Theoretical rotational–vibrational spectroscopy of $XY_3$–type molecules of industrial relevance

Phillip A. Coles

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Physics & Astronomy

University College London

September 24, 2019
I, Phillip A. Coles, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.
For my parents, to whom I owe everything.
Complete list of publications


List of associated publications

Chapter 3:


Chapter 4:

Abstract

This thesis presents research carried out as part of the ExoMol project in partnership with Servomex Ltd. The overarching aim has been to compute high-accuracy line lists for molecules of astronomical and industrial relevance, namely $^{14}\text{NH}_3$ and $^{75}\text{AsH}_3$. These line lists are to be used in high accuracy spectroscopic studies, atmospheric spectral retrievals, and to inform decisions regarding the development of new in situ gas analysers.

A high accuracy spectroscopic potential energy surface (PES) for $^{14}\text{NH}_3$ has been produced by refinement of a recently published ab initio surface to carefully chosen experimental data. The resulting energy level predictions represent a 5–10 times improvement over the previous best predictions computed as part of the ExoMol project. Several new ab initio dipole moment surfaces (DMSs) were analysed, but were found to be generally inferior to an older surface available. Using the new PES and older DMS, a room temperature $^{14}\text{NH}_3$ line list was produced for wavenumbers between 0 and 20 000 cm$^{-1}$.

Exploratory NH$_3$ measurements were performed at 1392 nm using second harmonic wavelength modulation spectroscopy. The measurements aimed to identify the key NH$_3$ absorption features that might interfere with trace moisture detection in high purity NH$_3$ used in the development of GaN based diode lasers. Line positions were derived for the strongest NH$_3$ lines in this region, with an estimated uncertainty of 0.05 cm$^{-1}$.

A room temperature $^{75}\text{AsH}_3$ line list is produced for wavenumbers between 0 and 7000 cm$^{-1}$, and suitable for use up to 300 K. The required PES and DMS were produced by fitting analytic expressions to a grid of nuclear geometries and dipole moments generated from electronic structure calculations. The PES is then refined to experimental data data to improve accuracy. Final line positions and intensities are suitably accurate for industrial modelling purposes.
Impact Statement

The work presented in this thesis represents an important advancement in the spectroscopy of the molecules ammonia and arsine. These molecules are highly poisonous, and are involved in various industrial processes, for example, semiconductor manufacturing, combustion, power generation, smelting, and many others. The development of sensors to detect ammonia and arsine concentration in industry is therefore important to increase product yield, and to prevent the escape of these molecules which harms the environment and nearby human populace. The two theoretical line lists generated in this thesis predict the absorption line positions and intensities of ammonia and arsine with unprecedented accuracy, and can be used to inform the development of sensors for industrial trace detection applications. Both line lists were specifically requested by Servomex Ltd., who are the UK’s leading producers of gas analysers.

Both molecules have been observed in the atmospheres of Jupiter and Saturn, and ammonia has also been observed in the interstellar medium, atmospheres of brown dwarf stars, and cometary coma. Therefore, it is important for accurate and complete line lists to be available to astronomers, in particular those who are interested in the characterisation of exoplanets.

From a quantum chemistry and nuclear motion point-of-view, arsenic is the heaviest atom for which there now exists an associated variationally computed molecular line list. Therefore the tests and comparisons of quantum chemistry methods performed here should be of fundamental use to chemists and spectroscopists interested in investigating similar systems in future.

Within the spectroscopy community, ammonia is one of the molecules of key interest. Assignment of experimentally measured spectra often requires accurate theoretical line lists, and so the ammonia line list, which represents a large improvement on what was previously available, should be widely used by experimental spectroscopists. The energy level predictions of ammonia calculated here are also of use to ongoing projects, such as MARVEL, which aims to compile, analyse and evaluate all available experimentally measured transitions of ammonia. The
energy level predictions reported in this thesis have already been used to identify a significant amount of incorrect data which had gone unnoticed in the MARVEL ammonia database.
Acknowledgements

I would like to extend my deepest thanks to Prof. Jonathan Tennyson, for taking me under his wing and providing me with a great deal of support over the past four years. Thanks also go to Prof. Sergey Yurchenko and Dr. Richard Kovacich for their technical help throughout my PhD, and to James Hobby and Servomex Ltd. for making this project possible.
# Contents

1 Introduction 25

1.0.1 Spectroscopic databases .............................................. 26
1.0.2 ExoMol ................................................................. 27
1.0.3 Thesis overview ...................................................... 29

2 Theoretical background 31

2.1 Solving the electronic Schrödinger equation ...................... 35

2.1.1 Hartree-Fock ......................................................... 36
2.1.2 Configuration Interaction . ........................................... 37
2.1.3 Coupled Cluster ....................................................... 38
2.1.4 One electron basis sets .............................................. 40
2.1.5 Explicitly correlated coupled cluster ......................... 43
2.1.6 Relativistic effects .................................................. 45

2.2 TROVE ................................................................. 49

2.2.1 Kinetic energy operator ............................................. 51
2.2.2 Coordinates .......................................................... 52
2.2.3 Potential energy function ......................................... 54
2.2.4 Rovibrational basis ................................................ 55
2.2.5 Refinement ............................................................. 57

2.3 Simulating absorption spectra ......................................... 59

2.3.1 Line strength and selection rules ............................... 59
2.3.2 Dipole moment surface ............................................ 62
2.3.3 Line intensities and absorption cross-sections ............. 64

2.4 f/wms ................................................................. 67
## 3 Ammonia ($^{14}$NH$_3$)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>71</td>
</tr>
<tr>
<td>3.2 Quantum labels and symmetry</td>
<td>73</td>
</tr>
<tr>
<td>3.3 Potential energy surface</td>
<td>76</td>
</tr>
<tr>
<td>3.3.1 Computational details</td>
<td>77</td>
</tr>
<tr>
<td>3.3.2 Experimental data and weights included in the refinement</td>
<td>79</td>
</tr>
<tr>
<td>3.3.3 Refined parameters</td>
<td>82</td>
</tr>
<tr>
<td>3.3.4 Equilibrium structure and rotational energies</td>
<td>83</td>
</tr>
<tr>
<td>3.3.5 Rovibrational term values</td>
<td>84</td>
</tr>
<tr>
<td>3.4 C2018 line list</td>
<td>96</td>
</tr>
<tr>
<td>3.4.1 Overview</td>
<td>96</td>
</tr>
<tr>
<td>3.4.2 Computational details</td>
<td>96</td>
</tr>
<tr>
<td>3.4.3 Dipole moment surface</td>
<td>99</td>
</tr>
<tr>
<td>3.4.4 Results and discussion</td>
<td>105</td>
</tr>
<tr>
<td>3.5 Assignment of the 7400-8000 cm$^{-1}$ region</td>
<td>113</td>
</tr>
<tr>
<td>3.5.1 Ground state combination differences</td>
<td>114</td>
</tr>
<tr>
<td>3.5.2 Assignments and derived upper state energies</td>
<td>117</td>
</tr>
<tr>
<td>3.5.3 Discrepancies with HITRAN</td>
<td>118</td>
</tr>
<tr>
<td>3.6 Measurement of the 7169–7195 cm$^{-1}$ region</td>
<td>120</td>
</tr>
<tr>
<td>3.6.1 Overview</td>
<td>121</td>
</tr>
<tr>
<td>3.6.2 Experimental setup</td>
<td>122</td>
</tr>
<tr>
<td>3.6.3 Calibration</td>
<td>123</td>
</tr>
<tr>
<td>3.6.4 Simulation of H$_2$O second harmonic spectrum</td>
<td>126</td>
</tr>
<tr>
<td>3.6.5 Dilute NH$_3$ spectrum</td>
<td>129</td>
</tr>
<tr>
<td>3.6.6 &gt; 99% NH$_3$ spectrum</td>
<td>132</td>
</tr>
<tr>
<td>3.6.7 Determination of H$_2$O concentration</td>
<td>138</td>
</tr>
<tr>
<td>3.6.8 Discussion and future work</td>
<td>139</td>
</tr>
<tr>
<td>3.7 Conclusion</td>
<td>141</td>
</tr>
</tbody>
</table>

## 4 Arsine ($^{75}$AsH$_3$)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>143</td>
</tr>
<tr>
<td>4.2 Tunneling and molecular symmetry group</td>
<td>144</td>
</tr>
<tr>
<td>4.3 Potential energy surface</td>
<td>146</td>
</tr>
</tbody>
</table>
# List of Figures

1.1 Flow chart of the procedure typically used to generate an ExoMol line list, taken from Ref. [260]. .......................................................... 28

2.1 Reference configuration of a general XY$_3$ molecule in the molecule-fixed axis system, the origin lies as the centre of mass. Figure taken from Ref. [306]. 53

3.1 Difference between $J = 0 − 10$ MARVEL term values under 6300 cm$^{-1}$ and those of C2018 and BYTE. .......................................................... 94

3.2 Difference between $J = 0 − 10$ MARVEL term values above 6300 cm$^{-1}$ and those of C2018 and BYTE. Only MARVEL levels derived from 3 or more transitions are shown. .......................................................... 94

3.3 Basis set convergence of $J = 20$ (E$'$ symmetry) energies as ($J = 0$)—contracted basis set threshold $\varepsilon$ is increased from 26 000 to 32 000. The difference $E_{\varepsilon=x} - E_{\varepsilon=x+2000}$, is displayed for $x = 26 000, 28 000, 30 000$ vs the energies computed using $\varepsilon = 32 000$ cm$^{-1}$ .................................................. 98

3.4 Dimensions of the E-symmetry matrices (squares) and the corresponding number of eigenvalues below 23 000 cm$^{-1}$ (circles). ......................... 99

3.5 Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 0–1200 cm$^{-1}$ region. ................................. 101

3.6 Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 1200–2150 cm$^{-1}$ region. ................................. 101

3.7 Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 2100–2900 cm$^{-1}$ region. ................................. 102

3.8 Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 2900–3700 cm$^{-1}$ region. ................................. 102
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.9</td>
<td>Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 4000–5300 cm$^{-1}$ region.</td>
<td>103</td>
</tr>
<tr>
<td>3.10</td>
<td>Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 6300–7000 cm$^{-1}$ region.</td>
<td>103</td>
</tr>
<tr>
<td>3.11</td>
<td>Cross-sections calculated using the C2018 PES and DMS-B/DMS-001 dipole moment surfaces. Lines have been convoluted with a Gaussian profile with HWHM = 0.5 cm$^{-1}$.</td>
<td>104</td>
</tr>
<tr>
<td>3.12</td>
<td>Overview of the C2018/DMS-B line list compared to HITRAN 2016.</td>
<td>105</td>
</tr>
<tr>
<td>3.13</td>
<td>Overview of strongly absorbing regions as given in HITRAN 2016 compared to the theoretical predictions of the C2018/DMS-B line list.</td>
<td>106</td>
</tr>
<tr>
<td>3.14</td>
<td>Comparison of the predicted C2018/DMS-B line intensities with the experimental values of HITRAN 2016.</td>
<td>107</td>
</tr>
<tr>
<td>3.15</td>
<td>Comparison of the C2018/DMS-B and BYTe line lists with HITRAN 2016 for three small windows within the 0–5500 cm$^{-1}$ range.</td>
<td>108</td>
</tr>
<tr>
<td>3.16</td>
<td>Comparison of the C2018/DMS-B and BYTe line lists with HITRAN 2016 for two expansions of the 6300–7000 cm$^{-1}$ region.</td>
<td>109</td>
</tr>
<tr>
<td>3.17</td>
<td>Synthetic $J = 0 - 20$ spectrum computed at 298.15 K compared to PNNL for the 5700–6200 cm$^{-1}$ region.</td>
<td>110</td>
</tr>
<tr>
<td>3.18</td>
<td>Comparison of the simulated (C2018/DMS-B, C2018/DMS-001 and BYTe) and observed [273] spectra of NH$_3$ at $T = 293$ K for 7400–8600 cm$^{-1}$ region, with expansions of the 7660–7760 cm$^{-1}$ region (middle row) and 8200–8300 cm$^{-1}$ region (bottom row).</td>
<td>111</td>
</tr>
<tr>
<td>3.19</td>
<td>Comparison of the simulated (C2018/DMS-B and BYTe) and observed [18] spectra of NH$_3$ at $T = 296$ K for 9000–10400 cm$^{-1}$ region, with expansions of the 9295–9329.5 cm$^{-1}$ region (first row from top), the 9720–9785 cm$^{-1}$ region (first row from bottom), and the 10080–10125 cm$^{-1}$ region (bottom row).</td>
<td>112</td>
</tr>
<tr>
<td>3.20</td>
<td>Agreement between energy levels derived from our assignments and the values predicted by the C2018 energies list for 6 vibrational bands. Differences between the observed and calculated term values, $E_{\text{obs}} - E_{\text{calc}}$, are given in units of cm$^{-1}$.</td>
<td>118</td>
</tr>
</tbody>
</table>
3.21 Sample combination difference pair that were reassigned during our analysis. Lines convoluted with a Gaussian profile HWHM=0.06.

3.22 Photodetector output for a single scan covering the 1392.5335 nm H$_2$O line (top), and the corresponding etalon signal (left) and tuning rate function fit (right).

3.23 Tuning rate coefficients used in Eq. (3.4).

3.24 Second harmonic signal of the 1392.5335 nm H$_2$O line measured for a range of current modulation amplitudes (right), and the corresponding central peak maxima as a function of modulation amplitude (left).

3.25 Comparison of the measured and synthetic 2$f$ H$_2$O spectra using the determined values of linewidth and modulation index.

3.26 Expanded view of the 7190–7192 cm$^{-1}$ region of our scan, recorded at 3% and > 99% NH$_3$ concentration, in comparison with a simulated second harmonic H$_2$O spectrum. Two probable water features and 3 NH$_3$ features are identified.

3.27 Complete scan of the 7186–7181 cm$^{-1}$ region measured using the 3% NH$_3$ gas sample (middle panel, black) and > 99% NH$_3$ gas sample (bottom panel, red), in comparison with a simulation of the second harmonic spectrum of H$_2$O (upper panel, blue). Dashed black lines indicate NH$_3$ peaks, grey lines indicate H$_2$O peaks.

3.28 Complete scan of the 7181–7195 cm$^{-1}$ region measured using the 3% NH$_3$ gas sample (middle panel, black) and > 99% NH$_3$ gas sample (bottom panel, red), in comparison with a simulation of the second harmonic spectrum of H$_2$O (upper panel, blue). Dashed black lines indicate NH$_3$ peaks, grey lines indicate H$_2$O peaks.

3.29 Direct absorption signal of the 1392.5335 nm H$_2$O line at 0.1 bar pressure.

3.30 Direct absorption signal of the 1392.5335 nm H$_2$O line at 1 bar pressure.

4.1 One dimensional cuts of the relativistic corrections for the ($r_1 = r_2 = r_3 =$ 1.51 Å; $\alpha_1 = \alpha_2 = 92.1^\circ$; 50 $\leq \alpha_3 \leq$ 140$^\circ$) bond angle and ($r_1 = r_2 = 1.51$ Å; $1.2 \leq r_3 \leq 2.2$ Å; $\alpha_1 = \alpha_2 = \alpha_2 = 92.1^\circ$) bond length displacements.
4.2 Agreement between observed \( J = 1 - 6 \) term values \( E_{\text{obs}} \) and the calculated values of this work \( E_{\text{calc}} \) using our refined PES and the EBSC. The \( 2\nu_2, \nu_2 + \nu_4, 2\nu_1 \) and \( \nu_3 \) bands (upper plot) were taken from \[267\]; the \( 2\nu^0_4 \) and \( 2\nu^2_4 \) bands (upper plot) were taken from \[268\]; the \( 2\nu^0_0 \) and \( 2\nu^2_4 \) bands (middle plot) were taken from \[297\]; the \( 3\nu^1_1, 3\nu^3_1, \nu_1 + 2\nu^3_2 \) and \( \nu_1 + 2\nu^0_3 \) bands (bottom plot) were taken from reference \[276\].

4.3 The partition functions \( Q_{\text{max}} \) of AsH\(_3\) at different temperatures versus the maximum \( J \) value used in Eq. (2.72).

4.4 Overview of complete \( J = 0 - 30 \) line list computed at 296 K.

4.5 Overview of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 0–7000 cm\(^{-1}\) region.

4.6 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 800–1150 cm\(^{-1}\) region.

4.7 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 935–965 cm\(^{-1}\) region.

4.8 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 2000–2250 cm\(^{-1}\) region.

4.9 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 2160–2190 cm\(^{-1}\) region.

4.10 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 2920–3260 cm\(^{-1}\) region.

4.11 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 3110–3140 cm\(^{-1}\) region.

4.12 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 4035–4285 cm\(^{-1}\) region.

4.13 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 4100–4130 cm\(^{-1}\) region.

4.14 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 5000–5300 cm\(^{-1}\) region.

4.15 Expansion of synthetic \( J = 0 - 30 \) spectrum computed at 298.15 K compared to PNNL for the 6000–6400 cm\(^{-1}\) region.
5.1 Ratio of $Q_{\text{CoYuTe}}/Q_0$ for temperatures $10 - 1600$ K, where $Q_{\text{CoYuTe}}$ is the partition function computed with the CoYuTe energies with $E_{\text{max}} = 11000$ cm$^{-1}$ and $J_{\text{max}} = 43$, and $Q_0$ is the high temperature partition functions of [238]. .......................................................... 177

5.2 MARVEL line list, computed using energies from the updated MARVEL database in conjunction with the C2018 line intensities. ....................... 178
List of Tables

3.1 Character table of $D_{3h}$ ......................................................... 74
3.2 Parametrisation of the primitive functions generated as solutions to 1D Schrödinger equations ......................................................... 78
3.3 Convergence of the vibrational energy levels (in units of cm$^{-1}$) with increasing polyad $P_{\text{max}} = 28, ..., 40$ computed using the PES by Polyansky et al. [199], compared to the MARVEL experimentally derived values [3] ................................................................. 79
3.4 Structural constants of our PES compared to previous theoretical calculations and experiment ................................................................. 84
3.5 Accuracy of calculated rotational term values up to $J = 30$ when compared to the empirical MARVEL values [3]. $\sigma_{\text{rms}}$ refers to the root-mean-square deviation and $\Delta$ refers to the $E_{\text{MARV}} - E_{\text{calc}}$ wavenumber differences of $K = J$ and $K = 0$ states. Units of $\Delta$ and $\sigma_{\text{rms}}$ are cm$^{-1}$ ................................................................. 85
3.6 Root-mean-square deviation statistics for the complete list of paired MARVEL–C2018 levels under 7555 cm$^{-1}$. Band centres and RMS statistics are in units of cm$^{-1}$. $N_{\text{states}}$ refers to the number of paired states included in the comparison. The value before the / refers to all MARVEL states, and the value after the / refers to those states derived from 3 or more transitions. Vibrational labels are taken from MARVEL ................................................................. 89
3.7 Vibrational band centre labelling comparisons between different data sources for 6300–7000 cm$^{-1}$ region ................................................................. 93
3.8 Comparison of the no-BODC C2018 vibrational term values (cm$^{-1}$) computed using TROVE and GENIUSH ................................................................. 95
List of Tables

3.9 Comparison of calculated intensities to the experimentally derived values by Vander Auwera and Vanfleteren [273] above 7000 cm\(^{-1}\). Line positions \(v_{\text{obs}}\) and their obs.–calc. differences are given in cm\(^{-1}\), and intensities are given in cm\(^{-1}\)/molecule cm\(^{-2}\). Assignments are those of this work, detailed in section 3.5. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 113

3.10 Examples of the GSCD process for 6 different derived upper states. Upper state term values \(E'_{\text{obs}}\) and \(E'_{\text{calc}}\), observed line positions \(v_{\text{obs}}\), and the difference between the observed and calculated line positions \(v_{o-c}\), are all given in units of cm\(^{-1}\). Units of intensity are cm\(^{-1}\)/molecule cm\(^{-2}\). \(\langle E'_{\text{obs}} \rangle\) is the averaged experimental term value. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116

3.11 An overview of the assignments from this work. Only the \(E\)-symmetry band origin for the \(v_2+v_3+2v_4\) band is shown, although our assignments include \(A_1\) and \(A_2\)-symmetry vibrational states as well. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117

3.12 Measured H\(_2\)O transition wavenumbers (in units of cm\(^{-1}\)) compared to the values given in HITRAN. \(\Delta\) refers to the difference (in units of cm\(^{-1}\)) between the two peak positions that were averaged to derive each transition wavenumber. \(s^2/\sigma^2\) is the estimated signal-to-noise ratio, where \(s\) is the peak amplitude and \(\sigma\) is the standard deviation in the no absorbing region. . 131

3.13 Measured NH\(_3\) transition wavenumbers (in units of cm\(^{-1}\)). \(\Delta\) refers to the difference (in units of cm\(^{-1}\)) between the two peak positions that were averaged to derive each transition wavenumber. \(s^2/\sigma^2\) is the estimated signal-to-noise ratio, where \(s\) is the peak amplitude and \(\sigma\) is the standard deviation in the no absorbing region. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 131

3.14 Measured NH\(_3\) transition wavenumbers (cm\(^{-1}\)), derived from the \(> 99\%\) NH\(_3\) measurements \(v_{\text{meas}}\), and averaged over both the 3% and \(> 99\%\) concentration scans \(\langle v_{\text{meas}} \rangle\). Also shown are peak standard deviations \(\sigma_{\text{std}}\), total number of scans \(N_{\text{scan}}\), and averaged relative signal amplitudes \(\langle s \rangle\) measured at \(> 99\%\) NH\(_3\) concentration . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
List of Tables

3.15 Measured H$_2$O transition wavenumbers (in units of cm$^{-1}$) compared to the values given in HITRAN. $\Delta$ refers to the difference (in units of cm$^{-1}$) between the two peak positions that were averaged to derive each transition wavenumber. $s^2/\sigma^2$ is the estimated signal-to-noise ratio, where $s$ is the peak amplitude and $\sigma$ is the standard deviation in the no absorbing region.

4.1 Equilibrium energies (in units of Hartree) calculated at the CCSD(T) level of theory using different basis sets and Hamiltonians.

4.2 Differences between experimentally derived band centres and our calculated values computed using all-electron DKH and pseudopotential-F12 based PESs. All numerical values are term values given in units of cm$^{-1}$.

4.3 Experimental and predicted structural constants of $^{75}$AsH$_3$.

4.4 Differences between calculated rotational term values, in cm$^{-1}$, and the hyperfine resolved values of [253] which we averaged using the spin statistical weights.

4.5 Agreement between our calculated energy levels and those derived from experiment. All calculations used our refined PES, AsH$_3$-CYT18. $J = 0$ comparisons are before employing the EBSC, and $J = 1 \sim 6$ comparisons are afterwards. Term values and their RMS statistics are given in cm$^{-1}$.

4.6 Comparison of observed and calculated band intensities. Column 1 refers to the local mode quantum numbers assigned by TROVE, where sym is the total symmetry. The units of intensity are $10^{-18} \text{ cm}^{-1}/(\text{molec cm}^{-2})$. The value under $/ \sim$ is the total intensity of the bands with the same quantum numbers $n_1n_2n_3$.

4.7 Comparison of calculated and observed [62] line positions (cm$^{-1}$) and intensities ($\text{cm}^{-1}/(\text{molecules cm}^{-2})$) belonging to the $\nu_1$ and $\nu_3$ bands.
Chapter 1

Introduction

Rotational vibrational spectroscopy is the study of how light interacts with the internal dynamics of a molecule. These interactions manifest through the emission or absorption of light, which results in a change of rotational and vibrational behaviour – the rovibrational state – of the molecule. These rovibrational states are quantized yet numerous, and unique to each molecular species, as are the allowed transitions between them. Thus, by observing the discrete frequencies of light absorbed or emitted by molecular gas, we might discern its constituents.

This phenomena has made rovibrational spectroscopy a popular technique for industries where in situ gas analysis is desirable. Moreover, as the consequences of climate change become more apparent, spectroscopy presents a key tool with which to monitor greenhouse gas emissions. Turning our attention away from earth, to the field of astronomy where the number of confirmed exoplanets is set to soon pass 4000, focus is being turned to their characterisation. This recent drive to find and characterise exoplanets mean a number of space and land based observatories, e.g., JWST, ARIAL, SPHERE@VLT, GPI@GEMINI and EPICS@ELT etc. have been built to measure molecular spectra over a wide wavelength range. These missions will generate a huge amount of data that can only be interpreted using spectroscopy.

The production of spectroscopic line lists - that is, lists of transition wavenumbers and intensities - for application to industry, environmental monitoring, and astronomy, is therefore an important activity. In order to be universally applicable, these line lists should be complete over the vast temperature ranges observed
in space, and contain accurate line-by-line information. Ideally all such line lists would be constructed from line positions and intensities derived purely from experimental sources, which can typically achieve levels of accuracy that is orders of magnitude better than the predictions of \textit{ab initio} calculations. However, achieving this through experiment alone is extremely difficult because of i) the sheer volume of data necessary to accurately model opacities of exoplanet atmospheres, for which billions of lines may be required, ii) the need for lower state energy assignments, which are necessary to determine the correct temperature dependence, iii) the difficulty for experimentalists in obtaining absolute line strengths, iv) the general need for completeness, which requires coverage over large wavelengths, and even at room temperature laboratory data is far from complete, v) the need for quantum assignments, which provide the information necessary to accurately model line broadening. All of these problems are exacerbated at high temperatures where the distribution of molecular state populations, according to Maxwell-Boltzmann statistics, results in molecular spectra becoming extremely rich and complex.

1.0.1 Spectroscopic databases

A number of spectroscopic databases exist that aggregate lists of spectroscopic parameters derived from experimental measurements, for astronomical and terrestrial applications alike. The most notable of these are HITRAN [91], GEISA [121], JPL [191], CDMS [78] and PNNL [232].

The HITRAN (High Resolution Transmission) database is arguably the most widely accessed database of spectroscopic parameters, including line positions, intensities, quantum state assignments, broadening parameters, partition functions etc. As of the 2016 edition it contained line-by-line data for 49 molecules and their isotopologues, out of which a significant proportion were partially or fully assigned quantum numbers. However, it is not without problems. For example, the 2012 edition was found to display labelling inconsistencies for 2529 $^{14}$NH$_3$ lines, which in some cases resulted in the breaking of ortho-para selection rules [71]. Although this was later corrected by Down \textit{et. al.} [71], it highlights the danger of becoming overly reliant on the HITRAN database, even at room temperature. GEISA is an al-
ternative source of line-by-line information, which is designed to facilitate radiative transfer calculations and contains some different molecules to HITRAN.

The Jet Propulsion Laboratory (JPL) spectral line catalogue aims to compile and analyse transitions from the literature, using in-house effective-Hamiltonian fitting software, and with that make accurate line predictions with full quantum assignments. The Cologne Database for Molecular Spectroscopy (CDMS) was designed to complement JPL, and uses the same procedure with a focus on molecular species which may be observed in astronomical spectra.

The Pacific Northwest National Laboratory (PNNL) infrared database is a library of low-resolution composite spectra measured for a large number of molecules and compounds, which extends at least over the wavenumber range 600 – 6500 cm$^{-1}$ for each molecule. Although it contains several times the number of entries that HITRAN does, it contains no line-by-line data, and spectra are recorded at three temperatures as standard. Nevertheless, it is an extremely valuable source of information when HITRAN is insufficient.

All of the aforementioned databases are aimed at modelling cool spectra, only HITEMP [216], which uses a combination of first principles calculations and room temperature line lists from HITRAN, has the capability to model hot objects, and so far only for 5 species. Therefore, to meet the demand from astronomers for complete, high-temperature line lists with full quantum state assignments, the ExoMol project [258] was formed.

1.0.2 ExoMol
The aim of the ExoMol project is to compute hot infra-red molecular line lists for molecules of key astronomical importance. The project has already successively computed a large number of ‘hot’ line lists for molecules such as CH$_4$ [301,310], NH$_3$ [302], PH$_3$ [237], CH$_2$O [6], SO$_2$ [270], H$_2$O [15], H$_2$S [12], H$_2$O$_2$ [4] and many others. These line lists can be extremely large, over 10 billion lines in some cases, all of which are fully assigned upper and lower state quantum numbers and energies. For this purpose, a suite of computer programs have been developed [259, 260]; these are Duo [309] and Level [217] for modelling diatomics, DVR3D [257]
for triatomics, and TROVE \[313\] for polyatomics of four or more atoms, which includes a recent extension to treat linear molecules \[47\].

Figure 1.1: Flow chart of the procedure typically used to generate an ExoMol line list, taken from Ref. \[260\].

Although each program has been specifically developed to treat different molecular systems, they all solve the nuclear Schrödinger equation variationally, and all share the same general approach to the construction of molecular line lists, which is outlined in Fig. 1.1. The process begins with the construction of a potential energy surface (PES) and dipole moment surface (DMS) using \textit{ab initio} electronic structure calculations to generate electronic energies, which are fit to suitable analytic expressions. Rovibrational energies and wavefunctions are then computed and used as a basis with which to empirically refine the PES. This is in contrast with the DMS, where there is evidence that \textit{ab initio} surfaces are more accurate than those fitted to experimental data \[156\]. Using the refined PES, nuclear motion calculations are performed to compute rovibrational energies and wavefunctions up to a level of (energetic and rotational) excitation such as to ensure completeness up to a specified temperature. Finally, the wavefunctions are used in conjunction with the DMS to compute lists of Einstein-A coefficients, which can be used to calculate
intensities or cross-sections using the in-house program ExoCross [300].

The coverage provided by the ExoMol line lists means that they have been widely adopted by astronomers (e.g. [9, 42, 63, 85, 149, 192–194, 218, 274]), and in certain cases the ExoMol line lists have even been used to demonstrate systematic deficiencies in recent high-level experiments [316]. The general approach employed by the ExoMol project, therefore, is well tested, and in this thesis we apply it to the molecules ammonia (NH₃) and arsine (AsH₃).

1.0.3 Thesis overview

This thesis aims to improve the general spectroscopic situation regarding the molecules ammonia (¹⁴NH₃) and arsine (⁷⁵AsH₃) in the infrared and visible wavelength regions, with specific applications to astronomical spectra, high-resolution spectroscopic studies, and industrial monitoring. I do this using a combination of theoretical, empirical and experimental techniques.

Chapter 2 reviews the important theoretical concepts and methods pertaining to the construction of molecular line lists from first principles. This includes an overview of the electronic structure methods relevant to the work reported here. That is, the methods with which we solve the electronic Schrödinger equation in order to construct a potential energy surface for the molecule in question. Secondly our chosen method for solving the nuclear motion problem is described, as implemented in the program suite TROVE. This includes a brief overview of the kinetic energy operator, coordinate system and rotational-vibrational basis. A theoretical discussion of the computation of line intensities and the generation of synthetic spectra is then given. Finally, a brief theoretical discussion of the experimental method employed in this work is given.

Chapter 3 has predominantly been adapted from the work presented in Ref. [49]. It details the key spectroscopic advances that have been achieved with regards to ammonia. Foremost, the development of a new high-accuracy potential energy surface capable of assisting the critical evaluation of the current set of experimentally derived energy levels. This is performed by refinement of a new high-accuracy ab initio surface with experimental data, performed using TROVE. With this re-
fined surface, in conjunction with a well-established dipole moment surface, room
temperature line list calculations are performed using TROVE for wavelengths ex-
tending into the visible. The line list is seen to be substantially more accurate than
the previous room temperature line list developed as part of the ExoMol project,
and is subsequently used in the assignment of spectra in the near-infra-red. Finally,
experimental measurements of a number of weak lines in a spectral region of rele-
vance to industry are performed. Several line positions are derived, with the aim of
assisting the development of new ammonia gas analysers.

Chapter 4 has predominantly been adapted from the work presented in Ref.
[48]. It details the production of a new theoretical line list for arsine from first-
principles. The current state of literature is reviewed and the currently available
theoretical and experimental data sets are found to be particularly lacking line in-
tensities. Two high-level electronic structure calculation methods are compared,
and the method discerned to be superior is used to generate a full-dimensional po-
tential energy surface and dipole moment surface for the electronic ground state.
The potential energy surface is refined to experimental data, and line list calcula-
tions are performed for room temperature using the refined potential energy surface
and an \textit{ab initio} dipole moment surface in conjunction with TROVE. The line list
displays excellent agreement with cross-sections provided in the PNNL database.
Chapter 2

Theoretical background

The aim of this section is to provide an overview of the general approach used to construct molecular line lists that was exploited in this thesis, and introduce the key elements of electronic structure and nuclear motion theory that are relevant to the construction of NH$_3$ and AsH$_3$ line lists from first principles.

In the construction of molecular line lists we are concerned with solving the rovibronic (rotational-vibrational-electronic) Schrödinger equation

$$\hat{H}_{rve} \Phi_{rve} = E_{rve} \Phi_{rve},$$

(2.1)

where the general spin-free rovibronic molecular Hamiltonian $\hat{H}_{rve}$ (also known as the Coulomb Hamiltonian) for $n$ electrons of mass $m_e$ and $N$ nuclei of mass $m_N$, in the absence of external electric or magnetic fields, is given by

$$\hat{H}_{rve} = -\frac{\hbar^2}{2m_e} \sum_{i=1}^{n} \nabla_i^2 - \frac{\hbar^2}{2M_N} \sum_{A=1}^{N} \nabla_A^2 - \frac{e^2}{4\pi\varepsilon_0} \sum_{i=1}^{n} \sum_{A=1}^{N} \frac{Z_A}{|r_A - r_i|}$$

$$+ \frac{e^2}{4\pi\varepsilon_0} \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{1}{|r_i - r_j|} + \frac{e^2}{4\pi\varepsilon_0} \sum_{A=1}^{N} \sum_{B>A}^{N} \frac{Z_A Z_B}{|R_A - R_B|}.$$

(2.2)

In truth, a complete description of the molecular Hamiltonian requires additional treatment of spin-spin, spin-orbit and hyperfine effects. Omission of these contributions is justified by the knowledge that i) energy level splittings due to hyperfine interactions are of order $\sim MHz$ [50], which are only observable at sub-doppler temperatures [266]; ii) spin-spin and spin-orbit couplings are not important for the
closed shell species considered here.

Equation 2.2 represents a $3(n + N) - 3$ dimensional problem, and it is impossible to solve analytically for any system larger than two particles ($n + N = 2$). To progress any further the Born–Oppenheimer (BO) approximation [31], a concept that is arguably the most important in all of molecular physics, must be invoked.

The Born-Oppenheimer approximation postulates that the nuclei remain stationary over time scales comparable to that of the electron motion, and therefore the electronic contribution to $\Phi_{\text{rve}}(r, R)$ is unaffected by the nuclear kinetic energy term in Eq. (2.2). In this paradigm the rovibronic wavefunction is separated into the product of electronic $\Phi_{\text{elec}}(r; R)$ and nuclear $\Phi_{\text{nucl}}(R)$ components, given by

$$\Phi_{\text{rve}}(r, R) = \Phi_{\text{elec}}(r; R)\Phi_{\text{nucl}}(R) \quad (2.3)$$

where $\Phi_{\text{elec}}(r; R)$ depends explicitly on the electron positions and parametrically on nuclei positions, and $\Phi_{\text{nucl}}(R)$ depends only on the nuclei positions. The procedure for obtaining $\Phi_{\text{elec}}(r; R)$ and $\Phi_{\text{nucl}}(R)$, as first outlined by Born and Oppenheimer [31], is as follows: first, the nuclei are clamped stationary in a particular configuration, and the wavefunction and energy of a specific electronic state (typically the lowest-lying) are obtained by solving the “clamped–nuclei” problem

$$\hat{H}_{\text{elec}}(r; R)\Phi_{\text{elec}}(r; R) = E_{\text{elec}}(R)\Phi_{\text{elec}}(r; R), \quad (2.4)$$

with corresponding Hamiltonian

$$\hat{H}_{\text{elec}} = -\frac{\hbar^2}{2m_e} \sum_{i=1}^{n} \nabla_i^2 - \frac{e^2}{4\pi\varepsilon_0} \sum_{i=1}^{n} \sum_{A=1}^{N} \frac{Z_A}{|R_A - r_i|} + \frac{e^2}{4\pi\varepsilon_0} \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{1}{|r_i - r_j|} + \frac{e^2}{4\pi\varepsilon_0} \sum_{A=1}^{N} \sum_{B>A}^{N} \frac{Z_A Z_B}{|R_A - R_B|}. \quad (2.5)$$

The electronic energies $E_{\text{elec}}(R)$ are nuclear geometry dependent, and act to raise or lower the total internal energy of the molecule depending on the nuclear configuration. Therefore the continuous spectrum of $E_{\text{elec}}(R)$ plays the role of the potential...
energy surface (PES) in which the nuclei move $E_{\text{elec}}(\mathbf{R}) = V(\mathbf{R})$ in the second, nuclear motion, Schrödinger equation

$$\hat{H}_{\text{rv}}(\mathbf{R})\Phi_{\text{rv}}(\mathbf{R}) = E_{\text{rv}}(\mathbf{R})\Phi_{\text{rv}}(\mathbf{R}),$$  \hspace{1cm} (2.6)

where the rovibrational Hamiltonian is written as

$$\hat{H}_{\text{rv}} = -\frac{\hbar^2}{2M_N} \sum_{A=1}^{N} \nabla_A^2 + V(\mathbf{R}).$$  \hspace{1cm} (2.7)

Solving Eqs. (2.4) and (2.6) requires making additional approximations, and will be discussed separately in Sections 2.1 and 2.2.

Some 27 years after Born’s original publication, Born and Huang published a different approach [30], which fixes some of the less satisfactory artefacts of the BO approximation. In the BH approach it is assumed that the clamped nuclei Hamiltonian has been solved for all possible nuclear geometries, and at each geometry there are infinitely many solutions $\Phi_{\text{elec}}^{j}(\mathbf{r};\mathbf{R})$ that are orthonormal and form a complete set. This completeness means that the exact rovibronic wavefunction at each geometry can be written as a linear expansion in terms of $\Phi_{\text{elec}}^{j}(\mathbf{r};\mathbf{R})$ with the geometry dependent coefficients $\chi^{i}(\mathbf{R})$, that is

$$\Phi_{\text{rvb}} = \sum_{i=1}^{\infty} \chi^{i}(\mathbf{R})\Phi_{\text{elec}}^{j}(\mathbf{r};\mathbf{R}),$$  \hspace{1cm} (2.8)

where the coefficients $\chi^{i}(\mathbf{R})$ play the role of the nuclear wavefunctions, and the summation is over all electronic states. The full rovibronic Hamiltonian is then applied to Eq. (2.8), multiplied on the left by $(\Phi_{\text{elec}}^{j})^*$ and integrated over all electronic coordinates to yield the following set of coupled differential equations for the coefficients $\chi^{i}(\mathbf{R})$

$$[\hat{T}_N + E_{\text{elec}}^{i}(\mathbf{R})] \chi^{i}(\mathbf{R}) + \sum_{j=1}^{\infty} \hat{A}^{ij}(\mathbf{R})\chi^{j}(\mathbf{R}) = E\chi^{i}(\mathbf{R}),$$  \hspace{1cm} (2.9)

where we have converted to atomic units in which the electron mass ($m_e$) = Planck’s
constant \( \hbar \) = the electron charge \( e \) = 1. \( E \) is the energy of a specific rovibronic state, \( \hat{T}_N \) is the nuclear kinetic energy operator, \( E_{\text{elec}}^i(\mathbf{R}) \) is the electronic potential energy surface for the \( i^{\text{th}} \) electronic state, and the vibronic coupling matrix elements \( \hat{\Lambda}^{ij}(\mathbf{R}) \) are defined by

\[
\hat{\Lambda}^{ij}(\mathbf{R}) = \sum_{\alpha=1}^{N-1} \left( \langle \Phi_{\text{elec}}^j(\mathbf{r}; \mathbf{R}) | \hat{T}_N(\mathbf{R}) | \Phi_{\text{elec}}^i(\mathbf{r}; \mathbf{R}) \rangle - \frac{1}{M_\alpha} \langle \Phi_{\text{elec}}^j(\mathbf{r}; \mathbf{R}) | \vec{v}_\mathbf{R} | \Phi_{\text{elec}}^i(\mathbf{r}; \mathbf{R}) \rangle \cdot \vec{v}_\mathbf{R} \right),
\]

where the sum is over \( N-1 \) nuclei rather than \( N \), as we are in the frame of reference of the molecular centre of mass. The above set of equations can be thought of as the full solution to the rovibronic Schrödinger equation, and the BO approximation is recovered upon neglecting \( \hat{\Lambda}^{ij}(\mathbf{R}) \). If only the diagonal terms \( \hat{\Lambda}^{ii}(\mathbf{R}) \) are kept, then we call this the adiabatic approximation. It corresponds to the case in which the nuclei move in a single electronic potential, and the corresponding correction to the PES is known as the Born-Oppenheimer Diagonal Correction (BODC), which is mass dependent and so must be separately computed for each molecular isotopomer. Note that the second term in Eq. (2.10) does not contribute to the BODC.

