyses. We assume both the H$_2$O and N$_2$ molecules rotate freely in space. Then, in order to provide quantitative measures with respect to all possible combinations of the orientations $\Omega_1$ and $\Omega_2$, how often specified values of $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ occur, we calculate probabilities with which $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ take these values. The calculations are carried out with $10^8$ random selections of the orientations of H$_2$O and N$_2$ with a 0.5 cm$^{-1}$ resolution for the specified values $V_0$. For example, the probability of $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R) = V_0$ represents a chance with which $V_{\text{atom-atom}}(\Omega_1,\Omega_2,R)$ has values between $V_0 - 0.5$ cm$^{-1}$ and $V_0 + 0.5$ cm$^{-1}$. 
Similarly, we calculate the probabilities for \( V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R) \) as well. We present these probabilities associated with four different distances \( R = 3.6, 3.8, 4.0, \) and \( 4.3 \text{ Å} \) in Fig. 3a and Fig. 3b. As shown by the figure, values of \( V_{\text{atom-atom}} \) are more likely to be small and negative in very narrow regions. Departing from its most probable values, the probabilities decrease very quickly as the value differences increase. In contrast, the distributions of \( V_{dq} + V_{qq} \) are wider and more flat without peaks.

Figure 3a. A plot to show probabilities of distribution of \( V_{\text{atom-atom}} \) and \( V_{dq} + V_{qq} \) over their possible values. The probabilities associated with \( R = 3.44, 3.60, 3.80, 4.00, \) and \( 4.30 \text{ Å} \) are calculated from \( 10^8 \) random selections of the orientations of \( \text{H}_2\text{O} \) and \( \text{N}_2 \) with a \( 0.5 \text{ cm}^{-1} \) resolution of potential values. The results of \( V_{\text{atom-atom}} \) are plotted by black, red, green, blue, and aqua solid curves, respectively, and those of \( V_{dq} + V_{qq} \) are given by five dot-dashed lines.
Figure 3b. The same as Figure 3a except providing more details of the distributions in the range from 25 to 500 cm\(^{-1}\).

As shown by the figure, the profiles of distributions for \(V_{\text{atom}}(\Omega_1, \Omega_2, R)\) exhibit high peaks. The locations of the peaks at the negative side of the abscissa axis represent the most probable values of \(V_{\text{atom}}(\Omega_1, \Omega_2, R)\). Two sides of these peaks are not symmetric. The left wings are very sharp and the right wings are more gradual. The limits of the left and right wings for a specified \(R\) are the corresponding minimum and maximum values of \(V_{\text{atom}}(\Omega_1, \Omega_2, R)\) shown in Fig. 2. By comparing the distributions at different distances, one can find that as \(R\) increases, both the heights and widths of the peak increases and the right wings decrease quickly. Based on these features, one can conclude that \(V_{\text{atom}}(\Omega_1, \Omega_2, R)\) more likely only takes small and negative values within the wider ranges predicted by Fig. 2. In addition, as \(R\) increases, there are fewer and fewer chances to have large magnitudes. In contrast, the probability distributions for \(V_{dq}(\Omega_1, \Omega_2, R) + V_{qq}(\Omega_1, \Omega_2, R)\) are almost symmetric and have flat plateaus with relatively narrow edges. As \(R\) increases, the plateau becomes higher and narrower. This means that \(V_{dq}(\Omega_1, \Omega_2, R) + V_{qq}(\Omega_1, \Omega_2, R)\) have a uniform chance to take values in most parts of allowed regions in Fig. 2. In the remaining parts a chance is quickly diminished to zero as the value reaches the maximum or minimum limits. Concerning the interesting question about which one is the major component in the range from \(R = r_{\text{c-min}}\) to \(R = 4.3\) Å, one can roughly claim that \(V_{\text{atom}}(\Omega_1, \Omega_2, R)\) has a good chance to have larger magnitudes than \(V_{dq}(\Omega_1, \Omega_2, R) + V_{qq}(\Omega_1, \Omega_2, R)\) at \(R = 3.6\) Å. The chance is reduced significantly at \(R = 3.8\) Å and it becomes almost impossible at \(R = 4.0\) Å and beyond.

In addition, we provide some numerical results related to Figs. 2, 3a, and 3b in Table 1. In the second and third columns of the table, we list maxima and minima of \(V_{dq}(\Omega_1, \Omega_2, R) + V_{qq}(\Omega_1, \Omega_2, R)\) for \(R = 3.44, 3.60, 3.80, 4.00,\) and \(4.30\) Å, respectively. In the fourth and fifth ones, maxima and minima of \(V_{\text{atom}}(\Omega_1, \Omega_2, R)\) are given. In the column labeled as \(\text{Sum P}_{v_{ab}<0}\), we provide the probabilities with which \(V_{\text{atom}}(\Omega_1, \Omega_2, R)\) has negative
values. Finally, in the last column labeled as $P_{V_{aa}>MaxV_{dq}}$, we list the probabilities with which values of $V_{atom\text{-}atom}(\Omega_1,\Omega_2,R)$ are larger than the maxima of $V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R)$.

It is worth mentioning that the above discussions are based on the assumption that with respect to all orientations of H$_2$O, the accessibilities are equal. The same assumption also applies for N$_2$. However, when one deals with specified states for H$_2$O and N$_2$, chances in occupying orientations are not uniform and they are determined by the corresponding wave functions. As a result, one has to use a more sophisticated method by introducing absolute squares of the products of the wave functions of H$_2$O and N$_2$ as the weighting functions in the averaging processes to find the probability distributions applicable for these states. But, we will not pursue this here. Given the fact that when one calculates the half-width for a H$_2$O line of interest, not only its initial and final states are involved, but also many other states (for example, see those labeled by the summation indices $j'$ and $\tau'$ in Eq. (8)) participate. Meanwhile, with respect to N$_2$ its all states are taken into account. Therefore, we believe the general trends presented above are, more or less, applicable here.

<table>
<thead>
<tr>
<th>R</th>
<th>Max $V_{dq\text{qq}}$</th>
<th>Min $V_{dq\text{qq}}$</th>
<th>Max $V_{aa}$</th>
<th>Min $V_{aa}$</th>
<th>Sum of $P_{V_{aa}&lt;0}$</th>
<th>Sum of $P_{V_{aa}&gt;MaxV_{dq}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.44 Å</td>
<td>317 cm$^{-1}$</td>
<td>-316 cm$^{-1}$</td>
<td>13813 cm$^{-1}$</td>
<td>-86 cm$^{-1}$</td>
<td>46.9 %</td>
<td>28.9 %</td>
</tr>
<tr>
<td>3.60 Å</td>
<td>260 cm$^{-1}$</td>
<td>-258 cm$^{-1}$</td>
<td>5165 cm$^{-1}$</td>
<td>-79 cm$^{-1}$</td>
<td>58.8 %</td>
<td>17.6 %</td>
</tr>
<tr>
<td>3.80 Å</td>
<td>205 cm$^{-1}$</td>
<td>-203 cm$^{-1}$</td>
<td>1588 cm$^{-1}$</td>
<td>-72 cm$^{-1}$</td>
<td>74.4 %</td>
<td>7.23 %</td>
</tr>
<tr>
<td>4.00 Å</td>
<td>165 cm$^{-1}$</td>
<td>-162 cm$^{-1}$</td>
<td>487 cm$^{-1}$</td>
<td>-67 cm$^{-1}$</td>
<td>88.8 %</td>
<td>1.74 %</td>
</tr>
<tr>
<td>4.30 Å</td>
<td>121 cm$^{-1}$</td>
<td>-118 cm$^{-1}$</td>
<td>56 cm$^{-1}$</td>
<td>-53 cm$^{-1}$</td>
<td>98.8 %</td>
<td>0 %</td>
</tr>
</tbody>
</table>

In summary, we can consider $\sigma$ (= 3.83 Å), one of the LJ parameter mentioned above, as a threshold in comparing $V_{atom\text{-}atom}(\Omega_1,\Omega_2,R)$ and $V_{dq}(\Omega_1,\Omega_2,R) + V_{qq}(\Omega_1,\Omega_2,R)$. As long as $R < \sigma$, both of them are main components of the total potential. For $R > \sigma$, the former becomes a minor one and as $R$ increases further beyond 4.3 Å, the former’s effects are very weak or even negligible.

