



Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

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

A new relational database structure and online interface for the HITRAN database

Christian Hill^{a,*}, Iouli E. Gordon^b, Laurence S. Rothman^b, Jonathan Tennyson^a^a Department of Physics and Astronomy, University College London, Gower Street, WC1E 6BT London, UK^b Harvard-Smithsonian Center for Astrophysics, Atomic and Molecular Physics Division, Cambridge, MA 02138, USA

ARTICLE INFO

Article history:

Received 1 March 2013

Received in revised form

15 April 2013

Accepted 24 April 2013

Available online 10 May 2013

Keywords:

Line-lists

Radiative transfer

Databases

HITRAN

ABSTRACT

A new format for the HITRAN database is proposed. By storing the line-transition data in a number of linked tables described by a relational database schema, it is possible to overcome the limitations of the existing format, which have become increasingly apparent over the last few years as new and more varied data are being used by radiative-transfer models. Although the database in the new format can be searched using the well-established Structured Query Language (SQL), a web service, HITRANonline, has been deployed to allow users to make most common queries of the database using a graphical user interface in a web page. The advantages of the relational form of the database to ensuring data integrity and consistency are explored, and the compatibility of the online interface with the emerging standards of the Virtual Atomic and Molecular Data Centre (VAMDC) project is discussed. In particular, the ability to access HITRAN data using a standard query language from other websites, command line tools and from within computer programs is described.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The HITRAN (*High Resolution Transmission*) molecular spectroscopic database was established in the 1970s (though it has its origins in a pair of monographs published in the mid-1960s) to meet the needs of early atmospheric transmission simulation codes [1]. Since then it has become the recognised international standard compilation of line-transition parameters as well as infrared and UV cross sections designed for terrestrial and planetary atmospheric modelling and simulation. It is available for free by anonymous FTP download (see <http://www.cfa.harvard.edu/HITRAN/requestnew.html>). Since the 2004 release of HITRAN [2], the line-transition data have been provided in a text-based format in which each transition is represented by a single 160-character record of fixed-width fields. These fields comprise identifying integer labels for the molecule and isotopologue, quantum number assignments and degeneracies for the upper and lower states, lower state energy, and the parameters listed in Table 1.

Each of the parameters listed in Table 1 also has associated with it integer index pointers representing a reference and estimated uncertainty.

This fixed-width format for the database is little different in concept from the original, publically-available HITRAN database released on magnetic tape [3] and the machine-readable punchcards that preceded it. However, as the quantity and variety of experimental and theoretical data relating to line-transition parameters in particular has increased, the limitations of this format as a way to store and search the HITRAN database have become clear. The 2004 release of HITRAN expanded the record format (where a record represents an individual line transition) to 160 from the 100 characters of the format that preceded it, but even this expanded format is unable to hold the range of currently-available relevant data that is necessary to fully parameterise each transition. The more important limitations are listed in the next section.

2. Limitations of the current HITRAN format

The fixed format introduced in HITRAN 2004 has the following limitations.

* Corresponding author.

E-mail address: christian.hill@ucl.ac.uk (C. Hill).

Table 1

Key parameters in the HITRAN format since the 2004 release. The temperature dependent parameters, namely S , γ_{air} , γ_{self} and δ_{air} are all given at 296 K.

Parameter	Description
$\tilde{\nu}$	Vacuum wavenumber (cm^{-1})
S	Intensity weighted by isotopologue abundance ($\text{cm}^{-1}/\text{molec cm}^{-2}$)
γ_{air}	Air-broadened half-width ($\text{cm}^{-1} \text{atm}^{-1}$)
γ_{self}	Self-broadened half-width ($\text{cm}^{-1} \text{atm}^{-1}$)
n_{air}	Temperature-dependence exponent for γ_{air}
δ_{air}	Air pressure-induced line shift ($\text{cm}^{-1} \text{atm}^{-1}$)