The case in which all off-diagonal elements \( \hat{\Lambda}^{ij}(\mathbf{R}) \) are also kept is known as the nonadiabatic case. It allows for the nuclei to tunnel between electronic surfaces, and it can be shown that its contribution to \( \hat{\Lambda}^{ij}(\mathbf{R}) \) is inversely proportional to the energy difference between electronic states \( i \) and \( j \). Thus, it becomes important when electronic surfaces lie close to one another. Computation of the nonadiabatic correction is extremely expensive because it requires the calculation of potential energy surfaces for multiple electronically excited states. Alternative methods have been developed to implicitly treat nonadiabatic effects by scaling terms in the nuclear kinetic energy operator [36, 226], which have worked well for certain diatomics, water [226] and ammonia [112]. In this thesis, all work has been carried out in the Born-Oppenheimer approximation.
2.1 Solving the electronic Schrödinger equation

According to the BO approximation, our starting point for generating rovibrational wavefunctions must be a suitable potential energy surface, which is produced by solving the clamped nuclei Schrödinger equation given in Eqs. (2.4) and (2.5) at a series of nuclear geometries. The resulting electronic energies are then fit to an analytic expression to provide a continuous description of the potential the nuclei move in. Conventionally it is this analytic expression, rather than the computed electronic energies, that is known as the PES. For the applications in this thesis we are solely concerned with the energetic values, however it should be noted that the electronic wavefunctions can be used to calculate important molecular properties such as dipole moments, multipole moments and polarizability.

Equation (2.4) is a $3n$ dimensional problem, which is impossible to solve analytically for $n > 2$. A large number of methods have been developed to approximately solve the clamped nuclei Schrödinger equation over the years, with no one method emerging as superior for every application. For the most accurate results, one must make a suitable choice of method based not only on the physics of the molecule, but also the particular physical and chemical properties under investigation, all bearing in mind the associated computational cost. Some of the more widely used methods are density functional theory (DFT), coupled-cluster (CC), Møller-Plesset perturbation theory (MP2), configuration interaction (CI), and multi-reference configuration interaction (MRCI). Only CC and CI are discussed in the following sections, for a review of the DFT methodology the reader is directed to [126], for MP2 see [138], for MRCI see [233].

No matter which method is chosen, with the exception of DFT all rely on the same initial procedure as a starting point. Namely, the $3n$—body problem is reduced to $n$ coupled equations in three-dimensions which are solved numerically via a self-consistent field (SCF) approach. This procedure was first proposed by Hartree [95], and later reformulated by Fock [83]. Although the resulting Hartree-Fock wavefunctions are not accurate enough for high-resolution spectroscopic applications, they usually give a good approximation of the total electronic energy.
### 2.1.1 Hartree-Fock

As an initial approximation, the electronic wavefunction is separated into the sum of products of individual electron spin-orbitals, as defined by the \( n \)-electron Slater determinant.

\[
\Phi^0_{\text{elec}}(r_1, \sigma_1, r_2, \sigma_2, \ldots) = \frac{1}{\sqrt{n!}} \left| \begin{array}{cccc}
\phi_1(r_1)\alpha(\sigma_1) & \phi_2(r_1)\alpha(\sigma_1) & \cdots & \phi_m(r_1)\beta(\sigma_1) \\
\phi_1(r_2)\alpha(\sigma_2) & \phi_2(r_2)\alpha(\sigma_2) & \cdots & \phi_m(r_2)\beta(\sigma_2) \\
\vdots & \vdots & \ddots & \vdots \\
\phi_1(r_n)\alpha(\sigma_n) & \phi_2(r_n)\alpha(\sigma_n) & \cdots & \phi_m(r_n)\beta(\sigma_n)
\end{array} \right|
\]

(2.11)

This form is necessary to fulfill the fundamental property of wavefunctions to change symmetry upon interchange of two fermions. The spin–orbitals are each comprised of a spin component \( \alpha(\sigma) \) or \( \beta(\sigma) \), which has \( z \)-axis projected eigenvalues of \( \pm \frac{1}{2} \hbar \) for an electron, and a spatial component \( \phi_k(r_i) \), which corresponds to the \( i^{th} \) electron, located at position \( r_i \), in orbital \( k \).

In accordance with the variational principle, we seek the set of spin–orbitals that minimise the Rayleigh ratio. Applying this condition to \( \Phi^0_{\text{elec}} \) results in the Hartree-Fock equations for the spin-orbitals. Although not shown here, there are numerous textbooks that perform step-by-step derivations of the HF equations, two examples being [11,206]. The Fock equation for an individual spin orbital involves a one electron term and a two-electron term. The one-electron term is the one-electron KE plus the electron-nucleus repulsion. The two-electron term represents the Coulomb repulsion between the electron and the average field of the \( n - 1 \) other electrons (hence why HF is referred to as a ‘mean field’ theory), plus a purely non-classical exchange term.

Because the Fock equation for a spin-orbital depends on the spin-orbitals of the \( n - 1 \) other electrons, the Hartree-Fock equations must be solved iteratively. This is the self-consistent field (SCF) approach, whereby a trial set of spin-orbitals are initially used to construct the Fock operator and solve the HF equations, the resulting spin-orbital solutions are then used to construct a revised Fock operator and the process is repeated until there is no change in spin-orbitals from one iteration to the
A modification to the Hartree-Fock equations to allow for the treatment of molecules was presented by Roothaan and Hall in 1951 [93][213]. The spatial components of the molecular spin-orbitals $\phi_k(r)$ are written as linear combinations of atomic orbitals $\chi_{\mu}(r)$

$$
\phi_k(r) = \sum_{\mu} c_{\mu k} \chi_{\mu}(r),
$$

(2.12)

which transforms the problem into a standard matrix eigenvalue equation which, given a set of fixed basis functions $\chi_{\mu}$ with initial coefficients $c_{\mu k}$, must again be solved iteratively for the optimal set of coefficients $c_{\mu k}$. Common choices of basis functions $\chi_{\mu}$ are Gaussian-type orbitals and Slater-type orbitals, which are both discussed in Section 2.1.4. Of course the accuracy of the SCF solution can be improved by increasing the number of basis functions employed. In the limit an infinitely large basis set, one is said to have reached the Hartree-Fock limit.

The major problem with HF is that electron-electron interactions are treated in an average way. In general, the electrons will be further apart than described by the HF solutions, and so the HF energy is the upper bound to the exact energy. The difference between the exact electronic energy and the HF energy is called the electron correlation energy, and accounts for, in a very broad sense, the individual electron-electron interactions. As mentioned previously there are a huge number of methods aimed at recovering electron correlation effects, some of which will be discussed in the following sections.

### 2.1.2 Configuration Interaction

The configuration-interaction (CI) is the most conceptually simple extension of Hartree-Fock theory aimed at recovering electron correlation, and is covered in most quantum chemistry textbooks. In particular Ref. [123] provides a thorough exposition accessible to the non-expert.

According to the Roothaan-Hall equations for molecular orbitals (MOs), a basis of $N_b$ atomic orbitals will produce $N_b$ molecular orbitals. For a system of $N$ electrons, the Slater determinant consisting of the $N/2$ (allowing for spin) lowest...
energy MOs is the HF energy. These $N/2$ lowest energy orbitals are known as *occupied orbitals* and the remaining $N_b - (N/2)$ orbitals are known as *virtual orbitals* (or *unoccupied orbitals*). A much better approximation to the ground state wavefunction can be formulated by adding to the HF wavefunction a linear combination of Slater determinants where a number of occupied orbitals have been replaced by virtual orbitals. A CI calculation considering only single electron excitations is referred to by the acronym CIS, for one and two simultaneous electron excitations it is CISD, for one to three simultaneous excitations CISDT etc. The improved ground state energy can then be found by diagonalising the electronic Hamiltonian matrix in this basis. It should be noted that due to orthogonality, CIS presents no improvement over a standard HF calculation.

Two truncations limit the accuracy of a CI calculation. The first is the number of atomic orbitals (AOs) used to construct the MOs. The second is the number of Slater determinants used to account for (static) electron correlation. In the limit of infinitely many AOs and infinitely many Slater determinants, the resulting wavefunction is the true ground state wavefunction of the system. However, from a computational standpoint the scaling of CI is very poor. With an AO basis set of size $M_{\text{basis}}$, CISD scales as $M_{\text{basis}}^6$, CISDT scales as $M_{\text{basis}}^8$ and CISDTQ as $M_{\text{basis}}^{10}$ [123]. For medium sized molecules, CISD typically recovers 80-90% of the correlation energy.

A major problem with truncated CI is that it is not size consistent nor size extensive. The former refers to the correct behaviour of the energy as the system is gradually stretched to dissociation. The latter refers to the correct (linear) scaling of a method with number of electrons [16], and implies that errors do not increase as more electrons are added to the calculation. Modifications to CI were developed to tackle this lack of size extensivity, but were later shown to be simply variations on the coupled cluster method [58].

### 2.1.3 Coupled Cluster

The use of coupled cluster (CC) is nowadays generally preferred over CI due to its guaranteed size extensivity. The basics of CC are covered in most quantum
chemistry textbooks at some level (e.g., [123]), for a detailed review the reader is
directed to [17]. As with CI, Slater determinants for electronically excited states are
introduced into the ground state wavefunction via an excitation operator

\[ T = T_1 + T_2 + T_3 + \ldots + T_{N_{\text{elec}}} \] (2.13)

where the \( T_i \) operator acting on the HF wavefunction \( \Phi^0 \) generates a summation of all \( i^{\text{th}} \) excited Slater determinants

\[ T_1 \Phi^0 = \sum_{i} \sum_{a} t_i^a \Phi^a_i \] (2.14)

\[ T_2 \Phi^0 = \sum_{i<j} \sum_{ab} t_{ij}^{ab} \Phi_{ij}^{ab} \] (2.15)

and \( t_i^a \) and \( t_{ij}^{ab} \) are referred to as the Slater determinant amplitudes. Conceptually these differ from CI coefficients in that they represent the strength of an excitation process itself, rather than the weight of an excited state determinant. The CC wavefunction is defined as

\[ \Psi_{\text{CC}} = \exp(T \Phi^0) = (1 + \frac{1}{2} T^2 + \frac{1}{6} T^3 + \ldots) \Phi^0 \] (2.16)

which, if all terms up to \( T_{N_{\text{elec}}} \) are included, is equivalent to full CI. Practically, the excitation operator must be truncated at some point, with the most common representation of \( T \) for medium sized systems being \( T = T_1 + T_2 \). The corresponding method is thus referred to as CCSD and scales as \( M_{\text{basis}}^6 \) [123].

The key feature of CC is that excited determinants beyond the truncation order of \( T \) appear in the wavefunction. For example, in the case of \( T = T_1 + T_2 \) there will be contributions from triply, quadruply etc. excited determinants (due to products of \( T_1 \) and \( T_2 \)), with amplitudes which are products of \( t_i^a \) and \( t_{ij}^{ab} \). These excitations are formally linked to the many-body perturbation theory (MBPT) diagrammatic representation of CC (see Ref. [17]), and are the reason CC is size extensive.

Going beyond CCSD becomes computationally infeasible for all but the small-
Chapter 2. Theoretical background

Est systems as CCSDT scales as $M_{\text{basis}}^8$, which is more computationally demanding than CISDT [123]. Various attempts have been made to approximate, and incorporate into CCSD, the unaccounted for triple excitations’ contribution (i.e. from $T_3$) to the wavefunction. By far the most successful of these has been the CCSD(T) method [204, 230], which adds a triples contribution calculated using 4th and 5th order Møller-Plesset (MP) perturbation theory in conjunction with the CCSD amplitudes [204, 278], to the CCSD results (for a perspective see ref. [243]). For most molecular systems CCSD(T) recovers the CCSDT energies to within a few hundred $\mu E_h$, with a much improved scaling of $M_{\text{basis}}^7$ [230]. This makes CCSD(T) the method of choice for most quantum chemistry calculations. However, it should be noted that for systems where the HF wavefunction is not a good approximation to the true wavefunction, such as during dissociation, CC becomes extremely slow to converge and methods such as MRCI are preferable.

2.1.4 One electron basis sets

So far we have not addressed the nature of the spatial component of the one-electron spin-orbitals $\chi(r)$ (also known as atomic orbitals, although they are not solutions to an atomic Schrödinger equation) used to construct the Fock matrix. These are the fundamental building blocks of the molecular wavefunction, and so a substantial amount of work has gone into developing them as basis sets over the years. Each basis set is specifically designed to model a particular physical property and/or complement a post-HF method, therefore no one is optimal in all circumstances, and indeed an incorrect choice of basis set can yield nonsensical results.

Two types of basis functions are commonly used in quantum chemistry, these are Slater-type orbitals [235] (STOs) and Gaussian-type orbitals [33] (GTOs). Although STOs display the correct long and short range behaviour whereas GTOs do not, GTOs are used almost exclusively due to the ease at which integrals can be computed. For an excellent review of Gaussian basis sets see [104], for a broader overview of basis sets in quantum chemistry see [170].
In cartesian coordinates GTOs take the following functional form

\[ G_{\zeta, \ell_x, \ell_y, \ell_z}(x, y, z) = N x^{\ell_x} y^{\ell_y} z^{\ell_z} e^{-\zeta r^2} \]  

(2.17)

where \( \ell_x + \ell_y + \ell_z \) determines the type of orbital (e.g., \( \ell_x + \ell_y + \ell_z = 1, 2, 3 \) for s,p,d- orbitals), \( N \) is a normalisation constant and \( \zeta \) is a parameter. A single gaussian function is known as a *primitive*. So that a closer resemblance to Slater-type functions might be obtained, linear combinations of *primitives* which are centred on the same atom are taken to form *contracted* functions

\[ \chi_k = \sum_{\ell_x, \ell_y, \ell_z} d_{k, \ell_x, \ell_y, \ell_z}^k G_{\zeta, \ell_x, \ell_y, \ell_z}(x, y, z) \]  

(2.18)

A linear combination of contracted functions then forms the spatial component of the atomic spin-orbitals. The advantage of this contraction, over simply using the primitives, is that fewer expansion coefficients need to be calculated during the HF-SCF procedure, saving computer resources.

The primitive expansion coefficients and exponentials in Eqs. (2.17) and (2.18) are optimised based on the chemical properties under investigation. For general applications, including spectroscopy, this involves minimising the electronic energy. Approaches for optimising the exponents are discussed in ref. [190], and for determining the coefficients \( d_{k, \ell_x, \ell_y, \ell_z}^k \) the two main schemes are *segmented* contraction and *general* contraction, which are outlined in refs. [73] and [203], respectively.

The number of contracted functions used to construct each spin-orbital is reflected in the nomenclature single-zeta (SZ), double-zeta (DZ), triple-zeta (TZ) etc. for 1,2,3,... contracted functions. Typically this refers only to the valence electrons, as single-zeta functions is sufficient for a description of the core electrons. Additional polarizing functions of higher angular momentum (than the lowest occupied orbital) may also be added to account for molecular bonding and electron correlation effects. The presence of these, and the nature of the valence-only treatment, is denoted by the affixes ‘P’ and ‘V’, to form the acronym VnZP, where \( n \) is referred to as the *cardinal number*. Physical interpretations of each basis set component are
Many different families of contracted basis sets exist in the literature, for an overview of the most popular ones see the review by Jensen [122] and the references therein. For spectroscopic applications the correlation-consistent set by Dunning [74], Peterson and co-workers (denoted cc-pV\(n\)Z for correlation-consistent, polarised, valence-only \(n\)-zeta) are generally preferred because they are well-developed and provide a straightforward, systematic route for extrapolating the electronic energy to the complete basis set (CBS) limit [79] (for a complete bibliography of correlation consistent basis sets see [102]). Dunning [74] observed that when additional higher angular-momentum primitives were added to the spin-orbitals, their contributions to the CISD correlation energy fell into well-defined bands, with each polarising function within a band contributing roughly equally to the correlation energy. Therefore, polarising functions should be added in a specific sequence that systematically recovers the largest contributions to the correlation energy first. Moreover, in order for the error on the core energy to not exceed that of the polarising functions, the core basis is increased simultaneously.

Many extensions to the cc- scheme have been developed over the years, and were reviewed by Peterson in 2007 [184]. These include augmentations with tight functions [186, 290] to account for core electron correlation (denoted with the ‘C’ and ‘wC’ affixes), or diffuse functions [128] to better describe the wavefunction at long range (denoted with the ‘aug-’ prefix). Inclusion of relativistic effects at the Douglas-Kroll-Hess [100] level was seen to require modifications to the conventional cc- method [64], whereas for calculations of Mass-velocity and Darwin terms standard cc- basis sets suffice. To model heavy elements, pseudopotentials (PPs) are often used, which necessitate their own specifically developed basis sets. Some of the first basis sets used in PP calculations are those by [24]. A more detailed discussion of relativistic pseudopotentials is given in Section 2.1.6.

Finally, to complement the recent advancements in explicitly correlated (F12/R12) methods [132, 139, 263], a number of cc-F12/R12 optimised basis sets have been developed, largely by Peterson, Hill and co-workers [105, 185, 188].
though conventional cc- basis sets were found to work reasonably well for this purpose \cite{285}, they were designed partly to account for the Coulomb cusp at short inter-electron distances \cite{127}, which is explicitly accounted for in the F12/R12 ansatz. Therefore, F12/R12 optimised basis sets focus on recovering the Hartree-Fock and long-range correlation energies \cite{185}. Although the resulting cc-F12 basis sets are slightly larger than their standard cc- counterparts, here basis set convergence is far quicker.

Most elements have cc-pV\textsubscript{n}Z basis sets, and modifications thereof, available at the online repository \cite{103}. A more diverse selection from various families is offered at the basis-set-exchange repository \cite{224}.

### 2.1.5 Explicitly correlated coupled cluster

Expansion of the electronic wavefunction in terms of one-electron orbitals may produce results of sufficient accuracy for spectroscopic applications if suitably large basis sets are used. However, convergence to the CBS limit is slow \cite{79,131}, a feature that is attributed to the GTOs inability to produce the correct wavefunction cusp as two electrons coalesce \cite{127,264}. This behaviour can be accounted for by including inter-electron distance terms ($r_{12}$) in the wavefunction, a technique that was first employed by Hylleraas in 1929 \cite{118}. Unfortunately the penalty for this is the necessary calculation of three and four electron integrals which is computationally unfeasible, and until the seminal work by Kutzelnigg and co-workers \cite{132,139,263} restricted the use of inter-electron distance ($r_{12}$) terms to the wavefunctions of light atoms and very small molecules. Since then, the F12 ansatz has been applied to MP2, CCSD and MRCI \cite{234}, and there have been several variations on the F12 approximations (for reviews see \cite{97,135}).

The general ansatz for the CCSD-f12 method is

$$\psi_{\text{CCSD-F12}} = \exp(\hat{T}_1 + \hat{T}_2 + \hat{R}) |\Phi_0\rangle$$

(2.19)

where the cluster operators $\hat{T}_1$ and $\hat{T}_2$ are the standard CCSD operators. The geminal operator $\hat{R}$ takes the same form as $\hat{T}_2$ and accounts for excitations from occupied
orbits $i, j, ...$ into the formally complete set of virtual orbitals $\alpha, \beta, ...$ that are orthonormal to both the HF reference function and the CCSD excitations. The $F12$ configurations used to augment the $\Psi_{CCSD}$ wavefunction are represented as

$$|\Psi_{ij}^{mn}\rangle = F_{\alpha\beta}^{mn} |\Phi_0\rangle$$

(2.20)

$$F_{\alpha\beta}^{mn} = \langle mn | F12 \hat{Q}_{12} | \alpha\beta \rangle$$

(2.21)

where $\hat{Q}_{12}$ is the operator that ensures the $F12$ orbitals are orthogonal to the HF configuration and the singly and doubly excited CCSD configurations. Explicit dependence on the inter-electron distance $r_{12}$ is introduced through the correlation factor, which is simply the distance $r_{12}$ for R12 methods, or more generally some function $f(r_{12})$ in F12 methods. Slater type functions are most commonly used, and take the general form

$$f(r_{12}) = -\frac{1}{\gamma} \exp(-\gamma r_{12})$$

(2.22)

where $\gamma$ is a length parameter in the approximate range $\sim 1.0 - 2.0 \, a_0^{-1}$. Cusp conditions are used to determine the amplitudes $t_{ij}^{mn}$ of the F12 configurations $|\Psi_{ij}^{mn}\rangle$ [254, 255]. The resulting explicitly correlated terms drastically improve the description of the wavefunction at small $r_{12}$ by effectively negating the probability of finding any two electrons close to one another.

A complication arising from augmenting the CCSD wavefunction with F12 terms is the appearance of additional two, three and even four electron matrix elements in the Hamiltonian matrix. Fortunately, evaluation of the many (three and four) electron integrals directly can be circumvented through a resolution of the identity [132, 139, 263], whereby the products of Gaussian basis functions are approximated as a linear expansion in terms of an auxiliary basis set (ABS). In most implementations the orbital basis set (OBS) is used in conjunction with an optimised complementary auxiliary basis set (CABS) for this purpose [272], which together are denoted OptRI. The two-electron integrals are approximated using density fit-
2.1. Solving the electronic Schrödinger equation

Solving the electronic Schrödinger equation (DF), which is procedurally identical to (RI) except the latter acronym is usually reserved for the many electron integrals in R12/F12. Separate ABSs for the Fock and exchange integrals are used compared to all other two-electron integrals. The former is suffixed JKFIt, and the latter MP2Fit. Each ABS works best if pre-optimised for the method to which it is applied, and so the user must be prudent when considering the numerous choices existing in the literature. For HF Coulomb and exchange integrals (JKFIT) the auxiliary basis sets by Weigand [280–282] are almost uniformly used for all elements. For density fitting of the standard 2-electron integrals (MP2FIT), the ABSs by [283] and [96] are recommended for light and medium mass atoms, and for the post-3d elements where a relativistic pseudopotential (PP) approach is utilised the correspondingly optimised basis by [106] is recommended. For resolution of the identity (RI) of the many-electron integrals, the basis set developed by [298] is recommended for light elements, and that by [106] for the post-3d elements when a PP based approach is utilised.

Currently most quantum chemistry software packages employ the CCSD(T)-F12 implementation of Adler et al. [2, 133] due to its simplicity and effectiveness. As of yet, no way has been found to include the perturbed triples’ contribution in the F12 treatment, and so it is computed in exactly the same way as conventional CCSD(T) calculations and scaled to reduce the basis set error [133].

2.1.6 Relativistic effects

Relativistic effects are important for particles with speeds approaching that of light. In heavy atoms, electrons close to the nucleus have average velocities of approximately \( \bar{v} = Z \) a.u., where \( Z \) is the nuclear charge and the speed of light \( c \approx 137 \) a.u. Therefore it is clear that for high accuracy spectroscopic applications inclusion of relativistic effects is essential. Pyykkö [201] summarises the consequences of relativistic effects observed throughout the periodic table. These are (1) the relativistic shrinking and stabilization of \( s \) and \( p \) orbitals \( O(Z^4 c^{-2}) \) [21], (2) the spin-orbit splitting of the \( p, d, \) etc. orbitals, and (3) the radial expansion and destabilisation of the valence \( d \) and \( f \) orbitals \( O(Z^2 c^{-2}) \) [225]. The third effect is indirectly caused by the core contraction increasing the nuclear screening of the slower moving \( d \) and
Chapter 2. Theoretical background

$f$ orbitals, that reside further from the nucleus due to their increased angular momentum. A review of relativistic effects in atoms and molecules was published by Pyykkö in 1988 [200], for a more recent overview see [168].

For light molecules, relativistic corrections can usually be calculated with sufficient accuracy by the Pauli Hamiltonian mass-velocity (MV) and Darwin (D1) corrections [242]. The mass-velocity term is always negative and corrects the kinetic energy of the system, the one-electron Darwin term is always positive and corrects the Coulomb attraction. However, the Pauli Hamiltonian cannot be used variationally, which is a serious drawback for heavy elements, in which the relativistic modification to the wavefunction is significant.

A more complete description of relativistic many-electron systems subject to a nuclear potential is given by the Dirac Hamiltonian [68, 75]

\[
\hat{H}_{DCB} = \sum_i h_i + \sum_{i<j} h_{ij}
\]

(2.23)

where the one-particle Dirac Hamiltonian is

\[
h_i = c\alpha \cdot \mathbf{p} + \beta c^2 + V_n
\]

(2.24)

and $h_{ij}$ is the electron-electron interaction. In the simplest approximation, and the one most commonly used for chemical purposes, $h_{ij}$ refers only to the non-relativistic Coulomb repulsion. More rigorous descriptions are provided by the frequency-dependent Breit interaction, or its more approximate Gaunt interaction [8], neither of which will be discussed here. Analogously to the non-relativistic Hartree-Fock equations outlined earlier, the Dirac-Hartree-Fock (DHF) equations can be formulated by requiring the stationarity (not minimisation, for reasons we will see shortly) of the Dirac equation, and solved iteratively using a self-consistent field procedure [123].

The one-electron Hamiltonian in Eq. (2.24) is $4 \times 4$ in structure, consisting of the standard momentum operator $\mathbf{p}$, a $2 \times 2$ matrix $\alpha$ of Pauli spin matrices $\sigma_i$, a $2 \times 2$ matrix of identity matrices $\mathbf{I}$, and a Coulomb-type external potential $V$. In
this structure the wavefunctions are four-component spinors, with individual components corresponding to positive energy (electronic) spin-up and spin-down, and negative energy (positronic) spin-up and spin-down solutions. However, the presence of negative solutions brings with it two practical problems. Firstly, there are infinitely many negative energy solutions, which Dirac interpreted as the creation of positron-electron pairs, and are not useful for chemistry applications as they have excitation energies far greater than those of typical core and valence electronic excitations (1 MeV as opposed to 100 eV [101]). Following from this is the fact that due to these positronic solutions, the desired electronic solution is no longer the global minimum. In order to prevent collapse of the variational (DHF) calculation, a careful balance (the so-called kinetic balance condition) between the large (electronic) and small (positronic) component basis set must obeyed. Unfortunately this requires a huge number of basis functions to represent the small component, and the result is that the majority of computational expense goes on computing integrals that do not contribute to the overall energy.

Several schemes for decoupling the large and small components of the Dirac spinors through the application of unitary transformation were subsequently proposed. The main three being the Foldy-Wouthuysen transformations [84], the Douglas-Kroll-Hess (DKH) approach [70, 100], and the X-operator techniques [20, 22]. In particular the DKH approach has been developed by Hess and co-workers to the point where it is a viable computational tool, and since implemented in several quantum chemistry packages including MOLPRO [287].

Reiher and Wolf [208, 209, 288] were the first to realise an algorithm for the analytic derivation of the DKH Hamiltonian to arbitrary order. Moreover, they show that the expansion order necessary to achieve ‘exact’ decoupling, i.e., infinite order decoupling, can be predetermined without the need for any quantum chemistry calculations. In practice, for ordinary chemical problems the DK2 approach gives satisfactory results. However, for optimum use either 4th-order expansion or infinite order expansion is recommended [209], this is because the DKH Hamiltonian is slightly dependent on the chosen unitary parametrisation beyond 4th-order, with this
dependence diminishing only in the case of an infinite order expansion. Although the decoupling scheme presented by Reiher and Wolf \[208, 209, 288\] is the most rigorous treatment of relativistic effects available, a major drawback of the approach is that the DKH Hamiltonian does not commute with the F12/R12 correlation functions [27], thus cannot be used optimally with F12/R12 methods. Therefore, in heavy elements where the treatment of electron correlation and relativistic effects are necessary to achieve a reasonable level of accuracy, alternative approaches may be preferable.

Another popular approach to the treatment of scalar relativistic effects in heavy atoms is that of effective core potentials (ECPs) owing to low computational expense. Thorough reviews of ECPs are given by [40, 69] and the references therein, below only an overview of their construction is presented. Here we consider only relativistic ECPs, the non-relativistic case involves only subtle differences (e.g., replacing DHF with standard HF).

The general procedure for formulating an ECP is as follows: initially, high quality all-electron Dirac-Hartree-Fock (DHF) orbitals are generated by employing, for example, the one-electron Dirac Hamiltonian with an appropriate two-electron operator (e.g., the non-relativistic Coulomb repulsion term), and a SCF procedure. A valence-electron-only molecular Hamiltonian is constructed where we are free to chose the size of the core. This choice may not reflect the ‘true’ core orbitals, and if our definition of the core is smaller than the ‘true’ core (which is always preferable), the ‘true’ outer core electrons (that are not included in our definition) are treated explicitly in valence space. This Hamiltonian takes the form

\[
\hat{H}_v = -\frac{1}{2} \sum_i n_v \nabla_i^2 + \sum_i n_v V_{PP}(i) + \sum_{i<j} \frac{1}{r_{ij}},
\]

which is inherently non-relativistic; all relativistic contributions are to be folded into the pseudopotential \(V_{PP}\). The eigenfunctions of this Hamiltonian are valence electron pseudo-orbitals, which are constructed on a grid, and must be smooth and nodeless in the core region and match the all-electron DHF orbitals in the valence region. The pseudopotential that satisfies these conditions is generally expressed
as the product of radial and angular components, where the radial components are typically linear combinations of Gaussians, and the angular components are spherical harmonic projection operators that ensure valence orbitals of different angular momenta ‘see’ different effective core potentials. Coefficients of the radial expansions are the parameters to be fit, in a least-squares sense, to the all-electron orbital functions.

It has been found that ECPs can yield almost identical results to all-electron calculations \cite{183,187,189}. Such high-accuracy is, in part, due to the use of post-Hartree-Fock reference data, whereby static core correlation effects can be introduced through multireference DHF \cite{166,167}. Finally, as with the DKH Hamiltonian, ECPs do not commute with the F12 correlation function. However, alternative treatments have been found to work well \cite{26,286}. Therefore in cases where electron-correlation is expected to dominate over relativistic effects, ECPs may be preferable to all-electron calculations, so long as suitably optimised F12 basis sets are available.

2.2 TROVE

Having gained the tools necessary to solve the electronic Schrödinger equation and thus generate a potential energy surface from first principles, we may now turn our attention to solving the nuclear motion problem.

For all nuclear motion calculations presented in this thesis I use the variational program TROVE \cite{313}. One of the great advantages of TROVE is its (approximate) treatment of the nuclear kinetic energy operator, which is represented as a series expansion about a chosen reference molecular configuration or configurations. This avoids the complex procedure of pre-deriving a unique Hamiltonian operator for each different molecular system under investigation, and in fact, allows new systems to be tackled with relative ease providing a description of the molecular symmetry group has been programmed. In the following section I provide a general overview of the TROVE methodology (see also Refs. \cite{313,314}), with specific focus on its application to XY$_3$-type molecules (see also Ref. \cite{306}).
The rotational-vibrational-translational Schrödinger equation for a system of $N$ nuclei in the Born-Oppenheimer approximation is

$$\left( -\frac{\hbar^2}{2} \sum_{i=1}^{N} \frac{1}{m_i} \nabla_i^2 + V \right) \Psi_{\text{trv}} = E_{\text{trv}} \Psi_{\text{trv}} \quad (2.26)$$

where nucleus $i = 1, 2, 3, \ldots, N$ has mass $m_i$ and the coordinates $(R_{iX}, R_{iY}, R_{iZ})$ are defined relative to the space-fixed $X, Y, Z$ axis, and $\nabla$ is the corresponding partial derivative operator. As written, the equation is not well suited to finding a solution for the internal state of the molecule. Firstly, the translation motion leads to a continuous energy spectrum that is not useful in spectroscopy, and Eq. (2.26) does not reflect the fact that this can be separated from the remaining rotational and vibrational motions. Secondly, the rotational and vibrational motions are only very weakly coupled. Therefore to facilitate a separation of the wavefunction into the product of rotational and vibrational components, we would like to embed a set of orthonormal axes that maximally separate these motions. This is the ‘molecule-fixed axis’ $(x, y, z)$, which can be found by fulfilment of the Eckart equations [76]

$$\sum_i \mathbf{r}_e^i \times \mathbf{r}_i = 0, \quad (2.27)$$

where $\mathbf{r}_e^i$ and $\mathbf{r}_i$ are the vectors defining the equilibrium and overall position of the $i^{th}$ nucleus respectively, in the $(x, y, z)$ axis. By differentiating the Eckart equations with respect to time, they are seen to minimise the angular momentum in the molecule fixed axis system. The molecule-fixed axes have their origin located at the nuclear centre of mass, and axes aligned along the principal axes of inertia in order to minimise off-diagonal components of the inertia tensor. The coordinate transform that rotates from the space-fixed to molecule-fixed axes (or vice-versa) is defined by the Euler angles $(\theta, \phi, \chi)$ and the direction cosine matrix. In the case of non-rigid molecules, that exhibit ‘contortional’ motion, an additional condition must be fulfilled [220]

$$\sum_{\alpha, i} m_i (a_i^\alpha - a^\alpha_{i\alpha}) a^\prime_{i\alpha} = 0, \quad (2.28)$$
where \( \alpha = x, y, z \), \( a_{i\alpha} = a_{i\alpha}(\rho) \) is the reference geometry of the \( i \)th nucleus, which is a function of the large amplitude coordinate \( \rho \), and \( a'_{i\alpha} = \partial a_{i\alpha}/\partial \rho \). Equation (2.28) is the Sayvetz equation, it defines contortion \( (\rho) \) so as to minimise contortion-vibration coupling.

Now an appropriate axis has been selected, we must choose a suitable set of internal vibrational coordinates \( \xi \). For our choice of \( \xi \), the only constraint is that they must unambiguously represent the internal degrees of freedom of the molecular vibrations, and as such, \( 3N - 6 \) (\( 3N - 5 \) for linear molecules) are required.

### 2.2.1 Kinetic energy operator

The molecular kinetic energy operator (KEO), expressed in terms of \( N \) generalised coordinates \( \Xi = (T_X, T_Y, T_Z, \theta, \phi, \chi, \xi_1, \ldots, \xi_{3N-6}) \) and generalised momenta \( \hat{\Pi} = (-i\hbar \partial/\partial \xi_1, \ldots, -i\hbar \partial/\partial \xi_{3N-3}, \hat{J}_x, \hat{J}_y, \hat{J}_z) \) covering all degrees of freedom, is expressed as

\[
\hat{T} = \frac{1}{2} \sum_{\lambda, \lambda'}^N \sum_{i=1}^{3N} p_{\lambda, \lambda'}^{\dagger} G_{\lambda, \lambda'}(\xi) p_{\lambda, \lambda'} + U(\xi),
\]

where \( p_{\lambda} \) is the momentum conjugate to coordinate \( \xi_{\lambda} \), \( \hat{J}_\alpha \) is the \( \alpha = x, y, z \) component of the total angular momentum, \( U(\xi) \) is a pseudopotential, and \( G_{\lambda, \lambda'} \) is a matrix of coefficients. Disregarding the translational motion, the above expression is split into three blocks: a \( (3N - 6) \times (3N - 6) \) block which is associated with the internal vibrational motion, a \( 3 \times 3 \) block which is associated with the overall rotational motion, and a \( (3N - 6) \times 3 \) block that describes the Coriolis coupling between the two motions. The kinetic energy matrix \( G_{\lambda, \lambda'} \) is the matrix of expansion coefficients that relate the products of the conjugate momenta to their associated contributions to the KEO. Each element of \( G_{\lambda, \lambda'} \) and \( U(\xi) \) is represented as a series expansion in terms of functions of the generalised coordinates.

The kinetic energy matrix \( G_{\lambda, \lambda'} \) has the form

\[
G_{\lambda, \lambda'} = \sum_{i=1}^{N} \sum_{\alpha=x, y, z} \frac{s_{\lambda, i\alpha} s_{\lambda', i\alpha}}{m_i}
\]

(2.30)
where \(x, y, z\) are the body-fixed axis, \(m_i\) is the mass of the \(i\)th nuclei \((i = 1, \ldots, N)\) and

\[
s_{\lambda, i\alpha} = \frac{\partial \Xi_{\lambda}}{\partial r_{i\alpha}} \tag{2.31}
\]

is the Jacobian coordinate transform. The pseudopotential \(U(\xi)\) is also a function of \(s_{\lambda, i\alpha}\) and the reader is directed to Ref. [313] for the exact analytic expression.

The inverse matrix \(t = s^{-1}\), has \(t_{\lambda, i\alpha}\) elements [236]

\[
t_{i\alpha, n} = \frac{\partial r_{i\alpha}}{\partial \xi_n} \quad (n = 1 \ldots 3N - 6) \quad \text{(vibration)}, \tag{2.32}
\]

\[
t_{i\alpha, \beta} = \sum_{\gamma = x, y, z} \epsilon_{\alpha\beta\gamma} r_{i\gamma} \quad (\beta = x, y, z) \quad \text{(rotation)}, \tag{2.33}
\]

\[
t_{i\alpha, \beta} = \delta_{\alpha\beta} \quad (\beta = x, y, z) \quad \text{(translational)}, \tag{2.34}
\]

where \(r_{i\alpha}\) denotes a Cartesian component \((\alpha = x, y, z)\) of an \(i\)th nucleus, \(\epsilon_{\alpha\beta\gamma}\) is the Levi-Civita symbol, and \(t_{i\alpha, n}, t_{i\alpha, \beta}\), and \(t_{i\alpha, \beta}\) are identified with the \(3N - 6\) vibrational, three rotational, and three translational coordinates, respectively. The matrix \(s\) is obtained by inverting \(t\), which is a non-trivial problem to solve analytically. Instead \(s\) and \(t\) are expanded in terms of functions of the internal coordinates \(g(\xi)\), and the resulting system of equations is solved through a recursive numerical scheme [306, 313].

### 2.2.2 Coordinates

In practice, the above scheme used to construct the KEO can be applied to any set of vibrational coordinates, and in order to simplify the procedure of obtaining \(s_{\lambda, i\alpha}\) we choose linearised versions of geometrically defined internal coordinates \(S_n = (\Delta r_1, \Delta r_2, \Delta r_3, S_{4a}, S_{4b}, \rho)\) where \(S_{4a} = \frac{1}{\sqrt{6}}(2\alpha_{23} + \alpha_{12} + \alpha_{13})\) and \(S_{4b} = \frac{1}{\sqrt{2}}(\alpha_{13} - \alpha_{12})\) are symmetrised linear combinations of the inter-bond angles.

The linearised coordinates \(S_n^l = (\Delta r_1^l, \Delta r_2^l, \Delta r_3^l, S_{4a}^l, S_{4b}^l, \rho)\) are given as linear combinations of the Cartesian displacements \(d_{i\alpha}\) from the equilibrium (reference) configuration as [306]

\[
\Delta r_k^l = \sum_{\beta = x, y, z} \frac{a_{k\beta} - d_{4\beta}}{r_e} (d_{k\beta} - d_{4\beta}), \tag{2.35}
\]
2.2. TROVE

Figure 2.1: Reference configuration of a general XY₃ molecule in the molecule-fixed axis system, the origin lies as the centre of mass. Figure taken from Ref. [306].

\[
\Delta \alpha_{kl} = \frac{1}{r^2 \sin \alpha_e} \sum_{\beta=x,y,z} \left[ [(a_{l\beta} - a_{4\beta}) - \cos \alpha_e (a_{k\beta} - a_{4\beta})](d_{k\beta} - d_{4\beta}) 
+ [(a_{k\beta} - a_{4\beta}) - \cos \alpha_e (a_{l\beta} - a_{4\beta})](d_{l\beta} - d_{4\beta}) \right],
\]

(2.36)

where \(d_{i\alpha}\) is the Cartesian displacement of the \(i^{th}\) nuclei from its reference position, and \(a_{k\beta}\) is the cartesian position \((\beta = x,y,z)\) of the \(k^{th}\) nuclei in the molecule’s reference configuration, in terms of the equilibrium N–H bond length and \(\rho\). These set of linearised coordinates along with \(\rho\) form our set of vibrational coordinates \([\Xi = (\xi_1, \ldots, \xi_{3N-6}) = (\Delta r'_1, \Delta r'_2, \Delta r'_3, S_{4a}', S_{4b}', \rho)]\) used to construct the molecular kinetic energy operator, and the corresponding conjugate momenta are therefore \([\hat{\Pi} = (p_{l1}', p_{l2}', p_{l3}', p_{4a}', p_{4b}', \hat{J}_x, \hat{J}_y, \hat{J}_z, \hat{J}_\rho)]\).