3. Three line categories

In order to analyze uncertainties associated with theoretically calculated half-widths, we would like to concentrate the recent work by Gamache and Laraia et al. [7] because this work represents their recent theoretical achievements. Based on our numerical tests, we have found that one can group 1639 H$_2$O lines of the pure rotational band listed in HITRAN into three categories according to their half-width values. The first category consists of lines with large half-widths, the second with intermediate and the third with small half-widths. It turns out that this division coincides with a division of lines according to their convergence behaviors. In fact, lines belonging to the first category converge very quickly as the cut-off order increases so that one does not need to worry about their convergence problems. Lines in the second and third categories converge progressively more slowly. In the following, we select sample lines from each of these three categories and explain why they exhibit quite different convergence behaviors.
3.1. Lines belonging to the first category

3.1.1. Two sample lines in category 1

In the fitting procedure to optimize the parameters $\sigma_{ij}$ and $\varepsilon_{ij}$ of $V_{\text{atom-atom}}(\Omega_1, \Omega_2, R)$, Gamache and Laraia et al. have selected two lines (i.e., $6_{1,6} \leftarrow 5_{2,3}$ at 22 GHz and $3_{1,3} \leftarrow 2_{2,0}$ at 183 GHz) and made comparisons between theoretically calculated values and measurements [7]. According to our division, these two lines belong to the first category because their half-width values are relatively large (their air-broadened half-widths listed in HITRAN are 0.0942 and 0.1014 cm$^{-1}$ atm$^{-1}$, respectively). It is well known that air-broadened half-widths are smaller than N$_2$-broadended ones because the molecule O$_2$, a minor component of the air, is less effective than N$_2$, the major component of the air, to cause the line broadening.

Based on the same potential model with the optimized values of $\sigma_{ij}$ and $\varepsilon_{ij}$, we repeat their calculations for these two lines, but with more choices of the cut-off orders, the numbers of correlations, and the trajectory models. In Table 2, we present our calculated results for these two lines at $T = 296$ K and 220 K, respectively. In the table, the letter c means correlations, p means the “parabolic” trajectory model, and e means the “exact” trajectory model. As shown in the table, the calculated half-widths with the lowest 8-th order cut-off and only including 20 correlations are already converged. In addition, differences between values derived from the “parabolic” trajectory model and the more accurate “exact” trajectory model are very small. However, one can not simply assume that the applicability of these conclusions can be extended to all other lines. It turns out that they are valid only for lines with large half-width values.

<table>
<thead>
<tr>
<th>Line</th>
<th>T</th>
<th>8th,20c,p</th>
<th>14th,20c,p</th>
<th>20th,20c,p</th>
<th>20th,38c,p</th>
<th>20th,88c,p</th>
<th>20th,88c,e</th>
</tr>
</thead>
<tbody>
<tr>
<td>$6_{1,6} \leftarrow 5_{2,3}$</td>
<td>296</td>
<td>0.097612</td>
<td>0.097370</td>
<td>0.097371</td>
<td>0.097689</td>
<td>0.097780</td>
<td>0.097067</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>0.120034</td>
<td>0.119745</td>
<td>0.119747</td>
<td>0.120148</td>
<td>0.120265</td>
<td>0.119164</td>
</tr>
<tr>
<td>$3_{1,3} \leftarrow 2_{2,0}$</td>
<td>296</td>
<td>0.108771</td>
<td>0.108661</td>
<td>0.108660</td>
<td>0.108763</td>
<td>0.108775</td>
<td>0.107656</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>0.134452</td>
<td>0.134365</td>
<td>0.134364</td>
<td>0.134469</td>
<td>0.134480</td>
<td>0.132510</td>
</tr>
</tbody>
</table>

Next, we would like to present more detailed analyses for the line of $3_{1,3} \leftarrow 2_{2,0}$ here because features of the other line of $6_{1,6} \leftarrow 5_{2,3}$ are very similar to that presented here. First of all, in order to explain with respect to the first-kind of cut-off, why calculated half-widths converge so quickly, we plot $\text{Re}(S_2)$, $\exp(-\text{Re}(S_2))$, and $r_c \frac{V^2}{V^2} (1 - e^{-\text{Re}(S_2)})$ derived from including 20, 38, and 88 correlations and using the 20-th order cut-off in Fig. 4. The latter is the integrand of the expression for the half-width given by Eq. (1). We note that $S_2$ is a function of $r_c$ and for simplicity, this argument is omitted here.

As shown by the figure, there are significant differences among values of $\text{Re}(S_2)$ derived from including 20 correlations and those obtained with more correlations in the range of $r_c < 3.83$ Å. After $r_c > 3.83$ Å, the differences diminish. Concerning $\text{Re}(S_2)$ alone, one can conclude that $\text{Re}(S_2)$ derived from 20 correlations is not converged, but those from 38 correlations are well converged and those from 88 correlations are completely converged. On the other hand, by looking at three blue curves representing $\exp(-\text{Re}(S_2))$ whose values have been multiplied by 1000 in the figure, one can find that
the values for \( r_c < 3.83 \text{ Å} \) are less than 0.007 which means all three factors of 1 - \( \exp(-\text{Re}(S^2)) \) are always close to 1. Finally, one can find that the integrands represented by three red lines are almost identical. An essential point is in Eq. (1) besides the lower integration limit \( r_{c,mim} \), it is the integrand itself, but not \( \text{Re}(S^2) \) that directly determines calculated half-width values. In fact, as long as values of \( \text{Re}(S^2) \) are large enough, they would completely diminish the role played by \( \text{Re}(S^2) \) itself in determining calculated half-widths. Therefore, a quick convergence of the integrand definitely means a good convergence of the half-width. In contrast, a poor convergence of \( \text{Re}(S^2) \) for \( r_c < 3.83 \text{ Å} \) does not inevitably cause a poor convergence of the half-width. For the line of interest here, it turns out that all these calculated \( \text{Re}S^2 \) have large values in the low \( r_c \) region. Thus, their large differences appearing in this region have no impact on the results. Meanwhile, in the region beyond 3.83 Å, their profiles are similar. As a result, calculated half-widths with the different numbers of correlations vary only very slightly. The plot demonstrates that with respect to the first kind of cut-off, the calculated half-width of this line with 20 correlations is well converged. In addition, this example exhibits different convergence behaviors for \( \text{Re}(S^2) \) and for the half-width and how different convergence behaviors could happen.

Figure 4. A plot to show with respect to the first-kind of cut-off, why calculated half-widths for the line of \( 3_{1,3} \leftarrow 2_{2,0} \) converge very quickly. Values of \( \text{Re}S^2 \) derived from including 20, 38, and 88 correlation functions and using the 20-th order cut-off are plotted by black solid, dashed, and dot-dashed lines, respectively. We also plot \( \exp(-\text{Re}(S^2)) \), and \( r_c \frac{V^2}{V^2} (1 - \exp(-\text{Re}(S^2)) \) (i.e., the integrand of the half-width \( \gamma \)) in the same figure. The former are represented by thin blue curves and the latter by red curves. We note that in order to show \( \exp(-\text{Re}(S^2)) \) more clearly in the figure, their values have been multiplied by 1000 in the plotting.
In order to explain with respect to the second-kind of cut-off, why calculated half-widths converge so quickly, we plot these three terms derived with the 8-th, 14-th, and 20-th order cut-offs in Fig. 5. The patterns of Fig. 5 are very similar to those given in Fig. 4. Thus, all conclusions drawn from Fig. 4 are applicable here. In the present case, values of \( \exp(-\text{Re}(S_2)) \) are at least less than 0.008 in \( r_c < 3.83 \) Å and become less than 0.0016 for \( r_c < 3.75 \) Å. We note that there are slight differences among \( r_{c, \text{min}} \), the minimum values of \( r_c \), associated with the different cut-offs. With the ‘parabolic’ trajectory model, the value of \( r_{c, \text{min}} \) is determined by the LJ parameters. Because the cut-off introduced in evaluating \( V_{\text{atom-atom}}(R) \) would affect a determination of its isotropic component, and thus on the LJ parameters in the modeling, and consequently on the value of \( r_{c, \text{min}} \) from Eq. (19).

![Figure 5](image.png)

Figure 5. The same as Figure 4 except results are derived from the 8-th, 14-th, and 20-th order cut-offs. The calculations are based on including 20 correlations and the “parabolic” trajectory model. The plot demonstrates that with respect to the second kind of cut-off, the calculated half-width of this line with 8-th order cut-off is well converged.

3.1.2. Comments on the selection

Before proceeding with our analyses of other lines belonging to the second and third categories, we would like to comment on Gamache and Laraia’s choice of selecting these two lines \( 6_{1,6} \leftarrow 5_{2,3} \) and \( 3_{1,3} \leftarrow 2_{2,0} \) in optimizing the short-range potential \( V_{\text{atom-atom}}(R) \). In fact, based on the same arguments used to explain the different convergence behaviors between \( \text{Re}(S_2(r_c)) \) and the half-width, one can judge whether the selection of these two lines is a good choice to optimize the short-range site-site model.