- Numerical precision:** The fixed-width fields in the current format specify the numerical parameter data to a pre-determined precision (for example, a 12-character field for the transition wavenumber, two decimal places in a 4-character field for n_{air})—as data is measured to greater accuracy, a way to represent this data in the HITRAN is needed.
- Line-broadening by collision partners other than ‘air’ and ‘self’:** Modern satellite remote-sensing missions require knowledge of the broadening of CO_2 lines by water vapour to correctly retrieve CO_2 amounts in the tropics [4]. Also, for the modelling of planetary atmospheres, parameters describing line-broadening by, CO_2 and H_2 for example, are needed; the current format does not have room for more parameters. The temperature-dependence of broadening parameters given in HITRAN is an exponent in a power law that has been determined only in a very small range of temperatures, normally spanning 20–50 K intervals around room temperature value. Extrapolation of the broadening parameter value to lower of higher temperatures using the same exponent may not give accurate results. It may be desirable to have different exponent values given for different temperature ranges.
- Non-Voigt lineshapes:** As the resolution and sensitivity of experimental and observational apparatuses has increased, it has become clear that the Voigt line profile does not adequately describe the absorption of radiation over a wide range of relevant temperatures and pressures, see Refs. [5,6] for example. Alternative line shapes, to account for Dicke narrowing and speed-dependence effects for example, require additional parameters beyond those that can be stored in the current HITRAN format.
- Quantum numbers:** The current HITRAN format allows for each state involved in a transition to be specified with up to 30 characters. Especially for larger molecules with many vibrational modes, this can be insufficient space for a full specification of a state’s quantum numbers, labels and symmetries. As the HITRAN database expands, more such labels are needed to accommodate polyatomics with many vibrational modes, torsional modes, open-shell species complex angular momentum coupling, etc.
- Data provenance:**
 - The original literature references for data in the HITRAN database are identified for the six parameters in Table 1 by integer IDs in the range (1–99), referenced in a supplementary document (`ref-table.pdf`). This document is difficult to maintain and, for inexperienced users, unwieldy to use, which leads many authors using HITRAN to cite the database but not the providers of the original data.
 - In between the quadrennial releases of HITRAN, updates and patches are released on the HITRAN website for users to download—that is, the most up-to-date version of the database is “pulled” by users rather than “pushed” to them automatically. Thus, different users may effectively be using different versions of the database, making it difficult to compare and reproduce research in which HITRAN has been used.
- Data consistency:** Since the lower-state energy, E'' , is recorded in a field for each transition in the current HITRAN format and because several transitions may originate from the same lower state, values in this field are necessarily repeated in the database. Furthermore, because the transition data collected in the database for a single species typically comes from many different sources in different spectral regions, there are inconsistencies in these reported values of E'' ; that is, the same lower state can be reported with different energies. This usually does not cause problems for radiative-transfer (RT) codes, since the small differences in E'' values are not significant in the calculation. Similar considerations apply to the upper state energies which are given implicitly through the relation $E' = \tilde{\nu} + E''$.
- Isotopologue IDs:** The fixed-field for the integer ID identifying the isotopologue of each species in the database is one character wide, so (starting at 1 for the most abundant isotopologue), a maximum of nine isotopologue forms can be referred to in this field. This limit has been reached in HITRAN2008 for the carbon dioxide molecule. Although alternative characters (for example, letters) could be used, such ad hoc solutions are not sustainable in the longer term; in this case, the proposed ad hoc solution preserves the existing format at the expense of breaking the RT codes (which expect to read an integer in this field) it is supposed to be serving.
- Line mixing:** It has been understood for several years that the effect of line mixing on the absorption due to molecular transitions must be properly treated in the accurate modelling of radiative transfer for many molecules (e.g. [7]). As more line mixing data become available, there is a need to integrate these data more

closely with the transitions they apply to. Currently, the HITRAN format implements a single-character ‘flag’ field indicating that such data are available (and can be downloaded separately from the HITRAN FTP site); it would be desirable to provide line mixing data automatically, if required, when the database is searched.

Finally the implementations of HITRAN up to now have not time-stamped their data, making it very difficult to precisely reconstruct previous data downloads.

3. Relational database—structure

The relational database schema described in this paper removes the above limitations and establishes a flexible and extensible way of storing and searching line-transition data. In particular, the structure of the database allows an arbitrary number of parameters to be stored as IEEE-754 double precision floating point numbers for each transition, separates the description of molecular states from the transitions themselves (‘normalisation’), and assigns a timestamp to each transition so that the state of the database at any date in the past may be recreated.