For molecules with large amplitude motion the Hamiltonian is expanded on a grid of equidistant points of the large-amplitude coordinate. In the case of inverting ammonia the large amplitude coordinate is \(\rho\) and the aforementioned expansions of the KEO in terms of the \(3N - 7\) small amplitude coordinates occurs on a grid of \(\rho\) values. Typically each expansion is taken to 6th order, and a grid of \(\sim 1000\) \(\rho\) values used. The approach of introducing a non-rigid reference configuration to treat floppy molecules was first introduced by Hougen, Bunker, and Johns [110] and it is the approach we follow in our treatment of NH₃.
2.2.3 Potential energy function

In order to complete our construction of the molecular Hamiltonian the potential energy surface \( V \) is represented as a polynomial expansion in terms of a suitable set of internal coordinates. For \( \text{XY}_3 \)-type molecules such as \( \text{NH}_3 \) and \( \text{AsH}_3 \) we choose an analytical expression of the following form

\[
V(\xi_1, \xi_2, \xi_3, \xi_4, \sin(\bar{\rho})) = V_e + V_0 \sin(\bar{\rho}) + \sum_i F_i \sin(\bar{\rho}) \xi_i + \sum_{i \leq j} F_{ij} \sin(\bar{\rho}) \xi_i \xi_j + \ldots
\] (2.37)

which is given in terms of the internal coordinates

\[
\xi_k = 1 - \exp\left[-a(r_k - r_{eq})\right], \quad (k = 1, 2, 3), \quad \xi_4 = (2\alpha_1 - \alpha_2 - \alpha_3)/\sqrt{6}, \quad \xi_5 = (\alpha_2 - \alpha_3)/\sqrt{2}, \quad \sin\bar{\rho} = \frac{2}{\sqrt{3}} \sin[(\alpha_1 + \alpha_2 + \alpha_3)/6].
\] (2.38) (2.39) (2.40) (2.41)

In Eq. (2.37)

\[
F_{ij...} \sin(\bar{\rho}) = \sum_{s=0}^{N} f_{ij...}^{(s)} \left[ \sin(\rho_{eq}) - \sin(\bar{\rho}) \right]^s
\] (2.42)

and \( r_k \) is the \( \text{N-H}_k \) bond length, \( \alpha_j \) is the \( j^{\text{th}} \) H-N-H bond angle (opposite to the \( j^{\text{th}} \) bond), \( r_{eq} \) is the equilibrium value of \( r_k \), \( a \) is a molecular (Morse) parameter, and \( \rho_{eq} \) is the equilibrium value of the umbrella coordinate \( \bar{\rho} \). \( V_0 \) represents the pure inversion potential and \( f_{ij...}^{(s)} \) are parameters to be fit to the grid of \textit{ab initio} electronic energies. This form has been used successfully for previous high resolution studies of \( \text{NH}_3 \) [302] and \( \text{PH}_3 \) [303].

Written in terms of our chosen coordinates, our analytic expression for the potential is not compatible with the KEO, which is defined in terms of linearised coordinates. Furthermore, our current representation of the potential is isotope independent, whereas the linearised coordinates are defined relative to a non-rigid
2.2. TROVE

reference configuration \((a_{ij}(\rho))\) in Eqs. \((2.35)\) and \((2.36)\) which is isotope dependent. The coordinates therefore undergo a three-step transformation into the linearised coordinates \(\{\xi_n^l, \rho\}\) (see \([306]\)), and the coefficients \(F_{ij}^l(\rho)\) are calculated numerically via a 4-point finite differences procedure on the same grid of \(\rho\) values used to numerically construct the KEO. The potential is then re-expanded on this grid. Usually an \(8^{th}\)-order expansion in \(\xi_n^l, \rho\) is sufficient to ensure accurate nuclear motion calculations.

2.2.4 Rovibrational basis

Following the the construction of a numerical representation of the Hamiltonian operator, a suitable rovibrational basis set must be chosen in order to perform our variational calculation and solve the nuclear Schrödinger equation.

The vibrational basis set in TROVE is constructed, at the most elementary level, as products of 1-dimensional primitive functions

\[
|\upsilon\rangle = \prod_v |\upsilon_v\rangle = \phi_{\upsilon_1}(\xi_1)\phi_{\upsilon_2}(\xi_2)\cdots\phi_{\upsilon_{N-6}}(\xi_{N-6}),
\]

(2.43)

where \(\phi_{\upsilon_i}(\xi_i)\) each depend on one internal coordinate. These are obtained as solutions to 1D-Schrödinger equations

\[
H_n^{1D} = -\frac{\hbar^2}{2} \frac{\partial}{\partial \xi_n} G_n^{1D} \frac{\partial}{\partial \xi_n} + U_n^{1D}(\xi_n) + V_n^{1D}(\xi_n)
\]

(2.44)

which are formed by freezing all coordinates in the full nuclear Hamiltonian, except \(\xi_n\), at their equilibrium values. In this thesis I use functions of stretching, bending and large amplitude (inversion) coordinates. For the stretching and inversion functions the Numerov Cooley approach is used to solve the 1D Schrödinger equation numerically on a grid of points. For the bending functions, harmonic oscillators are used.

In principle the full dimensional Hamiltonian could be diagonalised in the primitive bases, however it is far more computationally efficient to use symmetry to our advantage. Ro-vibrational wavefunctions that transform irreducibly according
to one of the representations of the molecular symmetry (MS) group do not interact with one another in diagonalisation of the Hamiltonian, and so in a symmetry adapted basis the Hamiltonian is block diagonal, greatly reducing computational cost.

The approach to symmetrizing our basis is the same as described in Ref. [314]. Summarily, we first define subspaces of coordinates that are transformed into one another by the MS group operations. For each subspace of coordinates, a reduced Hamiltonian is formed by vibrationally averaging the full-dimensional vibrational Hamiltonian over all primitive functions from the other subspaces. This reduced Hamiltonian is diagonalised in the basis of (products of) primitive functions relating to our chosen subspace. Each of the resulting eigenfunctions can be classified under one of the irreducible representations of the MS group, to discern which, a ‘symmetry sampling’ procedure is performed (see Ref. [314]). The process is repeated for each coordinate subspace resulting in a symmetry adapted vibrational basis set.

The rotational wavefunctions can be symmetrized analytically simply based off the rotational quantum numbers. The total ro-vibrational basis used to diagonalise the rotational-vibrational Hamiltonian is formed by taking products of the symmetrised vibrational and rotational basis functions from different subspaces as

$$\Psi_{\lambda_0}^{(0)} \otimes \Psi_{\lambda_1}^{(1)} \otimes \Psi_{\lambda_2}^{(2)} \otimes \ldots \otimes \Psi_{\lambda_L}^{(L)}$$

where \((0), (1), \ldots (L)\) denotes the subspace, and \(\Gamma\) denotes the irreducible representation of the eigenfunction, indexed by \(\lambda\), generated from that subspace. Projection operators are then applied to the resulting (reducible) representations generated by these products to convert them into irreducible representations. Our final symmetrised ro-vibrational basis can be written in the compact notation

$$|\nu, J, K, m, \tau_{\text{rot}}\rangle_{\Gamma} = \prod_{\nu} |\nu\rangle \times |J, K, m, \tau_{\text{rot}}\rangle^{\Gamma}$$

However, in almost all cases TROVE does not solve the full rotational-
vibrational Schrödinger exactly as described above. Rather, it uses the symmetrisation procedure to first generate symmetry adapted vibrational basis functions with which it solves the purely vibrational problem. The resulting vibrational eigenfunctions are then saved to disk, along with the $G_{\alpha\beta}$ and $G_{\lambda\alpha}$ matrix elements of the rotational and Coriolis contributions to the KEO in Eq. (2.29) (see also Eqs. (27) and (28) of Ref. [304]). These vibrational eigenfunctions $\Psi_{J=0,i}^{\Gamma}$ form the vibrational basis with which we perform $J > 0$ calculations. Multiplication of $\Psi_{J=0,i}^{\Gamma}$ with rigid symmetric rotor eigenfunctions, then forms the ($J = 0$)-contracted basis. Here, a final symmetrisation procedure is performed to reduce the rotational-vibrational basis functions into irreducible representations.

The final ($J = 0$)-contracted ro-vibrational basis set takes the form

$$\left|\psi_{i,J,K,m,\tau_{\text{rot}}}^{\Gamma}\right> = \left[\left|\psi_{J=0}^{\Gamma}\right>|J,K,m,\tau_{\text{rot}}\right>\right]^{\Gamma},$$

(2.47)

which, since states with different $J$ do not mix, is block-diagonal in $J$ and total symmetry $\Gamma$. Diagonalisation of the full Hamiltonian matrix, i.e., the matrix representation of Eq. (2.29) plus Eq. (2.37), in this basis, and therefore solving the Schrödinger equation variationally, is a standard matrix eigenvalue problem. Our final molecular wavefunctions then take the form

$$\left|\psi_{i,J,K,m,\tau_{\text{rot}}^{\Gamma}}\right> = \sum_{\nu,K,\tau} c_{\nu,K,\tau}^{J,K,m,\tau_{\text{rot}}^{\Gamma}} \left|\psi_{\nu,J=0,\Gamma_{\text{vib}}}^{\Gamma_{\text{rot}}^{\Gamma}}\right>|J,K,m,\tau_{\text{rot}}^{\Gamma},$$

(2.48)

where $c_{\nu,K,\tau}^{J,K,m,\tau_{\text{rot}}^{\Gamma}}$ are coefficients, and the largest amplitude coefficient labels the ro-vibrational state.

### 2.2.5 Refinement

The potential energy surface (PES) is the foundation of all nuclear motion calculations. Within the limitations imposed by convergence errors due to basis set expansion, KEO expansion and PES re-expansion, it effectively determines the accuracy of all subsequent ro-vibrational energy levels and transitions. A gold standard in computational molecular spectroscopy is the calculation of ro-vibrational energy
levels with so-called ‘spectroscopic’ accuracy, i.e., sub-wavenumber (< 1 cm$^{-1}$), although there is an increasing demand for accuracy to far exceed this. For small systems this may be achieved through purely $ab$ initio means. However, for larger molecules electronic structure calculations become more expensive, and so it is common for the analytic representation of the PES to be empirically adjusted using experimental data.

The procedure implemented in TROVE has been reported in Ref. [303]. Briefly, a small correction term $\Delta V$ is added to the $ab$ initio surface $V$ such that the modified potential is expressed as

$$V' = V + \Delta V = V + \sum_{ijk \ldots} \Delta f_{ijk \ldots} \left\{ \{ \xi_1 \xi_2 \xi_3 \ldots \}^A \right\}, \quad (2.49)$$

where the correction term takes the form of Eq. (2.37), except $f_{ijk \ldots}$ have been replaced by the adjustable parameters $\Delta f_{ijk \ldots}$, and the superscript $A$ indicates totally symmetric permutations of the coordinates $\{ \xi_1 \xi_2 \xi_3 \ldots \}$. The new ‘perturbed’ Hamiltonian thus takes the following form

$$H = T + V + \Delta V = H_0 + \Delta V. \quad (2.50)$$

Assuming the unperturbed Hamiltonian $H_0$ has been diagonalised and we have its eigenvalues $E_{0,n}^{J,\Gamma}$ and eigenvectors $\psi_{0,n}^{J,\Gamma}$, the energy derivatives with respect to the adjustable parameters $\Delta f_{ijk \ldots}$ can be calculated by taking the matrix elements

$$\frac{\partial E_{0,n}^{J,\Gamma}}{\partial \Delta f_{ijk \ldots}} = \langle \psi_{n}^{J,\Gamma} | \{ \xi_1 \xi_2 \xi_3 \ldots \}^A | \psi_{n}^{J,\Gamma} \rangle, \quad (2.51)$$

where we have used the Helmann-Feyman theorem [82]. Providing a suitably reliable set of experimental energies $E_n^{obs}$ has been compiled, we may now employ a standard nonlinear least-squares fitting algorithm, such as Newton-Gauss, to adjust the parameters $\Delta f_{ijk \ldots}$ in equation (2.49) so as to minimise the sum of squared
Simulating absorption spectra

2.3 Simulating absorption spectra

2.3.1 Line strength and selection rules

An electromagnetic wave may induce an oscillating electric (or magnetic) moment in a molecule, leading to absorption of a photon if certain resonance conditions are met. The amplitude of this moment is the transition moment $\langle f | \mu_A | i \rangle$ between an initial state $i$ and a final state $f$, where $\mu_A$ is the electric dipole moment vector in the space fixed $A = XYZ$ axis system. The transition moment is related to an important quantity known as the line strength, that determines the probability of the molecule making the transition $f \leftarrow i$ as follows

$$S(f \leftarrow i) = g_{ns} \sum_{m_i m_f A} \sum_{X,Y,Z} |\langle \Phi'_{int} | \mu_A | \Phi''_{int} \rangle|^2.$$  \hspace{1cm} (2.53)

The summation over $m$, which is the projection of $J$ onto the space fixed $Z-$axis, appears because in the absence of an applied external field the energy does not depend on the space-fixed molecular orientation. The multiplication factor $g_{ns}$ is the nuclear spin statistical weight, which appears because we have assumed the
wavefunction does not depend on nuclear spin, and so states with the same nuclear spin are degenerate. In order for the line strength in Eq. (2.53) not to vanish, the integrand must contain the totally symmetric representation, and so the symmetries of the total internal wavefunctions for the upper and lower states, \( \Phi'_{\text{int}} \) and \( \Phi''_{\text{int}} \), must satisfy [35]

\[
\Gamma'_{\text{int}} \otimes \Gamma''_{\text{int}} \supset \Gamma^*.
\] (2.54)

Here \( \Gamma^* \) is the symmetry of the molecular dipole moment, which must not change sign for any permutation operation, but change sign upon any inversion or permutation-inversion operation. This is the first of the rigorous selection rules governing electric dipole transitions.

As before, the rotational and vibrational motions of the internal wavefunction are effectively decoupled by using the molecule fixed \( xyz \)–axis representation, defined by the Eckart-Sayvetz conditions [76][220]. Therefore in order to evaluate the line strength in Eq. (2.53) it is necessary to represent the space-fixed dipole moment function \( \mu_A \) in the molecule fixed axis system, i.e., in terms of the Euler angles \( (\phi, \theta, \chi) \) and internal coordinates. This transformation is performed using spherical tensor algebra [317], the details of which are given in Ref. [35]. Here only the result is provided. Expanding the rovibrational wavefunction in terms of vibrational \( |V\rangle \) and rotational and \( |JKm_\tau_{\text{rot}}\rangle \) basis functions

\[
|\Phi_{\text{rovib}}\rangle = \sum_{VK_{\text{rot}}} C_{VK_{\text{rot}}} |V\rangle |JKm_\tau_{\text{rot}}\rangle
\] (2.55)

the final (Born-Oppenheimer) equation for the line strength is

\[
S(f \leftarrow i) = g_{ns} \sum_{m_f m_i \sigma = -1} \frac{1}{1} \sum_{V'K'_{\tau_{\text{rot}}} V''K''_{\tau_{\text{rot}}} \mu_{m_1}} C_{V'K'_{\tau_{\text{rot}}} \mu_{m_1}} C_{V''K''_{\tau_{\text{rot}}} \mu_{m_1}}
\times \sum_{\sigma = -1}^{1} \langle V' | (\Phi'_{\text{elec}} | \mu_{m_1}^{(1, \sigma')} ) | \Phi''_{\text{elec}} | V'' \rangle
\times \langle J'K'm_f \tau'_{\text{rot}} | D_{\sigma_{\tau} \sigma_1}^{(1)} (\phi, \theta, \chi) | J''K''m_i \tau''_{\text{rot}} \rangle^2,
\] (2.56)
where the space-fixed and molecule-fixed dipole moments have been written in spherical tensor notation as per Ref. [35]. The rotational matrix with elements $D^{(1)}_{\sigma'\sigma}(\phi, \theta, \chi)$, as given by [317], can be evaluated using the Clebsch-Gordan series

$$
\langle J'K'm_f\tau'_\text{rot}|[D^{(1)}_{\sigma'\sigma}(\phi, \theta, \chi)]^*|J''K''m_i\tau''_\text{rot}\rangle
= (-1)^{k'+m_f'}\sqrt{(2J'+1)(2J''+1)}\begin{pmatrix} J'' & 1 & J' \\ k'' & \sigma' & -k' \end{pmatrix} \begin{pmatrix} J'' & 1 & J' \\ m'' & \sigma & -m' \end{pmatrix}
$$

(2.57)

By the general properties of the $3j$–symbols, the line strength will vanish unless

$$
\Delta J = J' - J'' = 0, \pm 1
$$

(2.58)

This is the second of the rigorous selection rules for rovibronic transitions in the absence of spin. Furthermore, individual matrix elements will vanish unless

$$
\Delta K = K' - K'' = 0, \pm 1
$$

(2.59)

The electronic matrix elements in Eq. (2.56) can be calculated directly via expectation values or approximated numerically. In this thesis only transitions within the electronic ground state, which transforms according to the totally symmetric representation, are considered. Thus for the vibrational matrix elements to be non-zero the symmetry requirements are

$$
\Gamma'_{\text{vib}} \otimes \Gamma''_{\text{vib}} \supset \Gamma(\bar{\mu}_m)
$$

(2.60)

The dipole moment $\bar{\mu}_m$ can be related to $\bar{\mu}_\alpha$ by the relationships given by Ref. [35]. Under the $D_{3h}/C_{3v}$ group operations $\bar{\mu}_c$ transforms with $A''_2/A_1$ symmetry and $(\bar{\mu}_x, \bar{\mu}_y)$ transform together with $E'/E$ symmetry, which can be deduced simply by acknowledging that $\mu_\alpha$ has the same symmetry properties as the translational coordinates. Vernacular within the spectroscopic community labels a transition to be
parallel if the vibrational matrix elements of the operator \( \bar{\mu}_z \) are dominant, and perpendicular if the vibrational matrix elements of the coupled \( (\bar{\mu}_x, \bar{\mu}_y) \) operators are dominant.

As a final consideration it is important to note the role of rovibrational coupling when evaluating the line strength. Although previously we treated the rotational and vibrational components of the wavefunction as being separable through the appropriate choice of a coordinate system, in truth, Coriolis coupling and centrifugal distortion spoil \( K \) as a good quantum number and \( \Gamma_{\text{vib}} \) as a good symmetry label. For this reason Eqs. (2.59) and (2.60) are not rigorous selection rules.

### 2.3.2 Dipole moment surface

The dipole moment may be represented as the first derivative of the electronic energy with respect to the external electric field \( \varepsilon_\alpha \) in the limit of vanishingly small \( \varepsilon_\alpha \),

\[
\mu_\alpha = - \left( \frac{dE}{d\varepsilon_\alpha} \right)_{\varepsilon_\alpha = 0},
\]

which can be approximated using the numerical finite differences procedure

\[
\frac{dE}{d\varepsilon_\alpha} \approx \frac{E(\varepsilon_\alpha) - E(-\varepsilon_\alpha)}{2\varepsilon_\alpha},
\]

where \( E(\varepsilon_\alpha) \) is the electronic energy in the presence of a small electric field orientated along the space fixed \( \alpha = X, Y, Z \) axis. Thus six electronic energy calculations are required to generate all three cartesian components.

In order for the matrix elements in Eq. (2.53) to be evaluated numerically, the body fixed \( xyz \)–components of the dipole moment, as defined by the Eckart and Sayvetz conditions [76, 220], must be represented as a function of the vibrational coordinates. A suitable analytic expression for \( \mu_\alpha \) can be found by computing an \( n \)-dimensional grid (where \( n \) is the number of coordinate degrees of freedom) of dipole moment values, to which we least-squares fit the coefficients of, typically, a polynomial expansion in terms of geometrically defined coordinates. Generally a fourth-order expansion can be expected to obtain a sufficiently reliable representation of the dipole moment. For the purpose of computing the dipole moment values,
one of the many ab initio electronic structure packages e.g. MOLPRO [287], can be used. In such packages the dipole moment is usually returned as function of coordinates \(x'y'z'\), which are body-fixed with the origin and axes defined by the \(Z\)–matrix. For the \(XY_3\) systems discussed in this thesis this corresponds to the origin placed on the nitrogen atom in the case of \(NH_3\), and the arsenic atom in the case of \(AsH_3\).

In addition to these two coordinate representations \((x,y,z)\) and \((x',y',z')\), a more general expression that is independent of the choice of molecule fixed axis system is used. This is the symmetrized molecular bond (SMB) representation [304]. The process of transforming between the three representations is described in detail in Refs. [304, 312]; below only the analytic expression in the SMB representation is given because it is arguably the most intuitive.

In the SMB representation the electronically averaged dipole moment \(\bar{\mu}\) is constructed as symmetrized projections onto the molecular bonds with the dipole moment components \((\bar{\mu}_{\Gamma^1}, \bar{\mu}_{\Gamma^2a}, \bar{\mu}_{\Gamma^2b})\) in the molecule fixed axis system given by 4\(^{th}\) order polynomial expansions

\[
\bar{\mu}_{\Gamma^1}(\chi_1, \chi_2, \chi_3, \chi_4a, \chi_4b, \chi_6) = \cos \theta \left[ \mu_0^{\Gamma^1} + \sum_i \mu_i^{\Gamma^1} \chi_i + \sum_{i,j} \mu_{ij}^{\Gamma^1} \chi_i \chi_j \right], \tag{2.63}
\]

\[
\bar{\mu}_{\Gamma^2a}(\chi_1, \chi_2, \chi_3, \chi_4a, \chi_4b, \chi_6) = \mu_0^{\Gamma^2a} + \sum_i \mu_i^{\Gamma^2a} \chi_i + \sum_{i,j} \mu_{ij}^{\Gamma^2a} \chi_i \chi_j \tag{2.64}
\]

\[
\bar{\mu}_{\Gamma^2b}(\chi_1, \chi_2, \chi_3, \chi_4a, \chi_4b, \chi_6) = \mu_0^{\Gamma^2b} + \sum_i \mu_i^{\Gamma^2b} \chi_i + \sum_{i,j} \mu_{ij}^{\Gamma^2b} \chi_i \chi_j \tag{2.65}
\]

where \(\Gamma^1, \Gamma^2a\) and \(\Gamma^2b\) are the irreducible components \(A''_1, E'_a\) and \(E'_b\) of \(D_{3h}\), or \(A_2\),
E_a and E_b for C_{3v}. The coordinates \( \chi_i \) are

\[
\chi_k = \Delta r_k \exp\left(-\beta \Delta r_k^2\right), \quad (k = 1, 2, 3), \tag{2.66}
\]

\[
\chi_{4a} = (2\alpha_1 - \alpha_2 - \alpha_3)/\sqrt{6}, \tag{2.67}
\]

\[
\chi_{4b} = (\alpha_2 - \alpha_3)/\sqrt{2}, \tag{2.68}
\]

\[
\chi_6 = \sin \rho e - \sin \rho, \tag{2.69}
\]

\[
\sin \rho = \frac{2}{\sqrt{3}} \sin \left[\left(\alpha_1 + \alpha_2 + \alpha_3\right)/6\right], \tag{2.70}
\]

where \( \mu_{ij...}^\Gamma \) are the expansion parameters to be fit, in the least squares sense, to the \textit{ab initio} points, and \( \Delta r_k = r_k - r_{eq} \). The dipole moment components \( (\mu_{\Gamma_{2a}}^\Gamma, \mu_{\Gamma_{2b}}^\Gamma) \) are transformed as linear combinations of each other by the symmetry group operations and so transform together as \( E \)-symmetry. The relationship between these parameters is given by Yurchenko et al. [312]. For this reason the parameters \( (\mu_{ij...}^\Gamma_{2a}, \mu_{ij...}^\Gamma_{2b}) \) must be fit simultaneously and \( \mu_{ij...}^\Gamma \) are fitted separately. For an extensive discussion on the SMB representation of the dipole moment function the reader is directed to [304].

### 2.3.3 Line intensities and absorption cross-sections

For a beam of monochromatic light of frequency \( \tilde{\nu} \) passing through a sample of gas, and resonant with a rotational-vibrational transition, the intensity of absorbed radiation can be related to the line strength through the following equation

\[
I(f \leftarrow i) = \frac{8\pi^3 N_A \tilde{\nu} \exp\left(-E''/k_B T\right)[1 - \exp\left(-hc\tilde{\nu}/kT\right)]}{(4\pi\varepsilon_0)^3hcQ} \times S(f \leftarrow i), \tag{2.71}
\]

where \( E' \) and \( E'' \) are the upper and lower state energies respectively, \( T \) is the absolute temperature, \( k_B \) is the Boltzmann constant and \( h \) is Planck’s constant. In thermal equilibrium the occupied energy levels follow a Boltzmann distribution, and so the probability of a lower state being occupied is accounted for by the factor \( \exp(-E''/kT)/Q(T) \), where the partition function \( Q(T) \) is given by

\[
Q = \sum_w g_w \exp(-E_w/kT). \tag{2.72}
\]
In the above equation, $E_w$ is the energy and $g_w$ is the total degeneracy of state $w$. Equation 2.71 is most commonly given in units of $\text{cm}^{-1}/(\text{molecule/cm}^2)$, which is interpreted as energy absorbed per column density of molecules.

A spectroscopic line list refers to a list of transition frequencies and associated quantities from which intensities of the form of Eq. (2.71) can be calculated, such as Einstein $A$-coefficients. Although a single line is effectively a delta function centred on $\tilde{\nu}$ with amplitude $I(f \leftarrow i)$, when measured experimentally an absorption line will not be limited to a single frequency but will instead be ‘broadened’ over a range by various physical phenomena. The mechanisms for this broadening, and the profile with which they convolute the line intensity are as follows:

- **Natural lifetime broadening** – Inherent uncertainty of the upper state energy proportionate to the state lifetime due to Heisenberg’s time-energy uncertainty principle $\Delta E \sim \hbar/\Delta \tau$. It is responsible for the range of photon frequencies being emitted for a single transition. The resulting profile is Lorentzian and sufficiently narrow to be neglected in all but the most sensitive experiments. For example, for short-lived rotational-vibrational states with lifetimes of order $\Delta \tau \approx 10^{-2}$, one can expect broadening of order $10^{-9} \text{ cm}^{-1}$ which, for most spectroscopic investigations, is several orders of magnitude less than the two dominant contributions to the linewidth, which are discussed below.

- **Doppler broadening** – Thermal translational motion of the molecule relative to the experimental rest frame results in incident light appearing Doppler shifted by frequency $\Delta \nu = v_0(u_T/c)$ in the frame of the molecule, where $u_T$ is the velocity along the line of sight and $v_0$ is the incident light frequency. The corresponding normalised Doppler profile is Gaussian and expressed in terms of the Doppler half-width

$$\Gamma_D = \frac{v_0}{c} \sqrt{\frac{2\ln(2)kT}{m}}, \quad (2.73)$$

where $m$ is the molecular mass, $T$ is temperature and $k$ is the Boltzmann
constant. The Doppler profile takes the form
\[ F_D(\nu - \nu_0) = \sqrt{\frac{\ln(2)}{\pi}} \frac{1}{\Gamma_D} \exp\left( -\ln(2) \left( \frac{\nu - \nu_0}{\Gamma_D} \right)^2 \right). \] (2.74)

- Collisional broadening – Perturbation interactions between colliding molecules can reduce the effective state lifetimes if the interval between molecular collisions is much less than the natural lifetime. According to Heisenberg’s uncertainty principle this will act to increase the uncertainty on the upper and lower state energies and thus the frequency of the emitted photon. Additionally, the central transition frequency \( \nu_0 \) may be shifted with respect to the collision-free case by frequency \( \Delta \). As with natural lifetime broadening, the collisional broadening profile can be described by a Lorentzian line shape with half-width at half-maximum (HWHM) \( \Gamma_C \). This profile is given by
\[ F_L(\nu - \nu_0) = \frac{1}{\pi (\nu - \nu_0 - \Delta)^2 + \Gamma_C^2}. \] (2.75)

Calculation of \( \Gamma_C \) is far more involved than that of \( \Gamma_D \), and its value is well known to display a strong dependence on \( \Delta J \Delta K(J'', K'') \) and temperature. The current best theoretical predictions of \( \Gamma_C \) and \( \Delta \) are based on semi-empirical techniques that incorporate various corrections to the Robert-Bonamy or Anderson theories [38, 108, 210, 318]. Alternatively, the HWHM values can be measured experimentally for a variety of rotationally excited transitions and over a range of temperatures, and fit to suitable functional forms [120, 130].

Compilations of collisional broadening coefficients and temperature exponents have been performed for the most common terrestrial broadeners for limited values of \( J \) and \( K \), and are available in various databases [91, 311].

The most commonly used approximation to the line shape is a convolution of the Doppler and Lorentzian broadening profiles, which is referred to as the Voigt profile [114]. Although the Voigt profile has proven to be sufficiently accurate for
most spectroscopic applications, it suffers from two main deficiencies: The first is a characteristic W-shaped residual when fit to high-precision absorption measurements [125,150], which has motivated the development of several post-Voigt models which add additional parameters to the profile [256]. The second is the lack of a true analytic representation, which can make its computation for a large number of spectral lines expensive as convoluting the two profiles involves, for example, performing a fast Fourier Transform (FFT) and inverse fast Fourier Transform (IFFT) for each line. To save on computational expense, approximations are often used rather than the full Voigt profile [154,214].

Each of the aforementioned line profiles are normalised so that integration of an isolated absorption peak yields the underlying absolute line intensity. Inversely, to generate synthetic spectra from a theoretical line list, lines must first be convoluted with the relevant profiles on a (preferably dense) grid of frequencies. The resulting representation of the spectral features is referred to as one of absorption cross-sections, which are usually measured in units of cm$^2$/molecule. Cross-sections can be related to the oft directly measured quantities transmittance $T$ and absorbance $A$ by the Beer-Lambert law:

$$T = \frac{F_{ir}}{F_{in}} = e^{-\sum_{i=1}^{N} n_i L \sigma_i},$$

(2.76)

$$A = -\log_{10} T,$$

(2.77)

where $F_{in}$ and $F_{ir}$ are the incident and transmitted radiation fluxes respectively, $N$ is the number of absorbing species, $n_i$ is the number density of absorbers of species $i$, $L$ is the optical path length, and $\sigma_i$ is the absorption cross-section of the $i^{th}$ absorbing species.

### 2.4 $2f\text{wms}$

Wavelength-modulation spectroscopy (WMS) is a derivative form of spectroscopy that is commonly applied for trace species measurements (e.g., Refs. [72,111,246,292]) and measurements in harsh environments (e.g., Refs. [80,152,202,212]), due
Chapter 2. Theoretical background

... to its improved sensitivity and signal-to-noise ratio (by a factor of 2–100 [211]) when compared to direct absorption spectroscopy [109, 151] for a multitude of reasons [146]. An excellent source covering WMS from a theoretical point-of-view is that of Ref. [248], here only a brief summary key features is presented.

A coherent optical beam (e.g., from a tunable laser diode) emitting at a sinusoidally modulated wavelength is tuned across an absorption feature by ramping the DC laser injection current, and thus the laser wavelength and intensity, which results in an optical field expressed by

\[ E(t) = E_0 \exp\left[i(\omega_0 t + M \sin(\omega_m t))\right], \]  

(2.78)

where \( E_0 \) is the optical field amplitude, \( \omega_0 \) is the DC carrier frequency, \( M \) is the modulation amplitude, and \( \omega_m \) is the modulation frequency. As the carrier frequency is swept over a spectroscopic absorption feature the modulated signal will scan back-and-forth over the absorption peak, which leads to harmonic components in the photodetector signal. A lock-in amplifier then shifts the desired harmonic component to the DC through the process of phase mixing and the application of trigonometric identities [229]. There, a low-pass filter is applied to reject all frequencies outside the filter bandwidth, isolating the harmonic component. The entire process of phase mixing and filtering is known as demodulation, and if this occurs for the \( n \)th harmonic, the method is known as \( n \)th−harmonic wavelength modulation spectroscopy (\( n \)F\( n \)WMS, where \( n = 1, 2, \ldots \) for the first, second, etc. harmonic).

Typically in the literature the chosen values of \( M \) and \( \omega_m \) are split into two regimes. The first is referred to for historical reasons as wavelength modulation spectroscopy (WMS), and corresponds to one of large \( M \) and small \( \omega_m \), i.e., \( M >> 1 \) and \( \omega_m << \Gamma \), where \( \Gamma \) is the linewidth of the absorption feature of interest. The second is referred to as frequency modulation spectroscopy (FMS), and corresponds to one of small \( M \) and large \( \omega_m \) i.e. \( M \lesssim 1 \) and \( \omega_m >> \Gamma \). However, often the two extremes are referred to under the umbrella term of WMS.

In the limit of a small \( M \) the \( n \)th harmonic signal is proportional to the \( n \)th-derivative of the line shape of the absorbing feature. As \( M \) is increased the main ef-
fect is to distort the observed signal [66], losing the second derivative (see Fig. 3.24 for an example). The resulting broadening of the signal is known as *modulation broadening*. Generally, the optimum value of $M$ is considered to be the one that maximises the signal amplitude, and for Gaussian, Lorentzian and Voigt lineshape models this occurs for $M \approx 2.2$ [212]. However, this increase in amplitude must be balanced against the modulation broadening so as not to cause lines to overlap unnecessarily. In practical applications, modulation of the laser intensity also distorts the observed signal [151], and introduces asymmetry in the signal side lobes. This latter effect can difficult to model correctly as it requires intimate knowledge of the diode-laser parameters [221].
Chapter 3

Ammonia ($^{14}$NH$_3$)

### 3.1 Introduction

Ammonia (NH$_3$) is ubiquitous throughout the Universe. It was the first polyatomic molecule to be discovered in the interstellar medium (ISM) some 40 years ago [45], and has since been discovered in the atmospheres of gas giants Jupiter [89, 137] and Saturn [117-241], cometary coma [14, 29] and Y [142, 222] and T dwarfs [39, 148]. This prevalence, combined with its complex vibrational motion make it an important tool for astronomers to probe a variety of physical conditions such as temperature [162] and H$_2$ density [107]. In the atmospheres of extrasolar giant planets, reactions between H$_2$ and N$_2$ favour the form of NH$_3$ at low temperatures (and higher pressures), whose dependencies suggest that the outer atmospheres of planets with large orbital radii will contain significant quantities of ammonia [245].

Terrestrially, the chemical potential energy gradients concerned with nitrogen redox reactions are used by some microorganisms as a main source of energy to support life. Assuming the same processes may used by biological systems present on exoplanets, detection of atmospheric ammonia, along with several other gases, is speculated to a possible biosignature [231].

Ammonia exists as a trace species in the earth’s atmosphere where, due to its toxic nature, its presence must be closely monitored [250]. Global production of ammonia stems predominantly from livestock, fertilizers, fuel burning, natural vegetation, and the ocean [32, 77]. In coal-fired plants its use in NO$_x$ removal (via
selective catalytic reduction \cite{37} from post-combustion gases often leads to the escape of excess unconverted ammonia - so-called 'ammonia slip' - if the ambient temperature falls outside the range of the reaction \cite{157}.

These emissions react with other air pollutants, and thereby help to form fine particulate matter that causes respiratory and coronary diseases in the human populace, shortening lifespan \cite{13}. It is considered to be a significant precursor to the 2013 smog formation in Eastern China \cite{153}. Moreover, ammonia contributes to acid decomposition and eutrophication which may harm natural ecosystems, and emissions from agricultural sources severely impact nitrogen use though the food production chain, which indirectly contributes to greenhouse-gas emissions and water pollution \cite{249}. For these reasons the development of high accuracy in-situ sensors to monitor ambient ammonia concentration is necessary to minimise human health risks. These sensors rely on accurate and extensive laboratory data to infer gas properties, which may not exist within the desired parameters.

For both terrestrial and astronomical applications, it is therefore necessary to use \textit{ab initio} calculations when experimental coverage fails. In the former case such calculations provide the additional opacity that results from millions, if not billions, of lines that are not experimentally known, but are crucial for accurate spectral retrievals. In the latter case, \textit{ab initio} calculations can be used to inform decisions when purchasing the components required to make laboratory measurements to guide the development of new sensors, thus reducing cost.

So far the main provider of information of the near infrared spectrum of ammonia has been theoretical line lists computed with variational nuclear motion programs \cite{112,113,302,304}. In this context the important line lists are BYTe \cite{302} and HSL-pre3 \cite{115,247}.

BYTe is a variationally computed line list for $^{14}\text{NH}_3$ covering transitions from 0 to 12 000 cm$^{-1}$ and temperatures up to 1500 K. It provides the main source of ammonia opacity data used in spectral retrieval models \cite{181}, and has been used to assign ammonia lines in a number of high resolution measurements extending from 500–11 000 cm$^{-1}$. However, it provides no coverage of the visible spectrum of
3.2 Quantum labels and symmetry

Ammonia is a pyramidal tetratomic molecule consisting of one nitrogen atom at the top of the pyramid, and three hydrogen atoms at the three corners of the base. The location of the nitrogen nucleus may be on either side of the plane formed by the three hydrogen nuclei, and so the molecule can exist in two different configurations. These are defined by whether the three protons, labelled 1, 2 and 3, are seen to be ordered 1-2-3 or 1-3-2 when counted in a clockwise direction by an observer sitting on...
top of the nitrogen nucleus. The barrier between these two configurations is roughly 1800 cm$^{-1}$, and at room temperature, over the timescale of a typical experiment it is possible for the molecule to tunnel between them, leading to observable splittings in the spectrum. The appropriate complete nuclear permutation inversion (CNPI) symmetry group classification of NH$_3$ is therefore that of D$_{3h}$ \cite{35}, for which the character table is given in table 3.1 below.

Table 3.1: Character table of D$_{3h}$

<table>
<thead>
<tr>
<th></th>
<th>(123)</th>
<th>(12)</th>
<th>(12)$^*$</th>
<th>(123)$^*$</th>
<th>(12)$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1'$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_2'$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$E'$</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>$A_1''$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$A_2''$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>$E''$</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>-2</td>
<td>1</td>
</tr>
</tbody>
</table>

The labelling of rovibrational states of NH$_3$ has been discussed extensively by Down et al. \cite{71} who suggests the following 13 useful quantum numbers to uniquely define each state:

$$
\nu_1, \nu_2, \nu_3, \nu_4, L_3, L_4, L, i, J, K, \Gamma_{\text{rot}}, \Gamma_{\text{vib}}, \Gamma_{\text{tot}}
$$

where $\nu_i$ ($i = 1, 2, 3, 4$) are the vibrational normal mode quantum numbers; $L_3 = |l_3|$, $L_4 = |l_4|$, $L = |l_3 + l_4|$ and $K = |k|$. Here, $l_3$ and $l_4$ are the vibrational angular momentum quanta associated with $\nu_3$ and $\nu_4$; $J$ is the rotational angular momentum quantum number describing the rotation of the body fixed axis relative to the space fixed axis; $k = -J, ..., J$ is the projection of $J$ onto the body fixed $z$ axis; $i$ is the inversion parity; and $\Gamma_{\text{rot}}, \Gamma_{\text{vib}}$ and $\Gamma_{\text{tot}}$ are the rotational, vibrational and total symmetries, respectively, of the rovibrational state in the molecular symmetry group D$_{3h}$ \cite{35}. In practice, only 12 labels are needed, these consist of the set given in Eq. (3.1), with the omission of $i$, and $L_3, L_4, L$ replaced by $l_3, l_4, l$.

The symmetry species of the vibrational coordinates can be found by fixing a set of Cartesian axes to the centre of each nucleus in the molecule. The matrix
3.2. Quantum labels and symmetry

representation generated by considering the effect of each group operation in $D_{3h}$ on these four sets of axes spans all vibrational, translational and rotational motion. Neglecting the latter two, which span $\Gamma_{\text{rot+trans}} = A'_2 \oplus A''_2 \oplus \bar{E}' \oplus \bar{E}''$, the symmetry species of the vibrational coordinates are found to be $\Gamma_{\text{vib}} = A'_1 \oplus A''_2 \oplus 2E'$ which relate to the symmetric stretch, inversion, degenerate stretch and degenerate bend respectively.

The wavefunctions are assigned quantum numbers according to the Herzberg convention, so that the symmetric stretch $\nu_1$, symmetric bend $\nu_2$, degenerate stretch $\nu_3$ and degenerate bend $\nu_4$ have quantum numbers $\nu_1$, $\nu_2$, $\nu_3$, $\nu_4$ respectively. This is known as the normal mode representation. It is convention in the spectroscopic community to define $\nu_2$ in terms of the label $\nu_{\text{inv}}$, which corresponds to the total number of nodes in the inversion wavefunction. The two labels are related through the inversion parity $i$ by the relation $\nu_{\text{inv}} = 2\nu_2 + i$, where $i$ takes values 1 (odd) or 0 (even) depending on whether the inversion wavefunction changes sign upon tunelling through the barrier.