As shown by Figs. 4 and 5, changes of the cut-off choices could yield significant differences for calculated \( \text{Re}(S_2(r_c)) \) for \( r_c < 3.83 \) Å. The smaller \( r_c \) is, the larger the differences are. However, it is variations of values of \( V_{\text{atom-atom}}(R) \) resulting from different cut-off choices that directly causes these differences. Indeed, values of \( V_{\text{atom-atom}}(R) \) could vary significantly in short distances as the cut-offs vary. The shorter \( R \) is,
the larger the variations are. The similarity between the patterns of $V_{atom-atom}(R)$ and $Re(S_2(r_c))$ exhibit the intrinsic link between them. Of course, this is expected because $Re(S_2(r_c))$ is determined by the potential $V_{tot}(R)$ and for short distances $V_{atom-atom}(R)$ becomes its major part.

Now we come to our point. With respect to changes of $Re(S_2(r_c))$ what matters is not sources causing the variations of $V_{atom-atom}(R)$, but the variations of $V_{atom-atom}(R)$ themselves. As long as variations of $V_{atom-atom}(R)$ in short $R$ distances happen with the same scale, no matter if they are caused by the cut-offs, by adopting different parameters $\sigma_{ij}$ and $\epsilon_{ij}$ or by something else, changes of $Re(S_2(r_c))$ in the small $r_c$ region should keep the same amount. On the other hand, $Re(S_2(r_c))$ in the large $r_c$ region are mainly determined by the long-range interactions because the latter become dominant components in the large $R$ distances. There is no room for $V_{atom-atom}(R)$ itself and its variations to play a role at all. Given the fact that lines with large half-widths are always associated with relatively large $Re(S_2(r_c))$ in the entire range and, in general, their magnitude peaks occur in the small $r_c$ region, one can conclude that the effects from $V_{atom-atom}(R)$ on theoretically calculated half-widths are very small. In fact, as shown by Table 2, there is almost no effect on the calculated half-widths by varying the two cut-offs in the calculations. The same must be true if one varies the parameters $\sigma_{ij}$ and $\epsilon_{ij}$ of the site-site model.

In summary, one can conclude that the choice of these two lines to optimize the site-site model is poor. The calculated half-widths do not depend sensitively on the parameters $\sigma_{ij}$ and $\epsilon_{ij}$ at all. As a result, not only is one not able to find optimum values, but one may also unintentionally obtain values for them outside of physically acceptable limits.

3.1.3. What really happens in the fitting process

Before completing our comments on line selection to optimize the site-site potential, we have to answer a question about what really happens in optimization practices. In contrast with what we have claimed above, one can obtain different half-width values for the two lines by using different parameters $\sigma_{ij}$ and $\epsilon_{ij}$. In other words, by changing values of $\sigma_{ij}$ and $\epsilon_{ij}$, theoretically calculated half-widths of these two lines can be changed too. In our opinion, what mainly happens in the fitting calculations are changes of the isotropic part of the potential. By varying the parameters $\sigma_{ij}$ and $\epsilon_{ij}$ of the site-site model, they change the isotropic interaction as well. The latter is modeled by a L-J model with two parameters $\sigma$ and $\epsilon$. Meanwhile, because $V_{dq}$ and $V_{qq}$ are purely anisotropic interactions, they do not make contributions to the isotropic part at all. On the other hand, with the RB formalism, the isotropic potential plays an important role in determining the calculated half-widths. For example, if one selects the “parabolic” model and uses Eq. (1) to derive half-widths, the lower integration limit $r_{c,min}$ and the “apparent” velocity $v_c$ are functions of $\sigma$ and $\epsilon$. Therefore, we expect that in the fitting process the main source of changes is $\sigma$ and $\epsilon$, but not the $\sigma_{ij}$ and $\epsilon_{ij}$. There are other evidences to support our claim, but we will not present them here.

3.2. Lines in the third category

As an example, we choose the line $17_{2,15} \leftarrow 16_{1,16}$ whose air-broadened half-width has been measured by different research groups \cite{17,18}. Its measured air-broadened half-widths are small: the value provided by Rinsland et al. is 0.0207 cm$^{-1}$ atm$^{-1}$ and the value
by Toth is 0.0208 cm\(^{-1}\) atm\(^{-1}\). Based on the same potential model used above, we calculate the \(\text{N}_2\) broadened half-width for \(T = 296\) K and 220 K with different combinations of the two kinds of cut-offs, and present results (in units of cm\(^{-1}\) atm\(^{-1}\)) in Table 3. As shown by the table, the calculated value with the 8-th order cut-off and including 20 correlations is not converged. In fact, the half-width at \(T = 296\) K derived from the 8-th order and 20 correlations is 13.1 % less than that from the 20-th order and 20 correlations and 29.4 % less than that from the 20-th order and 88 correlations. With respect to the first cut-off, one must include 88 correlations. For the second cut-off, one has to at least use the 14-th order cut-off. Furthermore, by making adjustments from \(\text{N}_2\) to air by dividing by the factor 1.09, one can compare calculated values with the measurements. Thus, the calculated air-broadened half-width from the 8-th order and including 20 correlations is about 0.0237 cm\(^{-1}\) atm\(^{-1}\) and that from the 20-th order and 88 correlations becomes 0.0336 cm\(^{-1}\) atm\(^{-1}\). It appears that the former agrees better with the measurements while adding more correlations leads to larger differences. If one only looked at this, one could claim that higher cut-offs yield errors and they must be abandoned. Of course, this specious argument is false.

Table 3 Calculated \(\text{N}_2\)-broadened half-widths (in cm\(^{-1}\) atm\(^{-1}\))

<table>
<thead>
<tr>
<th>Line</th>
<th>(T)</th>
<th>8th,(20c,p)</th>
<th>14th,(20c,p)</th>
<th>20th,(20c,p)</th>
<th>20th,(38c,p)</th>
<th>20th,(88c,p)</th>
<th>20th,(88c,e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(17_{2,15} \leftrightarrow 16_{1,16})</td>
<td>296</td>
<td>0.0258387</td>
<td>0.0294149</td>
<td>0.0297244</td>
<td>0.0331365</td>
<td>0.0365911</td>
<td>0.0316113</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>0.0256283</td>
<td>0.0296385</td>
<td>0.0300268</td>
<td>0.0346675</td>
<td>0.0400757</td>
<td>0.0338328</td>
</tr>
<tr>
<td>(9_{9,0} \leftrightarrow 8_{6,3})</td>
<td>296</td>
<td>0.0521262</td>
<td>0.0529479</td>
<td>0.0530399</td>
<td>0.0581423</td>
<td>0.0584993</td>
<td>0.0583407</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>0.0605984</td>
<td>0.0617843</td>
<td>0.0619207</td>
<td>0.0692880</td>
<td>0.0697956</td>
<td>0.0696941</td>
</tr>
</tbody>
</table>

In order to explain why calculated half-widths for the line \(17_{2,15} \leftrightarrow 16_{1,16}\) converge so slowly, we plot \(\text{Re}(S_2)\), \(\exp(-\text{Re}(S_2))\), and \(r_c \frac{1}{\sqrt{V_c^2 \left(1 - e^{-\text{Re}(S_2)}\right)}}\) derived from including 20, 38, and 88 correlations in Fig. 6. The calculations are carried out with the 20-th order cut-off and the “parabolic” trajectory model. By comparing Fig. 6 with Fig. 4, the magnitudes of \(\text{Re}(S_2)\) for this line are more than one order smaller than those for the line of \(3_{1,3} \leftrightarrow 2_{2,0}\) and, in addition, they decrease more quickly as \(r_c\) increases. In fact, after \(r_c > 3.83\) Å their magnitudes are close to zero. By comparing three black curves representing \(\text{Re}(S_2)\) derived from including different numbers of the correlations, one can conclude that one curve differs from the others significantly. This means that calculated \(\text{Re}(S_2)\) with fewer correlations are not converged. By looking at the three blue curves representing \(10 \times \exp(-\text{Re}(S_2))\) in Fig. 6, one can see that values of \(\exp(-\text{Re}(S_2))\) increase from 0 as \(r_c\) increases from \(r_{c,\text{min}}\) and they become close to an asymptotic value 1 as \(r_c > 3.83\) Å. Finally, one can find that the magnitudes of integrands represented by three red curves decrease very quickly and are close to zero for \(r_c > 3.83\) Å. In contrast with Fig. 4, differences among these three red curves are significant.
Figure 6. The same as Figure 4 except for the line of $17_{2,15} \leftarrow 16_{1,16}$ and multiplication of $\exp(-\text{Re}(S_2))$ by 10.