A relational database is a collection of linked tables (relations), each consisting of a set of rows (tuples) and columns (fields, attributes). Each row in a given table is associated with a unique integer *primary key* and relationships between tables are established by reference to this key: two tables are linked when a *foreign key* field in one is associated with the primary key field in another. Searching for a set of data matching a specified set of constraints is expedited by indexing the relevant attributes within the relational database management system (RDBMS) used to implement the database.

The principal tables and their relationships in the schema are depicted in Fig. 1. The fields in each table are listed, with examples, in Appendix A.

In a fully normalised implementation of the database, each state of each isotopologue would appear only once in the *State* table, and two transitions involving the same lower (or upper) state would both refer to it by its primary key integer ID. In fact, a sufficiently complete *State* table could be used to calculate the partition sum (at least for terrestrial atmospheric temperatures).

The number of parameters that may be specified for a transition is not limited by the database structure in the same way that it is in the existing HITRAN format: parameters are linked to their parent *Transition* table though a foreign-key–primary-key relationship instead of being listed in fields within the *Transition* table. In this way, provided parameters for different models of line-broadening and coupling can be distinctly and unambiguously named, they can be included in the relational database without having to choose one in defining the database schema. For example, the air-broadened Lorentzian half-width component of the Voigt lineshape is identified by the label `gamma_air`; the first-order (Rosenkranz) line coupling coefficient [8] for self-broadened lines is `y_self`. Given that the RT community have yet to agree on a single model to account accurately

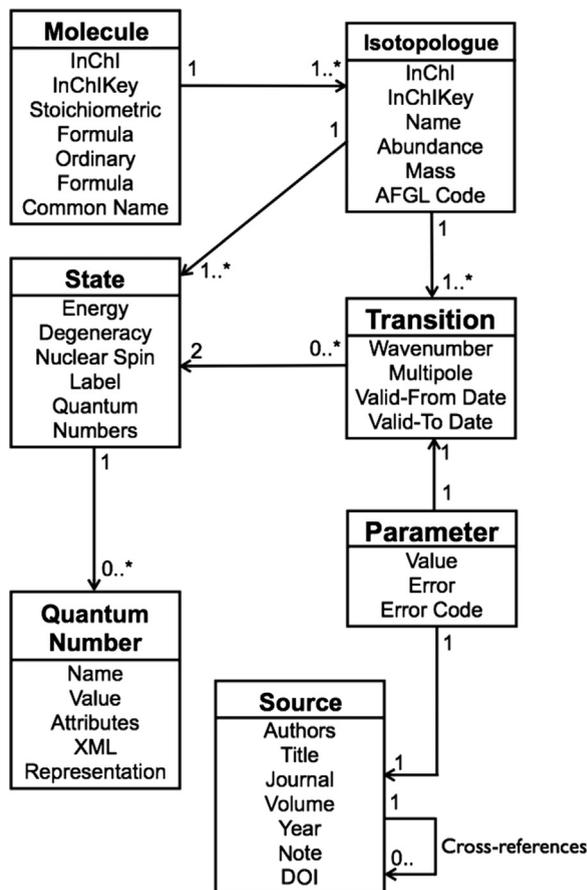


Fig. 1. Abbreviated UML diagram depicting the relationships between tables in the HITRAN database and their principal attributes. A fuller description of each table is given in Appendix A.

and efficiently for these effects, this approach has obvious advantages.

As an example, the spectral line parameters fitted to high-resolution measurements of the $2\leftarrow 0$ band of $^{12}\text{C}^{16}\text{O}$ by Malathy Devi et al. [9] have been imported into the database, as summarised in Table 2. Note that although it is feasible to treat the subdiagonal (or superdiagonal) elements of the relaxation matrix, representing coupling between pairs of neighbouring lines, as parameters in the conventional sense, a new data structure is necessary to represent the entire relaxation matrix for more accurate modelling of line coupling. Such data have yet to be fully integrated in the relational database schema, and are provided in supplementary files (including the Fortran code necessary to use it).

4. Relational database—implementation

The relational database model described above has been implemented using MySQL Server 5.5 and the InnoDB database engine. The line-transition data for the HITRAN 2008 release and subsequent updates have been parsed into this data model, and can be queried directly using Structured Query Language (SQL) with a local installation of the database, or online using the

Table 2New parameters included in the HITRAN relational database model for the 2–0 band of $^{12}\text{C}^{16}\text{O}$.