The symmetry species of the vibrational wavefunctions are discussed in Chapters 12.3 and 15.4.1 of Ref. [35]: $|\nu_1\rangle$ has symmetry $A'_1$ and $|\nu_2\rangle$ has symmetry $[A''_2]^{\nu_{\text{inv}}}$. Thus, states with even $\nu_{\text{inv}}$ are totally symmetric in $D_{3h}$ and states with odd $\nu_{\text{inv}}$ have the same symmetry as the inversion coordinate $\rho$ which, in the molecule’s reference configuration, is the angle between the N-H bonds and the z-axis. The two doubly degenerate vibrations $\nu_3$ and $\nu_4$ are usually modelled as 2D isotropic harmonic oscillators $|\nu_i\rangle$ ($i = 3, 4$) (see Chapter 11.3.2 of Ref. [35]) and so give rise to the additional quantum numbers $l_i$, where $l_i = -\nu_i, -\nu_i + 2, ..., 0, ..., \nu_i - 2, \nu_i$. These are associated with the symmetry of $|\nu_i\rangle$ as follows: $l_i = 0$ corresponds to $A''_1$ symmetry; $l_i = \pm 3n$ ($n = 1, 2, ...$) to both $A'_1$ and $A'_2$; and $l_i \neq \pm 3n$ ($n = 0, 1, 2, ...$) to $E'$. The total symmetry of the vibrational wavefunction is then found by taking the product of symmetries of the individual vibrational mode wavefunctions.

The rotational wavefunctions $|J, K, m, \tau_{\text{rot}}\rangle$ (where $\tau_{\text{rot}}$ is the rotational parity) are constructed as symmetrised linear combinations of rigid rotor wavefunctions [312]. By considering the effect of the rotation operations given in Table
12-1 of Ref [35] on the rotational wavefunctions (see Chapter 12.2 of Ref [35]), \( |J, K, m, \tau_{\text{rot}}\rangle \) can be shown to have the following symmetries: \( A_1^{\dagger} \) for \( K = 0 \) and \( J \) even; \( A_2^{\dagger} \) for \( K = 0 \) and \( J \) odd; \( A_1^{\dagger} \) for \( K = 3n \) \( (n = 1, 2, \ldots) \) and \( \tau_{\text{rot}} = 0; A_2^{\dagger} \) for \( K = 3n \) and \( \tau_{\text{rot}} = 1; E_a^{\dagger} \) for \( K \neq 3n \) and \( \tau_{\text{rot}} = 0; E_b^{\dagger} \) for \( K \neq 3n \) and \( \tau_{\text{rot}} = 1. \) Here, \( \dagger = ' \) if \( K \) is even, and \( \dagger = '' \) if \( K \) is odd.

### 3.3 Potential energy surface

A number of PESs for \( \text{NH}_3 \) currently exist in the literature, the most recent and notable of these are Y2010 [303], HSL-pre3 [112, 247], and the \textit{ab initio} surface by Polyansky et. al. [320]. A brief summary of each provided below.

Y2010 was produced by empirical refinement of the \textit{ab initio} surface reported in Ref. [315] using the procedure outlined in Section 2.2.5. It was optimised for use in TROVE, employing a \( P_{\text{max}} = 28 \) vibrational basis set and a KEO expansion at 6th-order, and PE re-expansion at 8th-order. Typical accuracy is sub-wavenumber for predicted term values under 6000 cm\(^{-1}\), and as much as several wavenumbers thereafter. It has been used to calculate infrared \( \text{NH}_3 \) line lists [302, 304], which have been used in a wide number of applications ranging from the assignment of astronomical spectra [39] to high precision studies [3].

HSL-pre3 has only been partially reported in the literature [247]. Its predecessor, reported in Ref. [112], includes a BODC correction and approximate treatment of non-adiabatic effects through a scaling of the nuclear kinetic energy operator [226]. The resulting rovibrational energy levels of HSL-pre3 are reported to achieve accuracy of 0.1 cm\(^{-1}\) or better for term values under 7500 cm\(^{-1}\) [115], and are therefore substantially more accurate than BYTe. The HLS-2 energy levels and transitions lists, which extend to \( J = 10 \), have been used to re-assign a number of lines in HITRAN and identify erroneous transitions below 7000 cm\(^{-1}\).

Polyansky et al. [199] report an \textit{ab initio} PES calculated at the MRCI level of theory in the aug-cc-pCVQZ and the aug-cc-pCV5Z basis sets, with extrapolation to the complete basis set limit. Relativistic corrections were generated as the expectation value of the MVD1 operator for the CASSCF wavefunctions in the aug-
3.3. Potential energy surface

cc-pwCV5Z basis set. The adiabatic correction was computed using CCSD in the cc-pwCVTZ basis set and the program CFOUR [88]. A grid of 22 494 nuclear geometries with energies below $\hbar c \cdot 20 000 \text{ cm}^{-1}$ were used in the final fit, for which the analytic expression given in Eq. (2.37) was used. Nuclear motion calculations were performed using a version of TROVE that had been specifically adapted to use curvilinear coordinates [295], in conjunction with a large vibrational basis set constrained by a maximum polyad of $P_{\text{max}} = 40$ (see Section 3.3.1). Expansion of the KEO and re-expansion of the PE function were taken to 6th and 8th-order respectively. Using this model, the computed vibrational term values are generally accurate to within 1–2 cm$^{-1}$ up to 6000 cm$^{-1}$. Comparisons with the early assignments of the visible spectrum by Coy and Lehmann [143] showed discrepancies between 2 and 15 cm$^{-1}$ for the stretching overtones up to 18 000 cm$^{-1}$. It is worth noting, however, that these levels of accuracy can be reached only using a very large rotation-vibration basis set that becomes computationally unmanageable for $J \gg 0$ calculations.

Our approach to generating a new ‘spectroscopic’ PES for NH$_3$, as employed in the construction of line lists as part of the ExoMol project, follows the refinement procedure outlined in section 2.2.5. Our starting point is provided by the surface by Polyansky et al. [199], which is refined using energy levels taken from the 2015 MARVEL database [3].

### 3.3.1 Computational details

Our vibrational basis set used in the refinement is constructed as symmetrised linear combinations of the 1D primitive functions $\phi_{n1}(r_1^j)$, $\phi_{n2}(r_2^j)$, $\phi_{n3}(r_3^j)$, $\phi_{n4}(\xi_4^j)$, $\phi_{n5}(\xi_5^j)$, $\phi_{n6}(\hat{\rho})$. Here $n_i$ are principle quantum numbers, the coordinates $(r_1^j, r_2^j, r_3^j, \xi_4^j, \xi_5^j)$ are linearised versions of $(r_1, r_2, r_3, \xi_4, \xi_5)$ and $\phi_i(\xi_i)$ are solutions to the corresponding 1D Schrodinger equations (discussed in Section 2.2.4). The parameters used in computing $\phi_i(\xi_i)$ are listed in Table 3.2. The rotational basis is constructed as linear combinations of spherical harmonics.

The maximum vibrational excitation allowed by the product of the primitive functions $\phi_{n_i}(\xi_i^j)$ is limited by the polyad number $P$. The polyad number is defined
Table 3.2: Parametrisation of the primitive functions generated as solutions to 1D Schrödinger equations.

<table>
<thead>
<tr>
<th>Basis</th>
<th>Borders</th>
<th>No. grid points</th>
<th>Range solns.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{n_i}(r_i)$ (i = 1, 2, 3)</td>
<td>-0.4, 2.0</td>
<td>2000</td>
<td>0, 8</td>
</tr>
<tr>
<td>$\phi_{n_i}(\xi_i)$ (i = 4, 5)</td>
<td>-1.91, 1.91</td>
<td>9000</td>
<td>0, 34</td>
</tr>
<tr>
<td>$\phi_{n_i}(\rho)$</td>
<td>-1.91, 1.91</td>
<td>1000</td>
<td>0, 34</td>
</tr>
</tbody>
</table>

in terms of the principle quantum numbers $n_i$, and represents the total quanta of vibrational excitations in terms of the lowest energy fundamental. In the case of NH$_3$, it is written as

$$P = 2(n_1 + n_2 + n_3) + n_4 + n_5 + \frac{n_6}{2}.$$  \hspace{1cm} (3.2)

Only combinations of primitive functions with $P \leq P_{\text{max}}$ are thus included in the variational calculation.

Owing to the computer resources and software available to us our vibrational basis set was limited to $P_{\text{max}} = 34$, using linearised coordinates. The most computationally expensive step necessary for the refinement procedure is the generation and symmetrisation of the Hamiltonian matrix elements of each term in the correction potential. This is required in order to calculate the Hellmann-Feynman derivatives that constitute the Jacobian matrix necessary for the least-squares fitting, and is the primary factor that limits our basis set size. At $P_{\text{max}} = 34$ this process requires 54Gb of RAM per correctional term, of which there are 304 to allow for every correctional term to vary. Extending the basis set to $P_{\text{max}} = 36$ increases the memory requirements to 67 Gb per correctional term, which exceeds the available of RAM of a standard compute node on the Darwin HPC cluster where the calculations were performed.

Ideally, we would pick a value of $P_{\text{max}}$ such that our vibrational eigenfunctions are converged for at least all energies up to the highest value included in the refinement. However, our convergence testing (see Table. 3.3) indicates a $P_{\text{max}}$ of 34 provides stretching eigenvalues converged to within 0.1 cm$^{-1}$ only up to about 7000 cm$^{-1}$ (also see Fig. 1 of Ref. [199]), compared to the highest experimentally derived term value included in the refinement, that was $\sim$18 000 cm$^{-1}$. In this case
Table 3.3: Convergence of the vibrational energy levels (in units of cm$^{-1}$) with increasing polyad $P_{\text{max}} = 28, \ldots, 40$ computed using the PES by Polyansky et al. \cite{199}, compared to the MARVEL experimentally derived values \cite{3}.

<table>
<thead>
<tr>
<th></th>
<th>obs.</th>
<th>$\Delta E_{P28}$</th>
<th>$\Delta E_{P32}$</th>
<th>$\Delta E_{P34}$</th>
<th>$\Delta E_{P36}$</th>
<th>$\Delta E_{P40}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A1'$</td>
<td>6796.73</td>
<td>-1.74</td>
<td>-1.66</td>
<td>-1.65</td>
<td>-1.63</td>
<td>-1.62</td>
</tr>
<tr>
<td>$A2''$</td>
<td>6795.31</td>
<td>-1.56</td>
<td>-1.43</td>
<td>-1.4</td>
<td>-1.4</td>
<td>-1.38</td>
</tr>
<tr>
<td>$E''$</td>
<td>6609.66</td>
<td>-1.88</td>
<td>-1.34</td>
<td>-1.17</td>
<td>-1.13</td>
<td>-0.98</td>
</tr>
<tr>
<td>$E'$</td>
<td>6608.83</td>
<td>-1.33</td>
<td>-1.06</td>
<td>-1.04</td>
<td>-0.95</td>
<td>-0.87</td>
</tr>
<tr>
<td>$E'$</td>
<td>6666.10</td>
<td>-0.6</td>
<td>-0.30</td>
<td>-0.26</td>
<td>-0.19</td>
<td>-0.12</td>
</tr>
<tr>
<td>$E''$</td>
<td>6677.95</td>
<td>-1.63</td>
<td>-1.34</td>
<td>-1.25</td>
<td>-1.23</td>
<td>-1.17</td>
</tr>
<tr>
<td>$E'$</td>
<td>6677.23</td>
<td>-1.43</td>
<td>-1.29</td>
<td>-1.27</td>
<td>-1.24</td>
<td>-1.20</td>
</tr>
<tr>
<td>$E''$</td>
<td>6850.70</td>
<td>-1.52</td>
<td>-1.40</td>
<td>-1.37</td>
<td>-1.36</td>
<td>-1.34</td>
</tr>
<tr>
<td>$E'$</td>
<td>6850.20</td>
<td>-1.49</td>
<td>-1.44</td>
<td>-1.44</td>
<td>-1.42</td>
<td>-1.41</td>
</tr>
<tr>
<td>$E''$</td>
<td>9738.84</td>
<td>-10.93</td>
<td>-5.80</td>
<td>-4.69</td>
<td>-4.35</td>
<td>-3.59</td>
</tr>
<tr>
<td>$E'$</td>
<td>9738.15</td>
<td>-6.39</td>
<td>-4.06</td>
<td>-3.69</td>
<td>-3.08</td>
<td>-2.56</td>
</tr>
<tr>
<td>$E''$</td>
<td>9689.72</td>
<td>-12.6</td>
<td>-7.41</td>
<td>-5.74</td>
<td>-5.14</td>
<td>-4.11</td>
</tr>
<tr>
<td>$E'$</td>
<td>9689.84</td>
<td>-5.88</td>
<td>-5.81</td>
<td>-4.57</td>
<td>-3.66</td>
<td>-2.88</td>
</tr>
<tr>
<td>$E''$</td>
<td>9642.32</td>
<td>-13.82</td>
<td>-7.54</td>
<td>-5.55</td>
<td>-4.78</td>
<td>-3.97</td>
</tr>
<tr>
<td>$E'$</td>
<td>9639.65</td>
<td>-9.2</td>
<td>-5.34</td>
<td>-4.62</td>
<td>-3.57</td>
<td>-2.97</td>
</tr>
<tr>
<td>$E''$</td>
<td>12628.2</td>
<td>-36.39</td>
<td>-23.55</td>
<td>-18.71</td>
<td>-16.61</td>
<td>-14.93</td>
</tr>
<tr>
<td>$E'$</td>
<td>12675.5</td>
<td>-43.3</td>
<td>-22.51</td>
<td>-18.21</td>
<td>-12.91</td>
<td>-12.60</td>
</tr>
<tr>
<td>$A1'$</td>
<td>15447.38</td>
<td>-60.18</td>
<td>-19.62</td>
<td>-14.25</td>
<td>-12.56</td>
<td>-7.51</td>
</tr>
<tr>
<td>$A1'$</td>
<td>15450.82</td>
<td>-36.38</td>
<td>-14.16</td>
<td>-13.68</td>
<td>-6.60</td>
<td>-5.59</td>
</tr>
<tr>
<td>$E''$</td>
<td>15448.7</td>
<td>-59.02</td>
<td>-18.57</td>
<td>-14.63</td>
<td>-11.61</td>
<td>-6.53</td>
</tr>
<tr>
<td>$A2''$</td>
<td>18109.18</td>
<td>-182.92</td>
<td>-74.80</td>
<td>-36.76</td>
<td>-24.16</td>
<td>-11.51</td>
</tr>
<tr>
<td>$A1'$</td>
<td>18109.47</td>
<td>-152.78</td>
<td>-33.78</td>
<td>-25.2</td>
<td>-15.95</td>
<td>-0.91</td>
</tr>
<tr>
<td>$E''$</td>
<td>18107.56</td>
<td>-187.22</td>
<td>-75.70</td>
<td>-38.36</td>
<td>-25.91</td>
<td>-13.48</td>
</tr>
<tr>
<td>$E'$</td>
<td>18109.47</td>
<td>-156.05</td>
<td>-33.86</td>
<td>-27.15</td>
<td>-16.16</td>
<td>-1.22</td>
</tr>
</tbody>
</table>


3.3. Potential energy surface

the error due to using an incomplete basis set is absorbed into the refined potential, and our results at higher energies are only reproducible using TROVE with a specific set of model input parameters. Our expansion of the kinetic energy operator and re-expansion of the potential function in terms of linearised coordinates, we take to 6th and 8th order respectively. The effect of this has been documented in the past \cite{199}, and is typically less than 0.1 cm$^{-1}$ for energies under $h\epsilon \cdot 10 000$ cm$^{-1}$.

The final contracted vibrational ($J = 0$) basis set used in the refinement comprised all eigenfunctions of the $J = 0$ Hamiltonian which correspond to the energies below $h\epsilon \cdot 20 000$ cm$^{-1}$.

3.3.2 Experimental data and weights included in the refinement

Our primary source of experimental data was the MARVEL (measured active rotation-vibration energy levels) \cite{87} study of ammonia by Al Derzi et al. \cite{3}, which
contains the most accurate and most complete list of experimentally-derived energies available for NH$_3$. From this source we initially included 543 carefully selected states ranging from 0–7254 cm$^{-1}$ with $J \leq 8$. Because the quality of the refinement depends crucially on the accuracy of the experimental data, we assessed the reliability of each MARVEL state prior to the refinement based on the number of transitions it was involved in, and whether it followed the expected $J - K$ dependence within the vibrational band. During the refinement, any states that did not behave similarly to the rest of the band were removed, as they tended to degrade the quality of the refinement. In total we identified 81 MARVEL states with $J \leq 15$ for which no suitable partner could be found in our calculated energies list, and a further 80 that displayed uncomfortably large residuals of between 0.5 and 4.0 cm$^{-1}$. Residuals of this magnitude are substantially larger than would be expected given the accuracy of the remaining theoretical energy levels. However, each vibrational band, $J$ and $K$ value must be considered separately, as well as the sources from which the MARVEL energies were derived. The MARVEL energies flagged at this point as erroneous were later removed or re-labelled in an updated version of the MARVEL database.

Between 7555 cm$^{-1}$ and 11 000 cm$^{-1}$ the only existing experimental assignments are those of Barton et al. [18,19]. Due to complexity of spectra in this region only some of their assignments could be confirmed by ground state combination differences (GSCDs), and in this case there is always the possibility of misassignment. Even if a GSCD partner is found, such dense spectra will contain false positives that fall within the tolerance ranges of the assignment. This is especially true if the line positions and intensities of the ab initio calculations are of dubious accuracy, and the particular GSCD partner is a medium strength or weak line. We therefore decided to take a cautious approach, preferring to use only 34 energies from the $\nu_2 + 2\nu_3$ band of Barton et al. [19], which we found to be reliable, and perform our own tentative assignments using the ab initio PES of Polyansky et al. [199] and intermediate versions of the refinement. This provided an additional 105 energy levels with wavenumbers ranging from 7584 to 10512 cm$^{-1}$, although most of these
levels were derived from only one experimental transition via a visual comparison between line lists and so cannot be considered validated.

It is well known that the inclusion of $J > 0$ states in the refinement is necessary to optimise the equilibrium geometry [223], and it is preferable for each $J$ to contain $K$ sublevels ranging from $0 - J$ to constrain both the $B_c$ and $C_c$ oblate top rotational constants. We included states with $J = 0 - 4, 6, 8$ in the refinement, however, to save on computational costs only $A'_2$ and $A''_2$ were used for $J > 3$. Enough experimental data fell within these criteria to provided sampling of all important normal mode directions along our PES. Several high energy band centres at 12 000, 15 000 and 18 000 cm$^{-1}$ were also included from the early work of Coy and Lehmann [56] which, at the time of performing our calculations, were the only assigned spectra above 12 000 cm$^{-1}$ available in the literature. We also note the unassigned 5$\nu_1$ absorption bands of NH$_3$ by Giver et al. [90].

As high energy data is scarce it is possible for the refined PES to assume unphysical shapes far from equilibrium. We therefore constrained the refinement to the original, low-weighted $ab\ initio$ points following the simultaneous fit approach by Yurchenko et al. [305], with the motivation that our refined surface not deviate substantially from the $ab\ initio$ surface. Initially these consisted of the same 20 000 points between 0 and 20 000 cm$^{-1}$ that were used to fit the $ab\ initio$ PES [199], but we had problems with holes and double minima appearing above 20 000 cm$^{-1}$ but below dissociation at 40 000 cm$^{-1}$. To prevent this, a further 13 000 points were generated between 20 000 and 50 000 cm$^{-1}$ using the $ab\ initio$ PES, and included in the refinement. The energy-dependent scheme of Ref. [199] was used to allocate weights below dissociation, and a constant weight of 0.00065 was used above. All $ab\ initio$ point weights were then multiplied by a constant factor that started at $1 \times 10^{-4}$ and was decreased, in factors of 10, to $1 \times 10^{-7}$ as the refinement progressed.

We used the weighting scheme $w_i \sim 1/\sigma_i^2$ for all MARVEL states, where $\sigma_i$ is the standard error of the $i^{th}$ energy level, as this information is provided in the MARVEL database. This is known to be the optimum weighting structure for a
general least-squares fit. For states derived from the Kitt Peak spectrum, weights were distributed uniformly, with a slight energy dependence to reduce the importance of very high energy states. On the last five iterations of the Newton-Gauss algorithm, once improvements with each iteration began to stagnate, the weights were changed so that each energy $E_J^\Gamma$ had a weight of $J$, and $J = 0$ had a weight of 1. This served to improve the fit for those bands containing fewer, or less accurately known, experimental energies. The robust weighting method by \cite{277} was used to adjust the fitting weights on the-fly.

3.3.3 Refined parameters

We could usefully vary 176 (out of 304) parameters up to fifth order in the refinement, excluding the Born-Oppenheimer diagonal correction term by Polyansky et al. \cite{199} which we kept fixed. Variation of linear terms allowed us to simulate the effects of optimising the equilibrium geometry without resorting to a separate Newton-Gauss style procedure, which cannot be performed concurrently with the PES refinement, and severely changes the vibrational structure.

Refined parameters higher than second order generally differed quite substantially from their starting value. However, the harmonic terms remained consistent, with the zero-order-inversion stretching and bending terms ($f_{11}^{(0)}$ and $f_{44}^{(0)}$) changing only by 0.45% and 0.06% respectively. Coupling terms $f_{12}^{(0)}$ and $f_{14}^{(0)}$ showed larger changes of 10.7% and 2.3%, and harmonic terms that were first order in the inversion coordinate $f_{11}^{(1)}$, $f_{44}^{(1)}$, $f_{12}^{(1)}$ and $f_{14}^{(1)}$ changed by 17%, 21%, 11% and 2%, respectively. This level of change in the low-order parameters is acceptable, as instabilities in low-order parameters might suggest the linear dependency is too high for the parameter set. Higher-order parameters predominantly influence high energy regions of the potential, and as the parameter order increases so does the degree of non-linearity. In truth, this high level of non-linearity means that large changes in higher order parameters can be offset by the response of other parameters, and therefore there are a large number of possible combinations of high-order parameters that can produce similarly shaped potentials at lower energies. Most importantly, the energy difference between the refined and \textit{ab initio} PES’s is al-
ways less than 10% that of the ab initio PES above its zero-point energy (ZPE) for grid points under 50 000 cm$^{-1}$. This was confirmed by evaluating both PES’s on a random grid of 50 000 points with borders at $0.6 \leq r_1 \leq r_2 \leq r_3 \leq 1.58$ Å and $30^\circ \leq \alpha_1 \leq \alpha_2 \leq \alpha_3 \leq 140^\circ$. Further reassurance that our PES does not suffer from any unphysical deformities is provided by nuclear motion calculations performed using our no-BODC refined PES and the GENIUSH nuclear motion program [164], carried out by Dr. Csaba Fábi. A large grid of one million points was used and by nature of the DVR approach any deep holes that exist within the grid borders are manifested in the zero point energy, which was seen to take unreasonable values if holes were present.

3.3.4 Equilibrium structure and rotational energies

The various structural parameters of our PES C2018 are given in Table 3.4 along with other theoretically predicted and experimentally derived values. Our equilibrium structure is very similar to that predicted by the NH3-Y2010 and HSL-2 PESs, with our bond angle, $\alpha_{eq}$, slightly larger than both, and our bond length, $r_{eq}$, roughly half way between the two. Of the three sets of theoretically predicted structural constants, the HSL-2 PES displays closest agreement with the experimentally determined values. Larger discrepancies with experiment are particularly noted for the C2018 values of $\alpha_{eq}$ and $\Delta E$(barrier), indicating that the C2018 bending potential could be improved.

Pure rotational energies are highly sensitive to changes in equilibrium geometry, and so accurate rotational energies are therefore indicative of an accurate equilibrium geometry. Table 3.5 compares our rotational energies to those given in MARVEL for states up to $J = 30$. Although MARVEL conventionally includes only experimentally measured transitions, for the Table 3.5 comparison the high-accuracy effective-Hamiltonian predictions by Pearson et. al. [182] and Yu et. al. [299] have have also been incorporated. This extends the largest value of $K$ from 20 to 34. For $J = 0 – 10$ our predictions show excellent agreement with the empirical values. The small differences of order 0.001 cm$^{-1}$ are likely to be inside the values which are determined due to beyond Born-Oppenheimer (BO) effects which in
Chapter 3. Ammonia ($^{14}$NH$_3$)

general are only considered when potential energy curves undergo (avoided) crossings, i.e., when two electronic states approach degeneracy (the reader is directed to Chapter 2, in particular Equations 2.8–2.10 and related text, for further discussion).

As $J$ increases, the agreement deteriorates, and beyond $J = 20$ significant discrepancies appear. These are as much as several wavenumbers for $J = 30$ states with large $K$ values, and as $K$ decreases the agreement rapidly improves. Intuitively this can be explained by the centrifugal force flattening and stretching the molecule, which will be more distorting for rotation about the primary axis. Both the inversion and stretching potentials therefore strongly couple to the equilibrium geometry at high rotational excitation.

Almost identical systematic deviations exist with the BYTe line list, for which the $J = 30$ rotational term values differ from our calculations by at most 0.1 cm$^{-1}$. Although contrary to conclusion of Zobov et. al. [321], this may suggest both PESs share a similar systematic offset in the inversion potential at high energies. One solution would be to include higher order $\nu_2$ overtones in the refinement, however, in order to sample the same energetic region as a $(J,K) = (30,0)$ rotational state ($\sim 8600$ cm$^{-1}$) we would require empirical energies from the $9\nu_2$ and $10\nu_2$ bands. Alternatively we may include very high $J$ states in the refinement, which is extremely computationally demanding using the current procedure. For example, simply saving the necessary Jacobian matrices for a $J = 30, A_2'$ block would require 10s of Tb of disk space, which is far beyond our current computational resources.

Table 3.4: Structural constants of our PES compared to previous theoretical calculations and experiment.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{eq}$/Å</td>
<td>1.010794</td>
<td>1.0109285</td>
<td>1.010668</td>
<td>1.01101</td>
<td>1.01139</td>
<td>1.0116</td>
</tr>
<tr>
<td>$\alpha_{eq}$/°</td>
<td>106.7894</td>
<td>106.7468</td>
<td>106.7489</td>
<td>106.75</td>
<td>107.17</td>
<td>106.68</td>
</tr>
<tr>
<td>$r_{SP}$/Å</td>
<td>0.993882</td>
<td>0.9943827</td>
<td>0.9942537</td>
<td>0.9942537</td>
<td>0.99460</td>
<td></td>
</tr>
<tr>
<td>$\Delta E$(barrier) / cm$^{-1}$</td>
<td>1775.17</td>
<td>1766.83</td>
<td>1784.66</td>
<td>1784.66</td>
<td>1786.8</td>
<td></td>
</tr>
</tbody>
</table>

3.3.5 Rovibrational term values

Rovibrational energy level calculations were performed up to $J = 12$ using the C2018 PES in conjunction with the variational nuclear motion program TROVE
3.3. Potential energy surface

Table 3.5: Accuracy of calculated rotational term values up to \( J = 30 \) when compared to the empirical MARVEL values \([3]\). \( \sigma_{\text{rms}} \) refers to the root-mean-square deviation and \( \Delta \) refers to the \( E_{\text{MARV}} - E_{\text{calc}} \) wavenumber differences of \( K = J \) and \( K = 0 \) states. Units of \( \Delta \) and \( \sigma_{\text{rms}} \) are cm\(^{-1}\).

<table>
<thead>
<tr>
<th>( J )</th>
<th>( \sigma_{\text{rms}} )</th>
<th>( \Delta(K = J) )</th>
<th>( \Delta(K = 0) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.001</td>
<td>-0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>5</td>
<td>0.002</td>
<td>-0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>6</td>
<td>0.002</td>
<td>-0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>7</td>
<td>0.001</td>
<td>-0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>8</td>
<td>0.001</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>9</td>
<td>0.002</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>10</td>
<td>0.004</td>
<td>0.002</td>
<td>0.005</td>
</tr>
<tr>
<td>11</td>
<td>0.006</td>
<td>0.002</td>
<td>0.009</td>
</tr>
<tr>
<td>12</td>
<td>0.009</td>
<td>-0.001</td>
<td>0.015</td>
</tr>
<tr>
<td>13</td>
<td>0.014</td>
<td>-0.004</td>
<td>0.020</td>
</tr>
<tr>
<td>14</td>
<td>0.018</td>
<td>-0.011</td>
<td>0.029</td>
</tr>
<tr>
<td>15</td>
<td>0.026</td>
<td>-0.024</td>
<td>0.038</td>
</tr>
<tr>
<td>20</td>
<td>0.176</td>
<td>-0.576</td>
<td>0.130</td>
</tr>
<tr>
<td>30</td>
<td>3.322</td>
<td>-11.353</td>
<td>0.546</td>
</tr>
</tbody>
</table>

\([313]\). Basis set and Taylor series truncations were the same as discussed in Section 3.3.1, although this time we removed the energy cut-off (of \( \hbar c \cdot 20000 \text{ cm}^{-1} \)) in our vibrational basis set. The result is a huge increase in computational demand, but we found it necessary in order to converge all rotationally excited states belonging to the stretching overtones at 18 000 cm\(^{-1}\).

To assess the accuracy of our list of energies we first compare with MARVEL levels under 7555 cm\(^{-1}\). Beyond this, the only empirical energies available in the literature, apart from those we determined using our own PES (see Section 3.5 and Ref. \([320]\)), are from the works by Barton et al. \([19]\), which we will discuss separately in Section 3.5. In the following discussion we compare our predicted energy levels to those in an unpublished update of MARVEL, that have been deemed the most reliable.

Figure 3.1 displays \( E_{\text{MARV}} - E_{\text{calc}} \) energy residuals for \( J \leq 10 \) states under 6300 cm\(^{-1}\) with BYTe residuals (\( E_{\text{MARV}} - E_{\text{BYTe}} \)) included for comparison in grey. The \( y \)-axis range has been restricted to \( \pm 1 \text{ cm}^{-1} \) for illustrative purposes, although a number BYTe residuals fall outside this range. Table 3.6 provides the associ-
ated root-mean-square (rms) deviation statistics as a function of $J$ and MARVEL vibrational label. Overall agreement is excellent, with our rms values $\sigma_{\text{rms}}$ generally staying below 0.1 cm$^{-1}$ except in a few cases which will be discussed below. Unlike BYTe, we do not utilise the empirical basis set correction (EBSC) [304], whereby the calculated vibrational band centres are replaced by their experimental counterparts. Despite this, all our vibrational term values below 6300 cm$^{-1}$ fall within 0.01–0.04 cm$^{-1}$ of the MARVEL empirical values. Similar accuracy is expected for the remaining band origins for which only $J > 0$ MARVEL data exists, with the only likely exception being the $(\nu_2 + 2\nu_4^2)^s$ (where $s/a$ denotes whether the vibrational wavefunction is symmetric or asymmetric with respect to inversion through the planar configuration) band for which, at the time of refinement, data were extremely limited. Whilst no experimentally derived band centre could be found in the literature, we judge from our $J = 1$ comparisons that our predicted band centre is roughly 0.1 cm$^{-1}$ larger than the true value.

A relatively smooth increase in energy residuals with $J$ is observed for most vibrational bands, and despite excluding states above $J = 8$ from the refinement the $J = 9, 10$ residuals behave in the same systematic way as those for $J = 0 – 8$. This speaks for the predictive power of the refinement, and reassures us that our calculations can safely be extended to higher rotational excitations. Larger increases in residuals are observed for the $2\nu_2 + \nu_3$ band, for which data only became available in the final stages of the refinement, and the $(\nu_1)^s$ band, for which the $J = 10$ rms error of 0.106 cm$^{-1}$ is still at least five times smaller than our predecessor BYTe. Note that below 6300 cm$^{-1}$ there is little difference in rms errors between MARVEL states derived from 1 or 2 transitions and those derived from 3 or more. However, given the accuracy of our $J = 0 – 10$ calculations for the $2\nu_4^0$ and $2\nu_4^2$ bands it is possible that the MARVEL $J = 12$ energies, derived from only one transition, may not be correct.

Further suspicious levels are those belonging to the $2\nu_2 + \nu_4$, $2\nu_2 + 2\nu_4$ and $2\nu_2 + 3\nu_4$ bands. Inclusion of these bands in the refinement severely damaged the quality of surrounding energies, and so they were omitted. Comparing our pre-
dictions for the $2\nu_2 + \nu_4$ with HSL-pre3 we find slightly better agreement, and for the 14 $J = 0 - 8$ states present in MARVEL the rms deviation between C2018 and HSL-pre3 is only 0.272 cm$^{-1}$. Considering the general lack of experimental data assigned to these bands, it remains unclear whether this is a problem of theory or experiment.

Above 6300 cm$^{-1}$ we noticed several conflicts between our vibrational labels and those given in the literature (HSL-pre3, MARVEL, HITRAN) (see Table 3.7), associated with the $\nu_1 + \nu_3$, $\nu_3 + 2\nu_4^2$ and $\nu_5 + 2\nu_4^0$ bands. For this reason we simply use the MARVEL labels for our comparisons in Figure 3.2 and Table 3.6 and provide our own vibrational labels and band origins separately in Table 3.7. To avoid duplication of the $\nu_1 + \nu_3$ band centre in our energies list, we suggest using the labels $(\nu_3 + 2\nu_4^0)^{s/a}$ for the 6608 and 6609 cm$^{-1}$ vibrational term values so that our labelling scheme is identical to that of HSL-pre3. Experimentally derived vibrational band centres are provided in MARVEL for 9 out of the 14 symmetric and asymmetric bands present within the 6500–6900 cm$^{-1}$ region, with which our calculations typically agree to within 0.01–0.07 cm$^{-1}$. Svoboda et. al. [251] also recently measured 7 band centres within this region using intensity comparisons between low temperature spectra recorded at 20 K and 80 K. Their derived band centres are to be considered the most accurate to-date, and we note good agreement for all but the $2\nu_2 + 3\nu_4^1$ band which displays a 2.63 cm$^{-1}$ discrepancy with our own work (and that of HSL-pre3), as well as a 3.88 cm$^{-1}$ discrepancy with the work by Lees et. al. [141] indicating the measurements by Lees may have been misassigned. Finally Sung et al. [247], who contribute the majority of lines in HITRAN between 6300–7000 cm$^{-1}$, also suggest a unique list of labels based on BYTe and HSL-2 along with estimated band centers.

Figure 3.2 shows the wavenumber differences between MARVEL and C2018, and MARVEL and BYTe, for 350 states above 6300 cm$^{-1}$. Discrepancies between C2018 and MARVEL are typically less than 0.2 cm$^{-1}$, whereas for BYTe the discrepancies are substantially larger. Although the $y$–axis range has been restricted to $\pm 2$ cm$^{-1}$, approximately one-third of BYTe energy residuals are between 2.0 cm$^{-1}$
and 4.0 cm$^{-1}$. To provide the most reliable comparison possible in Fig. 3.2, energies derived from only 1 or 2 transitions are not shown. Their associated rms deviations are, however, still included in Table 3.6 and show comparable levels of agreement to those derived by 3 or more transitions, except for the (2$\nu_3^0$)$^s$ where the RMS is roughly a factor of three larger. Such minor differences in most cases support the notion that the current set of MARVEL energy levels are reliable. For the (2$\nu_3^0$)$^s$ band it is possible that state mixings are responsible for the larger difference. A handful of C2018 levels display residuals larger than 0.5 cm$^{-1}$ when compared to MARVEL. In all instances using the alternative vibrational labels of (4$\nu_4$)$^a$, ($\nu_1 + 2\nu_2 + \nu_4$)$^s$ or (4$\nu_2 + \nu_4$)$^s$ as given by C2018 could explain this issue, and to reflect this these states have been omitted in Table 3.6. Considering that no experimental data from any of these bands was sampled in the refinement, a reduction in the accuracy of our predictions is unsurprising. There are likely also to be issues with perturbations due to resonance interactions between vibrational states which have not been correctly represented in our PES; see the recent work on this by Mizus et al. [169].
Table 3.6: Root-mean-square deviation statistics for the complete list of paired MARVEL−C2018 levels under 7555 cm$^{-1}$. Band centres and RMS statistics are in units of cm$^{-1}$. N$_{\text{states}}$ refers to the number of paired states included in the comparison. The value before the / refers to all MARVEL states, and the value after the / refers to those states derived from 3 or more transitions. Vibrational labels are taken from MARVEL.