Figure 7. The same as Figure 5 except for the line of $17_{2,15} \leftarrow 16_{1,16}$ and multiplication of $\exp(-\text{Re}(S_2))$ by 10.

Similarly, we also plot these three terms derived with the 8-th, 14-th, and 20-th order cut-offs and including 20 correlations in Fig. 7. The patterns of Fig. 7 are very similar to that given by Fig. 6. Thus, all conclusions drawn from Fig. 6 are applicable here again.
The only difference is that with respect to the second kind of cut-off the plots in Fig. 7 demonstrate results derived from the 14-th cut-off are well converged and those from the 20-th cut-off are completely converged.

In summary, from analyzing Figs. 6 and 7, one can draw two conclusions. The first is that the dominant contributions to the calculated half-width come from those collisions whose closest distances $r_c$ are less than 3.83 Å. The second is that varying values of $V_{\text{atom-atom}}$ could effectively cause significant variations of the calculated half-widths. We note here again that variations of the values of $V_{\text{atom-atom}}$ could be caused either by adopting different cut-offs or by changing the parameters $\sigma_{ij}$ and $\epsilon_{ij}$. Thus when one studies lines in the third category such as $17_{2,15} ← 16_{1,16}$, one must be careful to check the convergence of calculated results. In addition, one can consider lines in this category as possible candidates to optimize the short-range site-site model.

### 3.3. Lines in the second category

Finally, we choose the line of $9_{9,0} ← 8_{6,3}$ whose air-and N$_2$-broadened half-widths have been measured by Toth [18]. The measured values are moderate and they are 0.0525 and 0.0608 cm$^{-1}$ atm$^{-1}$, respectively. Based on the same potential model, we calculate the N$_2$-broadened half-width for $T = 296$ K and 220 K and present results in Table 3. By looking at these values, one can find that the convergence behavior of this line is better than $17_{2,15} ← 16_{1,16}$, but is worse than $3_{1,3} ← 2_{2,0}$. The half-width at $T = 296$ K derived from the 8-th order and 20 correlations is 1.7% less than that from the 20-th order and 20 correlations and it is 10.9% less than that from the 20-th order and 88 correlations. This demonstrates a general trend that the smaller the half-width, the worse the convergence. Meanwhile, agreement between the calculated half-widths and the measurements for $9_{9,0} ← 8_{6,3}$ are not as bad as for $17_{2,15} ← 16_{1,16}$. It is also apparent that the agreement improves by adopting higher cut-offs, but one line does not mean too much.

Similarly, we show how calculated Re$(S_2)$, exp(-Re$(S_2)$), and $r_c \frac{V^2}{\epsilon} (1 - e^{-\text{Re}S_2})$ for this line vary as one adopts different choices of the cut-offs in Figs. 8 and 9. By comparing Fig. 8 with Figs. 4 and 6 and also comparing Fig. 9 with Figs. 5 and 7, one can find that Re$(S_2)$ of the line $9_{9,0} ← 8_{6,3}$ derived from different choices of the cut-offs also differ. However, their magnitudes are in between those of the lines $3_{1,3} ← 2_{2,0}$ and $17_{2,15} ← 16_{1,16}$. As a results effects from Re$(S_2)$ to calculated the half-widths in the region of $r_c < 3.83$ Å are less damped than $3_{1,3} ← 2_{2,0}$, but more than $17_{2,15} ← 16_{1,16}$. This implies that $V_{\text{atom-atom}}$ plays a more important role for $9_{9,0} ← 8_{6,3}$ than $3_{1,3} ← 2_{2,0}$, but less important for $17_{2,15} ← 16_{1,16}$. We will not repeat detailed discussions here because their profiles are in between those for the other two lines.
Figure 8. The same as Figure 4 except for the line of 9,0 ← 8,3 and multiplying exp(-Re(S_2)) by 10.

Figure 9. The same as Figure 5 except for the line of 9,0 ← 8,3 and multiplying exp(-Re(S_2)) by 10.

### 3.4. Summary of analyzing lines in the three categories

#### 3.4.1. Comparisons among lines in the three categories
In order to make comparisons among lines belonging to different categories more clearly, we present calculated integrands in Eq. (1) by three sets of curves in Fig. 10. Results corresponding to $3_{1,3} \leftarrow 2_{2,0}$, $9_{9,0} \leftarrow 8_{6,3}$, and $17_{2,15} \leftarrow 16_{1,16}$ are plotted by the black, the red, and the green curves, respectively. Meanwhile, for each of these lines values derived from the lowest choice of the cut-offs (i.e., the 8-th order and 20 correlations) and from the highest one (i.e., the 20-th order and 88 correlations) are represented by a dot-dashed and a solid curves. As shown by the figure, the magnitudes for these three lines reach their maxima at $r_c = r_{c,\text{min}}$ and these are almost the same. As $r_c$ increases, their magnitudes decrease with different rates. The $3_{1,3} \leftarrow 2_{2,0}$ line decreases at the slowest rate, while $17_{2,15} \leftarrow 16_{1,16}$ the fastest and its magnitude almost become zero at $r_c = 3.83$ Å. Meanwhile, the rate of $9_{9,0} \leftarrow 8_{6,3}$ is intermediate. Because they are integrands of the calculated half-widths, these patterns are consistent with the magnitudes of their half-widths. On the other hand, by looking at differences between their results obtained from different cut-offs, one can find that these three lines have different convergence behaviors from the best, to intermediate, and to worst, respectively.

![Figure 10. A plot to show comparisons among calculated $r_c V_f c / V^2 (1 - e^{-R_c S_f})$ for the three lines $3_{1,3} \leftarrow 2_{2,0}$, $9_{9,0} \leftarrow 8_{6,3}$, and $17_{2,15} \leftarrow 16_{1,16}$. For each of these lines, we plots two results derived from the lowest choice of the cut-offs (i.e., the 8-th order and 20 correlations) and from the highest one (i.e., the 20-th order and 88 correlations) by a dot-dashed and a solid curves, respectively. Meanwhile, the curves corresponding to the different lines are distinguished by three colors black, red, and green.](image-url)
Figure 11. A plot to show percentage contributions to calculated half-widths at T = 296 K from collisions whose closest distances ranging from \( r_{c,\text{min}} \) to \( r_c \). Results derived for the three lines \( 3_{1,3} \leftarrow 2_{2,0}, 9_{9,0} \leftarrow 8_{6,3}, \) and \( 17_{2,15} \leftarrow 16_{1,16} \) are represented by black, red, and green curves, respectively. In order to remove convergence errors, all the calculations are carried out with the 20-th order and including 88 correlations.

Finally, in order to provide a quantitative measure of how many contributions to the calculated half-widths come from specified collisions, one can introduce percentage contributions from collisions whose closest distances are in ranges from \( r_{c,\text{min}} \) to \( r_c \). The mathematical definition of this measure is \( \gamma(r_{c,\text{min}}, r_c)/\gamma(r_{c,\text{min}}, +\infty) \) % where the two arguments of \( \gamma \) are the lower and upper integration limits in Eq. (1) and the result is a function of \( r_c \). We have calculated them at T = 296 K with the “parabolic” trajectory model for all these three lines and we present the results in Fig. 11. In order to remove convergence errors, all the calculations are carried out with the 20-th order and including 88 correlations. Because the order of these three lines is arranged according to their half-widths from the largest to the smallest, one can conclude from the figure that the smaller the half-width, the larger the percentage contributions come from nearly head-on collisions. In fact, for the line of \( 17_{2,15} \leftarrow 16_{1,16} \) the dominant part of contributions (i.e., 92.6 %) comes from collisions with \( r_c \) from \( r_{c,\text{min}} \) (about 3.436 Å derived with the potential model used in the present study and the “parabolic” trajectory model) to 3.83 Å. The contributions from the same collisions become 75.6 % for \( 9_{9,0} \leftarrow 8_{6,3} \) and are reduced to 42.1 % for \( 3_{1,3} \leftarrow 2_{2,0} \). It is worth mentioning that the 92.6 % percentage contributions resulting from collisions in a very narrow region between 3.436 Å and 3.83 Å reflect an important fact that when one calculates half-widths for lines belonging to the same categories \( 17_{2,15} \leftarrow 16_{1,16} \), it is extremely crucial to model the short-range interactions and nearly head-on collision trajectories well. Any errors in modeling these could easily influence the calculated results and could thus lead to large uncertainties. As a result, we would like to make two comments for lines belong to the category here. The first is that

![Percentage Contributions](image)
the theoretically calculated results in the literature are more likely unreliable because they were derived from adopting lower cut-offs. The second is that even using our method to solve the convergence challenge, one must also describe nearly head-on collision trajectories accurately. At present, we know that the “exact” model is better than the “parabolic” one but is still based on the assumption that the trajectory is determined by the isotropic part of potential only. Unfortunately, one could expect that the anisotropic interaction can significantly affect the trajectories, and the closer the distance is, the larger the effect. So far, no significant progress has been made in this area, and one has to keep in mind that this challenge remains.