<table>
<thead>
<tr>
<th>Band</th>
<th>Band center</th>
<th>J = 1−8</th>
<th>J = 10</th>
<th>J = 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N$_{\text{states}}$</td>
<td>$\sigma_{\text{rms}}$</td>
<td>N$_{\text{states}}$</td>
</tr>
<tr>
<td>$(v_2)^r$</td>
<td>932.4170</td>
<td>40/40</td>
<td>0.021/0.021</td>
<td>10/10</td>
</tr>
<tr>
<td>$(v_2)^a$</td>
<td>968.1253</td>
<td>40/40</td>
<td>0.010/0.010</td>
<td>11/11</td>
</tr>
<tr>
<td>$(2v_2)^r$</td>
<td>1597.4674</td>
<td>40/40</td>
<td>0.008/0.008</td>
<td>10/10</td>
</tr>
<tr>
<td>$(2v_2)^a$</td>
<td>1882.1847</td>
<td>40/40</td>
<td>0.012/0.012</td>
<td>11/11</td>
</tr>
<tr>
<td>$(v_4)^r$</td>
<td>1626.2803</td>
<td>80/80</td>
<td>0.006/0.006</td>
<td>21/21</td>
</tr>
<tr>
<td>$(v_4)^a$</td>
<td>1627.3867</td>
<td>80/80</td>
<td>0.011/0.011</td>
<td>21/21</td>
</tr>
<tr>
<td>$(3v_2)^r$</td>
<td>2384.1711</td>
<td>40/40</td>
<td>0.009/0.009</td>
<td>10/10</td>
</tr>
<tr>
<td>$(3v_2)^a$</td>
<td>2895.5336</td>
<td>40/40</td>
<td>0.015/0.015</td>
<td>11/11</td>
</tr>
<tr>
<td>$(v_2 + v_4)^r$</td>
<td>2540.4980</td>
<td>80/80</td>
<td>0.019/0.019</td>
<td>21/21</td>
</tr>
<tr>
<td>$(v_2 + v_4)^a$</td>
<td>2586.1435</td>
<td>80/80</td>
<td>0.018/0.018</td>
<td>20/19</td>
</tr>
<tr>
<td>$(2v_2 + v_4)^r$</td>
<td>3189.3784</td>
<td>9/0</td>
<td>0.659/</td>
<td>3/0</td>
</tr>
<tr>
<td>$(2v_2 + v_4)^a$</td>
<td>3502.5812</td>
<td>5/0</td>
<td>0.405/</td>
<td>2/0</td>
</tr>
<tr>
<td>$(2v_2 + v_4)^r$</td>
<td>3215.9471</td>
<td>36/36</td>
<td>0.026/0.026</td>
<td>4/4</td>
</tr>
<tr>
<td>$(2v_2 + v_4)^a$</td>
<td>3217.5812</td>
<td>27/27</td>
<td>0.032/0.032</td>
<td>1/0</td>
</tr>
<tr>
<td>$(2v_4)^r$</td>
<td>3240.1535</td>
<td>74/72</td>
<td>0.044/0.039</td>
<td>14/11</td>
</tr>
<tr>
<td>$(2v_4)^a$</td>
<td>3241.5812</td>
<td>71/69</td>
<td>0.031/0.031</td>
<td>8/7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>$(v_1)^s$</td>
<td>3336.1098</td>
<td>40/40</td>
<td>0.017/0.017</td>
<td>6/6</td>
</tr>
<tr>
<td>$(v_1)^a$</td>
<td>3337.0801</td>
<td>3337.0971</td>
<td>40/40</td>
<td>0.008/0.008</td>
</tr>
<tr>
<td>$(v_3)^s$</td>
<td>3443.6359</td>
<td>3443.6294</td>
<td>80/80</td>
<td>0.013/0.013</td>
</tr>
<tr>
<td>$(v_3)^a$</td>
<td>3443.9984</td>
<td>3443.9878</td>
<td>80/80</td>
<td>0.014/0.014</td>
</tr>
<tr>
<td>$(4v_2)^s$</td>
<td>3462.4101</td>
<td>12/0</td>
<td>0.099/</td>
<td>4/0</td>
</tr>
<tr>
<td>$(v_2 + 2v_4^0)^s$</td>
<td>4115.8584</td>
<td>3/0</td>
<td>0.062/</td>
<td></td>
</tr>
<tr>
<td>$(v_2 + 2v_4^0)^a$</td>
<td>4173.0903</td>
<td>11/3</td>
<td>0.036/0.022</td>
<td>1/0</td>
</tr>
<tr>
<td>$(v_2 + 2v_2^2)^s$</td>
<td>4135.8304</td>
<td>24/18</td>
<td>0.094/0.095</td>
<td></td>
</tr>
<tr>
<td>$(v_2 + 2v_2^2)^a$</td>
<td>4192.9435</td>
<td>20/11</td>
<td>0.049/0.035</td>
<td></td>
</tr>
<tr>
<td>$(v_1 + v_2)^s$</td>
<td>4294.5397</td>
<td>40/40</td>
<td>0.035/0.035</td>
<td>6/1</td>
</tr>
<tr>
<td>$(v_1 + v_2)^a$</td>
<td>4320.0052</td>
<td>4320.0306</td>
<td>40/40</td>
<td>0.047/0.047</td>
</tr>
<tr>
<td>$(v_3 + v_2)^s$</td>
<td>4410.9568</td>
<td>4410.9151</td>
<td>76/64</td>
<td>0.028/0.027</td>
</tr>
<tr>
<td>$(v_3 + v_2)^a$</td>
<td>4435.4577</td>
<td>4435.4465</td>
<td>70/56</td>
<td>0.028/0.027</td>
</tr>
<tr>
<td>$(2v_2 + 2v_0^0)^a$</td>
<td>5093.5549</td>
<td>4/2</td>
<td>0.305/0.304</td>
<td></td>
</tr>
<tr>
<td>$(2v_2 + 2v_2^2)^s$</td>
<td>4773.8184</td>
<td>1/0</td>
<td>0.022/</td>
<td></td>
</tr>
<tr>
<td>$(2v_2 + 2v_2^2)^a$</td>
<td>5113.2536</td>
<td>1/1</td>
<td>1.585/1.585</td>
<td></td>
</tr>
<tr>
<td>$(3v_2^2)^r(E)$</td>
<td>4799.2215</td>
<td>1/0</td>
<td>0.015/</td>
<td>1/0</td>
</tr>
<tr>
<td>$(3v_2^2)^a(E)$</td>
<td>4801.4128</td>
<td>1/0</td>
<td>0.038/</td>
<td></td>
</tr>
<tr>
<td>$(3v_2^2)^r(A_2)$</td>
<td>4840.8899</td>
<td>1/0</td>
<td>0.087/</td>
<td></td>
</tr>
<tr>
<td>$(v_1 + v_4)^s$</td>
<td>4955.7216</td>
<td>4955.7561</td>
<td>65/24</td>
<td>0.032/0.031</td>
</tr>
<tr>
<td>$(v_1 + v_4)^a$</td>
<td>4956.8717</td>
<td>39/15</td>
<td>0.038/0.030</td>
<td>5/1</td>
</tr>
<tr>
<td>$(v_1 + 2v_2)^s$</td>
<td>5000.2486</td>
<td>2/2</td>
<td>0.124/0.124</td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>Energy (cm⁻¹)</td>
<td>J/C</td>
<td>A/C</td>
<td>S/C</td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>(ν₃ + ν₄)(^{(A₂)})</td>
<td>5052.0195</td>
<td>5/3</td>
<td>0.032/0.017</td>
<td></td>
</tr>
<tr>
<td>(ν₃ + ν₄)(^{(E)})</td>
<td>5052.6032</td>
<td>58/45</td>
<td>0.043/0.044</td>
<td>2/0</td>
</tr>
<tr>
<td>(ν₃ + ν₄)(^{(A₁)})</td>
<td>5052.6641</td>
<td>3/2</td>
<td>0.029/0.005</td>
<td></td>
</tr>
<tr>
<td>(ν₃ + ν₄)(^{(E)})</td>
<td>5053.2343</td>
<td>26/16</td>
<td>0.063/0.048</td>
<td></td>
</tr>
<tr>
<td>(ν₃ + ν₄)(^{(A₂)})</td>
<td>5067.7243</td>
<td>7/3</td>
<td>0.106/0.069</td>
<td></td>
</tr>
<tr>
<td>(ν₃ + ν₄)(^{(A₁)})</td>
<td>5067.7812</td>
<td>12/8</td>
<td>0.057/0.061</td>
<td></td>
</tr>
<tr>
<td>(4ν₂ + ν₄)(^{(A₁)})</td>
<td>4530.6138</td>
<td>2/1</td>
<td>0.261/0.288</td>
<td></td>
</tr>
<tr>
<td>(4ν₂ + ν₄)(^{(E)})</td>
<td>5104.9370</td>
<td>1/1</td>
<td>0.017/0.017</td>
<td></td>
</tr>
<tr>
<td>(2ν₂ + ν₃)(^{(E)})</td>
<td>5144.9353</td>
<td>8/0</td>
<td>0.205/</td>
<td>2/0</td>
</tr>
<tr>
<td>(2ν₂ + ν₃)(^{(A₁)})</td>
<td>5352.9840</td>
<td>12/0</td>
<td>0.117/</td>
<td></td>
</tr>
<tr>
<td>(ν₁ + 2ν₂ + ν₄)(^{(E)})</td>
<td>5897.8022</td>
<td>1/0</td>
<td>0.004/</td>
<td></td>
</tr>
<tr>
<td>(ν₁ + 2ν₂ + ν₄)(^{(A₁)})</td>
<td>5937/34</td>
<td>0.075/0.078(^i)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ν₁ + 2ν₂ + ν₄)(^{(E)})</td>
<td>5973/31</td>
<td>0.134/0.115(^j)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ν₁ + ν₃)(^{(E)})</td>
<td>67/57</td>
<td>0.129/0.105(^k)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ν₁ + ν₃)(^{(A₁)})</td>
<td>65/60</td>
<td>0.085/0.078(^l)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ν₃ + 2ν₂)(^{(E)})</td>
<td>59/46</td>
<td>0.105/0.095</td>
<td>1/0</td>
<td>0.188/</td>
</tr>
<tr>
<td>(ν₃ + 2ν₂)(^{(A₁)})</td>
<td>62/48</td>
<td>0.078/0.077</td>
<td>1/0</td>
<td>0.048/</td>
</tr>
<tr>
<td>(ν₃ + 2ν₂)(^{(E)})</td>
<td>5/4</td>
<td>0.068/0.058</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ν₃ + 2ν₂)(^{(A₁)})</td>
<td>2/1</td>
<td>0.127/0.164(^m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2ν₁ + ν₄)(^{(E)})</td>
<td>26/18</td>
<td>0.266/0.090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2ν₁ + ν₄)(^{(A₁)})</td>
<td>25/19</td>
<td>0.065/0.052</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2ν₁ + ν₄)(^{(E)})</td>
<td>17/11</td>
<td>0.046/0.052</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(2ν₂)ᵣ

17/11  0.080/0.095

i) I have omitted a state with residual -0.618 cm⁻¹ from the comparison, as C2018 labels it as (4ν₄)ᵣ.
j) I have omitted a state with residual -0.692 cm⁻¹ from the comparison, as C2018 labels it as (ν₁ + 2ν₂ + ν₄)ᵣ.
k) I have omitted a state with residual -0.593 cm⁻¹ from the comparison, as C2018 labels it as (ν₁ + 2ν₂ + ν₄)ᵣ.
l) I have omitted a state with residual -0.573 cm⁻¹ from the comparison, as C2018 labels it as (ν₁ + 2ν₂ + ν₄)ᵣ.
m) I have omitted a state with residual 1.215 cm⁻¹ from the comparison, as C2018 labels it as (4ν₂ + ν₄)ᵣ.
Table 3.7: Vibrational band centre labelling comparisons between different data sources for 6300–7000 cm$^{-1}$ region.

<table>
<thead>
<tr>
<th>Band #</th>
<th>$\Gamma_{\text{tot}}$</th>
<th>$s/a$</th>
<th>This work</th>
<th>HSL-pre3 [115, 247]</th>
<th>MARVEL [3]</th>
<th>Svoboda et. al. [251]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$E^s$</td>
<td>6556.3877</td>
<td>$v_1 + 2v_4^2$</td>
<td>6556.3990</td>
<td>$v_1 + 2v_4^2$</td>
<td>6556.4218</td>
</tr>
<tr>
<td>1</td>
<td>$E''$</td>
<td>6557.9091</td>
<td>6557.9065</td>
<td>6557.9306</td>
<td>6557.9305</td>
<td>6557.9305</td>
</tr>
<tr>
<td>2</td>
<td>$E'$</td>
<td>6666.0662</td>
<td>$v_3 + 2v_4^2$</td>
<td>6666.1946</td>
<td>$v_3 + 2v_4^2$</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>$E''$</td>
<td>6665.7778</td>
<td>6665.7915</td>
<td>–</td>
<td>–</td>
<td>6666.0907</td>
</tr>
<tr>
<td>3</td>
<td>$E'$</td>
<td>6608.7819</td>
<td>$v_1 + v_3$</td>
<td>6608.7773</td>
<td>$v_3 + 2v_4^0$</td>
<td>6608.8218</td>
</tr>
<tr>
<td>3</td>
<td>$E''$</td>
<td>6609.6907</td>
<td>6609.7031</td>
<td>6609.7352</td>
<td>6609.7536</td>
<td>6609.7536</td>
</tr>
<tr>
<td>4</td>
<td>$E'$</td>
<td>6677.4899</td>
<td>$v_1 + v_3$</td>
<td>6677.4125</td>
<td>$v_1 + v_3$</td>
<td>6677.4317</td>
</tr>
<tr>
<td>4</td>
<td>$E''$</td>
<td>6678.1829</td>
<td>6678.1141</td>
<td>6678.3103</td>
<td>6678.3098</td>
<td>6678.3098</td>
</tr>
<tr>
<td>5</td>
<td>$A''_1$</td>
<td>6795.2933</td>
<td>$2v_3^0$</td>
<td>6795.2529</td>
<td>$2v_3^0$</td>
<td>6795.3382</td>
</tr>
<tr>
<td>5</td>
<td>$A''_2$</td>
<td>6796.7741</td>
<td>6796.9054</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>$E'$</td>
<td>6850.2303</td>
<td>$2v_2^3$</td>
<td>6850.1524</td>
<td>$2v_2^3$</td>
<td>6850.2449</td>
</tr>
<tr>
<td>6</td>
<td>$E''$</td>
<td>6850.6830</td>
<td>6850.5680</td>
<td>6850.6550</td>
<td>6850.6557</td>
<td>6850.6557</td>
</tr>
<tr>
<td>7</td>
<td>$E'$</td>
<td>6314.0913</td>
<td>$2v_2 + 3v_4^1$</td>
<td>6313.1898</td>
<td>$2v_2 + 3v_4^1$</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>$E''$</td>
<td>6680.4035</td>
<td>6680.3736</td>
<td>–</td>
<td>–</td>
<td>6683.0364</td>
</tr>
</tbody>
</table>
Figure 3.1: Difference between $J = 0 - 10$ MARVEL term values under 6300 cm\(^{-1}\) and those of C2018 and BYTe.

Figure 3.2: Difference between $J = 0 - 10$ MARVEL term values above 6300 cm\(^{-1}\) and those of C2018 and BYTe. Only MARVEL levels derived from 3 or more transitions are shown.

Our choices regarding basis set, and expansion of the Hamiltonian all con-
tribute to the reproducibility of our results using other nuclear motion programs, meaning our results at higher energies are only reproducible using TROVE with a specific set of model input parameters. Nevertheless, with a converged basis set we expect the only source of variation between programs to be our expansion of the Hamiltonian. At lower energies the reproducibility of our results is illustrated by comparing TROVE calculations with ones performed using the nuclear motion program GENIUSH. Table 3.8 compares the 16 lowest lying vibrational term values calculated using our refined PES in the absence of the Born-Oppenheimer diagonal correction (BODC), in conjunction with TROVE and GENIUSH. There are small discrepancies no greater than 0.281 cm$^{-1}$ which we attribute to i) our truncation of the kinetic energy operator expansion after 6th order; and ii) our re expansion of the potential in terms of linearised coordinates.

As a final testament to the improvement our PES presents, it is worth mentioning its use in computing a hyperfine resolved rotation-vibration line list for $^{14}$NH$_3$ [50]. Comparisons of the predicted saturation dip line shapes with the sub-Doppler spectroscopic measurements ($\nu_1 + \nu_3$ band) by [266], showed that several lineshape features that could not be explained in Refs. [266][294] could be resolved,
owing predominantly to our higher quality wavefunctions.

### 3.4 C2018 line list

In the following section I report two new room temperature line lists for $^{14}$NH$_3$, henceforth referred to as C2018/DMS-B (or simply ‘C2018’) and C2018/DMS-001, which take their names from the PES C2018 they both employ (Section 3.3). The former represents a general-purpose line list for use in any high resolution spectroscopic investigation involving $^{14}$NH$_3$, the latter is designed to complement C2018 for specific wavenumber ranges, and should only be used secondarily. The C2018 line list represents a substantial improvement over BYTe in terms of line position accuracy, frequency coverage, and completeness at room temperature and below. It represents a precursor to a ‘hot’ line list referred to as CoYuTe, that is currently nearing completion. Our line lists have already been used to assign a large number of transitions in the near infrared and the visible, and to identify ammonia absorption features in the visible spectrum of Jupiter.

#### 3.4.1 Overview

The C2018 line list is a complete list of rovibrational transition wavenumbers, Einstein-A coefficients, energy levels and quantum numbers that fully characterises the allowed electric dipole transitions within the ground electronic state of $^{14}$NH$_3$ for wavenumbers in the range 0–19 000 cm$^{-1}$, with a small additional extension to 20 000 cm$^{-1}$. It consists of 312 273 690 transitions between 2 927 322 energy levels up to 23 000 cm$^{-1}$. In computing Einstein-A coefficients and intensities, no threshold has been applied, meaning the only source of missing opacity is due to convergence issues inherent in the variational approach. Both line lists are available to download from the ExoMol website (www.exomol.com), where they are provided in the ExoMol format.[262]

#### 3.4.2 Computational details

The vibrational and rotational basis sets employed here are the same as described in Sections 3.3.1 and 3.3.5. The line list is intended to be complete for temperatures up to 300 K and wavenumbers up to 19 000 cm$^{-1}$, with a small extension
3.4. C2018 line list

to 20 000 cm\(^{-1}\). By consideration of the contribution of the Boltzmann factor \(\exp(-E''/kT)\) to the line intensity in Eq. (2.71), we see it is sufficient to consider only those lower states with energies \(E'' \leq h\nu \cdot 4000\) cm\(^{-1}\), for which the Boltzmann factors are \(> 4 \times 10^{-9}\). By similar arguments, the range of the rotational excitations can be safely limited by \(J = 20\). Allowing for all possible transitions that occur from states with \(E'' \leq 4000\) cm\(^{-1}\) and fall within the wavenumber range \(0 \sim 19 000\) cm\(^{-1}\), the maximum upper state energy must necessarily extend to \(h\nu \cdot 23 000\) cm\(^{-1}\). In the extended range of \(19 000 \sim 20 000\) cm\(^{-1}\) we are omitting the relatively small number of transitions that will occur from lower states within the range \(3000 \sim 4000\) cm\(^{-1}\) to upper states within the range \(23 000 \sim 24 000\) cm\(^{-1}\).

The primitive basis-set was 1 732 500 elements employing a maximum polyad of \(P = 34\) and energy truncation at \(h\nu \cdot 40 000\) cm\(^{-1}\). This resulted in contracted \(J = 0\) wavefunctions with energies up \(h\nu \cdot 30 000\) cm\(^{-1}\). Consequently the vibrational basis set employed here is over double the size of the basis used in construction of the BYTe line list \([302]\). Computation of rotationally excited states results in a \((2J + 1)\) factor increase in the size of the rotational-vibrational basis. Therefore in order to facilitate the calculation of states with \(J\) up to 20 (and eventually \(J = 43\) in the CoYuTe line list) an energy cut-off of \(h\nu \cdot 32 000\) cm\(^{-1}\) was applied to the \((J = 0)\)-contracted basis. This value was chosen as a balance between computational cost and the convergence of energies at \(h\nu \cdot 23 000\) cm\(^{-1}\) as \(J\) was increased. Note that only states with \(E^{J,K}_{\text{rot}} > h\nu \cdot 2000\) cm\(^{-1}\), i.e., for rotational excitations \(J > 13\), will be impacted. The convergence of the \((J = 0)\)-contracted basis with increasing energy cut-off, for \(J = 20\), is displayed in Fig. 3.3. It is necessary to emphasise the more important contribution to the overall accuracy of the rovibrational energies is that of polyad truncation, which we would have ideally extended to \(P_{\text{max}} > 40\).

Nuclear motion calculations were performed using TROVE on the Darwin and COSMOS high performance computing (HPC) facilities in Cambridge, UK. Each of the computing nodes on the Darwin cluster provides 16 CPUs and a maximum of 64 Gb of RAM, with a wall clock limit of 36 hours. COSMOS provides 7.3 Gb per CPU and 8 CPUs per node, with a maximum (standard) job size of 448 Gb
Figure 3.3: Basis set convergence of \(J = 20\) (\(E'\) symmetry) energies as \((J = 0)\)–contracted basis set threshold \(\epsilon\) is increased from 26,000 to 32,000. The difference \(E_{\epsilon=x} - E_{\epsilon=x+2000}\), is displayed for \(x = 26,000, 28,000, 30,000\) vs the energies computed using \(\epsilon = 32,000\) cm\(^{-1}\).

and a wall clock limit of 12 hours. Since multiple nodes can be accessed by a single user at any time, multiple computations could be carried out simultaneously. Our approach to constructing and diagonalising the Hamiltonian matrix for NH\(_3\) in TROVE is the same as used by Underwood et. al. [269] for SO\(_3\), which involves three steps. Firstly the Hamiltonian matrix is calculated and saved to disc. It is then diagonalised separately for each \(J\) and \(\Gamma_{\text{tot}}\) using an MPI-optimized version of the eigensolver PDSYEVD. Finally, TROVE reads the eigenvectors and eigenvalues and converts them into a readable format.

Construction of the Hamiltonian matrices for each \(J\) and \(\Gamma_{\text{tot}}\) was performed on the COSMOS HPC. In total, for states with \(J = 1 - 20\) and symmetry blocks \((\text{A}'_2, \text{A}''_2, E', E'')\) this step took 329 hours real-time (5212 CPU hours), and required a maximum of 221 Gb of RAM for the most expensive calculation, which corresponded to the \(E'\) symmetry block of the \(J = 20\) Hamiltonian matrix. The process was then moved to Darwin for diagonalisation, which took 272 hours (real-time), and for which the largest matrix to be diagonalised \((J = 20, E'\) block) had 222,697 rows (see Fig. 3.4) and required the use of 24 parallel nodes.
3.4. **C2018 line list**

![Graph](image)

Figure 3.4: Dimensions of the E-symmetry matrices (squares) and the corresponding number of eigenvalues below 23 000 cm\(^{-1}\) (circles).

Evaluation of the line strengths (see Eq. (2.53)) and corresponding Einstein-A coefficients was performed using the GAIN-MPI [5] program on the Wilkes2 GPU cluster at Cambridge. Each GPU node contains \(4 \times \text{Nvidia P100 16 GB GPUs}\). With this program we were able to calculate approximately 22 000 transitions per second using up to 10 parallel nodes.

### 3.4.3 Dipole moment surface

Several accurate DMSs currently exist for NH\(_3\). The most well established of these is the surface developed by Yurchenko *et al.* [304], which was used to construct the BYTe line list [302] and is known to be reliable for transition wavenumbers up to 12 000 cm\(^{-1}\). Alternatively, 3 new DMSs have recently been constructed with the intent that one of them be used to construct the new line list for \(^{14}\text{NH}_3\). The 4 DMSs are briefly summarised as follows:

The DMS by Yurchenko *et al.* [304] (referred to henceforth as DMS-B), was computed at the CCSD(T)/aug-cc-pVTZ level of theory in the frozen core approximation using a numerical finite difference procedure with an added external dipole field of 0.005 au. The *ab initio* surface was calculated on a 6-dimensional grid of 50 000 nuclear geometries, all of which were used in the final fit to the analytic ex-
pression given by Eqs. (2.63–2.65). All calculations were performed using Molpro package [287].

The more recently produced surfaces all employed internally contracted MRCI in the full valence reference space comprising 8 electrons in 7 orbitals, with the aug-cc-pwCVQZ basis. The same 6-dimensional grid of 50 000 nuclear geometries as used by [304] (DMS-B) were used although only 10 782 points were computed successfully. Various fits were performed by systematically removing points based on their agreement with the fitted analytic expression, these are henceforth referred to as DMS-001, DMS-0001 and DMS-00001. In the case of DMS-001, the final fit included 9498 points that were reproduced with an unweighted rms error of 0.0009 D.

To determine the best choice of DMS for the C2018 line list, several comparisons were performed. For each DMS a truncated \( J \leq 10 \) line list was generated using TROVE in conjunction with the C2018 PES. To simulate absolute line intensities we use the expression given by Eq. (2.71), as implemented in the Exocross program [300]. Einstein-A coefficients \( A_{if} \) were calculated using the program GAIN, which are related to the linestrength \( S(f \leftarrow i) \) in Exocross [300] through the equation

\[
A_{if} = \frac{64\pi^4}{(4\pi\epsilon_0)^3h(2J_f + 1)} \frac{\nu_{if}^3}{(2J_f + 1)} S(f \leftarrow i),
\]

where \( J_f \) is the final state rotational quantum number, \( h \) is Planck’s constant and \( \nu_{if} \) is the transition wavenumber. The nuclear spin statistical weights \( g_{ns} \) for \( ^{14}\text{NH}_3 \) are [0,12,6,0,12,6] for states of \( [A_1',A_2',E',A_1''',A_2''',E''] \) symmetry, and the selection rules for rovibrational transitions are \( A_1' \leftrightarrow A_1''', A_2' \leftrightarrow A_2''' \) and \( E' \leftrightarrow E'' \). The room temperature partition function was calculated using the C2018 energies list to be \( Q(T = 296K) = 1725.2861 \), which is in good agreement with the value obtained by Sousa-Silva et.al. of 1725.2247 [238].

Line intensities for each of the 4 line lists were compared with the experimental values given in HITRAN 2016 [91] by matching transitions under 7000 cm\(^{-1} \). The intensity ratios \( I_{\text{calc}}/I_{\text{obs}} \) for the main absorption features are displayed in Figs 3.5. DMS-B is consistently seen to outperform the DMS-001, DMS-0001 and DMS-00001 surfaces, with several marked examples being the \( \nu_3 \), \( \nu_2 + \nu_3 \) and...
Figure 3.5: Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 0–1200 cm\(^{-1}\) region.

Figure 3.6: Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 1200–2150 cm\(^{-1}\) region.

\(2\nu_3^0/2\nu_3^2\) bands around 3500, 4400, and 6900 cm\(^{-1}\), respectively.

Additional visual comparisons between HITRAN 2016 and the 4 calculated spectra above 7000 cm\(^{-1}\) showed spurious intensities appearing in the DMS-0001
Figure 3.7: Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 2100–2900 cm$^{-1}$ region.

Figure 3.8: Comparison of line intensities computed using the four test DMSs with those of HITRAN 2016 in the 2900–3700 cm$^{-1}$ region.

and DMS-0001 results, and therefore they were ruled out. However, DMS-001 was seen to reproduce the $\nu_1 + \nu_2 + \nu_3$ and $\nu_2 + \nu_3 + 2\nu_4$ bands around 7600 cm$^{-1}$ significantly more accurately than DMS-B (see Section 3.4.4 for comparisons). It was
therefore decided that DMS-B would be used to construct the general purpose line list, however, due to the novelty of the 7400–11 000 cm\(^{-1}\) region, a complementary line list using the DMS-001 would be generated for use in investigations regarding
the 7400–7800 cm\(^{-1}\) region.

![Graph showing cross-sections calculated using the C2018 PES and DMS-B/DMS-001 dipole moment surfaces. Lines have been convoluted with a Gaussian profile with HWHM= 0.5 cm\(^{-1}\)](image)

Figure 3.11: Cross-sections calculated using the C2018 PES and DMS-B/DMS-001 dipole moment surfaces. Lines have been convoluted with a Gaussian profile with HWHM= 0.5 cm\(^{-1}\)

Beyond 10 400 cm\(^{-1}\) it is difficult to estimate the reliability of the DMS-B and DMS-001 based line lists owing to the lack of data between the regions studied in Refs. [18] and [320]. Furthermore, errors on line positions may be tens of cm\(^{-1}\) for bending and combination bands due to basis set convergence errors and lack of constraint during the refinement. These errors are due to less accurate wavefunctions, which will in turn affect the line intensities. The behaviour of the DMS-B and DMS-001 surfaces above 11 000 cm\(^{-1}\) are compared in Fig. 3.11. Here, both line lists have been convoluted with a Gaussian profile with HWHM of 5 cm\(^{-1}\) and, as before, the maximum rotational excitation included is \(J = 10\), meaning that a substantial contribution to the total opacity is be missing. Experience has shown that there should be an approximately exponential decrease in band intensity as frequency increases. We therefore expect that the number of points used in the DMS-001 fit are too few to accurately represent molecular dipole moment at energies higher than \(hc\cdot12 000\) cm\(^{-1}\), which has resulted in spurious intensities. For this reason full line list calculations are only performed for transitions up to 12 000 cm\(^{-1}\) for DMS-001.
3.4.4 Results and discussion

An overview of the C2018/DMS-B line list \((I > \times 10^{-28} \text{ cm}^{-1}/(\text{molecules cm}^{-2}))\) is presented in Fig 3.12 with HITRAN 2016 included for comparison. Over 2.5 million lines from the C2018 line list are displayed in Fig 3.12 which, when compared to the 65 828 included in HITRAN 2016, illustrates the completeness the line list provides.

Figure 3.12: Overview of the C2018/DMS-B line list compared to HITRAN 2016

An Expansion of each region of strong absorption is compared to HITRAN 2016 in Fig. 3.13 and shows excellent overall agreement. Several strong lines between 4400 and 4450 cm\(^{-1}\) are present in our line lists (and that of BYTe), but not in HITRAN. Absorption peaks that can be attributed to these lines (through comparison with our theoretical cross-sections) also appear in the NH\(_3\) spectra recorded at the Pacific Northwest National Laboratory (PNNL). Therefore they are almost certainly due to missing experimental data rather than an error in our DMSs.

Figure 3.14 displays the ratio of our predicted C2018/DMS-B line intensities
Figure 3.13: Overview of strongly absorbing regions as given in HITRAN 2016 compared to the theoretical predictions of the C2018/DMS-B line list.

\[(I_{\text{calc}} > 1 \times 10^{-26} \text{ cm}^{-1}/(\text{molecule cm}^{-2}))\] to the experimental values taken from HITRAN 2016 for 12862 lines in the 0–7000 cm\(^{-1}\) region. These were obtained by matching lower state energies, transition wavenumbers, \(J\), total symmetry and the sums of vibrational stretching and vibrational bending quanta between C2018 and HITRAN. Lower state term value difference thresholds were set to 0.1 cm\(^{-1}\) and transition wavenumber thresholds were dependent on the energy range based on our Section 3.3.5 comparisons, but ranged from 0.1 – 0.6 cm\(^{-1}\). Only results for DMS-B are shown, however, DMS-1 performed similarly. The experimental uncer-
3.4. C2018 line list

Figure 3.14: Comparison of the predicted C2018/DMS-B line intensities with the experimental values of HITRAN 2016

The uncertainty [91] varies substantially between different sources and we do not consider it here. The majority of lines fall within ±20% of experiment, and as expected, discrepancies between our intensities and those of HITRAN grow as lines get weaker, along with an increase in the general scattering of the intensity ratios. Several partially discernible bow-like structures appear to sprout from the main body of points (e.g., between $3 \times 10^{-24} < I_{\text{HITRAN}} < 1 \times 10^{-23} \text{ cm}^{-1}/(\text{molecule cm}^{-2})$), which may be artefacts which originate from effective Hamiltonian predictions. Such behaviour has been previously observed by Zak et al. [316] for CO$_2$, however, we do not attempt to analyse these further.

Figures 3.15 and 3.16 illustrate the accuracy of the C2018 line list and BYTe in comparison with HITRAN 2016. At wavenumbers below 6300 cm$^{-1}$ there are consistent improvements on line positions due to the C2018 PES reported in Section 3.3. Although BYTe energies are significantly less accurate than those of C2018, this improvement does not translate directly onto the accuracy of the predicted transitions due to the $\Delta J, \Delta K$ transition selection rules. However, in the region 6300–7000 cm$^{-1}$ the improvement is far more striking. In many cases, strong and medium lines present in HITRAN cannot be visually identified in BYTe, whereas matching...
Figure 3.15: Comparison of the C2018/DMS-B and BYTE line lists with HITRAN 2016 for three small windows within the 0–5500 cm$^{-1}$ range.
3.4. C2018 line list

Figure 3.16: Comparison of the C2018/DMS-B and BYTe line lists with HITRAN 2016 for two expansions of the 6300–7000 cm$^{-1}$ region.

between HITRAN and C2018 lines is, in most cases, straightforward. This level of agreement between the three line lists is typical for the entire 6300–7000 cm$^{-1}$ region.

We note that the 5700 – 6200 cm$^{-1}$ region is completely missing from HITRAN. In this region we predict strong absorption by the $\nu_2 + \nu_3 + \nu_4$ band (calculated band centre $s/a = 6012.8563/6036.5254$ cm$^{-1}$, calculated DMS-B band intensity $s/a = 4.841/4.840 \times 10^{-20}$ cm/molecule), the $\nu_1 + \nu_2 + \nu_4$ band ($s/a = 5897.8022/5930.8407$ cm$^{-1}$, $1.929/3.362 \times 10^{-21}$ cm/molecule) and the $(3\nu_2 + \nu_3)^s$ band (5856.0580 cm$^{-1}$, $5.752 \times 10^{-21}$ cm/molecule). A comparison with ($T = 25^\circ$C) spectra recorded at the Pacific Northwest National Laboratory (PNNL) in this region is presented in Fig. 3.17 where our line list has been temperature adjusted and convoluted with a Voigt profile with HWHM= 0.1 cm$^{-1}$ to match the linewidths of the PNNL absorption features. Overall qualitative agreement is very good. Numerical integration of the PNNL cross-sections in the range

...
5700 – 6200 cm\(^{-1}\) yields a value of 4.5338 cm\(^2\)/molecule, which is within 4% of the predicted C2018/DMS-B value of 4.7136 cm\(^2\)/molecule.

![Graph showing cross-sections](image)

Figure 3.17: Synthetic \(J = 0 – 20\) spectrum computed at 298.15 K compared to PNNL for the 5700–6200 cm\(^{-1}\) region.

Recently Vander Auwera and Vanfleteren \[273\] measured line intensities of the 7400–8600 cm\(^{-1}\) region with an estimated accuracy of 10% for most strong lines. Their measurements validated Barton et al.’s \[18, 19\] estimated uncertainty of 15% for lines weaker than \(1 \times 10^{-22} \text{ cm}^{-1}/\text{molecule cm}^{-2}\), but suggested that their uncertainty is significantly more for stronger lines. Our synthetic spectra, calculated using DMS-B and DMS-001, as well as that of BYTe, is temperature adjusted to 293 K and compared with their works in Figures 3.18 and 3.19. Numerical comparisons between our calculated intensities and those of Vander Auwera and Vanfleteren \[273\] for the strongest lines included in a series of assignments that we have performed (detailed in Section 3.5), are also provided in Table 3.9.

From Figures 3.18 and 3.19 we see that qualitative agreement is good and the dominant spectral features are reproduced well by our calculations, although there is a notable reduction in line position accuracy beyond 8000 cm\(^{-1}\) that could undoubtedly be improved with the availability of additional high energy assignments. Nevertheless, across the entire wavenumber range there is a substantial improvement when compared to BYTe. Of the bands covered by our assignments, DMS-001 reproduces intensities of the \(\nu_1 + \nu_2 + \nu_3\) and \(\nu_2 + \nu_3 + 2\nu_4\) bands excellently, often
Figure 3.18: Comparison of the simulated (C2018/DMS-B, C2018/DMS-001 and BYTe) and observed [273] spectra of NH$_3$ at $T = 293$ K for 7400–8600 cm$^{-1}$ region, with expansions of the 7660–7760 cm$^{-1}$ region (middle row) and 8200–8300 cm$^{-1}$ region (bottom row).

To within 10% of the experimental value, but severely underestimates the $\nu_2 + 2\nu_3^2$ band (see Table 3.9). In contrast, DMS-B reproduces the $\nu_2 + 2\nu_3^2$ band somewhat better, but consistently underestimates the $\nu_1 + \nu_2 + \nu_3$ and $\nu_2 + \nu_3 + 2\nu_4$ bands. We suspect that by employing a denser, more extensive grid of geometries in our fit, the underestimated intensities of DMS-001 will be resolved. However, in or-
Figure 3.19: Comparison of the simulated (C2018/DMS-B and BYTe) and observed [18] spectra of NH$_3$ at $T = 296$ K for 9000–10400 cm$^{-1}$ region, with expansions of the 9295–9329.5 cm$^{-1}$ region (first row from top), the 9720–9785 cm$^{-1}$ region (first row from bottom), and the 10080–10125 cm$^{-1}$ region (bottom row).
3.5 Assignment of the 7400-8000 cm\(^{-1}\) region

Currently the \(^{14}\text{NH}_3\) entry in the HITRAN 2016 database extends to 10300 cm\(^{-1}\), and contains 20 526 lines above 7000 cm\(^{-1}\). These are taken solely from the work by Barton et al. \cite{18,19}, who also assigned over 3000 transitions spanning 27 vibrational bands in the 7400–9900 cm\(^{-1}\) range employing a method using BYTe.
combination differences and the method of branches [198].

Our fits reported in Section [3.3] raised a number of issues with the current set of assignments in the near-infrared that are present in HITRAN. Furthermore, many strong lines remain unassigned due to the inaccuracy of BYTe line positions. Our improved PES and accompanying line lists should facilitate line assignments in the near-infrared region, and therefore decided to re-analyse the observed spectrum in the 7400-8000 cm$^{-1}$ region.

### 3.5.1 Ground state combination differences

769 lines from the 7400–8000 cm$^{-1}$ region of the Barton et al. line list [19] were assigned using ground state combination differences (GSCDs). The method involves first performing tentative assignments based on a visual comparison between experimental and theoretical line lists. Theoretical upper state energies of each assignment are then replaced with empirical values calculated by adding the MARVEL lower state energies to the observed line positions. A tentative assignment is then considered validated if one or more additional isolated transitions sharing the same upper state (as the tentative assignment) can then be identified in both theoretical and experimental line lists to within the measurement uncertainty.

Our tolerance for accepting GSCD partners was ±0.003 cm$^{-1}$, which was the same as used by Barton et al. [19] for their unblended assignments in this region. For line intensities, the ratio $I_{\text{obs}}/I_{\text{calc}}$ for the GSCD partners were required to fall within 2/3 and 3/2 times the ratio of $I_{\text{obs}}/I_{\text{calc}}$ for the manual assignment. This accounted for the variation in accuracy between different vibrational bands, in particular the underestimated $\nu_2 + 2\nu_3^2$ band discussed in Section [3.4.4].

Tentative assignments were performed using the online assignment tool Spectropedia [60]. As a measure of confidence, each tentative assignment was placed into one of three categories based on our confidence in the assignment, which generally depended on the individual line strength, how congested the spectral region was, if a PQR-branch pattern was visible etc. For the least reliable assignments (Flag C in Table [3.10]), three additional GSCD partners were required for it to be considered validated (i.e., 4 lines in total). More reliable tentative assignments required either one or two ad-
3.5. Assignment of the 7400-8000 cm\(^{-1}\) region

ditional GSCD partners (Flags A and B in Table 3.10), one being reserved only for strong, isolated lines. After applying these validation criteria only 284 of our 827 initial hand assignments were accepted into the final list. GSCD partners could be found for more transitions but these did not fulfil our minimum requirements to be retained.

Even despite stringent measures, the possibility of false partners must be considered. For weak lines with intensity \(I < 5.0 \times 10^{-24} \text{ cm}^{-1}/(\text{molecule cm}^{-2})\), there is an average of 0.022 lines per 0.006 cm\(^{-1}\) interval. Assuming a Poisson distribution, this translates to roughly a 1 in 47 chance of such a match being false. This probability reduces significantly for medium and strong lines (\(I > 5.0 \times 10^{-24} \text{ cm}^{-1}/(\text{molecule cm}^{-2})\)) that contribute \(\sim 2/3\) of our overall GSCD partners. For our assignments the overall standard deviation of our derived upper state term values is 0.0009 cm\(^{-1}\), and only in a few cases does the range of derived upper state term values within a GSCD set exceed 0.003 cm\(^{-1}\). This, combined with the additional assurances provided by intensity comparisons, practically eliminates the possibility of fortuitous matches.
Table 3.10: Examples of the GSCD process for 6 different derived upper states. Upper state term values $E'_{\text{obs}}$ and $E'_{\text{calc}}$, observed line positions $v_{\text{obs}}$ and the difference between the observed and calculated line positions $v_{\text{o-c}}$, are all given in units of cm$^{-1}$. Units of intensity are cm$^{-1}$/(molecule cm$^{-2}$). $\langle E'_{\text{obs}} \rangle$ is the averaged experimental term value.

<table>
<thead>
<tr>
<th>Flag (J′K′t′)</th>
<th>(J″K″t″)</th>
<th>Band</th>
<th>$v_{\text{obs}}$</th>
<th>$v_{\text{o-c}}$</th>
<th>$E'_\text{obs}$</th>
<th>$E'_\text{calc}$</th>
<th>$\langle E'_{\text{obs}} \rangle$</th>
<th>$I_{\text{obs}}$</th>
<th>$I_{\text{obs}}/I_{\text{calc}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (2,1,a)</td>
<td>(3,2,a)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7622.9614</td>
<td>0.0503</td>
<td>7728.1451</td>
<td>7728.0948</td>
<td>7728.1452</td>
<td>8.150×10$^{-23}$</td>
<td>0.970</td>
</tr>
<tr>
<td>A (2,1,a)</td>
<td>(2,2,a)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7682.5580</td>
<td>0.0505</td>
<td>7728.1453</td>
<td>7728.0948</td>
<td>7728.1452</td>
<td>4.911×10$^{-23}$</td>
<td>0.905</td>
</tr>
<tr>
<td>B (3,2,a)</td>
<td>(4,1,a)</td>
<td>$v_1 + 2v_2 + 2v_4^0$</td>
<td>7506.2948</td>
<td>0.0673</td>
<td>7701.9061</td>
<td>7701.8388</td>
<td>7701.9061</td>
<td>8.359×10$^{-24}$</td>
<td>1.094</td>
</tr>
<tr>
<td>B (3,2,a)</td>
<td>(3,1,a)</td>
<td>$v_1 + 2v_2 + 2v_4^0$</td>
<td>7585.6281</td>
<td>0.0675</td>
<td>7701.9063</td>
<td>7701.8388</td>
<td>7701.9061</td>
<td>2.334×10$^{-23}$</td>
<td>0.989</td>
</tr>
<tr>
<td>B (3,2,a)</td>
<td>(2,1,a)</td>
<td>$v_1 + 2v_2 + 2v_4^0$</td>
<td>7645.1972</td>
<td>0.0676</td>
<td>7701.9064</td>
<td>7701.8388</td>
<td>7701.9061</td>
<td>4.199×10$^{-23}$</td>
<td>1.096</td>
</tr>
<tr>
<td>C (6,1,a)</td>
<td>(5,1,a)</td>
<td>$v_1 + 2v_2 + 2v_4^0$</td>
<td>7757.0252</td>
<td>0.1115</td>
<td>8050.9935</td>
<td>8050.8820</td>
<td>8050.9935</td>
<td>1.440×10$^{-23}$</td>
<td>1.012</td>
</tr>
<tr>
<td>C (6,1,a)</td>
<td>(7,1,s)</td>
<td>$v_1 + 2v_2 + 2v_4^0$</td>
<td>7500.2346</td>
<td>0.1112</td>
<td>8050.9932</td>
<td>8050.8820</td>
<td>8050.9935</td>
<td>3.149×10$^{-24}$</td>
<td>1.223</td>
</tr>
<tr>
<td>C (6,1,a)</td>
<td>(6,1,s)</td>
<td>$v_1 + 2v_2 + 2v_4^0$</td>
<td>7638.3694</td>
<td>0.1117</td>
<td>8050.9937</td>
<td>8050.8820</td>
<td>8050.9935</td>
<td>1.842×10$^{-24}$</td>
<td>0.834</td>
</tr>
<tr>
<td>C (6,1,a)</td>
<td>(5,2,a)</td>
<td>$v_1 + 2v_2 + 2v_4^0$</td>
<td>7767.3768</td>
<td>0.1114</td>
<td>8050.9934</td>
<td>8050.8820</td>
<td>8050.9935</td>
<td>9.473×10$^{-24}$</td>
<td>0.916</td>
</tr>
<tr>
<td>A (1,0,s)</td>
<td>(2,1,s)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7620.2164</td>
<td>0.0830</td>
<td>7676.1551</td>
<td>7676.0721</td>
<td>7676.1550</td>
<td>5.574×10$^{-23}$</td>
<td>0.883</td>
</tr>
<tr>
<td>A (1,0,s)</td>
<td>(1,1,s)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7659.9820</td>
<td>0.0829</td>
<td>7676.1550</td>
<td>7676.0721</td>
<td>7676.1550</td>
<td>6.581×10$^{-23}$</td>
<td>0.946</td>
</tr>
<tr>
<td>B (7,2,s)</td>
<td>(8,3,s)</td>
<td>$v_2 + 2v_3^0$</td>
<td>7693.5363</td>
<td>0.1174</td>
<td>8372.8243</td>
<td>8372.7069</td>
<td>8372.8252</td>
<td>4.247×10$^{-24}$</td>
<td>2.133</td>
</tr>
<tr>
<td>B (7,2,s)</td>
<td>(7,3,s)</td>
<td>$v_2 + 2v_3^0$</td>
<td>7851.2037</td>
<td>0.1188</td>
<td>8372.8257</td>
<td>8372.7069</td>
<td>8372.8252</td>
<td>8.787×10$^{-24}$</td>
<td>1.713</td>
</tr>
<tr>
<td>B (7,2,s)</td>
<td>(6,3,s)</td>
<td>$v_2 + 2v_3^0$</td>
<td>7989.5070</td>
<td>0.1186</td>
<td>8372.8255</td>
<td>8372.7069</td>
<td>8372.8252</td>
<td>6.677×10$^{-24}$</td>
<td>2.026</td>
</tr>
<tr>
<td>C (6,2,s)</td>
<td>(7,1,s)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7498.2354</td>
<td>-0.0374</td>
<td>8048.9940</td>
<td>8049.0314</td>
<td>8048.9936</td>
<td>3.656×10$^{-24}$</td>
<td>0.825</td>
</tr>
<tr>
<td>C (6,2,s)</td>
<td>(7,4,a)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7552.3180</td>
<td>-0.0374</td>
<td>8048.9940</td>
<td>8049.0314</td>
<td>8048.9936</td>
<td>3.855×10$^{-24}$</td>
<td>1.050</td>
</tr>
<tr>
<td>C (6,2,s)</td>
<td>(6,1,s)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7636.3684</td>
<td>-0.0387</td>
<td>8048.9927</td>
<td>8049.0314</td>
<td>8048.9936</td>
<td>1.187×10$^{-23}$</td>
<td>0.919</td>
</tr>
<tr>
<td>C (6,2,s)</td>
<td>(5,1,s)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7755.0233</td>
<td>-0.0398</td>
<td>8048.9916</td>
<td>8049.0314</td>
<td>8048.9936</td>
<td>2.974×10$^{-23}$</td>
<td>1.208</td>
</tr>
<tr>
<td>C (6,2,s)</td>
<td>(5,2,a)</td>
<td>$v_1 + v_2 + v_3$</td>
<td>7765.3791</td>
<td>-0.0357</td>
<td>8048.9957</td>
<td>8049.0314</td>
<td>8048.9936</td>
<td>1.015×10$^{-24}$</td>
<td>1.122</td>
</tr>
</tbody>
</table>
3.5. Assignment of the 7400-8000 cm\(^{-1}\) region

3.5.2 Assignments and derived upper state energies

A summary of our final list of assignments is presented in Table 3.11. The empirical upper state energy of each GSCD set was calculated by averaging the observed line positions plus the MARVEL lower state energies. In total we assigned rovibra-
tional quanta for 769 transitions and upper state energies for 284 levels, spanning
an estimated 11 vibrational bands. Vibrational labels are taken from the leading
coefficient basis set contributor in our variational calculation, which were as low as
0.22 for some bands listed in Table 3.11, and indeed lower for the 2\(v_1 + v_2\) band
which was highly mixed. Rotationally excited states may also possess significantly
smaller leading coefficients. Thus, all vibrational labelling should be viewed as ten-
tative. Figure 3.20 compares the term values calculated using our new PES to those
empirically derived from our assignments. Residuals are seen to have a system-
atric dependence on rotational quanta, although this is partly obscured by mixing
between states in the \(v_1 + v_2 + v_3^1\) and \(v_1 + v_2 + 2v_4^0\) bands, possibly due to our
heavy reliance on these bands during the refinement. This systematic behaviour,
apart from highlighting deficiencies in the C2018 PES, further reassures us of the
reliability of our assignments.

Table 3.11: An overview of the assignments from this work. Only the \(E\)-symmetry
band origin for the \(v_2+v_3+2v_2^2\) band is shown, although our assignments include \(A_1\)
and \(A_2\)-symmetry vibrational states as well.

<table>
<thead>
<tr>
<th>Band</th>
<th>(s/a)</th>
<th>Center (cm(^{-1}))</th>
<th>(N_{\text{lines}})</th>
<th>(N_{\text{eners}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(v_1+v_2+v_3^1)</td>
<td>(s)</td>
<td>7656.6402</td>
<td>160</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>7673.4138</td>
<td>137</td>
<td>50</td>
</tr>
<tr>
<td>(v_2+v_3+2v_4^0)</td>
<td>(s)</td>
<td>7567.8000</td>
<td>90</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>7597.5458</td>
<td>98</td>
<td>36</td>
</tr>
<tr>
<td>(v_2+v_3+2v_4^2)</td>
<td>(s)</td>
<td>7605.9881((E))</td>
<td>35</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>7640.2230((E))</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>(v_1+v_2+2v_4^2)</td>
<td>(s)</td>
<td>7484.6131</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>7525.9480</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>(v_2+2v_3^2)</td>
<td>(s)</td>
<td>7854.3016</td>
<td>99</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>7864.1066</td>
<td>98</td>
<td>40</td>
</tr>
<tr>
<td>(2v_1+v_2)</td>
<td>(s)</td>
<td>7575.3439</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 3.20: Agreement between energy levels derived from our assignments and the values predicted by the C2018 energies list for 6 vibrational bands. Differences between the observed and calculated term values, $E_{\text{obs}} - E_{\text{calc}}$, are given in units of cm$^{-1}$.

3.5.3 Discrepancies with HITRAN

There is significant overlap between our work and that of Barton et al. [19] who previously assigned 230 of the lines in our current analysis. 177 of these, all assigned by combination difference (CD) bar one, agree with our assignments in at least $J$ and total symmetry, and so we consider these assignments validated. The 53 that disagree consist of 9 unblended CD lines, 17 blended CD lines, and 27 assigned using the method of branches [198]. The method of branches allows for
small systematic differences between theory and experiment within a vibrational band with increasing $J$. However, it fails if this dependency is large and there are gaps in experimental values which act to indicate the value of the next (in $J$ and $K$) $E_{\text{obs}} - E_{\text{calc}}$ discrepancy. Therefore occasional missassignments are not surprising.

For the 9 disagreeing CD states, 7 were previously assigned to the $\nu_1 + \nu_2 + 2\nu_4^2$ band for which BYTE reproduces the Kitt Peak spectral features particularly poorly.

Figure 3.21 shows absorption cross-sections of a sample CD pair that were reassigned during our analysis, in this case the experimental line at 7612.6690 cm$^{-1}$ (cross) was reassigned from $E''(3,3)^s \leftarrow E'(2,2)^s$ (hollow square) to $A''_2(1,1)^a \leftarrow A''_2(0,0)^a$ (filled circle), and its partner line at 7553.0483 cm$^{-1}$ (cross) was reassigned from $E''(3,3)^s \leftarrow E'(3,2)^s$ (hollow square) to $A''_2(1,1)^a \leftarrow A''_2(2,0)^a$ (filled circle). Here, our notation corresponds to $\Gamma_{\text{tot}}(J,K)^{s/a}$. For the purpose of clarity in Figure 3.21 lines are convoluted with a Gaussian profile, HWHM=0.06 cm$^{-1}$.

Whilst our assignments account for only 47.5% of the summed intensity for this region, our aim was not complete assignment. Rather, to demonstrate the predictive power of our line list in this newly charted region, and extract a reliable list of energies that can be used to refine future \textit{ab initio} calculations. We expect our line list to also be useful for assignments up to 10 400 cm$^{-1}$, although we have made no attempt to do so here.
3.6 Measurement of the 7169–7195 cm\(^{-1}\) region

High purity NH\(_3\) is the main nitrogen-bearing precursor used in the production of nitride based semi-conductors. Blue light emitting diodes and laser diodes, for which there is currently a high demand \cite{172, 173}, and which are based on GaN systems, rely on high purity NH\(_3\) that is free from trace impurities such as moisture at a ppb level. These impurities can greatly reduce device yield, which has economic and environmental consequences, and means that NH\(_3\) purity measurements are extremely important \cite{23, 291}. Diode-laser based sensors used for such applications commonly focus on the so-called ‘telecoms window’ due to an abundance of cheap laser diodes that were originally developed for use over wavelength ranges where optical fibers have small transmission losses. This low-loss region extends from 1260 nm to 1625 nm, which coincides with several strong absorption bands of H\(_2\)O, and with the 1.3 \(\mu\)m and 1.5 \(\mu\)m absorption bands of NH\(_3\). This, combined with other desirable features \cite{291}, means that these laser diodes are a popular choice for both NH\(_3\) and H\(_2\)O monitoring in industrial processes (e.g. Refs. \cite{111, 202, 246, 279, 292}).

Servomex is the leading UK gas analytics company who develop a wide variety of gas analysers for industrial use, including NH\(_3\) analysers for ammonia slip applications and trace multigas analysers for trace measurements. A particular interest of theirs is trace moisture detection in high purity NH\(_3\) at 1392.5(±2.5) nm, where they possess a tunable diode-laser. This wavelength falls into the E-band of the telecom window, and is commonly used for trace moisture detection owing to the strong 1392.5335 nm H\(_2\)O line \cite{111}. On the other hand, NH\(_3\) is almost completely uncharacterised at this wavelength, with data lacking from HITRAN and PNNL, meaning trace measurements of H\(_2\)O in high purity NH\(_3\) at this wavelength may be unknowingly corrupted by NH\(_3\) absorption.

The aim of this section, therefore, is to investigate NH\(_3\) absorption in the 1392 nm region, and its possible interference in trace moisture detection. From predictions by the C2018 line lists, NH\(_3\) absorption here is known to be extremely weak, with line intensities of order 1 \(\times\) 10\(^{-25}\) cm\(^{-1}\)/(molecule cm\(^{-2}\)) at most. How-
ever, moisture contamination must be detectable on a ppb level to be useful, and so the \( \text{H}_2\text{O} \) lines are likely to be even weaker. We therefore aim to characterise the strongest \( \text{NH}_3 \) lines that are likely to interfere most severely with measurements of moisture concentration. These \( \text{NH}_3 \) lines may also be useful in performing dual \( \text{NH}_3\text{–H}_2\text{O} \) measurements in future. As far as we are aware, this is the first \( \text{NH}_3 \) focussed study of this region.

All experiments reported here were performed at the Servomex headquarters in Crowborough UK, under the supervision of Dr. Richard Kovacich.

### 3.6.1 Overview

In this section I report a survey of \( \text{NH}_3 \) spectra in the 1390–1395 nm region, to this author’s knowledge for the first time, using second harmonic wavelength modulation spectroscopy employing a system based on a modified commercial laser gas analyser (Servomex Laser 3 plus) tunable laser diode.

Tunable laser diodes are lasers whose output wavelength may be adjusted, i.e. tuned, to discrete or continuous wavelength values within a specified range. This tuning is often achieved via changing the ambient temperature of the laser cavity using a thermoelectric cooler, and/or by changing the injection current of the laser diode itself. The combination of refractive index and bandgap energy changes within the gain medium due to heating, and semiconductor band filling effects due to the injection current, act together to modify the lasing wavelength. A discussion of the exact physical mechanisms for this are beyond the scope of this thesis, but the interested reader is directed to Ref. [86] for more information. Our system utilises both temperature and current based tuning of the wavelength to provide continuous coverage of the 1390 – 1395 nm region.

Our method is as follows: Coarse control of the laser wavelength is obtained by adjusting the external laser temperature, which, although a slow method of tuning the laser and generally inadequate for practical instrumentation and low frequency noise rejection, does allow wide tuning of the laser diode wavelength, and a method of exploiting this is discussed in this chapter. Once tuned to the feature of interest, much faster and finer tuning is obtained by the laser diode injection current, which
Chapter 3. Ammonia (\(^{14}\text{NH}_3\))

is tuned with a ramp waveform that in turn ramps the laser intensity and the laser wavelength across the absorption feature to record the line shape in detail.

In wavelength modulation spectroscopy an additional sinusoidal waveform is superimposed over the DC ramp current, which is synchronously detected by a lock-in amplifier. This significantly improves the noise rejection allowing highly sensitive measurement of weak gas absorption lines. A complete scan of the 1390 – 1395 nm range thus consists of repeatedly stepping the laser diode temperature to obtain the necessary wavelength tuning range, and at each step ramping the injection current across the absorption feature. The individual scans are then analysed and combined to form a complete picture of the spectrum. Calibration of the wavelength change with laser tuning was performed by separately directing the beam through an etalon glass, producing Fabry-Pérot interference peaks in the transmitted signal. Absolute absorption peak positions were then deduced by comparing \(\text{H}_2\text{O}\) features in the recorded spectrum to known \(\text{H}_2\text{O}\) lines in HITRAN. From our investigation, which is outlined in the following sections, a large number of \(\text{NH}_3\) absorption features were observed and several apparently isolated line centres of weak \(\text{NH}_3\) transitions are derived with an estimated accuracy of 0.05 cm\(^{-1}\).

### 3.6.2 Experimental setup

The light source was a tunable laser diode emitting at 1390 – 1395 nm, with a temperature tuning rate of approximately 0.1 nm/K operating at laser temperatures between 267.6 and 322.6 K. Fixed 2 mA amplitude, frequency modulation at 50 kHz was applied to the laser diode injection current, and the second harmonic signal was isolated using a lock-in amplifier system that is part of the gas analyser. For relative wavelength calibration, an etalon glass with free-spectral-range (FSR) of 0.0301 nm at 1392 nm (0.1555 cm\(^{-1}\)) was used. The procedure used to determine absolute wavelength calibration is described in Section 3.6.3. The average optical power output of the TLDS was 11 mW at approximately 70 mA, with an efficiency of 0.2 mW/mA and a lasing threshold current which increased by 0.2 mA/K.

Ammonia gas samples were available at 3% concentration with \(\text{N}_2\) buffer gas, as well as a more limited supply at > 99% concentration from a commercial \(\text{NH}_3\).
3.6. Measurement of the 7169–7195 cm\(^{-1}\) region

To reduce moisture contamination that is within the NH\(_3\) gas mixture bottles, these were passed through a calcium-oxide scrubber, and spaces outside the gas cell were purged with high purity nitrogen to prevent spectral interference from atmospheric moisture. Measurement gas samples were contained in a 1.019 m heated cell capable of temperatures up to 120°C.

NH\(_3\) is a highly toxic gas, and for safety reasons the experiment was housed in a fume cupboard with a solenoid operated gas valve to shut off the gas supply if the fume cupboard air flow stops. In addition the laboratory has an ambient NH\(_3\) sensor that is connected to a building alarm and external shut-off valve to the bottle, which is kept outside the building. To reduce the effects of pressure broadening and isolate the absorption features as much as possible, gas pressure was set to 0.1 bar for all measurements, which was monitored using a pressure gauge (GE Druck PDCR 330) and seen to be stable to within 1 mbar.

3.6.3 Calibration

The 2000-point resolution temporal photodetector output for a single scan covering the strong H\(_2\)O line at 1392.5335 nm (7181.1558 cm\(^{-1}\)), (occurring at scan point number \(\sim 935\)), and the corresponding (unfiltered) etalon signal are shown in the top and bottom-left panels respectively of Figure 3.22. A laser modulation ‘burst’ occurs during the first 700 scan points of the \(2f\) signal, which is normally used for compensation of beam obscuration that occurs in practical field measurements [136]. However, this is discarded in the experiment as the beam alignment is stable and unobscured, and the modulation burst cannot be located over an absorption line. The raw etalon signal (blue) of each scan was smoothed (orange), and each peak of the smoothed signal (circled) then indicates a wavelength change of \(1 \times FSR\). Note that any background signals should be filtered out of the \(2f\) signal by default. A function to relate the remaining scan points to laser wavelength change is obtained by first fitting a linear expression to the change in wavelength over the last 4–5 etalon fringe cycles, which occurs approximately linearly. An exponential term is then fitted to the residuals, which accounts for the initial non-linearity of the laser ramp. The linear and exponential function fits for the laser sweep over the strong
Chapter 3. Ammonia ($^{14}\text{NH}_3$)

Figure 3.22: Photodetector output for a single scan covering the 1392.5335 nm H$_2$O line (top), and the corresponding etalon signal (left) and tuning rate function fit (right).

H$_2$O line at 1392.5335 nm (7181.1558 cm$^{-1}$) are shown in the bottom-right panel of Figure 3.22. Summing these components gives a function of the form:

$$\Delta\lambda(n; T) = c_0 + c_3n + c_1e^{c_2n},$$

(3.4)

where $n = 1, 2, 3, \ldots, 2000$ is the scan point number, and the coefficients $c_0, c_1, c_2$ and $c_3$ are functions of laser (tuning) temperature and must be determined individually for each of the 56 scans that were needed to cover the complete 1390–1395 nm range. They are given as a function of laser temperature in Figure 3.23. Coefficient $c_3$ relates to the effective laser current tuning rate $\delta\lambda/\delta I$, which increases with laser temperature. The sawtooth pattern visible for $c_0$ occurs because the wavelength change per temperature step of 1 K is not an integer multiple of etalon fringe FSR. This must be adjusted to a ramp ($c_4$ in Figure 3.23) before being inserted into
3.6. Measurement of the 7169–7195 cm\(^{-1}\) region

Eq. (3.4). Note that the variation in coefficients \(c_1\) and \(c_2\) is small, and the linear fit residuals are in all cases < 0.005 nm, which corresponds to laser temperature control error of < 0.005 K.

![Graphs showing coefficients](image)

Figure 3.23: Tuning rate coefficients used in Eq. (3.4)

Once the scans had been linearised in \(\delta \lambda (n;i)\) (where the scan index \(i = 1,2,\ldots,56\) for laser temperatures \(T = 267.6, 268.6,\ldots, 322.6\) K), they were stitched together using the function

\[
\Delta \lambda (n;i) = \delta \lambda (n;i) + 0.0903(i - 1)a(i),
\]

(3.5)

where \(\Delta \lambda (n;i)\) is the wavelength of the \(i^{th}\) scan at scan point \(n\), relative to the first
scan \((i = 1)\), and \(a(i)\) is a small factor between 0.996 and 1.005. This function was obtained by attempting to overlap any shared absorption features that were present in neighbouring scans.

A final adjustment to compensate for the decrease in laser power output with increasing laser temperature was attempted, but ultimately discarded, due to the resulting amplification of the small amount of baseline drift which would also need to be corrected for. This power decrease occurs because laser diode threshold current is temperature dependent, and is as large as 21\% over the 55 K laser temperature range. Corrective treatment would have been necessary to employ algorithms used to infer absolute line strength \([99]\). However, to do so successfully would require more accurate modelling of modulation broadening and signal filtering, and given all the uncertainties of this measurement, it is not possible to determine the absolute line strength with any accuracy.

### 3.6.4 Simulation of H₂O second harmonic spectrum

The problem of moisture contamination from either atmospheric signal, or test gas mixtures impurities, is ubiquitous in high-resolution spectroscopy. Fortunately, however, H₂O line positions and intensities are well characterised, and an essentially complete line list for this region is available from the HITRAN 2016 database. To distinguish features of NH₃ absorption from those of H₂O a second harmonic model H₂O spectrum was generated using lines taken from the HITRAN database. This serves a secondary purpose in calibrating the absolute wavelength of the combined laser wavelength scans.

A general analytical expression for the second harmonic signal of a Lorentzian line shape, as a function of modulation amplitude, is given by \([10]\). Its ability to reproduce the experimentally measured values of fundamental 2\(f\) signal parameters, such as signal amplitude and signal height, are compared to Gaussian and Voigt lineshape based models, which have the disadvantage of requiring numerical integration, in Ref. \([207]\). Even at pressures as low as 0.02 bar the Lorentz approximation was found to be sufficient in all cases where \(m \geq 1\), which fall well within
3.6. Measurement of the 7169–7195 cm\(^{-1}\) region

our experimental parameterisation. The analytical expression \cite{10} is

\[ H_L^2(x, m) = \frac{4}{m^2} - \frac{\sqrt{2}}{m^2} \frac{(M + 1 - x^2)[(M^2 + 4x^2)^{1/2} + M]^{1/2} + 4x[M^2 + 4x^2]^{1/2} - M^{1/2}}{(M^2 + 4x^2)^{1/2}}, \]  

(3.6)

where the modulation index \(m\), normalised grid \(x\) and parameter \(M\) are defined as

\[ m = \frac{a}{\Delta \nu}, \]  

(3.7)

\[ x = \frac{\bar{\nu} - \nu_0}{\Delta \nu}, \]  

(3.8)

\[ M = 1 - x^2 + m^2, \]  

(3.9)

and \(a\) is the modulation amplitude, \(\Delta \nu\) is the line HWHM, \(\bar{\nu}\) is the grid of frequencies upon which \(H_L^2(x, m)\) is evaluated, and \(\nu_0\) is the line centre of a particular absorption feature. The unknowns \(m\) and \(\Delta \nu\) can be derived using methods that relate them to characteristic features of the measured 2\(f\) signal as the modulation amplitude \(a\) is varied. Note, however, that this simple model ignores imperfections in a practical WMS system like residual amplitude modulation of the laser diode \cite{221} and signal filtering. Therefore the model from Ref. \cite{10} is used only to give an estimation here.

We used the strong H\(_2\)O line (in N\(_2\) buffer gas) at 1392.5335 nm (7181.1558 cm\(^{-1}\)) to derive the parameters \(m\) and \(\Delta \nu\). The second harmonic signal is measured for a range of modulation current amplitudes 0.5 \(\leq a \leq 10\) (shown Figure 3.24 bottom panel), which are assumed to be proportional to the current modulation amplitude \(A\) applied to the laser diode \cite{207}. For an isolated line, \(m\) is therefore proportional to \(A\). It is well known that as \(m\) is varied, the maximum 2\(f\) central peak signal amplitude \(g\) occurs at \(m \approx 2\). Therefore by incrementally increasing \(A\) and measuring \(g\) for each value, \(m\) can be directly related to \(A\) as \(\frac{m}{A} = \frac{m_{\text{peak}}}{A_{\text{peak}}}\). Figure 3.24 (top panel) shows the measured values of \(g\) as a function of \(A\) for the strong H\(_2\)O line at 1392.5335 nm, where the maximum \(g\) occurs at \(A \approx 4\) mA. For our current modulation amplitude of 2 mA, this corresponds to a modulation index of \(m = 1.1\).
Figure 3.24: Second harmonic signal of the 1392.5335 nm H$_2$O line measured for a range of current modulation amplitudes (right), and the corresponding central peak maxima as a function of modulation amplitude (left).

Figure 3.25: Comparison of the measured and synthetic 2f H$_2$O spectra using the determined values of linewidth and modulation index.
3.6. Measurement of the 7169–7195 cm$^{-1}$ region

To deduce the effective line width $\Delta \nu$, consider the numerical relationship derived by Henningsen [99]

$$\frac{\Delta \nu_{2f}}{\Delta \nu} = (0.89m^2 + 1)^{1/2},$$

(3.10)

where $\Delta \nu_{2f}$ is the $2f$ signal width between the central peak maximum and one of the side-lobe minima. For a modulation amplitude of 2 mA ($m = 1.1$) Equation (3.10) evaluates to 1.44, and the signal width $\Delta \nu_{2f}$ is measured to be 0.0912 cm$^{-1}$, where we have used 0.0008 cm$^{-1}$ per scan point as derived from our calibration in Section 3.6.3. Using these values the line width $\Delta \nu$ is determined to be 0.063 cm$^{-1}$, which is slightly over double the value obtained by numerical integration of the first derivative signal (0.029 cm$^{-1}$) and the values given in HITRAN. This is most likely due to our filtering process, where we use a convolution filter rather than a simple low pass filter commonly used. Therefore some modification of the signal amplitude may occur due to changing signal width modulation.

Figure 3.25 compares the measured and synthetic $2f$ H$_2$O spectra in the 1392.2 – 1392.6 nm region, where there are three strong water lines. Clearly the synthetic H$_2$O $2f$ spectra reproduce shape and relative intensities of the measured second harmonic signals very well.

3.6.5 Dilute NH$_3$ spectrum

A scan of the 1392 nm region was first performed using 3% ammonia in N$_2$ buffer gas, at 21$^\circ$C and 0.1 bar. After calibration the total wavelength change was scaled by a factor of 0.986 so that the relative H$_2$O peak positions better matched those of the synthetic $2f$ H$_2$O spectra. The prominent 1392.5335 nm (7181.1558 cm$^{-1}$) H$_2$O line (shown Fig. 3.25) was used as the reference point from which our absolute laser diode wavelength was determined, and the measured and synthetic $2f$ signals were scaled so that the 1392.5335 nm peak maximum had an amplitude of 1. All signal amplitudes are therefore given as fractions of the 1392.5335 nm line peak.

The full 1392 nm region scan of the 3% NH$_3$ sample is presented in the middle panel of Figures 3.27 and 3.28. Peaks that fell within 0.05 cm$^{-1}$ of a known H$_2$O
line, and qualitatively agreed in both signal shape and amplitude with the corresponding synthetic H$_2$O 2$f$ signal [top panel of Figs. 3.27 and 3.28 also see Fig. 3.25] were attributed to H$_2$O absorption. In total, 14 water lines were identified, and are indicated by solid grey vertical lines in Figs. 3.27 and 3.28. Table 3.12 lists the derived H$_2$O transition wavenumbers (in units of cm$^{-1}$), which were calculated by averaging the measured H$_2$O peak positions, and compares our values with those of HITRAN. The difference ($\Delta$) between our two measurements of a given peak position (as derived from two overlapping scans), and the estimated signal-to-noise ratio (SNR) are also given. In estimating the SNR we have used the standard definition, that is, the ratio of the peak amplitude squared ($s^2$) to the squared standard deviation of the no absorption signal ($\sigma^2$). The standard deviation was approximated, in the absence of a true no absorption signal, by assuming the 7185.9389–7187.9420 cm$^{-1}$ and 7176.2652–7178.1512 cm$^{-1}$ (1391.2188–1391.6066 nm and 1393.1164–1393.4825 nm) regions are sufficiently weakly absorbing at 3% NH$_3$ concentration that any fluctuations in baseline can be safely considered as noise. This noise is due to weak optical interference effects in our system. Although the optics are designed to minimise this, some residual signal remains, which will slowly change with temperature and give the appearance of low frequency noise.

The derived H$_2$O transition wavenumbers are within 0.024 cm$^{-1}$ of the values given in HITRAN, although the discrepancies ($\Delta$) between multiple measurements of the same peak are as much as 0.021 cm$^{-1}$. This indicates that the derived coefficients $c_i(T)$ in Eq. (3.4) produce small inconsistencies in the calibration of each scan. Peaks that were not assigned to water were tentatively attributed to NH$_3$ absorption, as all other contaminant species in the NH$_3$ gas mixture are at negligible levels. In total, 8 NH$_3$ transitions were identified, which are indicated by dotted black vertical lines in Figs. 3.27 and 3.28 and are listed in Table 3.13. All transitions, apart from the 7191.380 cm$^{-1}$ line which was derived from only one peak, are derived from peak measurements that agree to within $\Delta = 0.017$ cm$^{-1}$. Given that the discrepancies with HITRAN are a similar magnitude, the uncertainties on $\nu_{\text{meas}}$
are likely to be somewhat larger. It must also be noted that, from the estimated SNR values, several NH$_3$ peaks effectively fall within the range of what we have defined as noise. Rather than being discarded they have been listed as transitions due to the fact that both overlapping signal shapes correlate with characteristic features of the 2$f$ line shape, and they appear more prominently in the $> 99\%$ NH$_3$ spectrum.

Table 3.12: Measured H$_2$O transition wavenumbers (in units of cm$^{-1}$) compared to the values given in HITRAN. $\Delta$ refers to the difference (in units of cm$^{-1}$) between the two peak positions that were averaged to derive each transition wavenumber. $s^2/\sigma^2$ is the estimated signal-to-noise ratio, where $s$ is the peak amplitude and $\sigma$ is the standard deviation in the no absorbing region.

<table>
<thead>
<tr>
<th>$v_{\text{meas}}$</th>
<th>$v_{\text{meas}} - v_{\text{HIT}}$</th>
<th>$\Delta$</th>
<th>$s^2/\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7194.785</td>
<td>0.021</td>
<td>-0.010</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>7194.117</td>
<td>0.024</td>
<td>-0.001</td>
<td>14.6</td>
</tr>
<tr>
<td>7190.721</td>
<td>-0.018</td>
<td>0.000</td>
<td>24.2</td>
</tr>
<tr>
<td>7189.340</td>
<td>-0.004</td>
<td>0.000</td>
<td>31.1</td>
</tr>
<tr>
<td>7185.591</td>
<td>-0.006</td>
<td>-0.001</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>7182.957</td>
<td>0.005</td>
<td>-0.003</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>7182.212</td>
<td>0.003</td>
<td>0.000</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>7181.162</td>
<td>0.006</td>
<td>-0.011</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>7180.404</td>
<td>0.004</td>
<td>-0.004</td>
<td>29.1</td>
</tr>
<tr>
<td>7179.761</td>
<td>0.009</td>
<td>-0.017</td>
<td>18.5</td>
</tr>
<tr>
<td>7175.994</td>
<td>0.007</td>
<td>-0.021</td>
<td>3.6</td>
</tr>
<tr>
<td>7174.138</td>
<td>0.001</td>
<td>-0.017</td>
<td>11.8</td>
</tr>
<tr>
<td>7172.699</td>
<td>-0.000</td>
<td>-0.009</td>
<td>3.9</td>
</tr>
<tr>
<td>7170.254</td>
<td>-0.024</td>
<td>-0.003</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>

Table 3.13: Measured NH$_3$ transition wavenumbers (in units of cm$^{-1}$). $\Delta$ refers to the difference (in units of cm$^{-1}$) between the two peak positions that were averaged to derive each transition wavenumber. $s^2/\sigma^2$ is the estimated signal-to-noise ratio, where $s$ is the peak amplitude and $\sigma$ is the standard deviation in the no absorbing region.

<table>
<thead>
<tr>
<th>$v_{\text{meas}}$</th>
<th>$\Delta$</th>
<th>$s^2/\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7191.795</td>
<td>0.000</td>
<td>27.8</td>
</tr>
<tr>
<td>7191.380</td>
<td>-</td>
<td>23.3</td>
</tr>
<tr>
<td>7175.797</td>
<td>0.004</td>
<td>5.3</td>
</tr>
<tr>
<td>7174.478</td>
<td>0.004</td>
<td>15.9</td>
</tr>
<tr>
<td>7171.972</td>
<td>0.006</td>
<td>0.6</td>
</tr>
<tr>
<td>7171.817</td>
<td>0.017</td>
<td>2.1</td>
</tr>
<tr>
<td>7171.279</td>
<td>0.016</td>
<td>0.8</td>
</tr>
<tr>
<td>7169.029</td>
<td>0.002</td>
<td>8.7</td>
</tr>
</tbody>
</table>
3.6.6  > 99% NH$_3$ spectrum

A scan of the 1392 nm region was next performed using the > 99% concentration ammonia gas sample at 21°C and 0.1 bar pressure. Calibration of the absolute wavelength was performed using the same method as for the 3% sample, including the same coefficients ($c_i$) in Eq. (3.4), and the same wavelength scaling factor (0.986). The measured $2f$ signal amplitudes were also scaled by the same factor applied to the 3% NH$_3$ spectra, so as to be presented as multiples of the 1392.5335 nm (7181.1558 cm$^{-1}$) peak amplitude recorded at 3% NH$_3$ concentration. In doing so the 1392.5335 nm peak amplitude (> 99% NH$_3$) was scaled to $\approx 1$. This indicates that, up until the halfway point in the scan, the H$_2$O concentration remained unchanged from the 3% measurements.

NH$_3$ lines were distinguished from H$_2$O in the same way as at 3% concentration. An example of two weak lines that were attributed to H$_2$O absorption is given in Fig. 3.26. In both cases the derived line centres (grey vertical lines) are within 0.011 cm$^{-1}$ of known H$_2$O lines (blue vertical lines). The 7190.742 cm$^{-1}$ peak amplitude is similar for the 3% and > 99% concentration measurements, whereas for the 7190.279 cm$^{-1}$ peak, the change in amplitude may indicate an overlapping NH$_3$ line.

Figure 3.26: Expanded view of the 7190–7192 cm$^{-1}$ region of our scan, recorded at 3% and > 99% NH$_3$ concentration, in comparison with a simulated second harmonic H$_2$O spectrum. Two probable water features and 3 NH$_3$ features are identified.
3.6. Measurement of the 7169–7195 cm\(^{-1}\) region

The bottom panels of Figs. 3.27 and 3.28 display an overview of our 1392 nm region scan at > 99% ammonia concentration, where we note the increased y-axis scale relative to the middle and upper panels. Unfortunately during the second half of the scan (7180 \(\rightarrow\) 7169 cm\(^{-1}\)) additional water contamination appears to have occurred. This is apparent from a large increase in the 2\(f\) signal of H\(_2\)O lines relative to the 3% NH\(_3\) scan. In the optically thin limit the amplitude of 2\(f\) signal is directly proportional to the concentration of the absorbing species. Therefore, by scaling the synthetic H\(_2\)O 2\(f\) spectra and directly comparing with the measured 2\(f\) signal, we estimate that the increase in peak amplitude at the end of the scan (7170.2577 cm\(^{-1}\) H\(_2\)O line) is consistent with roughly a factor of 40 increase in H\(_2\)O concentration. This contamination originates from the commercial NH\(_3\) liquification process, from which a small residual amount related to the NH\(_3\) dew point temperature remains trapped in the NH\(_3\) liquid.

Despite the increased water interference between 7180 and 7169 cm\(^{-1}\), a large number of NH\(_3\) absorption peaks are clearly visible in Figs. 3.27 and 3.28. Line centres of the 20 strongest measured NH\(_3\) transitions (\(s \geq 0.063\)) are indicated by dotted black vertical lines, and their derived transition wavenumbers and \(\Delta\) values are listed in Table 3.14. Also given are the transition wavenumbers averaged over both the 3% and > 99% NH\(_3\) concentration scans (\(\nu_{\text{meas}}\)), the standard deviation \(\sigma_{\text{std}}\) of the \(N_{\text{scan}}\) peaks used to determine each transition wavenumber, and the averaged relative peak amplitudes (\(s\)). All peaks attributed to NH\(_3\) absorption at 3% concentration are clearly identifiable in the pure NH\(_3\) spectrum. Twelve additional peaks that were not identified at 3% concentration are also listed in Table 3.14 with \(N_{\text{scan}} \leq 2\). Although there are clearly a large number of additional weaker lines that are omitted from our analysis, these generally display too poor signal-to-noise to be useful for quantitative analysis.

Table 3.15 compares our measurement of 12 H\(_2\)O line centres, measured at > 99% NH\(_3\) concentration, to the values given in HITRAN. Although 8 of our derived line centres are accurate to within 0.015 cm\(^{-1}\), there are discrepancies as large as 0.03 cm\(^{-1}\), which is in approximate agreement with the line position uncertainty.
Table 3.14: Measured NH\textsubscript{3} transition wavenumbers (cm\textsuperscript{-1}), derived from the > 99% NH\textsubscript{3} measurements \(v_{\text{meas}}\), and averaged over both the 3% and > 99% concentration scans \(\langle v_{\text{meas}} \rangle\). Also shown are peak standard deviations \(\sigma_{\text{std}}\), total number of scans \(N_{\text{scan}}\), and averaged relative signal amplitudes \(\langle s \rangle\) measured at > 99% NH\textsubscript{3} concentration.

<table>
<thead>
<tr>
<th>(v_{\text{meas}})</th>
<th>(\Delta)</th>
<th>(\langle v_{\text{meas}} \rangle)</th>
<th>(\sigma_{\text{std}})</th>
<th>(N_{\text{scan}})</th>
<th>(\langle s \rangle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7194.3469</td>
<td>-0.005</td>
<td>7194.3469</td>
<td>-</td>
<td>2</td>
<td>0.24</td>
</tr>
<tr>
<td>7192.9836</td>
<td>-0.016</td>
<td>7192.9836</td>
<td>-</td>
<td>2</td>
<td>0.14</td>
</tr>
<tr>
<td>7191.8014</td>
<td>-0.002</td>
<td>7191.7980</td>
<td>0.004</td>
<td>4</td>
<td>0.46</td>
</tr>
<tr>
<td>7191.3952</td>
<td>-0.006</td>
<td>7191.3902</td>
<td>0.009</td>
<td>3</td>
<td>0.21</td>
</tr>
<tr>
<td>7191.0549</td>
<td>-0.013</td>
<td>7191.0549</td>
<td>-</td>
<td>2</td>
<td>0.09</td>
</tr>
<tr>
<td>7189.0929</td>
<td>-0.009</td>
<td>7189.0929</td>
<td>-</td>
<td>2</td>
<td>0.17</td>
</tr>
<tr>
<td>7187.8169</td>
<td>-0.017</td>
<td>7187.8169</td>
<td>-</td>
<td>2</td>
<td>0.09</td>
</tr>
<tr>
<td>7186.6567</td>
<td>-0.001</td>
<td>7186.6567</td>
<td>-</td>
<td>2</td>
<td>0.07</td>
</tr>
<tr>
<td>7185.8336</td>
<td>-0.010</td>
<td>7185.8336</td>
<td>-</td>
<td>2</td>
<td>0.22</td>
</tr>
<tr>
<td>7184.7430</td>
<td>-0.001</td>
<td>7184.7430</td>
<td>-</td>
<td>2</td>
<td>0.18</td>
</tr>
<tr>
<td>7182.5099</td>
<td>-0.008</td>
<td>7182.5099</td>
<td>-</td>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>7181.7363</td>
<td>0.000</td>
<td>7181.7363</td>
<td>-</td>
<td>1</td>
<td>0.18</td>
</tr>
<tr>
<td>7180.6106</td>
<td>-0.010</td>
<td>7180.6106</td>
<td>-</td>
<td>2</td>
<td>0.10</td>
</tr>
<tr>
<td>7177.9878</td>
<td>0.000</td>
<td>7177.9878</td>
<td>-</td>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>7175.8044</td>
<td>-0.006</td>
<td>7175.8008</td>
<td>0.005</td>
<td>4</td>
<td>0.19</td>
</tr>
<tr>
<td>7174.4808</td>
<td>-0.001</td>
<td>7174.4794</td>
<td>0.002</td>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>7171.9691</td>
<td>-0.009</td>
<td>7171.9703</td>
<td>0.004</td>
<td>4</td>
<td>0.11</td>
</tr>
<tr>
<td>7171.8135</td>
<td>-0.019</td>
<td>7171.8155</td>
<td>0.011</td>
<td>4</td>
<td>0.29</td>
</tr>
<tr>
<td>7171.2802</td>
<td>-0.010</td>
<td>7171.2798</td>
<td>0.008</td>
<td>4</td>
<td>0.16</td>
</tr>
<tr>
<td>7169.0313</td>
<td>-0.002</td>
<td>7169.0299</td>
<td>0.002</td>
<td>4</td>
<td>0.31</td>
</tr>
</tbody>
</table>

measured in Section 3.6.6. A pragmatic approach to assigning uncertainties, therefore, would be to assume the absolute error on our transition wavenumbers does not reach a value considerably larger than this, and we suggest an absolute uncertainty of 0.05 cm\textsuperscript{-1} is reasonable.
Table 3.15: Measured H$_2$O transition wavenumbers (in units of cm$^{-1}$) compared to the values given in HITRAN. $\Delta$ refers to the difference (in units of cm$^{-1}$) between the two peak positions that were averaged to derive each transition wavenumber. $s^2/\sigma^2$ is the estimated signal-to-noise ratio, where $s$ is the peak amplitude and $\sigma$ is the standard deviation in the no absorbing region.