3.4.2. **Guidance to select lines to optimize potential models**

When people select lines to make comparisons between theoretically calculated half-widths and measurements, usually their main concern is about accuracies of the measurements. People prefer to choose important lines whose half-widths have been measured repeatedly and accurately. Besides this, it seems there is a lack of other theoretical considerations to guide people in making their choices. Analyses as described above could help theorists to make better choices.

From the theoretical point of view, how to select target lines should depend on one’s purpose. Thus, our suggestion is if one wants to optimize the short-range part of potential, one needs to select lines with small half-widths, or if one wants to optimize the long-range part, it is better to choose lines with large half-widths.

It is well known that models of the interaction potential between two molecules can be very complicated. Certain features of the potential may affect a given physical property, but other features would not. As a result, one has to combine all diagnostic means available to grasp its complexity. Even in measuring the same physical property, this general principle is also applicable. Therefore, it is better to select sample lines from all three categories and to optimize the long- and short-range potentials simultaneously.

4. Numerical calculations of the half-width

4.1. **Theoretically calculated N₂-broadened half-widths of H₂O lines in the pure rotational band**

After deriving the parameters $\sigma_{ij}$ and $\epsilon_{ij}$ from the fitting processes and obtaining the optimized site-site model for the H₂O – N₂ pair, Gamache and Laraia et al. have calculated half-widths for all other H₂O lines in the pure rotational band. In our opinion, for lines in the first category their calculated values are reliable. For lines in the second category their results are less reliable, and for lines in the third category their results are not reliable.

There are two arguments to support this claim. First, except for the isotropic part, the applicability of the anisotropic part of $V_{\text{atom-atom}}$ has not been tested in the fitting process. This statement looks strange because the applicability of the total of $V_{\text{atom-atom}}$ is, indeed, tested for the two lines selected by Gamache and Laraia et al. However, as we have explained above, in these tests effects from the anisotropic potential have been suppressed. This implies that the tests were not well done. How can one trust theoretically calculated half-widths for other lines whose significant contributions come from this untested anisotropic part of $V_{\text{atom-atom}}$? Thus, the larger the contributions from $V_{\text{atom-atom}}$ are, the less the reliable the results are. Secondly, based on our understanding of
the convergence problem, we are sure that many of the updated theoretical values in HITRAN are not converged. Because theoretically calculated half-widths and shifts play an important role in updating the HITRAN database, needless to say that non-converged results should not be used.

We have calculated N₂-broadened half-widths of all 1639 H₂O lines of the pure rotational band based on the same potential model used by Gamache and Laraia et al. In order to check the convergence problems related to the two kind of cut-offs and associated with two different trajectory models, we have performed a lot of numerical calculations by adopting different choices of the cut-offs and trajectory models and we present some of our results in the following.

4.1.1. Convergence check for the second kind of cut-off

In order to check convergence behaviors due to the second kind of cut-off, we present calculated half-widths of the 1639 H₂O line of the pure rotational band listed in HITRAN by selecting 8-th, 14-th, and 20-th order cut-offs in Fig. 12. The calculations are performed with including 88 correlations and adopting the “parabolic” trajectory model. We also calculate relative errors representing differences of half-width values from using the lower order cut-offs versus those from the highest one. These results are plotted in Fig. 13. In addition, we divide the 1639 lines into three groups according to their error ranges. For example, in comparison between results derived from 88 and 20 correlations, we find that there are 664 lines with the errors beyond 10 %, 403 lines within 4 – 10 %, and 572 lines below 4 %. We provide these general error statistics in Table 4. Both the maximum errors for the 8-th and 14-th order cut-offs happen for the line of 21₀,₂₁ ← 20₁,₂₀ located at 390.499 cm⁻¹ with the air-broadened half-width 0.0075 cm⁻¹ atm⁻¹ listed in HITRAN. Based on Figs. 12 and 13, one can conclude that for those lines with small half-width values, calculated results with 8-th order cut-off are not converged. Meanwhile, results from the 14-th order cut-off are well converged and those from the 20-th order are completely converged. The error plots in Fig. 13 clearly demonstrate that the division according to the relative errors into the three groups coincides with the division according to their half-width values into the three categories discussed above.
Figure 12. Comparisons between calculated N$_2$-broadened half-widths derived the 8-th, 14-th, and 20-th order cut-offs. These results are plotted by symbols +, $\times$, and $\Delta$, respectively. All the calculations are carried out from the “parabolic” trajectory model and by including 88 correlation functions. The 1639 H$_2$O lines of the pure rotational band are arranged according to the ascending order of their calculated half-widths from the 20-th order cut-off.

Figure 13. Relative convergence errors of the calculated half-widths resulting from adopting the 8-th and the 14-th order cut-offs versus calculated half-width values from the 20-th order cut-off. The errors associated with the 8-th order cut-off are represented
by symbols ∆ and that from the 14-th order are given by ×. The calculations are performed with including 88 correlations and adopting the “parabolic” trajectory model.

Table 4 Relative convergence errors for the second kind of cut-off

<table>
<thead>
<tr>
<th>Relative Errors</th>
<th>&lt; 4%</th>
<th>4 – 10 %</th>
<th>&gt; 10 %</th>
<th>Max. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-th &amp; para.</td>
<td>927 lines</td>
<td>415 lines</td>
<td>297 lines</td>
<td>42 %</td>
</tr>
<tr>
<td>14-th &amp; para.</td>
<td>1937 lines</td>
<td>2 lines</td>
<td>0 lines</td>
<td>4.2 %</td>
</tr>
</tbody>
</table>

It is worth mentioning here that these calculated relative errors correspond to the optimized potential model used in the present study only. We expect that if one uses different potential models in calculations, these results would vary somewhat, however, we believe that the general features would remain true. Similar claims are applicable for other cases presented later.

4.1.2. Convergence check for the first kind of cut-off

In order to check the convergence related to the first kind of cut-off, we present calculated N2-broadened half-widths for the 1639 lines by including 20, 38 and 88 correlations and adopting the “parabolic” trajectory model in Fig. 14. As shown in the Appendix, these three choices correspond to the values 2, 3, and 4 as the upper limit for L1 and to keep 2 as the upper limit for L2. With respect to the highest cut-offs (i.e., 20-th order and 88 correlations), we calculate relative errors for the results derived from the lower cut-offs. We do not present a figure to show the error distributions among all lines of interest here. Instead, in Table 5 we show the numbers of lines in the three error ranges. The maximum error (i.e., 70 %) for 20 correlations occurs for the line of $15_{15,0} \leftarrow 14_{14,1}$ located at 647.045 cm$^{-1}$ with air-broadened half-width 0.0029 cm$^{-1}$ atm$^{-1}$ given in HITRAN. Meanwhile, with respect to results derived from 38 correlations, the maximum error (i.e., 18 %) for 38 correlations occurs for the line of $20_{1,19} \leftarrow 20_{0,20}$ located at 364.066 cm$^{-1}$ with the air-broadened half-width 0.0075 cm$^{-1}$ atm$^{-1}$. As shown by the figure and the table, one can conclude that many lines calculated results with 20 correlations have convergence problems. In contrast, except for a few of the lines with small half-widths, results derived from 38 correlations are reasonably converged. Finally, we can reasonably assume that results with 88 correlations are well converged.
Figure 14. Comparisons between calculated N$_2$-broadened half-widths of all 1639 lines of the pure rotational band of H$_2$O derived by including 20, 38, and 88 correlation functions. These results are plotted by symbols $\triangle$, $\times$, and $\Delta$, respectively. All these calculations are performed based on the “parabolic” trajectory model and by adopting 20-th order cut-off. The 1639 H$_2$O lines are arranged according to the ascending order of the calculated half-widths by including 88 correlations.

Table 5 Relative convergence errors for the first kind of cut-off

<table>
<thead>
<tr>
<th>Relative Errors</th>
<th>&lt; 4%</th>
<th>4 – 10 %</th>
<th>&gt; 10 %</th>
<th>Max. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 cor. &amp; para.</td>
<td>572 lines</td>
<td>403 lines</td>
<td>664 lines</td>
<td>70 %</td>
</tr>
<tr>
<td>38 cor. &amp; para.</td>
<td>1479 lines</td>
<td>129 lines</td>
<td>31 lines</td>
<td>18 %</td>
</tr>
</tbody>
</table>

4.1.3. Comparison between results derived from different trajectory models

By comparing results derived from the “parabolic” and “exact” trajectory models, one is able to ascertain effects resulting from adopting a more accurate trajectory model. We compare calculated half-widths both obtained with the highest cut-offs to avoid any convergence errors. The comparison is presented in Fig. 15. As shown by the figure, in general, values of half-widths derived from the “exact” trajectory model are smaller than those from the “parabolic” model. We also present calculated relative errors in Table 6. The maximum error is 142 % for the line of 21$_{1,20}$ $\leftarrow$ 20$_{2,19}$ located at 408.325 cm$^{-1}$ in HITRAN. Based on these results, one can conclude that effects on the calculated half-widths from adopting the “exact” trajectory model are significant.
Figure 15. Comparisons between calculated N₂-broadened half-widths derived from the “parabolic” and the “exact” trajectory models. These results are plotted by symbols ×, and Δ, respectively. The calculations are carried out with the 20-th order cut-off and including 88 correlation functions.

Table 6 Relative errors due to adopting the “parabolic” trajectory model

<table>
<thead>
<tr>
<th>Relative Errors</th>
<th>&lt; 4%</th>
<th>4 – 10 %</th>
<th>&gt; 10 %</th>
<th>Max. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Parabolic” traj.</td>
<td>1085 lines</td>
<td>310 lines</td>
<td>244 lines</td>
<td>142 %</td>
</tr>
</tbody>
</table>

4.1.4. **Comparison between results obtained from the most accurate theoretical model and those commonly used in the literature**

In the present study, the most accurate and sophisticated method to repeat the same calculations is to adopt the 20-th order cut-off, to include 88 correlation functions, and to use the “exact” trajectory model. Meanwhile, theoretically calculated half-widths by Gamache and Laraia et al. were derived from the 8-th order cut-off, 20 correlation functions and based on the modified parabolic trajectory model. Because we don’t know exactly how the modified parabolic model was derived, we prefer to repeat their calculations with the same cut-offs, but with the original “parabolic” model. In addition, they use the original RB formalism and we use the modified RB one. As a result, our results derived from the 8-th order cut-off and 20 correlations do not match theirs exactly. But, comparisons presented here are adequate to show differences due to adopting the lower cut-offs and the simpler trajectory model versus the calculated half-widths using our most accurate method. We present the comparisons in Fig. 16. We also provide some numerical measures for the relative errors in Table 7. In this case, the maximum error (i.e., 68 %) occurs for the line of 16₁₅₂ ← 15₁₄₁ located at 672.672 cm⁻¹ with air-broadened half-width 0.0075 cm⁻¹ atm⁻¹ in HITRAN. As shown by the figure and the table, there are significant differences between these values for many lines. Thus, one can
conclude that results of Gamache and Laraia et al. [7] are not true values of the half-width which would be obtained without any artificial distortions from calculation errors.

Figure 16. Comparisons between calculated N$_2$-broadened half-widths derived from the most accurate and sophisticated method (i.e., to adopt 20-th order cut-off, to include 88 correlations, and to use the “exact” trajectory model) and those from the method used in the paper by Gamache et al. (i.e., with 8-th order, 20 correlations, and the “parabolic” trajectory model). These results are plotted by symbols Δ and ×, respectively.

Table 7 Relative errors associated with the current values by Gamache

<table>
<thead>
<tr>
<th>Relative Errors</th>
<th>&lt; 4%</th>
<th>4 – 10 %</th>
<th>&gt; 10 %</th>
<th>Max. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-th, 20c, parab</td>
<td>588 lines</td>
<td>415 lines</td>
<td>636 lines</td>
<td>68 %</td>
</tr>
</tbody>
</table>

It is worth mentioning that our claim does not mean that the values obtained from our most accurate method represent real half-width values because nobody has verified that the potential model used in the calculations is a “realistic” model which can represent the interaction between H$_2$O and N$_2$ well. In fact, we have repeatedly demonstrated that the current potential model not only suffers from the poorly selected line choices, but also from using the lower order cut-off, including fewer correlations, and adopting the simpler trajectory model. As a result, we do not believe that results derived from our most accurate method are able to predict all half-widths well. In addition, one has to keep in mind that the development of the method to do numerical calculations is completely based on the RB formalism. The latter contains several approximations which could limit its applicability in predicting accurate half-width values. Therefore, it is better to consider results derived from our most accurate and sophisticated method as “true results” within the modified RB formalism and the potential model without any artificial distortions.
4.1.5. Comparison between theoretically calculated half-widths and those listed in HITRAN

It is well known that calculated pressure broadened H\textsubscript{2}O half-widths sensitively depend on the interaction potential model between the H\textsubscript{2}O molecule and the bath molecule used in calculations. The importance and the necessity to derive converged values are obvious because only the converged results truly represent the potential model. The above argument is important because some positive conclusions to support the current potential model of the H\textsubscript{2}O – N\textsubscript{2} pair were drawn from comparisons between measurements and somehow “faulty results” representing this potential. Thus, these conclusions become suspect.

First of all, we present comparisons between calculated half-widths from the 8-th order cut-off, 20 correlations, and the “parabolic” trajectory model and air-broadened half-widths listed in the HITRAN database in Fig. 17. In plotting, we have transformed the N\textsubscript{2}-broadened half-widths to the air-broadened ones by multiply a factor of 1/1.09. From the figure, it appears as if the agreement between theoretically calculated results and the values listed in HITRAN is pretty good, although the good agreement may be specious.
Next we present a comparison between the theoretically calculated “true” half-widths derived from the 20-th order cut-off, including 88 correlations, and using the “exact” trajectory mode and those listed in HITRAN in Fig. 18. As shown by the figure, the agreement becomes poorer, especially for lines in the categories 2 and 3 (i.e., $\gamma < 0.075 \text{ cm}^{-1}/\text{atm}$). This is not surprising because we already know that values derived from the most accurate method are significantly larger than those obtained from the simplest method, especially for lines with small half-widths. We note that some data in HITRAN may be also be responsible for these discrepancies because many data come from theoretically calculations that Gamache’s group made a few of years ago. Among them, those with small values are not yet converged. In summary, these comparisons clearly demonstrate the necessity for a replacement of the current potential model.

4.2. Temperature dependence of the half-width

In atmospheric applications, knowledge of the line parameters at temperatures below 296 K is essential. Therefore, the HITRAN database provides a way how to derive these parameters at temperatures of interest from those listed for 296 K. One assumes its temperature dependence (T dependence) can be described as

$$\gamma(T) = \gamma(T_0) \times \left(\frac{T_0}{T}\right)^n,$$

where $T_0 = 296$ K and $n$ is the temperature exponent. We note that a positive $n$ means the T dependence is negative and a negative $n$ means the T dependence is positive. In HITRAN, a value of $n$ is assigned according to values of $|m|$ where $m = -j''$ for the P branch and $m = j'' + 1$ for the R branch. For example, $n = 0.78$ for $|m| = 0, 1, \text{ and } 2$; $n = 0.77$ for $|m| = 3; n = 0.73$ for $|m| = 4$; and so on. This implies that in HITRAN all lines are
assumed to have negative T dependences. Recently, several groups have questioned the correctness of this assignment formula because they have found that this simple assignment method does not always work well and could yield big errors [19,20]. In addition, based on theoretical calculations they claim that the temperature exponent n can be negative for many transitions [7]. In general, we also doubt the applicability of this simple assignment of n in HITRAN for some lines and we are cautiously open minded to accept possible negative n values. Given the fact that lines claimed to have negative n values are those with small half-widths, whose calculated values are more likely to suffer from convergence problems resulting both from adopting lower-order cut-offs and including fewer correlations, the calculated temperature exponent n have inevitably convergence problem also. Thus, one can not trust all claims drawn from previous theoretical calculations. Therefore, it is worthwhile for us to also investigate the T dependence.