<table>
<thead>
<tr>
<th>$\nu_{\text{meas}}$</th>
<th>$\nu_{\text{meas}} - \nu_{\text{HIT}}$</th>
<th>$\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7194.7938</td>
<td>-0.011</td>
<td>-0.008</td>
</tr>
<tr>
<td>7182.9606</td>
<td>0.010</td>
<td>-0.006</td>
</tr>
<tr>
<td>7182.2235</td>
<td>0.015</td>
<td>0.000</td>
</tr>
<tr>
<td>7181.1679</td>
<td>0.012</td>
<td>-0.014</td>
</tr>
<tr>
<td>7180.4177</td>
<td>0.018</td>
<td>-0.003</td>
</tr>
<tr>
<td>7179.7816</td>
<td>0.030</td>
<td>-0.009</td>
</tr>
<tr>
<td>7178.4574</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>7175.9949</td>
<td>0.008</td>
<td>-0.016</td>
</tr>
<tr>
<td>7174.1321</td>
<td>-0.005</td>
<td>-0.014</td>
</tr>
<tr>
<td>7173.7758</td>
<td>-0.002</td>
<td>-0.004</td>
</tr>
<tr>
<td>7172.6963</td>
<td>-0.003</td>
<td>-0.009</td>
</tr>
<tr>
<td>7170.2577</td>
<td>-0.020</td>
<td>-0.003</td>
</tr>
</tbody>
</table>
Figure 3.27: Complete scan of the 7186–7181 cm$^{-1}$ region measured using the 3% NH$_3$ gas sample (middle panel, black) and > 99% NH$_3$ gas sample (bottom panel, red), in comparison with a simulation of the second harmonic spectrum of H$_2$O (upper panel, blue). Dashed black lines indicate NH$_3$ peaks, grey lines indicate H$_2$O peaks.
Figure 3.28: Complete scan of the 7181–7195 cm$^{-1}$ region measured using the 3% NH$_3$ gas sample (middle panel, black) and > 99% NH$_3$ gas sample (bottom panel, red), in comparison with a simulation of the second harmonic spectrum of H$_2$O (upper panel, blue). Dashed black lines indicate NH$_3$ peaks, grey lines indicate H$_2$O peaks.
3.6.7 Determination of H$_2$O concentration

To quantify the H$_2$O contamination level increase observed over the course of the > 99% NH$_3$ scan, direct absorption measurements were performed on the strong 1392.5335 nm (7181.1558 cm$^{-1}$) H$_2$O line after both surveys were complete. The direct absorption signal (DAS), recorded at 21°C and 0.1 bar pressure, is given in Fig. 3.29 (left panel). The baseline was fit to a second-order polynomial (orange), and the absorbance was calculated for a photodetector DC signal of 902.5 mV. The absolute line intensity was taken from HITRAN and convoluted with a Lorentzian line profile. The Lorentzian HWHM was obtained as 0.029 cm$^{-1}$ from the recorded first-derivative signal, which was provided by the signal with zero modulation amplitude applied. This zero-modulation amplitude signal is high-pass filtered by the gas analyser electronics, which has a time constant much shorter than the signal variations, and the resulting signal is the first derivative of the absorption peak. The concentration was adjusted, and a value of 630 ppm was seen to reproduce the measured absorbance excellently, as shown in Fig. 3.29 (right panel). This is roughly 30× higher than the CaO moisture scrubber is specified to maintain (< 20 ppm), and may suggest it became saturated with moisture during the measurement campaign.

Figure 3.29: Direct absorption signal of the 1392.5335 nm H$_2$O line at 0.1 bar pressure.

Although we have no equivalent DAS measurement at the time of the 3% scan, one was recorded at 21°C and 1 bar pressure several days earlier. The measured
3.6. Measurement of the 7169–7195 cm\(^{-1}\) region

absorption signal and fitted Lorentzian, convoluted with the absolute line intensity, are shown in Fig. 3.30 (right panel). A HWHM of 0.1 cm\(^{-1}\) and concentration of 4 ppm were seen to reproduce the measured absorbance reasonably well. Although this difference in concentration is more drastic than observed over the course of the scan, it demonstrates instabilities in the current experimental setup that could be improved in future measurements.

![Graph showing direct absorption signal of the 1392.5335 nm H\(_2\)O line at 1 bar pressure.](image)

3.6.8 Discussion and future work

Line positions have been derived for 20 weak NH\(_3\) lines in the 7169–7195 cm\(^{-1}\) region. By comparison of the measured 2\(f\)-WMS spectrum with a simulation of the second harmonic spectrum of H\(_2\)O, line positions were seen to display errors as large as 0.03 cm\(^{-1}\). Several NH\(_3\) lines are noted as potentially interfering lines in trace moisture detection measurements. For future NH\(_3\) characterisation in this region the ‘strong’ line at 7191.798 cm\(^{-1}\) (by our measurements), and the weaker 7184.743 cm\(^{-1}\) and 7192.984 cm\(^{-1}\) lines are promising candidates. The latter two pose the advantage of being sufficiently isolated so as to avoid blending with adjacent absorption features at atmospheric pressure. Moreover, the presence of a non absorbing baseline either side of each feature permits the use of certain calibration free approaches to determining environmental conditions such as temperature or...
A number of factors in our current experimental setup could be modified to improve the accuracy and precision of our measurements. Firstly, the current laser tuning process should be replaced by one of slowly ramping the laser temperature across the entire wavelength range. This would circumvent the need to separately calibrate then stitch each of the individual scans; an error-prone process that is exacerbated by the degree of nonlinearity in the laser current tuning rate, that also varies with temperature. Secondly, the $2f$-WMS signal filtering should be improved so it does not distort the true signal amplitude, alongside better understanding of other instrument effects that affect the signal amplitude accuracy. This way absolute line strengths could be obtained. Thirdly, absolute measurement of the laser frequency could be obtained by sampling part of the laser beam into a laser wavemeter instrument, which would avoid the need for an etalon glass. Finally, a better NH$_3$ purifier, or higher purity (N5.0 or better) NH$_3$ gas bottle to reduce the moisture contamination. A multipass cell could also be considered to increase optical path length, however they are known to increase optical interference noise, so the entire path length benefit might not be obtained.

From the point-of-view of performing dual NH$_3$–H$_2$O measurements in future, the next two important steps are, firstly, the derivation of absolute line intensities, for example using the methods proposed by Henningsen [99] or Reiker [212], and secondly, the assignment of the observed NH$_3$ lines in this region. In my opinion, however, assignment of the transitions presented in this work would require considerable effort, and is likely impossible with the current state of experimental data. It may be possible using the C2018 line list(s) reported in Section [3,4], however, I have checked the C2018 predictions within the 7169–7195 cm$^{-1}$ region and noted several obstacles that would impede such an undertaking. These are as follows: i) the strongest predicted lines in this region are labelled $v_2 + 4v_4^2$, $3v_2 + 3v_4^3$, $v_2 + 4v_4^4$ and $v_1 + v_2 + 2v_4^2$, which are bands for which no level of accuracy can be guaranteed in C2018, and so line positions may be several wavenumbers in error
3.7 Conclusion

In this chapter I have presented an improved potential energy surface for $^{14}$NH$_3$, produced by empirical refinement of a high accuracy ab initio PES \cite{199} to experimentally derived energy levels \cite{19,87}, and a number of my own assignments. The resulting rovibrational energy levels reproduce the MARVEL experimentally derived values for states with $J = 0 - 10$ generally to within 0.10 cm$^{-1}$ under 6400 cm$^{-1}$, and 0.2 cm$^{-1}$ between 6400 and 7555 cm$^{-1}$.

Using a well established DMS \cite{307}, a new MRCI/aug-cc-pwCVQZ DMS, and the C2018 potential energy surface, room temperature line list calculations were performed for transitions between states with $J = 0 - 20$, in the wavenumber ranges 0–20 000 cm$^{-1}$, and 0–12 000 cm$^{-1}$ respectively. The predicted line positions and line intensities were seen to be a substantial improvement over those of the BYTe line list \cite{302}. Several strong bands in the 5700–6200 cm$^{-1}$ were noted missing from HITRAN2016.

Our line lists were used to assign 769 transitions in the 7400–8000 cm$^{-1}$ wavenumber range using ground state combination differences, and derive 284 upper state energies. Of our assignments, 230 lines were previously assigned \cite{19}, out of which we found 53 disagreed in at least one ‘good’ quantum number. Our stricter validation criteria and higher accuracy line list suggest that a handful of the previous assignments were incorrect. It is hoped that future analysis above 8000 cm$^{-1}$ can validate the current assignments present in HITRAN, and in turn inform future ab initio calculations. We note that the marked improvement in C2018 line positions over those of BYTe, for wavenumbers above 9000 cm$^{-1}$, is predominantly due to our inclusion of a mere 49 empirical energies in the 9000–10 500 cm$^{-1}$ region.
These are energies that we derived simply by visually matching lines in HITRAN to our own predictions as the refinement progressed, yet there is a resounding impact throughout the entire region that emphasises the importance of reliable experimental data when attempting to produce a high quality PES. In this context we note the recent use of our NH$_3$ line list by Irwin et al. [119] to study ammonia spectra in Jupiter at near-infrared and visible wavelengths, their comparisons suggest that the line list represents a significant improvement on what is currently available but that further work is need to improve the predicted line positions at wavelengths shorter than 1 μm.
Chapter 4

Arsine ($^{75}$AsH$_3$)

4.1 Introduction

Arsine (AsH$_3$) is a highly poisonous gas [180] which is the direct analogue molecular structure of ammonia (NH$_3$) and phosphine (PH$_3$). Like these two gases it has been detected the atmospheres of the gas giant planets Jupiter [175,176] and Saturn [25]. It may therefore be expected to be also present in the atmospheres of gas giant exoplanets.

Arsene is also important for industrial applications as high purity arsine is widely used in the semiconductor manufacturing industry, for example, in processing GaAs surfaces [44,129,134]. Given its highly poisonous nature, with an exposure limit value of 50 ppb mole concentration [98], the detection of AsH$_3$ escape at such levels is an important safety requirement in this industry [43]. It is also monitored in the polymer industry as trace level arsine impurity in ethylene and propylene monomer feedstock gases may contaminate the catalysts, resulting in reduced quality and yield of the polymer products [81].

Arsine is also a trace atmospheric pollutant due to emissions from various industrial processes, such as power generation and smelting [159,163]. Routine methods for arsine measurement in industry include gas chromatography, electrochemical sensors, colorimetric sensors, and Fourier-transform infrared spectroscopy. However, the development of high resolution laser spectroscopy based measurements is a growing area [53,244] for which detailed line lists are required to model...
the high resolution absorption spectra.

While there have been a number of studies of the infrared and microwave spectrum of arsine, there is no comprehensive line list for the system and there is a lack of information on the intensity of many bands. The situation for absolute line intensities is particularly dire, with existing data confined solely to the measurements reported by Dana et al. [62]. Previous attempts to model the global vibrational structure [195, 219], and rovibrational sub-structure [92, 147, 267, 268, 276, 296, 297], have focused predominantly on effective-Hamiltonians, which have limited predictive capability outside the fitted data. In addition, an \textit{ab initio} potential energy surface (PES) for AsH$_3$ was reported in 1995 by Breiding and Thiel in the form of the cubic anharmonic force field [34] using relativistic effective core potentials (ECPs).

Considering the unsuitability of the current state of AsH$_3$ data for either exoplanet modelling, which necessitates completeness, or industrial monitoring, which necessitates accuracy, we decided to construct a comprehensive line list for arsine which could be used for the applications mentioned above.

4.2 \textbf{Tunneling and molecular symmetry group}

Analogous to the inversion motion of NH$_3$ it is possible for the As atom to tunnel through the plane of the hydrogens so that the ordering of the H atoms (labelled 1, 2 and 3) is switched from (1,2,3) to (1,3,2) when counted sequentially during a clockwise rotation about the positive $z$–axis. The energy barrier to this inversion motion is about 14 100 cm$^{-1}$ (see Section 4.7.1), and so it is assumed no tunneling takes place on the time scale of a typical spectroscopic experiment. A semi-quantitative argument for this can be formulated by considering a 1D model of pseudoparticle trapped in a double-well potential [161]. The effective barrier strength is expressed in terms of the dimensionless parameter

$$\beta = \frac{\sqrt{mV_0x_0}}{\hbar}, \quad (4.1)$$
where $V_0$ is the barrier height, $m$ is the reduced mass of the pseudoparticle and $x_0 = 0.895 \, \text{Å}$ is half the distance between the two minima, which can be estimated as half the length of the arc that passes through the two minima and the saddle point, i.e., the three points $(r(\text{Å}), \alpha(\degree)) = [(1.52, 92.2), (1.47, 120), (1.52, -92.2)]$ (see Section 4.7.1). In the case of AsH$_3$, the reduced mass can be approximated by

$$m = \frac{3m_H(m_{As} + 3m_H \sin^2 \theta)}{3m_H + m_{As}}, \quad (4.2)$$

where $\theta$ is the angle between the As–H bond length at equilibrium, and planar configuration. The parameter $\beta$ can then be related to the ground state energy splitting $\Delta E$ in the deep potential regime by

$$\frac{\Delta E}{V_0} \approx 1.8977 \cdot \frac{1}{\beta} e^{-\sqrt{2} \beta}, \quad (4.3)$$

which is then related to the tunneling period $\tau$ (i.e. the length of time it takes the molecule to invert twice) by

$$\tau = \frac{h}{\Delta E}. \quad (4.4)$$

From the above equations it is clear that $\Delta E$ is extremely sensitive to the parameter $\beta$, which depends on molecular parameters that are not well known. Nevertheless, for $^{75}$AsH$_3$ the value obtained for $\beta$ is 49.7, which results in a tunneling time scale of $\tau \sim 10^{12}$ seconds. Applying the same model for NH$_3$, we obtain $\tau \sim 10^{-11}$ seconds with the corresponding tunneling splitting $\Delta E \sim 1 \, \text{cm}^{-1}$ which is the same order as the true value (0.79 cm$^{-1}$). This astronomically long $\tau$ value for AsH$_3$ will vary substantially for rotationally and vibrationally excited states, as the wavefunction ‘moves up’ the potential and experiences perturbative interactions. In fact, in the case of NH$_3$, the tunneling splitting has been found to display complex and sometimes irregular dependence on rotation and vibration [59]. Although our argument is far from rigorous, it presents a reasonable grounds for classifying AsH$_3$ in the molecular symmetry group $C_{3v}(M)$. We also note the work by Sousa et al. [240], who used variational nuclear motion calculations to predict the splittings
in the ground state and $\nu_2$ overtones of PH$_3$ ($V_0 \approx 12,000$ cm$^{-1}$). This was later searched for by Okuda and Sasada [178], who observed no splitting despite a spectral resolution of 150 kHz, and so we conclude it is unlikely to be observed in AsH$_3$ for some time.

4.3 Potential energy surface

4.3.1 Electronic structure calculations

Accurate modelling of heavy elements in quantum chemistry is made particularly challenging by increased relativistic effects, core-core electron correlation and core-valence electron correlation (see Sections 2.1.5 and 2.1.6). Unfortunately, for reasons discussed in Section 2.1.6, the most rigorous relativistic treatments (the Douglas-Kroll-Hess Hamiltonian) cannot be used in conjunction with explicitly correlated (F12/R12) methods. In the case of AsH$_3$, F12-pseudopotential and standard all electron DKH based approaches are possible, and the benefits of F12 must therefore be weighed against the penalty of introducing an additional scalar-relativistic approximation. Peterson [189] showed that complete basis set (CBS) extrapolated CCSD(T)/aug-cc-pwCV$n$Z-PP pseudopotential calculations performed almost identically to their DKH all-electron counterparts in a series of molecular benchmark calculations for post-3d main group elements, including the As$_2$, AsF, AsCl and AsN molecules. They go on to develop a new family of F12-specific cc-pV$n$Z-PP-F12 basis sets to be used at the CCSD(T)-F12 level, which yield accuracy comparable to the 2-3 times larger aug-cc-pwCV(n+2)Z-PP basis sets used at standard CCSD(T) level [106, 189]. Their pseudopotential-F12 optimised approach is the one followed in this work.

All electronic structure calculations were performed using MOLPRO [287] and employed the explicitly correlated coupled cluster method CCSD(T)-F12b [2, 133] with implicit treatment of scalar-relativistic effects via replacement of 10 core electrons with a pseudopotential (PP). Calculations were carried out in the frozen core approximation and utilized the correlation consistent quadruple-zeta, PP-F12 optimised basis set of Hill et al. [106] (cc-pVQZ-PP-F12) to represent the As electronic
wavefunction, and cc-pVQZ-F12 basis sets for the H atoms. Density fitting (DF) for the 2-electron (MP2FIT) and exchange term (JKFIT) integrals employed the cc-pVTZ-PP-F12/MP2Fit and def2-QZVPP/JKFIT basis sets, respectively, and for the resolution of the identity of the many-electron F12 integrals (OPTRI) we used the VTZ-PP-F12/OPTRI basis set. For the geminal exponent $\gamma$, a value of $1.4 \text{ a}_0^{-1}$ was used as recommended by Hill et al. All calculations were performed on the ground electronic state, which is sufficiently uncoupled from higher electronic excitations that non-adiabatic effects are expected to be very small \cite{7}. Furthermore, owing to the inverse mass dependence of the Born-Oppenheimer diagonal correction, adiabatic effects are expected to be far smaller than, say, for NH$_3$ and thus far outweighed by electron correlation and relativistic contributions.

**Table 4.1:** Equilibrium energies (in units of Hartree) calculated at the CCSD(T) level of theory using different basis sets and Hamiltonians.

<table>
<thead>
<tr>
<th>Basis</th>
<th>Energy/E$_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVQZ</td>
<td>-2236.17795527</td>
</tr>
<tr>
<td>AVQZ-DK</td>
<td>-2261.02359376</td>
</tr>
<tr>
<td>AVQZ-PP</td>
<td>-333.14700414</td>
</tr>
<tr>
<td>AV5Z</td>
<td>-2236.18098311</td>
</tr>
<tr>
<td>AV5Z-DK</td>
<td>-2261.02786843</td>
</tr>
<tr>
<td>AV5Z-PP</td>
<td>-333.14983002</td>
</tr>
</tbody>
</table>

Figure 4.1: One dimensional cuts of the relativistic corrections for the $(r_1 = r_2 = r_3 = 1.51$ Å; $\alpha_1 = \alpha_2 = 92.1^\circ$; $50 \leq \alpha_3 \leq 140^\circ$) bond angle and $(r_1 = r_2 = 1.51$ Å; $1.2 \leq r_3 \leq 2.2$ Å; $\alpha_1 = \alpha_2 = \alpha_2 = 92.1^\circ$) bond length displacements.

To initially qualify the importance of including scalar relativistic effects in our
calculations, relativistic corrections $\Delta E_X$ along 1-dimensional cuts of the potential energy surface PES were calculated (shown in Fig. 4.1 along with cuts through the AVQZ surface for reference). This was done by first shifting the potential energy curves by their respective energies at equilibrium, listed in Table 4.1, where we note the relatively small absolute energies of the ECP based calculations owing to their implicit treatment of 10 core electrons. We then have $\Delta E_X = E_{X_1} - E_{X_2}$ where $E_{X_1} = AVnZ$-PP/$AVnZ$-DK and $E_{X_2} = AVnZ$ for the pseudopotential all-electron calculations. Here, and in all subsequent DKH calculations, the DKH Hamiltonian has been expanded to 8th–order (DKH8) using optimal unitary parametrization. In Fig. 4.1 only quadruple-zeta ($n = 4$) results are presented as they were seen to differ by no more than 3 cm$^{-1}$ from the respective 5-zeta ($n = 5$) curves. Clearly the inclusion of scalar relativistic effects are important, and both approximations have a similar effect on the total energy. However, the pseudopotential approximation tends to raise the energy at stretched geometries and lower the energy at contracted geometries, relative to the all electron calculations. It is difficult to asses the effect of this difference within the Hill et al. [106] regime for a full dimensional PES. We therefore opted to generate a second 6D surface at the AVQZ-DKH8 level of theory (henceforth denoted AVQZ-DK), to provide a benchmark for our VQZ-PP-F12 based ab initio nuclear motion calculations. The results of these are presented in Section 4.4.

As mentioned in Section 2.1.6, beyond 4th–order expansion the DKH Hamiltonian depends slightly on the chosen parameterization of the unitary transformations applied [171]. Thus, additional DKH4 calculations were performed on a subset of 5000 DKH8 PES geometries. Between 4th and 8th–order the resulting electronic energies were seen to differ by less than $\hbar c \cdot 0.1$ cm$^{-1}$ above their equilibrium values. For the purpose of benchmarking the ECP based nuclear motion calculations, therefore, this dependency is not expected to be significant.

### 4.3.2 Nuclear geometry grid

Our grid of nuclear geometries was built by combining 1D–6D sub-grids. Our 1D grid consisted of a cut along the $r_1 = r_2 = r_3$ stretch with $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_{eq}$, and
a cut along the $\alpha_1 = \alpha_2 = \alpha_3$ bend with $r_1 = r_2 = r_3 = r_{eq}$. Each additional degree of freedom was then added by allowing either an additional As-H bond length or H-As-H bond angle to vary. Because this method causes the number of points to grow so rapidly, the range and intervals of $r$ and $\alpha$ were reduced with each increasing degree of freedom using the 1D cuts as a guide. This also helped to limit the range of electronic energies generated, as large distortions in geometry can lead to unnecessarily high values of energy that are not needed in the fit.

In order to ensure each grid point was fully unique we applied the $C_{3v}$ molecular symmetry group transformations prior to computing the electronic energy. If two grid points were transformed into one another, then one was discarded. Finally, any energetically sparse regions were filled by generating additional geometries that were estimated to fall within our desired range. The 1D cuts provided an initial guide to the electronic energy, then intermediate versions of our PES were used to more accurately choose geometries. Our final grid consisted of 39 873 nuclear geometries within the range $1.10 \leq r_i \leq 3.74$ Å and $37^\circ \leq \alpha_i \leq 130^\circ$, with electronic energies extending to $\hbar c \cdot 27 000$ cm$^{-1}$, although $\sim 38$ 000 of these were below $\hbar c \cdot 10$ 000 cm$^{-1}$ ($1.25 \leq r_i \leq 1.9$ Å and $60^\circ \leq \alpha_i \leq 126^\circ$). The additional points in the $\hbar c \cdot 10$ 000 cm$^{-1}$ to $\hbar c \cdot 27$ 000 cm$^{-1}$ energy range predominantly belonged to the As-H dissociative stretch, which is where holes commonly appear if the function is not suitably constrained at high energy. Grid points for our AVQZ-DK reference PES were chosen by randomly sampling 16 396 equally energetically distributed points from our VQZ-PP-F12 grid, which spanned the bond lengths $1.3 \leq r_i \leq 1.8$ Å, bond angles $65^\circ \leq \alpha_i \leq 130^\circ$ and energies below $\hbar c \cdot 15$ 000 cm$^{-1}$.

Each grid point computed at the CCSD(T)-F12b/cc-pVQZ-PP-F12 level took approximately 10-15 minutes to compute on UCL’s Legion computer cluster. This was increased to 20-30 mins for the DHKH Hamiltonian-based calculations, owing to the increased computational demand of explicitly treating the 10 core electrons.

### 4.3.3 Analytic representation

To represent the PES analytically we used the functional form given in Eq. (2.37), which is the same as used for NH$_3$ in Section 3.3. Points were given energy ($E_i$)
dependant weights \( (w_i) \)

\[
w_i = \frac{2}{1 + e^{2 \times 10^{-4} \times E_i}}
\]  

(4.5)

as used by Polyansky et al. \[199\]. We could usefully fit terms in the potential up to 5th order resulting in a root mean-square (RMS) deviation of 0.7 cm\(^{-1}\) for the 39 678 nuclear geometries. For our all-electron reference PES, the weighted RMS error increased to 1.2 cm\(^{-1}\), most likely due to the proportionally fewer points close to equilibrium. However, it should be noted that adding more points to the fit had little effect on the computed vibrational term values reported in Section 4.4.

### 4.4 Nuclear motion calculations

To calculate rovibrational energy levels we used the variational nuclear motion program TROVE. The general methodology of TROVE is discussed in Section 2.2, and so only the specific details relevant to AsH\(_3\) are discussed here.

Rovibrational basis functions were constructed as symmetrised linear combinations of 1D primitive-basis-function products

\[
|\nu, J, K, m, \tau_{\text{rot}}\rangle = [|J, K, m, \tau_{\text{rot}}\rangle |n_1\rangle |n_2\rangle |n_3\rangle |n_4\rangle |n_5\rangle |n_6\rangle]^{\Gamma_{ir}},
\]  

(4.6)

where the 1D stretching functions \(|n_1\rangle, |n_2\rangle, |n_3\rangle\) and bending functions \(|n_4\rangle, |n_5\rangle, |n_6\rangle\) are obtained by solving the corresponding one-dimensional Schrödinger equations using the Numerov-Cooley approach \[52, 177\] for the stretches, and 1D harmonic oscillator eigenfunctions for the bends. In the above equation, \(\Gamma_{ir}\) represents one of the irreducible representations of \(C_{3v}\) spanned by \(|\nu, J, K, m, \tau_{\text{rot}}\rangle\). A multi-step contraction scheme was employed to limit the vibrational, then rovibrational basis set size. This is outlined in the following paragraphs and in Section 2.2.4.

Owing to the structural similarities between AsH\(_3\) and other \(XY_3\)-type molecules which have been investigated in the past, variational calculations could be performed with relative ease once a PES and DMS had been constructed, and required only a few molecule specific parameters to be defined: atomic mass,
4.4. Nuclear motion calculations

Table 4.2: Differences between experimentally derived band centres and our calculated values computed using all-electron DKH and pseudopotential-F12 based PESs. All numerical values are term values given in units of cm$^{-1}$.

<table>
<thead>
<tr>
<th>Band</th>
<th>Sym.</th>
<th>Obs.</th>
<th>VQZ-PP-F12</th>
<th>AVQZ-DK</th>
<th>CalcPP-F12</th>
<th>CalcDK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_2$</td>
<td>A$_1$</td>
<td>906.752</td>
<td>904.812</td>
<td>905.058</td>
<td>1.940</td>
<td>1.694</td>
</tr>
<tr>
<td>$\nu_4$</td>
<td>E</td>
<td>999.225</td>
<td>994.460</td>
<td>994.132</td>
<td>4.765</td>
<td>5.093</td>
</tr>
<tr>
<td>2$\nu_2$</td>
<td>A$_1$</td>
<td>1806.149</td>
<td>1802.443</td>
<td>1802.451</td>
<td>3.706</td>
<td>3.698</td>
</tr>
<tr>
<td>$\nu_2 + \nu_4$</td>
<td>E</td>
<td>1904.115</td>
<td>1897.551</td>
<td>1897.465</td>
<td>6.564</td>
<td>6.650</td>
</tr>
<tr>
<td>2$\nu_4^{l=0}$</td>
<td>A$_1$</td>
<td>1990.998</td>
<td>1982.116</td>
<td>1981.574</td>
<td>8.882</td>
<td>9.424</td>
</tr>
<tr>
<td>$\nu_1$</td>
<td>A$_1$</td>
<td>2115.164</td>
<td>2108.659</td>
<td>2105.000</td>
<td>6.505</td>
<td>10.164</td>
</tr>
<tr>
<td>$\nu_3$</td>
<td>E</td>
<td>2126.432</td>
<td>2116.469</td>
<td>2112.542</td>
<td>9.963</td>
<td>13.890</td>
</tr>
<tr>
<td>$\nu_1 + \nu_2$</td>
<td>A$_1$</td>
<td>3013$^a$</td>
<td>3006.718</td>
<td>3002.875</td>
<td>6.3</td>
<td>10.1</td>
</tr>
<tr>
<td>$\nu_1 + \nu_4$</td>
<td>E</td>
<td>3102$^a$</td>
<td>3096.255</td>
<td>3084.866</td>
<td>12.7</td>
<td>17.1</td>
</tr>
<tr>
<td>2$\nu_1$</td>
<td>A$_1$</td>
<td>4166.772</td>
<td>4151.833</td>
<td>4143.187</td>
<td>14.939</td>
<td>23.585</td>
</tr>
<tr>
<td>$\nu_1 + \nu_3$</td>
<td>E</td>
<td>4237.700</td>
<td>4222.006</td>
<td>4214.312</td>
<td>15.694</td>
<td>23.388</td>
</tr>
<tr>
<td>2$\nu_3^{l=0}$</td>
<td>A$_1$</td>
<td>4247.720</td>
<td>4229.805</td>
<td>4221.816</td>
<td>15.970</td>
<td>24.408</td>
</tr>
<tr>
<td>2$\nu_3^{l=2}$</td>
<td>E</td>
<td>5041.541</td>
<td>5030.916</td>
<td>5029.916</td>
<td>15.5</td>
<td>26.1</td>
</tr>
<tr>
<td>2$\nu_1 + \nu_2$</td>
<td>A$_1$</td>
<td>5057$^a$</td>
<td>5041.541</td>
<td>5030.916</td>
<td>15.5</td>
<td>26.1</td>
</tr>
<tr>
<td>$\nu_1 + \nu_2 + \nu_3$</td>
<td>E</td>
<td>5057$^a$</td>
<td>5041.911</td>
<td>5030.920</td>
<td>15.8</td>
<td>26.1</td>
</tr>
<tr>
<td>2$\nu_1 + \nu_4$</td>
<td>E</td>
<td>5128$^a$</td>
<td>5111.286</td>
<td>5100.477</td>
<td>16.7</td>
<td>27.5</td>
</tr>
<tr>
<td>2$\nu_3^{l=0} + \nu_2$</td>
<td>A$_1$</td>
<td>5128$^a$</td>
<td>5113.615</td>
<td>5104.249</td>
<td>14.4</td>
<td>23.8</td>
</tr>
<tr>
<td>$\nu_1 + \nu_3 + \nu_4$</td>
<td>E</td>
<td>5158$^a$</td>
<td>5137.282</td>
<td>5127.176</td>
<td>20.7</td>
<td>30.8</td>
</tr>
<tr>
<td>$\nu_1 + \nu_3 + \nu_4$</td>
<td>A$_1$</td>
<td>5158$^a$</td>
<td>5137.555</td>
<td>5127.471</td>
<td>20.4</td>
<td>30.6</td>
</tr>
<tr>
<td>3$\nu_1$</td>
<td>A$_1$</td>
<td>6136.340</td>
<td>6116.822</td>
<td>6101.231</td>
<td>19.518</td>
<td>35.109</td>
</tr>
<tr>
<td>2$\nu_1 + \nu_3$</td>
<td>E</td>
<td>6136.330</td>
<td>6116.793</td>
<td>6101.192</td>
<td>19.537</td>
<td>35.138</td>
</tr>
<tr>
<td>$\nu_1 + 2\nu_3^{l=0}$</td>
<td>A$_1$</td>
<td>6275.830</td>
<td>6257.116</td>
<td>6243.540</td>
<td>18.714</td>
<td>32.290</td>
</tr>
<tr>
<td>$\nu_1 + 2\nu_3^{l=2}$</td>
<td>E</td>
<td>6282.350</td>
<td>6261.282</td>
<td>6247.600</td>
<td>21.068</td>
<td>34.750</td>
</tr>
<tr>
<td>3$\nu_3^{l=1}$</td>
<td>E</td>
<td>6294.710</td>
<td>6270.037</td>
<td>6256.059</td>
<td>24.673</td>
<td>38.651</td>
</tr>
<tr>
<td>3$\nu_3^{l=3}$</td>
<td>A$_1$</td>
<td>6365.950</td>
<td>6340.980</td>
<td>6327.902</td>
<td>24.970</td>
<td>38.048</td>
</tr>
</tbody>
</table>

$^a$ experimental uncertainties of Halonen et al. [94] are estimated to be 2 cm$^{-1}$ or more.

molecular symmetry group, Z-matrix, equilibrium parameters, and the definition of our 1-D grids upon which the wavefunctions are evaluated. As arsenic has only one stable isotope, $^{75}$As, all reported nuclear motion calculations were performed for $^{75}$AsH$_3$.

Within the limitations of our PES, the accuracy of our variational calculation is of course determined predominantly by i) the size of our nuclear-motion basis set; and ii) our Taylor-type expansion of the kinetic energy operator $\hat{T}$ and our re-
expansion of the potential function $V$ in terms of linearised coordinates. The former, we choose to restrict via the polyad number $P$, which, in the case of AsH$_3$ takes the form

$$P = 2(n_1 + n_2 + n_3) + n_4 + n_5 + n_6,$$

(4.7)

where $n_1 + n_2 + n_3$ is the total number of stretching quanta and $n_4 + n_5 + n_6$ is the total number of bending quanta, corresponding to the primitive functions $|n_i\rangle$ ($i = 1, ..., 6$) in Eq. 4.6. For our comparison of the VQZ-PP-F12 and AVQZ-DK based ab initio PESs we chose to include in our variational calculations all vibrational states with $P \leq P_{\text{max}} = 14$ as used previously for NH$_3$ and PH$_3$ [239][304]. This resulted in our vibrational eigenvalues converged to within 0.1 cm$^{-1}$ below 6000 cm$^{-1}$ for the stretches, and as much as 3 cm$^{-1}$ for the bending overtones. Our $\hat{T}$ and $V$ expansions we take to 6th and 8th order, respectively. Increasing these to 8th and 10th order changed the vibrational term values reported throughout this work by $< 0.1$ cm$^{-1}$ for the stretches, and $< 0.7$ cm$^{-1}$ for the bends. For highly excited bending overtones, such as the 5 and 6–quanta bends, the convergence error due to our $\hat{T}$ and $V$ expansions may be several wavenumbers.

Table 4.2 shows the 26 lowest-lying experimentally derived band centres compared to our calculations. Term values known to sub-wavenumber accuracy are taken from Sanzharov et al. [219]; the remaining eight bands are from the work by Halonen et al. [94] and have an estimated 2 cm$^{-1}$ uncertainty, although this may be larger for the 5050–5200 cm$^{-1}$ bands [155]. Using the VQZ-PP-F12 and AVQZ-DK PESs the experimentally derived values of the four fundamentals are reproduced to within 10 cm$^{-1}$ and 14 cm$^{-1}$, respectively. Whilst far from the accuracy achieved in previous studies of NH$_3$ and PH$_3$, our results are comparable to the achievements of Nikitin et al. in their recent ab initio study of GeH$_4$ [174], and we deem it reasonable considering the greater contribution of relativistic effects, core-core electron correlation and core-valence electron correlation associated with heavier atoms. For the overtones and combination bands the quality of our ab initio predictions steadily decreases in proportion to the error on the fundamentals, except for the $2\nu_4^{\ell=0}$ band which is independently examined in Section 4.7.2. Most
importantly, the VQZ-PP-F12 surface consistently and significantly outperforms the VQZ-DK surface. Given the factor of 2 reduction in computational time, this highlights the value of the work by Hill, Peterson and co-authors [106, 189].

4.5 Refinement

In order to achieve so-called ‘spectroscopic’ accuracy in our variational nuclear motion calculations the \textit{ab initio} VQZ-PP-F12 PES was refined to experimental data. For details of the refinement procedure, the reader is directed to Section 2.2.5.

Because As is heavier than N or P, the rotational energies of AsH$_3$ are more closely spaced than those of NH$_3$ and PH$_3$, and so more highly populated at room temperature. Particular attention was therefore paid to optimising the equilibrium bond lengths and bond angles. This optimisation was performed prior to the refinement by using the hyperfine resolved rotational energies of Tarrago et al. [253], which we averaged using the spin-statistical weights ($\{A_1,A_2,E\} = \{16,16,16\}$), and a Newton-Gauss style procedure with a step size of $\pm 0.002$ Å and $\pm 0.002$ rad. Although TROVE is capable of computing quadrupole-hyperfine effects [294], requiring only a quadrupole moment surface and electric field gradient tensor in addition to the PES and DMS, the resulting splittings are small (roughly a few MHz) and so not considered here.

For the full nonlinear least squares refinement we allowed for corrections to harmonic and certain cubic terms in our PES, and used 322 experimentally derived energies with $J \leq 6$ compiled from Refs. [253, 267, 268, 276, 296, 297]. These sampled the following vibrational bands: the fundamentals $\nu_1$, $\nu_2$, $\nu_3$, $\nu_4$; overtones $2\nu_1$, $2\nu_2$, $2\nu_3$, $2\nu_4$, $3\nu_1$, $3\nu_3$; and combination bands $\nu_1 + \nu_3$, $\nu_2 + \nu_4$, $2\nu_1 + \nu_3$, $\nu_1 + 2\nu_3$. Because we could find no rotationally excited states belonging to the $\nu_2$ and $\nu_4$ bands in the literature, only their band centres were included in the refinement. The vibrational band centres measured by Halonen et al. [94] were not included due to the large estimated uncertainty.

Weights of $w_n = 0.1$ were distributed to all experimentally derived rovibrational states except for pure rotational states, which were given weights of 1000.0,
and the $2\nu_4^{\prime}=0$ band of Yang et al. [297], for which we struggled to match experimental energies to our calculated energies owing to conflicting quantum labels, and so gave a weight of 0.0. These were adjusted on-the-fly using Watson’s robust fitting scheme [277]. A scaling factor of $k = 1 \times 10^{-4}$ was initially applied to the 39 678 ab initio points included in the fit. As the refinement progressed this was incrementally decreased to $1 \times 10^{-6}$ so as to reduce the relative contribution of the ab initio data. Care was taken throughout to ensure the refined PES did not deviate substantially from the ab initio surface, and we note that for all ab initio grid points, the energy difference between the refined and geometry optimised ab initio PES’s is less than 10% that of the ab initio PES above its zero-point energy (ZPE). Our final fitted PES is called AsH$_3$-CYT18 below.

### 4.6 Dipole moment surface

The electric dipole moment was approximated using a numerical finite-difference procedure as outlined in Section 2.3.2. A field strength of 0.002 a.u. was deemed sufficiently small to accurately approximate the first derivative without approaching numerical noise [51]. As with our PES, electronic structure calculations were carried out at the CCSD(T)-F12b level of theory with a cc-pVQZ-PP-F12 basis set for the arsenic atom, and cc-pVQZ F12 for the hydrogens. Due to the sevenfold increase in computational demand of the DMS over the PES, dipole moments were calculated on a reduced grid of 10 000 points, generated by randomly sampling our PES grid.

Our ab initio DMS was expressed analytically using the symmetrized molecular bond (SMB) representation (see Section 2.3.2). The final fit required 261 parameters and reproduced the ab initio data with an RMS difference of 0.0008 Debye for electronic energies up to $\hbar c \cdot 12$ 000 cm$^{-1}$, which is comparable to the level of numerical noise in the finite differences procedure.
4.7 Results

4.7.1 Structural parameters

Table 4.3 shows the various structural parameters of AsH$_3$ computed at different levels of theory, compared to those of our refined PES and those derived from experiment. *Ab initio* calculations of the equilibrium values of $r$ and $\alpha$ were performed using the geometry optimisation procedure in MOLPRO. Both VQZ-PP-F12 and AVQZ-DK level calculations are seen to somewhat overestimate $r_{\text{eq}}$ and $\alpha_{\text{eq}}$ when compared experiment, a feature that is exacerbated by the exclusion of relativistic effects altogether ($r_{\text{eq}}^{\text{AVQZ}} = 1.52375$ Å, $\alpha_{\text{eq}}^{\text{AVQZ}} = 92.5553^\circ$ and $r_{\text{eq}}^{\text{AVSZ}} = 1.523653$ Å, $\alpha_{\text{eq}}^{\text{AVSZ}} = 92.54910^\circ$). As expected, the effect of our equilibrium geometry adjustment results in equilibrium bond lengths and angles much closer to those of experiment [41]. This is reflected in the good agreement between our purely rotational energies and spin-statistics averaged hyperfine resolved rotational energies of Tarrago *et al.* [253] (see Table 4.4). There are small systematic residuals as large as 0.01 cm$^{-1}$, suggesting our treatment of the rotational motion could be improved by further tweaking the equilibrium parameters. However, doing so would undoubtedly spoil the vibrational accuracy so we decided against it.

<table>
<thead>
<tr>
<th>AsH$_3$-CYT18</th>
<th>VQZ-PP-F12</th>
<th>AVQZ-DK</th>
<th>AV5Z-DK</th>
<th>Exp [41]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{eq}}$ / Å</td>
<td>1.511394</td>
<td>1.520269</td>
<td>1.521481</td>
<td>1.520432</td>
</tr>
<tr>
<td>$\alpha_{\text{eq}}$ /°</td>
<td>92.04025</td>
<td>92.21595</td>
<td>92.17049</td>
<td>92.18705</td>
</tr>
<tr>
<td>$r_{\text{SP}}$ / Å</td>
<td>1.4688</td>
<td>1.4663</td>
<td>1.4670</td>
<td></td>
</tr>
<tr>
<td>$\Delta E$(barrier) / cm$^{-1}$</td>
<td>14495.</td>
<td>14187.</td>
<td>14171.</td>
<td></td>
</tr>
</tbody>
</table>

As yet, the inversion barrier height $\Delta E$(barrier) of AsH$_3$ remains unmeasured. The previous highest-level predictions are those by Schwerdtfeger *et al.* [228] in 1992, who calculated a value of 13079.3 cm$^{-1}$ at the second-order Møller-Plesset (MP2) level of theory. This is somewhat lower than our CCSD(T) values of just over 14 000 cm$^{-1}$, shown in Table 4.3. MP2 is the least computationally expensive *ab initio* method to account for electron correlation effects, and a number of disadvantages have been noted, e.g., in Ref. [57] and the references therein, that generally make coupled-cluster (CC) methods preferable nowadays. Aside from this, the dif-
ference in barrier heights between our own predictions and those of Schwerdtfeger et al. [228] may be explained by their use of a substantially smaller basis set and their neglect of relativistic corrections.