Fortunately, if one adopts the “parabolic” trajectory model, with our method to calculate \( \gamma \) at other temperatures of interest is not difficult because the correlation functions whose argument is defined as \( z = \frac{V_c t}{t_c} \) are common for all temperatures [5]. In other words, in order to calculate \( \gamma(T) \), one does not need to derive all these correlation functions again. However, if one selects the “exact” trajectory model, one has to recalculate the correlations for each of other temperatures of interest. We have calculated the half-width values for \( T = 220 \) K for the two lines of 6\(_{1,6} \leftarrow 5\_{2,3} \) and 3\(_{1,3} \leftarrow 2\_{2,0} \) belonging to the category 1 with several choices of the cut-offs and trajectory models and already provide these results together with those obtained for \( T = 296 \) K in Table 2. For other sample lines (i.e., 17\(_{2,15} \leftarrow 16\_{1,16} \) and 9\(_{9,0} \leftarrow 8\_{6,3} \)) in the categories 3 and 2, respectively, we have also listed corresponding results for \( T = 296 \) K and 220 K in Table 3. As shown by these tables, all lines in the categories 1 and 2 exhibit negative T dependences. For the line 17\(_{2,15} \leftarrow 16\_{1,16} \) in category 3, one can find that results obtained with the lowest cut-offs have a positive T dependence, but all other values from higher cut-offs indicate the T dependences are still negative. This implies that due to convergence errors, the prior claim is suspect.
Figure 19. The T-dependence of the half-width for the line of $16_{14,2} \leftarrow 15_{13,3}$ in the range from 220 K – 340 K derived from adopting different cut-offs and including different numbers of the correlation functions. The highest cut-off is the 20-th order and the largest number of the correlations is 88. Meanwhile, the lowest cut-off and the smallest number of the correlations are 8-th order and 20, respectively.

Furthermore, in the recent paper by Gamache and Laraia et al. [7] they have plotted theoretically calculated T dependences for three lines: $19_{7,13} \leftarrow 18_{6,12}$, $2_{1,1} \leftarrow 2_{0,2}$, and $16_{14,2} \leftarrow 15_{13,3}$. The potential model is the optimized one, the same used in the present study. Based on their plots, they claim that for the first two lines, values of $n$ are positive and for the last line, $n$ is negative. It is an easy job for us not only to repeat their calculations, but also to make a comprehensive convergence checks by using much higher cut-offs to guarantee the convergence. Based on our checks, we find that their $n$ value for $2_{1,1} \leftarrow 2_{0,2}$ are well converged, that for $19_{7,13} \leftarrow 18_{6,12}$ is reasonably converged, but their $n$ for $16_{14,2} \leftarrow 15_{13,3}$ is non-converged at all. We present the convergence check for the line of $16_{14,2} \leftarrow 15_{13,3}$ in Fig. 19 and do not present other two checks here. The calculations are carried out with the “parabolic” trajectory model and with several choices of the cut-offs including the highest one. As shown by the figure, results derived from the two lower cut-offs (i.e., the 8-th order plus 20 correlations and the 20-th plus 20 correlations) exhibit strong positive T dependences and those from the two higher ones (i.e., 20-th plus 39 correlations and 20-th plus 88 correlations) exhibit mild negative T dependence. Figure 19 demonstrates that the claim of a negative $n$ for $16_{14,2} \leftarrow 15_{13,3}$ is due to convergence errors. It is worth mentioning again that our results presented here are dependent on the potential model used in calculations. Because the current potential model is not good, one should not consider these results as reliable theoretical predictions.

5. Uncertainty analyses and conclusions
In order to investigate the uncertainties associated with the calculated half-widths, let’s consider individual uncertainties associated with contributions from certain types of collisions first. One can claim that, in general, contributions from nearly head-on collisions contain the largest uncertainties and those from glancing collisions have the smallest ones. Besides the fact that magnitudes of the former are larger than the later, there are other arguments to support this claim.

First of all, in comparison with a relatively well-established knowledge of the long-range potential, people do not know the short-range interaction very well. More specifically, precise expressions for the long-range multipole interactions are available and values of the multipole moments for H_2O and N_2 are well known. This is not true for the short-range interactions. Thus, there is lack of guidance how to select the short-range models, not only their parameters, but the functional forms themselves. Usually, one uses the LJ form in developing the site-site models. This preference does not result from physical considerations, rather mainly from its technical convenience in manipulating the spherical expansions for the LJ function in terms of the inverse powers of \( r_{ij} \). As demonstrated above, how to choose the LJ parameters could contain other problems. Furthermore, we have repeatedly shown that there could be large convergence errors, which mainly occur for nearly head-on collisions, because the cut-offs are not high enough to guarantee the convergence. Therefore, all these uncertainties are related to the short-range interactions.

When one calculates contributions from nearly head-on collisions, the short-range potential is the major or even the dominant component of the total potential. Thus, it must introduce uncertainties into nearly head-on collision processes as well. On the other hand, glancing collisions are not sensitive to the short-range potential. Based on these discussions it is obvious that the uncertainties due to the potential model mainly occur for the nearly head-on collisions and become less for the more glancing ones.

Next, we consider uncertainties associated with collision trajectories. As mentioned in Section 2.3, at present all the half-width calculations for the H_2O – N_2 pair are based on the assumption that the trajectories are governed by the isotropic potential only. In addition, one also assumes that the translational motion can be treated classically. Except for these two assumptions, the “exact” trajectory model does not have any others. Thus, it is the most accurate and physically consistent one among the trajectory models. Therefore, we prefer to consider this model here. As shown by Fig. 1, the “exact” trajectories would deviate dramatically as \( r_c \) varies slightly for the nearly head-on collisions but they would change very little for the glancing ones. This pattern indicates that when one depicts trajectories for the nearly head-on collisions from the isotropic interaction, the latter’s small uncertainties would be significantly enhanced.

Besides, there are more profound effects resulting from the basic assumption that the anisotropic interaction does not play any role in determining the trajectories. The drawback of this assumption is not only that couplings between the translation and internal motions are completely ignored, but also the trajectories derived from the potential are less accurate. At present, we are not able to estimate effects from ignoring the couplings, but we know how to roughly judge effects from the unsuitably depicting of trajectories. Given the fact that the whole interaction, as well as its anisotropic component, show their strong effects for nearly head-on collisions and have little effects on the glancing ones, one should respect that the uncertainty occurs for the former, but
not for the latter. In summary, one can conclude that large uncertainties in depicting the trajectory definitely exist and these uncertainties would inevitably affect calculated contributions from the nearly head-on collisions.

At this stage, we would like to point out that there are other uncertainties existing in the current RB formalism. One is that contributions from $S_3$, the third-order expansion of the $\hat{S}$ matrix, are not included in the expression for the half-width. Recently, there have been efforts to derive formulas for contributions from $S_3$, and even from $S_4$ [21]; however, these formulas are only applicable for cases where the potential model does not contain site-site interactions [22]. Further work on this subject is required, and even after theoretical formulas are available, to carry out practical calculations remains a big challenge. We expect that the coordinate representation could be a helpful tool in solving this problem. Therefore, at present nobody is sure that results calculated using only the first two terms $S_1$ and $S_2$ are converged or not. Another problem is associated with the assumption that with respect to states of the absorber molecule the resolvent operator appearing in the expression for the spectral density $F(\omega)$ is diagonal. This implies the line-couplings are not taken into account in calculations. Due to the importance for certain transitions (Q branches, band heads, etc.), there have been many recent studies on the line-coupling in other molecules. There has been some progress for the H$_2$O molecule, but more work on this subject is also required.

From the discussion presented above, one can conclude that large uncertainties are always associated with the nearly head-on collisions. Thus, with respect to contributions to the calculated half-widths of the line of interest, the determination of the contribution coming from the nearly head-on collisions is a good measure of the uncertainties. On the other hand, we have already shown in Fig. 11 that the largest share of contributions happens for lines in category 3 and the smallest for lines in category 1. Therefore, one can claim that the division of lines into different categories according to their achievable accuracies coincides with the division according to magnitudes of their half-width values introduced previously. More specifically, theoretically calculated half-widths can achieve the highest accuracy for lines with large half-widths and they could suffer the worst uncertainties for lines with small half-widths. Meanwhile, lines with moderate half-width values would contain moderate errors.

There are no sharp half-width boundaries to distinguish these three categories, but we would like to offer our suggestion. For example, according to the air-broadened half-width values listed in HITRAN, the three categories can be divided into greater than 0.075, between 0.075 and 0.045, and less than 0.045 (in units of cm$^{-1}$ atm$^{-1}$), respectively. With this division, among the 1639 lines in the pure rotational band of H$_2$O, there are 268 lines in category 1, 661 lines in category 2, and 710 lines in category 3. Similarly, there are also no sharp uncertainty boundaries either. But, we can provide uncertainty estimations for the accuracy of theoretically calculated N$_2$-broadened half-widths of the H$_2$O lines that one can conservatively achieve. In contrast with laboratory data where measured half-widths of weak lines are more likely to have larger uncertainties than stronger lines, this is not necessarily true for theoretical calculations. Therefore, the uncertainty estimations are applicable for both weak and strong lines in the same category. We believe that for lines in the category 1, one is able to attain 5 % accuracy, which is close to the 3 % set by HITRAN for strong lines and better than the 10 % set for weak lines. In contrast, for lines in the category 3 it is absolutely impossible to meet the
HITRAN requirements, even the 10 % level. We believe that the estimations for lines in category 3 should be much worse. (Our most optimistic guess is at least 20 – 30 %.) For lines in the category 2, our estimation of the uncertainty is between these two extremes.