The minimum energy path over the barrier reduces the As-H bond lengths to their so-called saddle-point value \( r_{SP} \) at planar geometry. Of this, the predicted value of 1.457 Å by Schwerdtfeger et al. is in reasonable agreement with our own (see Table 4.3). For comparison, the \( \text{NH}_3 \) barrier height is measured to be 1786.8 cm\(^{-1}\) occurring for \( r_{SP} = 0.99460 \) Å [205], and for \( \text{PH}_3 \) the calculated values of Sousa-Silva et al. [240] are currently the most reliable, predicting a value of 11 130 cm\(^{-1}\) at 1.3611 Å.

### 4.7.2 Rovibrational energies

Rovibrational energy level calculations were performed up to \( J = 30 \) using the \( \text{AsH}_3 \)-CYT18 PES in conjunction with the nuclear motion program TROVE. Model input parameters were kept the same as reported in Section 4.4, including our \( P_{\text{max}} = 14 \) vibrational basis. With a basis set of this size the vibrational Hamiltonian \( E \)-symmetry block has 2571 roots. Therefore, given the \( 2J+1 \) multiplication factor for rotationally exited states, it was necessary to perform additional basis set truncations to reduce computational cost. Firstly, our purely vibrational energies \( E_{\text{vib}}^i \) were truncated at \( \hbar c \cdot 12 \) 000 cm\(^{-1}\). These, upon multiplication with rigid symmetric rotor wavefunctions, form the basis for our full rovibrational calculation, which we term the \((J = 0)\)-contracted basis. Our second truncation, performed only once \( J \) exceeded 21, is therefore to remove all \((J = 0)\)-contracted eigenfunctions with energy greater than \( E_{\text{vib}}^i + E_{\text{rot}}^i = \hbar c \cdot 16 \) 000 cm\(^{-1}\), where \( E_{\text{rot}}^i \) are eigenvalues of a symmetric rigid rotor.

Our complete list of calculated energies is available from the ExoMol website (www.exomol.com), along with associated local mode quantum numbers \((n_1, n_2, n_3, n_4, n_5, n_6, \Gamma_{\text{vib}}, J, K, \Gamma_{\text{rot}}, \Gamma_{\text{tot}})\). Here \((n_1, n_2, n_3)\) are stretching quantum numbers, \((n_4, n_5, n_6)\) are bending quantum numbers, \( K \) is the projection of the total rotational angular momentum \( J \) onto the molecular axis of symmetry, and \((\Gamma_{\text{vib}}, \Gamma_{\text{rot}}, \Gamma_{\text{tot}})\) are the vibrational, rotational and total symmetry in \( C_{3v} \). The local
Table 4.4: Differences between calculated rotational term values, in cm$^{-1}$, and the hyperfine resolved values of [253] which we averaged using the spin statistical weights.

<table>
<thead>
<tr>
<th>J</th>
<th>K</th>
<th>Sym</th>
<th>Obs</th>
<th>Obs−Calc.ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>A$_2$</td>
<td>7.503018</td>
<td>-0.000368</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>E</td>
<td>7.249824</td>
<td>-0.000381</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>E</td>
<td>22.253842</td>
<td>-0.001287</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>E</td>
<td>21.494930</td>
<td>-0.000807</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>A$_2$</td>
<td>45.005427</td>
<td>-0.003161</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>E</td>
<td>44.753718</td>
<td>-0.002935</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>E</td>
<td>43.997254</td>
<td>-0.002248</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>A$_1$</td>
<td>42.732027</td>
<td>-0.001078</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>A$_2$</td>
<td>42.732025</td>
<td>-0.001077</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>E</td>
<td>74.742676</td>
<td>-0.005085</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>E</td>
<td>73.989211</td>
<td>-0.004375</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>A$_1$</td>
<td>72.729005</td>
<td>-0.003149</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>A$_2$</td>
<td>72.728988</td>
<td>-0.003184</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>E</td>
<td>70.955354</td>
<td>-0.001422</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>A$_2$</td>
<td>112.460919</td>
<td>-0.007935</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>E</td>
<td>112.211432</td>
<td>-0.007705</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>E</td>
<td>111.461647</td>
<td>-0.007010</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>A$_1$</td>
<td>110.207557</td>
<td>-0.005824</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>A$_2$</td>
<td>110.207557</td>
<td>-0.005824</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>E</td>
<td>108.442621</td>
<td>-0.004109</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>E</td>
<td>106.157387</td>
<td>-0.001816</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>E</td>
<td>157.148421</td>
<td>-0.010773</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>E</td>
<td>156.403009</td>
<td>-0.010103</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>A$_1$</td>
<td>155.156374</td>
<td>-0.008751</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>A$_2$</td>
<td>155.156164</td>
<td>-0.009170</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>E</td>
<td>153.401598</td>
<td>-0.007307</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>E</td>
<td>151.129706</td>
<td>-0.005092</td>
</tr>
</tbody>
</table>

Mode vibrational quantum numbers can be converted to the normal mode representation using symmetry rules (see Down et al. [71]), under the assumption that the total number of stretching and bending quanta are conserved between representations. We produced a list of calculated vibrational states that have been converted to the normal mode representation for $n_1 + n_2 + n_3 \leq 4$ and $n_4 + n_5 + n_6 \leq 4$. This covers all strong bands under 7000 cm$^{-1}$, and should aid any future labelling of AsH$_3$ spectra.

Table 4.5 compares the calculated $J = 0$ term values under 7000 cm$^{-1}$, com-
Table 4.5: Agreement between our calculated energy levels and those derived from experiment. All calculations used our refined PES, AsH$_3$-CYT18. $J = 0$ comparisons are before employing the EBSC, and $J = 1 - 6$ comparisons are afterwards. Term values and their RMS statistics are given in cm$^{-1}$.

<table>
<thead>
<tr>
<th>Band</th>
<th>Symmetry</th>
<th>$J = 0$</th>
<th>$J = 1 - 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Obs.</td>
<td>Calc.</td>
</tr>
<tr>
<td>$\nu_2$</td>
<td>A$_1$</td>
<td>906.752</td>
<td>906.109</td>
</tr>
<tr>
<td>$\nu_4$</td>
<td>E</td>
<td>999.225</td>
<td>998.833</td>
</tr>
<tr>
<td>$2\nu_2 + \nu_4$</td>
<td>A$_1$</td>
<td>1806.149</td>
<td>1806.161</td>
</tr>
<tr>
<td>$\nu_2$</td>
<td>A$_1$</td>
<td>1904.115</td>
<td>1904.046</td>
</tr>
<tr>
<td>$2\nu_4^{l=0}$</td>
<td>A$_1$</td>
<td>1990.998</td>
<td>1990.293</td>
</tr>
<tr>
<td>$2\nu_4^{l=2}$</td>
<td>E</td>
<td>2003.483</td>
<td>1997.315</td>
</tr>
<tr>
<td>$\nu_1 + \nu_2$</td>
<td>A$_1$</td>
<td>3013$^a$</td>
<td>3016.531</td>
</tr>
<tr>
<td>$\nu_1 + \nu_4$</td>
<td>E</td>
<td>3102$^a$</td>
<td>3100.438</td>
</tr>
<tr>
<td>$2\nu_1$</td>
<td>A$_1$</td>
<td>4166.772</td>
<td>4166.694</td>
</tr>
<tr>
<td>$\nu_1 + \nu_3$</td>
<td>E</td>
<td>4167.935</td>
<td>4167.877</td>
</tr>
<tr>
<td>$2\nu_3^{l=0}$</td>
<td>A$_1$</td>
<td>4237.700</td>
<td>4237.407</td>
</tr>
<tr>
<td>$2\nu_3^{l=2}$</td>
<td>E</td>
<td>4247.720</td>
<td>4247.842</td>
</tr>
<tr>
<td>$2\nu_1 + \nu_2$</td>
<td>A$_1$</td>
<td>5057$^a$</td>
<td>5040.690</td>
</tr>
<tr>
<td>$\nu_1 + \nu_2 + \nu_3$</td>
<td>E</td>
<td>5057$^a$</td>
<td>5040.799</td>
</tr>
<tr>
<td>$2\nu_1 + \nu_4$</td>
<td>E</td>
<td>5128$^a$</td>
<td>5129.956</td>
</tr>
<tr>
<td>$2\nu_3^{0} + \nu_2$</td>
<td>A$_1$</td>
<td>5128$^a$</td>
<td>5131.122</td>
</tr>
<tr>
<td>$\nu_1 + \nu_3 + \nu_4$</td>
<td>E</td>
<td>5158$^a$</td>
<td>5155.741</td>
</tr>
<tr>
<td>$\nu_1 + \nu_3 + \nu_4$</td>
<td>A$_1$</td>
<td>5158$^a$</td>
<td>5156.434</td>
</tr>
<tr>
<td>$3\nu_1$</td>
<td>A$_1$</td>
<td>6136.340</td>
<td>6136.846</td>
</tr>
<tr>
<td>$2\nu_1 + \nu_3$</td>
<td>E</td>
<td>6136.330</td>
<td>6136.859</td>
</tr>
<tr>
<td>$\nu_i + 2\nu_4^{l=0}$</td>
<td>A$_1$</td>
<td>6275.830</td>
<td>6275.814</td>
</tr>
<tr>
<td>$\nu_1 + 2\nu_3^{l=2}$</td>
<td>E</td>
<td>6282.350</td>
<td>6282.414</td>
</tr>
<tr>
<td>$3\nu_3^{l=1}$</td>
<td>E</td>
<td>6294.710</td>
<td>6294.695</td>
</tr>
<tr>
<td>$3\nu_3^{l=3}$</td>
<td>A$_1$</td>
<td>6365.950</td>
<td>6365.759</td>
</tr>
</tbody>
</table>

$^a$ experimental uncertainties of Halonen et al. [94] are estimated to be 2 cm$^{-1}$ or more.
4.7. Results

\( J = 1 - 6 \) states belonging to the \( 2\nu_4^\pm = 2 \) band, which fall within \( \pm 1.0 \) cm\(^{-1} \) of experiment, we strongly suspect the empirical band origin \(^{297}\) of 2003.483 cm\(^{-1} \) is incorrect. From a comparison of our \( J = 1, 2 \) energy residuals, we expect the true value to be closer to 1997.5 cm\(^{-1} \). Interestingly, for the bands at 3000 and 5000 cm\(^{-1} \) measured by Halonen \textit{et al.} \(^{94}\), all calculated \( J = 0 \) term values, except for the \( 2\nu_1 + \nu_2 \) and \( \nu_1 + \nu_2 + \nu_3 \) bands, fall within a few wavenumbers of experiment despite being omitted from the refinement. This illustrates the interpolative power of the refinement, and suggests that even bands not yet observed experimentally should be predicted with reasonable accuracy by our refined PES. Alternative matches for the 5057 cm\(^{-1} \) bands within our energies list would be the \( \nu_1 + 3\nu_4^1 \) (predicted at 5052.561 cm\(^{-1} \)) and \( \nu_3 + 3\nu_4^1 (A_2) \) (5052.758 cm\(^{-1} \)) bands. However, considering that these bands are not predicted to be observable at room temperature, the discrepancies are more likely due to deficiencies in our PES.

The residual differences between our calculated \( J = 0 \) term values and those of experiment can be removed from the final line list by utilising an empirical basis set correction (EBSC) \(^{304}\), whereby our calculated band centres are simply replaced by the corresponding experimental values. We employed the EBSC for all experimentally known bands taken from \(^{219}\), except the suspicious \( 2\nu_4^\pm = 2 \) band. Figure \(^{4.2}\) displays the difference between our calculated energies and those derived from experiment for states with \( J \leq 6 \) taken from Refs. \(^{253, 267, 268, 276, 296, 297}\) after employing the EBSC. The corresponding root-mean-square errors (\( \sigma_{\text{rms}}^{\text{ebsc}} \)), split by vibrational band, are shown in Table \(^{4.5}\). Although there is some deterioration in quality with \( J \), this is slow and systematic in most cases, reassuring us that our calculations can safely be extended to higher rotational excitations. Agreement for the \( 2\nu_2 \) and stretching bands is particularly pleasing, and all calculated \( J = 1 - 6 \) term values, bar those belonging to the \( 2\nu_3^2 \) band, are calculated to within \( \pm 0.2 \) cm\(^{-1} \) of the experimental values. Judging by the systematic offset of the \( 2\nu_3^2 \) band in Figure \(^{4.2}\) the experimental band centre used in the EBSC is most likely \( \approx 0.2 \) cm\(^{-1} \) lower than the true value. Slightly larger \( \sigma_{\text{rms}}^{\text{ebsc}} \) values are observed for the \( 2\nu_4^0, 2\nu_2^2 \) and \( \nu_2 + \nu_4 \) bands. Whereas the \( 2\nu_4^2 \) and \( \nu_2 + \nu_4 \) bands display clear \( J - K \)
Figure 4.2: Agreement between observed $J = 1 - 6$ term values $E_{\text{obs}}$ and the calculated values of this work $E_{\text{calc}}$ using our refined PES and the EBSC. The $2\nu_2$, $\nu_2 + \nu_4$, $2\nu_1$ and $\nu_3$ bands (upper plot) were taken from [267]; the $2\nu_4^0$ and $2\nu_4^2$ bands (upper plot) were taken from [268]; the $2\nu_1$ and $\nu_1 + \nu_3$ bands (middle plot) were taken from [297]; the $2\nu_3^0$ and $2\nu_3^2$ bands (middle plot) were taken from [296]; and the $3\nu_1^3$, $3\nu_3^3$, $\nu_1 + 2\nu_2^2$ and $\nu_1 + 2\nu_3^0$ bands (bottom plot) were taken from reference [276].
dependencies, it is difficult to discern any such trends for the $2\nu_4^0$ band, which was omitted from the refinement altogether. Two possible reasons for this are either corrupt experimental data, or perturbation interactions due to nearby states that are not correctly represented by our PES. Even so, the 0.207 cm\(^{-1}\) root-mean-square error is very reasonable.

### 4.7.3 Line intensity predictions

To simulate absolute absorption intensities we use the expression given in Eq. (2.71). The dipole moment operator transforms with $A_2$ symmetry in $C_{3v}$, and so the rigorous selection rules for rovibrational transitions are $A_1 \leftrightarrow A_2$ and $E \leftrightarrow E$. The spin statistical weights can be derived by first using the equations of Landau and Lifshitz to find the characters of the representation $\Gamma_{\text{sw}}^{\text{ve}}$ spanned by the rovibrational states, then reducing this character representation to its irreducible components (see [35] for details) whose coefficients are the spin–statistical weights. Given that $^{75}\text{As}$ has nuclear spin $I_{\text{As}} = \frac{3}{2}$ and H has nuclear spin $I_H = \frac{1}{2}$, the corresponding statistical weights are found to be \{16,16,16\} for states of \{A_1,A_2,E\} symmetry, and so the total degeneracy of a state is $g_w = 16(2J_w + 1)$.

![Figure 4.3: The partition functions $Q_{J_{\text{max}}}$ of AsH\(_3\) at different temperatures versus the maximum $J$ value used in Eq. (2.72)](image-url)
No calculated or experimentally derived values of the partition function could be found in the literature, so we computed partition function values for temperatures ranging from 10 to 500 K in intervals of 10 K. Fig. 4.3 illustrates the convergence of $Q$ as the rotational basis is increased from including only $J = 0$ states ($J_{\text{max}} = 0$), to all computed states with $J \leq 30$ ($J_{\text{max}} = 30$). In reality there will be additional contributions from our vibrational basis ($P_{\text{max}} = 14$) and PES, although these are difficult to quantify. The room temperature partition function was calculated to be $Q(T = 296) = 8250.2801$ using $J_{\text{max}} = 30$, which we estimate to be better than 99% converged.

Figure 4.4: Overview of complete $J = 0 – 30$ line list computed at 296 K.
Table 4.6: Comparison of observed and calculated band intensities. Column 1 refers to the local mode quantum numbers assigned by TROVE, where sym is the total symmetry. The units of intensity are $10^{-18} \text{cm}^{-1}/(\text{molec cm}^{-2})$. The value under / is the total intensity of the bands with the same quantum numbers $n_1n_2n_3$.

<table>
<thead>
<tr>
<th>$(n_1, n_2, n_3; \text{sym})$</th>
<th>band</th>
<th>band centre</th>
<th>$I_{\text{obs}}$</th>
<th>$I_{\text{calc}}$</th>
<th>$I_{\text{calc (this work)}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(100; $A_1$)</td>
<td>$v_1$</td>
<td>2115.164</td>
<td>11.4/44.1</td>
<td>10.7/40.4</td>
<td>11.2/44.9</td>
</tr>
<tr>
<td>(100; $E$)</td>
<td>$\nu_3$</td>
<td>2126.432</td>
<td>29.7</td>
<td>32.7</td>
<td>33.7</td>
</tr>
<tr>
<td>(200; $A_1$)</td>
<td>$2\nu_1$</td>
<td>4166.772</td>
<td>/0.618</td>
<td>0.157/0.427</td>
<td>0.225/0.722</td>
</tr>
<tr>
<td>(200; $E$)</td>
<td>$\nu_1 + \nu_3$</td>
<td>4167.935</td>
<td>–</td>
<td>0.270</td>
<td>0.497</td>
</tr>
<tr>
<td>(110; $A_1$)</td>
<td>$2\nu_3^{l=0}$</td>
<td>4237.700</td>
<td>–</td>
<td>0.0117/0.0123</td>
<td>0.0143/0.0163</td>
</tr>
<tr>
<td>(110; $E$)</td>
<td>$2\nu_3^{l=2}$</td>
<td>4247.720</td>
<td>–</td>
<td>0.000671</td>
<td>0.00201</td>
</tr>
<tr>
<td>(300; $A_1$)</td>
<td>$3\nu_1$</td>
<td>6136.340</td>
<td>/0.00989</td>
<td>0.00456/0.00656</td>
<td>0.00337/0.00548</td>
</tr>
<tr>
<td>(300; $E$)</td>
<td>$2\nu_1 + \nu_3$</td>
<td>6136.330</td>
<td>–</td>
<td>0.00200</td>
<td>0.00211</td>
</tr>
<tr>
<td>(210; $A_1$)</td>
<td>$\nu_1 + 2\nu_3^{l=0}$</td>
<td>6275.830</td>
<td>/0.00275</td>
<td>0.000112/0.00182</td>
<td>0.000734/0.00116</td>
</tr>
<tr>
<td>(210; $E$)</td>
<td>$\nu_1 + 2\nu_3^{l=2}$</td>
<td>6282.350</td>
<td>–</td>
<td>0.000104</td>
<td>0.0000741</td>
</tr>
<tr>
<td>(210; $E$)</td>
<td>$3\nu_3^{l=1}$</td>
<td>6294.710</td>
<td>–</td>
<td>0.000596</td>
<td>0.000356</td>
</tr>
<tr>
<td>(210; $E$)</td>
<td>$3\nu_3^{l=3}$</td>
<td>6365.950</td>
<td>–</td>
<td>0.0000650</td>
<td>0.0000934</td>
</tr>
<tr>
<td>(111; $A_1$)</td>
<td>$3\nu_1$</td>
<td>6365.950</td>
<td>–</td>
<td>0.0000650</td>
<td>0.0000934</td>
</tr>
</tbody>
</table>
Table 4.7: Comparison of calculated and observed line positions (cm\(^{-1}\)) and intensities (cm\(^{-1}/(\text{molecules cm}^{-2})\)) belonging to the \(\nu_1\) and \(\nu_3\) bands.

| \(J'\) | \(K'\) | Sym' | \(J''\) | \(K''\) | Sym'' | Band | \(v_{\text{obs}}\) \([62]\) | \(I_{\text{obs}}\) \([62]\) | \(v_{\text{calc}}\) | \(I_{\text{calc}}\) | \(\%|I_{\text{obs}} - I_{\text{calc}}|\) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 6 | E | 10 | 7 | E | \(\nu_3\) | 2051.894 | \(4.799 \times 10^{-20}\) | 2052.082 | \(4.992 \times 10^{-20}\) | 4.03 |
| 9 | 7 | E | 10 | 8 | E | \(\nu_3\) | 2052.548 | \(5.755 \times 10^{-20}\) | 2052.767 | \(6.079 \times 10^{-20}\) | 5.63 |
| 7 | 1 | A\(_2\) | 8 | 0 | A\(_1\) | \(\nu_3\) | 2064.460 | \(4.396 \times 10^{-20}\) | 2064.468 | \(4.773 \times 10^{-20}\) | 8.57 |
| 7 | 6 | E | 8 | 7 | E | \(\nu_3\) | 2067.961 | \(1.038 \times 10^{-19}\) | 2068.139 | \(1.110 \times 10^{-19}\) | 6.95 |
| 4 | 4 | E | 5 | 5 | E | \(\nu_3\) | 2090.433 | \(1.572 \times 10^{-19}\) | 2090.542 | \(1.699 \times 10^{-19}\) | 8.04 |
| 4 | 1 | A\(_1\) | 5 | 0 | A\(_2\) | \(\nu_3\) | 2088.098 | \(5.876 \times 10^{-20}\) | 2088.087 | \(6.423 \times 10^{-20}\) | 9.31 |
| 5 | 4 | E | 6 | 5 | E | \(\nu_3\) | 2082.601 | \(1.250 \times 10^{-19}\) | 2082.714 | \(1.320 \times 10^{-19}\) | 5.64 |
| 3 | 3 | E | 4 | 4 | E | \(\nu_3\) | 2097.659 | \(1.462 \times 10^{-19}\) | 2097.738 | \(1.571 \times 10^{-19}\) | 7.49 |
| 2 | 2 | E | 1 | 1 | E | \(\nu_3\) | 2140.716 | \(8.659 \times 10^{-20}\) | 2140.678 | \(9.453 \times 10^{-20}\) | 9.17 |
| 2 | 1 | A\(_1\) | 1 | 0 | A\(_2\) | \(\nu_3\) | 2141.069 | \(8.949 \times 10^{-20}\) | 2141.053 | \(9.612 \times 10^{-20}\) | 7.41 |
| 6 | 6 | E | 5 | 5 | E | \(\nu_3\) | 2168.331 | \(1.965 \times 10^{-19}\) | 2168.331 | \(2.093 \times 10^{-19}\) | 6.48 |
| 8 | 5 | E | 7 | 5 | E | \(\nu_1\) | 2172.196 | \(3.114 \times 10^{-20}\) | 2172.262 | \(3.301 \times 10^{-20}\) | 5.99 |
| 8 | 7 | E | 9 | 7 | E | \(\nu_1\) | 2045.190 | \(2.459 \times 10^{-20}\) | 2045.261 | \(2.326 \times 10^{-20}\) | 5.40 |
| 8 | 8 | E | 9 | 8 | E | \(\nu_1\) | 2045.319 | \(1.310 \times 10^{-20}\) | 2045.397 | \(1.328 \times 10^{-20}\) | 1.36 |
| 10 | 7 | E | 9 | 7 | E | \(\nu_1\) | 2185.605 | \(2.026 \times 10^{-20}\) | 2185.716 | \(2.114 \times 10^{-20}\) | 4.35 |
Line list calculations were performed using the AsH$_3$-CYT18 PES and the cc-pVQZ-PP-F12 DMS detailed in Section 4.6. Transitions involve states with energies up to $\hbar c \cdot 10 \ 500 \ \text{cm}^{-1}$, rotational excitation up to $J = 30$, and a maximum lower state energy of $\hbar c \cdot 3500 \ \text{cm}^{-1}$. The final line list consists of 3.6 million absorption lines in the range $0 - 7000 \ \text{cm}^{-1}$ with intensity greater than $1 \times 10^{-28} \ \text{cm}^{-1}/(\text{molecule cm}^{-2})$ at 296 K. An overview is presented in Figure 4.4.

Several sources of experimental absorption data exist for AsH$_3$. In the following paragraphs our intensity calculations are validated by comparison with only the most recent and reliable sources. For the first test of our absolute intensities we compare our calculated band intensities with those obtained by Zheng et al. [319], shown in Table 4.6. Zheng et al. produced a three-dimensional DMS based on results of density functional theory calculations, and compared the resulting absolute vibrational band intensities to the values obtained by direct integration of absorption spectra, which they provide with $20-40\%$ estimated uncertainty. Due to multiple bands overlapping only the combined intensity of bands with the same local mode quanta are presented in some cases. For the $\nu_1$ and $\nu_3$ fundamentals we compare well with experiment, reproducing the observed values within $2\%$ and $14\%$ respectively. Zheng et al. only provide the measured intensity of the sum of the $2\nu_1$ and $\nu_1 + \nu_3$ bands, for which we are stronger by $17\%$. No measurements of the weaker $2\nu_3^{l=0}$ and $2\nu_3^{l=2}$ bands are given, most likely due to difficulties resolving them without accurate theoretical line positions. Finally, for the three-quantum stretches, our calculated intensities are typically 2-3 times weaker than the measured values. However, it is difficult to estimate the reliability of these measurements, given the recorded spectrum is only medium-resolution ($\Delta \nu = 0.2 \ \text{cm}^{-1}$) and the bands are weak.

Dana et al. [62] measured absolute intensities of 387 lines belonging to the $\nu_1$ and $\nu_3$ bands. Their line measurements range from $2010 - 2235 \ \text{cm}^{-1}$ although they make no attempt to measure the $Q-$branch from $2110 - 2140 \ \text{cm}^{-1}$, presumably owing to the density of lines. Table 4.7 compares our calculated line positions and intensities to the experimentally measured values for 14 randomly selected strong
lines measured by Dana et al. In all cases our calculated intensity values are within ±10% of experiment, although there is a slight tendency to be higher. Nevertheless, this is reassuring given our 14% discrepancy with the ν3 band as measured by Zheng et al. [319].

Figure 4.5: Overview of synthetic J = 0 − 30 spectrum computed at 298.15 K compared to PNNL for the 0−7000 cm⁻¹ region.

The PNNL database provides a composite spectrum of pure AsH3 up to 6500 cm⁻¹ measured at 5, 25 and 50°C. For comparison, we generated synthetic T = 298.15 K spectra using a J = 0 − 30 line list convoluted with a Voigt profile with half-width at half-maxima (HWHM) of 0.09 cm⁻¹. Although linewidths are well known to depend upon the upper and lower state quantum numbers, the strongest dependency being J and K, as far as we know no such data exists for AsH3, and we found the value 0.09 cm⁻¹ reasonably approximated the PNNL linewidths on average. To convert the PNNL absorption spectra to cm²/molecule a multiplication factor of 9.28697 × 10⁻¹⁶ is necessary.

Figure 4.5 shows an overview of our synthetic spectrum compared to PNNL, with the key absorption features expanded in Figs. 4.6−4.15. Qualitative agreement is very good, particularly for the ν₁/ν₃ (see Figs. 4.8 and 4.9) and ν₂/ν₄ (see Figs. 4.6 and 4.7) fundamentals, and the 2ν₁/ν₁ + ν₃/2ν₃²/2ν₃² band system (shown Figs. 4.12
4.7. Results

Figure 4.6: Expansion of synthetic $J = 0 - 30$ spectrum computed at 298.15 K compared to PNNL for the 800–1150 cm$^{-1}$ region.

Figure 4.7: Expansion of synthetic $J = 0 - 30$ spectrum computed at 298.15 K compared to PNNL for the 935–965 cm$^{-1}$ region.

and 4.13). Note that despite only including the $\nu_2$ and $\nu_4$ band centres in the refinement, their rotational structures are reproduced well. In the 2920–3260 cm$^{-1}$ region (shown Figs. 4.10 and 4.11) the dominant sources of opacity are predicted
to be the strong $\nu_1 + \nu_4$ and $\nu_3 + \nu_4$ (A$_1$) (calculated band centre 3119.400 cm$^{-1}$) bands, and the slightly weaker $\nu_2 + \nu_3$ (3023.706 cm$^{-1}$) and $\nu_1 + \nu_2$ bands. Considering that no associated experimental energies were included in the refinement,
Figure 4.10: Expansion of synthetic $J = 0 - 30$ spectrum computed at 298.15 K compared to PNNL for the 2920–3260 cm$^{-1}$ region.

Figure 4.11: Expansion of synthetic $J = 0 - 30$ spectrum computed at 298.15 K compared to PNNL for the 3110–3140 cm$^{-1}$ region.

the level of agreement is satisfying. Above 5000 cm$^{-1}$ most absorption features are lost in the PNNL background noise; only the $2\nu_1 + \nu_4$, $\nu_2 + 2\nu_3^0$ and $\nu_1 + \nu_3 + \nu_4$ bands (our labelling) are clearly visible between 5000–5200 cm$^{-1}$ (see Fig. 4.14).
Figure 4.12: Expansion of synthetic $J = 0 - 30$ spectrum computed at 298.15 K compared to PNNL for the 4035—4285 cm$^{-1}$ region.

There is a tenuous absorption bump in PNNL at 5050 cm$^{-1}$ for which we appear to be offset by roughly 15 cm$^{-1}$, confirming our discrepancies with the $2\nu_1 + \nu_2$ and $\nu_1 + \nu_2 + \nu_3$ band centres measured by Halonen et al. [94] (see Table 4.5). In the
4.7. Results

Figure 4.14: Expansion of synthetic $J = 0 - 30$ spectrum computed at 298.15 K compared to PNNL for the 5000–5300 cm$^{-1}$ region.

Figure 4.15: Expansion of synthetic $J = 0 - 30$ spectrum computed at 298.15 K compared to PNNL for the 6000–6400 cm$^{-1}$ region.

6000 – 6400 cm$^{-1}$ region (see Fig. 4.15) the salient feature is the $3\nu_1$ and $2\nu_1 + \nu_3$ Q-branch at 6135 cm$^{-1}$, for which we clearly underestimate the intensity. Although in line with our comparisons with Zheng et al. [319] (see Table 4.6) it is difficult
to quantify this, or indeed draw any conclusions regarding the weaker $\nu_1 + 2\nu_3/3\nu_3$
bands, without additional high-resolution experimental data.

The largest source of error in our intensity calculations will undoubtedly be
the DMS. To improve on the CCSD(T)-F12b/cc-pVQZ-PP-F12 method by Hill et
al. [106], large CCSD(T)/aug-cc-pwCVnZ-DK ($n = 4, 5$) calculations would likely
be necessary (for an example, including additional post-CCSD(T) corrections see
Ref. [65]), which are currently computationally unmanageable for a full 6D surface.
Secondarily, the PES quality must be considered. Line intensities are inexorably
connected to the PES through the wavefunctions in Eq. (2.71), and accurate mod-
elling of intensity transfer between lines (so-called ‘intensity stealing’) relies on the
correct representation of rotation-vibration resonances in the PES. Therefore, from
a nuclear motion point-of-view, high-resolution measurements complete with line
intensities and quantum assignments, particularly for the 800 – 1200, 2900 – 3300
and 5000 – 5300 cm$^{-1}$ regions, would be most beneficial for future modelling.

4.8 Conclusion

We have produced the first full-dimensional PES and DMS for the arsine molecule.
Both PES and DMS were computed at the CCSD(T)-F12b/cc-pVQZ-PP-F12 level
of theory, with implicit treatment of scalar relativistic effects via replacement of
10 core electrons with a relativistic pseudopotential. A comparison with standard
CCSD(T)/aug-cc-pVQZ-DK based calculations employing the DKH8 Hamiltonian,
showed that CCSD(T)-F12b/cc-pVQZ-PP-F12 level theory resulted in significantly
more accurate nuclear motion calculations.

Geometry optimisation and empirical adjustment of harmonic and certain cu-
bic terms in the pVQZ-PP-F12 PES resulted in $J = 1 – 6$ rotational term values
with a root-mean-square error of 0.0055 cm$^{-1}$, and vibrational term values accu-
rate to within 1 cm$^{-1}$ for all reliably known experimental band centres under 6400
cm$^{-1}$. Utilising the empirical basis set correction scheme, 578 experimentally de-
derived ($J = 1 – 6$) rovibrational energies are reproduced with an RMS of 0.122 cm$^{-1}$.
Vibrational term value comparisons with eight approximately known band centres
showed that six agreed within 3.5 cm$^{-1}$ despite being omitted from the refinement. The remaining two displayed $\sim 16$ cm$^{-1}$ discrepancies, most likely due to deficiencies in our PES.

Rotational-vibrational line intensity calculations were performed using the refined PES and \textit{ab initio} DMS, in conjunction with variational nuclear motion calculations. The resulting line list, with full quantum assignments, extends to 7000 cm$^{-1}$ and is complete up to 300 K. Comparisons with multiple experimental sources show our intensity predictions to be reliable, in particular, good overall agreement with the main absorption features present in PNNL is noted.

As far as we know, arsenic is the heaviest element for which there exists an associated variationally-computed infrared molecular line list. Considering that the quantum chemistry methods employed here are available for most p-block main group elements \cite{106}, the outlook for studying similar systems in future is positive. Additionally, we note that the study of deuterated arsine could be conducted using the same adiabatic PES, although the AsH$_2$D and AsD$_2$H isomers belong to the molecular symmetry group C$_{2v}$(M) and so require a different symmetrisation procedure in TROVE. It is unlikely a hot line list will be produced in the same manner as NH$_3$, owing to the huge computational expense.
Chapter 4. Arsine ($^{75}$AsH$_3$)
Chapter 5

Summary and outlook

In this thesis I have reported new high-accuracy room temperature line lists for the molecules $^{75}\text{AsH}_3$ and $^{14}\text{NH}_3$, as well as exploratory measurements and analysis of $\text{NH}_3$ spectra in the weak 7169–7195 cm$^{-1}$ region.

In Chapter 3, a new ‘spectroscopic’ potential energy surface (PES) for $^{14}\text{NH}_3$ has been generated by refinement of a high accuracy $ab$ $initio$ PES to experimental data. The quality of the PES is reflected in the excellent agreement between the calculated rovibrational energy levels and the empirical values of MARVEL. The PES is used in conjunction with two different DMSs to generate room temperature line lists up to 12 000 and 20 000 cm$^{-1}$, including all transitions from lower states up to 4000 cm$^{-1}$, with total angular momentum up to $J = 20$. It is my hope that these line lists, and the C2018 energy levels list, will be useful for future spectroscopic investigations, particularly in the missing 5800–6200 cm$^{-1}$ window, and above 8000 cm$^{-1}$. It should be emphasised that the C2018 energies list is of comparable accuracy to the HSL-pre3 energies list produced by Xinchuan Huang and the NASA Ames group [112, 247]. The advantages that our energies list poses over their work are i) our energies list extends to $J = 20$, whereas theirs is truncated at $J = 10$, and ii) our list contains full normal mode quantum labels, as recommened by Ref. [71], that accompany each energy.

The second important achievement of Chapter 3 was the characterisation of $^{14}\text{NH}_3$ spectra in the 7169–7195 cm$^{-1}$ region, and the derivation of line positions for a number of possibly interfering lines in trace moisture measurements. Several
extensions to the study presented here have already been discussed in Section 3.6, although unfortunately it would appear that assignment of any of the measured lines is out of the question, for the moment at least.

In Chapter 4, I have presented the first fully-dimensional PES, DMS and line list calculations for $^{75}\text{AsH}_3$, and I expect my work regarding this molecule is essentially finished. The requirements of Servomex for the arsine line list were that line positions should be accurate to within 1 cm$^{-1}$, which is the scanning range of their gas analyser system, and line intensities should be accurate to within $\sim 40\%$. These criteria were to be met at room temperature for the stretching fundamentals and first overtones, which they clearly are. Any possibility of measuring arsine spectra at the Servomex headquarters is unlikely due its highly toxic nature. If there is a demand, there is the option to compute line lists for the deuterated species AsH$_2$D and AsD$_2$H, which have C$_2v$ symmetry, using the same adiabatic PES. Moreover, for ‘hot’ exoplanet applications the computation of a hot line list might be considered, although this would be one of the most computationally expensive line lists that has been attempted as part of the ExoMol project.

5.0.1 CoYuTe

Our next aim regarding $^{14}\text{NH}_3$ is the completion of the ‘hot’ line list CoYuTe, which uses the same PES and DMS employed in the construction of the C2018 line list. However, CoYuTe differs in several notable ways from C2018 and BYTe, and improves on both in terms of coverage and accuracy.

CoYuTe was constructed to cover wavenumbers up to 20 000 cm$^{-1}$, for temperatures up to 1500 K. To this end, transitions from all states with energies up to 11 000 cm$^{-1}$ above the ground state have been considered which involved computing rotational excitations up to $J = 43$; BYTe only considered lower state energies up to 8000 cm$^{-1}$ and $J \leq 36$. Comparisons with the high temperature partition functions of [238], which are presented in Fig. 5.1, suggest that these parameters are more than sufficient to cover temperatures up to 1500 K.

As with the C2018 line list, the maximum upper state energy considered in CoYuTe was 23 000 cm$^{-1}$, which means that the complete representation of the hot
Figure 5.1: Ratio of $Q_{\text{CoYuTe}}/Q_0$ for temperatures $10 - 1600$ K, where $Q_{\text{CoYuTe}}$ is the partition function computed with the CoYuTe energies with $E_{\text{max}} = 11\,000 \text{ cm}^{-1}$ and $J_{\text{max}} = 43$, and $Q_0$ is the high temperature partition functions of [238].

spectrum will be obtained for wavenumbers below $12\,000 \text{ cm}^{-1}$ but for wavenumbers above this value there will be some loss of opacity at higher temperatures. However, considering the C2018 PES was tuned to only 10 rovibrational levels above $12\,000 \text{ cm}^{-1}$, we cannot realistically justify an extension to higher energies. We note that to provide a similar level of completeness for transitions up to $20\,000 \text{ cm}^{-1}$ as we do at $12\,000 \text{ cm}^{-1}$, all upper states with energies up to approximately $\hbar c \cdot 31\,000 \text{ cm}^{-1}$ must be considered. Convergence properties of the $(J = 20)$ rovibrational term values at $23\,000 \text{ cm}^{-1}$ has already been illustrated in Fig.3.3. By the same argument, we expect it necessary to include energies approaching $\hbar c \cdot 40\,000 \text{ cm}^{-1}$ in our $(J = 0)$-contracted basis, which is approximately where NH$_3$ dissociates. This would, nevertheless, necessitate the diagonalisation of impractically large matrices, and require an appropriate global PES that remains physical at such energies. A possible way to ascertain the completeness of CoYuTe above $12\,000 \text{ cm}^{-1}$, would be to compute vibrational transition moments, which only rely on solving the $J = 0$ problem so are relatively inexpensive calculations to perform.

CoYuTe will also make use of the full list of empirically derived energy levels present in the $^{14}$NH$_3$ MARVEL database [3,87]. MARVEL currently represents the most complete, and most accurate, source of empirical energy levels available.
in the literature. However, a number of problems regarding the original set of energy levels and transitions were first noticed during the initial refinement of C2018 in Section 3.3 (and separately in Ref. [59]), and since then it has been an ongoing project to find and resolve any outstanding issues with the MARVEL dataset. Through detailed analysis and comparison of the full set of 4938 MARVEL energies with the current highest level theoretical predictions provided by the C2018 and HSL-pre3 [112, 247] energies lists, the reliability of every MARVEL level is now guaranteed. Moreover, the three energy levels lists are fully compatible, with a one-to-one correspondence between every MARVEL level, and every C2018 and HSL-pre3 level now possible.

These MARVEL empirically derived energies, which extend to 7555 cm\(^{-1}\), will ultimately replace the theoretical values in the CoYuTe line list, resulting in a large number of line positions in CoYuTe being predicted with experimental accuracy. To illustrate this contribution to CoYuTe, a \(T = 296\) K MARVEL line list, computed using the current (updated) \(J \leq 20\) MARVEL energies in conjunction with the C2018 line intensities, is shown in Fig.5.2. It is intended for the full list of semi-empirical plus theoretical CoYuTe lines to form the basis of the \(^{14}\)NH\(_3\) entry in the HITEMP database [216].
Bibliography


Bibliography


[71] M. J. Down, J. Tennyson, J. Orphal, P. Chelin, and A. A. Ruth. Analysis of an $^{18}$O and D enhanced water spectrum and new assignments for the HD$^{18}$O and