The above claim provides a theoretical uncertainty with which one is able to achieve with the current RB formalism. We note that the theoretical uncertainty here means one is sure that the results presented arise from a sound physical basis and the relative error from the true value is seldom beyond this limitation. It is worth mentioning that when we claim we do not believe the 10 % accuracy are achievable for lines in the category 3, that does not mean that values of the half-width of all these lines appearing in the literature contain at least 10 % errors. In fact, it could happen that some theoretically predictions may match the true values with smaller errors; rather we expect that many lines in this category contain errors larger than 10 %. The key point here is due to distortions occurring in calculations, the trustfulness built on a sound physics could be lost. For example, no matter how good the calculated results look like, if these values are not converge nobody would accept them as real representatives of the physics properties.

There is an urgent need to provide accurate theoretical results for line widths and shifts for many practical applications. It appears that one still needs to rely on the RB formalism to do calculations at present, because close-coupling calculations are not feasible for the H₂O – N₂ pair. However, we believe that a responsible way to address the problem is when one provides theoretically predicted values, one should provide uncertainties estimations as well.

Finally, we would like to make two suggestions. First, when one updates the half-width values in HITRAN using all sources of measurements and theoretical calculations, it is useful to distinguish lines according to their half-width values. It would be prudent if one weights the measurements more than the theoretical calculations for lines in the category 3 and vice versa for lines in the category 1. Secondly, when experimentalists determine their measurement priorities based on all their lab considerations, it would be nice to consider one more: due to the weakness suffered by theoretical calculations for lines in category 3, the measurements enjoy an extra reliability advantage.

Acknowledgements
Two of authors (Q. Ma and R. H. Tipping) acknowledge financial support from NASA under grants NNG06GB23G, NNX09AB62G, and FCCS-547. Q. Ma wishes to acknowledge financial support from the Biological and Environmental Research Program (BER), U.S. Department of Energy, Intergency Agreement No. DE-AI02-93ER61744. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. R. R. Gamache is pleased to acknowledge support of this research by the National Science Foundation through Grant No. ATM-0803135.

References
1. The latest HITRAN database is available at www.cfa.harvard.edu/hitran


Appendix
Assignments of \(L_1\), \(K_1\), \(K_1'\), and \(L_2\) for the Correlation Functions

There are six representations in defining wave functions for asymmetric top molecules. For the H\(_2\)O molecule, most of people use the IR representation because this representation is more suitable to diagonalize of the Hamiltonian of H\(_2\)O and to derive accurate eigenvalues and eigenvectors. However, concerning calculations of the half-widths the IIR and IIL representations have more advantages than the others because with these representations one is able to exploit the symmetry associated with the assignment that in the molecular-fixed frame, the two H atoms are symmetrically located. It is well known that the eigenvalues are independent of the choice of the representation but wave functions depend on it. By unitary transformations, one can transform the wave functions from one representation into another. Detailed discussions about the transformation and the two different conventions used to define asymmetric top wave functions in terms of the symmetric top wave functions will be given elsewhere. Therefore, without losing generality, we use the IIR representation to develop all formulas.

There are several symmetries available. Because the two H atoms are symmetrically located, \(K_1\) must be even in the coefficients \(u(L_1L_2L;K_1;R(t))\) of \(V(\tilde{R}(t))\) in Eq. (5) and \(u(L_1L_2L;K_1;R(t)) = u(L_1L_2L;-K_1;R(t))\). Then, due to the exchange symmetry between \(K_1\) and \(K_1'\) in the product of \(u(L_1L_2L;K_1;\tilde{R}(t' + t/2))\) and \(u(L_1L_2L;K_1';\tilde{R}(t' - t/2))\) appearing in Eq. (10), the correlation functions have the same exchange symmetry between \(K_1\) and \(K_1'\). As a result, some of the correlation functions become identical. Thus, the number of the correlations required to evaluate can be reduced significantly. For example, the four correlations labeled by (2 2 2 0), (2 2-2 0), ((2 2 2 0) and (2-2 2 0) are identical and the eight correlations labeled by (4 4 2 2), (4 4 -2 2), (4-4 2 2), (4-4-2 2), (4 2 4 2), (4-2-4 2), (4-2 4 2), and (4-2-4 4) are identical. We list all correlations with the upper limit 4 for \(L_1\) and the upper limit 2 for \(L_2\) in Table A. In the column labeled by \#C, we list the numbers for each of the correlations and in a column labeled by \#IC, we list the numbers of groups whose members share the same functions. As shown by the table, there are 88 individual correlations and 26 independent groups.

<table>
<thead>
<tr>
<th># C</th>
<th># IC</th>
<th>L(_1)</th>
<th>K(_1)</th>
<th>K(_1')</th>
<th>L(_2)</th>
<th># C</th>
<th># IC</th>
<th>L(_1)</th>
<th>K(_1)</th>
<th>K(_1')</th>
<th>L(_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td>17</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>52</td>
<td>18</td>
<td>4</td>
<td>-4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>53</td>
<td>19</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>54</td>
<td></td>
<td>4</td>
<td>0</td>
<td>-4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td>0</td>
<td>55</td>
<td></td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
<td>56</td>
<td></td>
<td>4</td>
<td>2</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>57</td>
<td></td>
<td>4</td>
<td>-2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>58</td>
<td></td>
<td>4</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>59</td>
<td></td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>2</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>60</td>
<td></td>
<td>4</td>
<td>-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>61</td>
<td></td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>62</td>
<td>4</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td>63</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td>2</td>
<td>64</td>
<td>21</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>2</td>
<td>65</td>
<td>4</td>
<td>4</td>
<td>-4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>66</td>
<td>4</td>
<td>-4</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>67</td>
<td>4</td>
<td>-4</td>
<td>-4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>68</td>
<td>22</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>0</td>
<td>-2</td>
<td>2</td>
<td>69</td>
<td>4</td>
<td>4</td>
<td>-2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>70</td>
<td>4</td>
<td>-4</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>71</td>
<td>4</td>
<td>-4</td>
<td>-2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>72</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>-2</td>
<td>2</td>
<td>0</td>
<td>73</td>
<td>4</td>
<td>2</td>
<td>-4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
<td>74</td>
<td>4</td>
<td>-2</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>75</td>
<td>4</td>
<td>-2</td>
<td>-4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>3</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>76</td>
<td>23</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>77</td>
<td>4</td>
<td>-4</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>3</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>78</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>79</td>
<td>4</td>
<td>0</td>
<td>-4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>80</td>
<td>24</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>3</td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td>81</td>
<td>4</td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>3</td>
<td>-2</td>
<td>2</td>
<td>2</td>
<td>82</td>
<td>4</td>
<td>-2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3</td>
<td>-2</td>
<td>-2</td>
<td>2</td>
<td>83</td>
<td>4</td>
<td>-2</td>
<td>-2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>13</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>84</td>
<td>25</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>35</td>
<td>3</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>85</td>
<td>4</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>86</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>3</td>
<td>0</td>
<td>-2</td>
<td>2</td>
<td>87</td>
<td>4</td>
<td>0</td>
<td>-2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>14</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>88</td>
<td>26</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>39</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>42</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>4</td>
<td>4</td>
<td>-4</td>
<td>0</td>
<td>41</td>
<td>4</td>
<td>-4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>4</td>
<td>-4</td>
<td>-4</td>
<td>0</td>
<td>42</td>
<td>4</td>
<td>-4</td>
<td>-4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>43</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>4</td>
<td>4</td>
<td>-2</td>
<td>0</td>
<td>45</td>
<td>4</td>
<td>-4</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>4</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
<td>46</td>
<td>4</td>
<td>-4</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>47</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>4</td>
<td>2</td>
<td>-4</td>
<td>0</td>
<td>48</td>
<td>4</td>
<td>2</td>
<td>-4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>4</td>
<td>-2</td>
<td>4</td>
<td>0</td>
<td>49</td>
<td>4</td>
<td>-2</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>4</td>
<td>-2</td>
<td>-4</td>
<td>0</td>
<td>50</td>
<td>4</td>
<td>-2</td>
<td>-4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>