Parameter	Description
n_{self}	Temperature exponent for the self-broadening coefficient, γ_{self}
δ_{self}	Self-induced pressure shift, referred to 1 atm, 296 K ($\text{cm}^{-1} \text{atm}^{-1}$)
δ'_{self}	Linear temperature dependence coefficient for δ_{self} ($\text{cm}^{-1} \text{atm}^{-1} \text{K}^{-1}$)
δ'_{air}	Linear temperature dependence coefficient for δ_{air} ($\text{cm}^{-1} \text{atm}^{-1} \text{K}^{-1}$)
SD_{air}	Speed-dependence parameter for the speed-dependent Voigt lineshape (air-broadened component)
SD_{self}	Speed-dependence parameter for the speed-dependent Voigt lineshape (self-broadened component)
Y_{air}	First-order (Rosenkranz) line coupling coefficient (air-broadened component) at 296 K ($\text{cm}^{-1} \text{atm}^{-1}$)
Y_{self}	First-order (Rosenkranz) line coupling coefficient (self-broadened component) at 296 K ($\text{cm}^{-1} \text{atm}^{-1}$)
$W_{J,J+1}^{\text{air}}$	Superdiagonal relaxation matrix element for line coupling (air-broadened component) at 296 K ($\text{cm}^{-1} \text{atm}^{-1}$)
$W_{J,J+1}^{\text{self}}$	Superdiagonal relaxation matrix element for line coupling (self-broadened component) at 296 K ($\text{cm}^{-1} \text{atm}^{-1}$)

HITRANonline interface described below. A typical query for transition data will return rows corresponding to the transition parameters, and references (foreign key IDs) to the corresponding states and data sources. The interfacing software (written in the Python programming language) parses the returned transitions rows into a user-defined output format, and makes the additional queries required to build an output file of states and a bibliography (in plain text, HTML or BibTeX format).

Although querying the database with the online interface is the most straightforward way to retrieve data, the power and flexibility of the SQL language makes advanced queries possible, for users with a local copy of the database (for security reasons, raw SQL queries cannot be made using the online service).

5. Relational database—data maintenance and validation

In addition to the software used to query HITRAN described below, a suite of programs has been developed to import updates and new data into the database. To ensure the integrity of the data, it is important to distinguish between transitions which are entirely new to the database and those which replace older data. In the latter case, in which a transition is updated, its original row is 'expired' by setting its `valid_to` field set to the current date, and its replacement enters the database with its `valid_from` field set to the current date. The SQL clause:

```
WHERE valid_from = 'YYYYMMDD' AND
      valid_to < 'YYYYMMDD'
```

then ensures that the same transition is not returned twice within a query, but that the former state of the database can be retrieved if required.

The relational database structure allows the validity of the HITRAN data to be assessed according to various criteria. For example, quantum numbers can be checked to ensure that they represent physically possible states, transition selection rules can be tested, missing or duplicated lines can be identified, and state degeneracies can be verified. The current state of the relational database is intended to be an exact representation of the most recent HITRAN release in its conventional format, and so contains the same inconsistencies and errors. Where two states (as identified by their quantum numbers) appear in the database with different energies these

Table 3Energies retrieved from the HITRAN2008 [10] database for the $(0, 0^+, 0, 0)(J = 6, K = 1)$ state of $^{14}\text{NH}_3$.

State ID	Energy (cm^{-1})
1376196	412.6243
1376643	412.6244
1384357	413.2378
1386816	412.6243
1389564	413.2
1389568	412.6
1392319	144.2585

are listed separately in the `State` table. For example, a single SQL query can list all of the `State` rows for $^{12}\text{C}^{16}\text{O}$ which are assigned the same quantum numbers, but have different energies. The $v = 0, J = 9$ state is found to appear three times, with energies 172.9779, 172.978 and 172.9781 cm^{-1} . Such inconsistencies may be relatively benign in that the differences do not exceed the implied uncertainty in E'' . However, a similar search for the $(0, 0^+, 0, 0)(J = 6, K = 1)$ state of $^{14}\text{NH}_3$ reveals seven different energies with one clear outlier indicating a mistake in the database (Table 3). The absorption intensity at temperatures below 296 K is likely to be seriously overestimated for the line with lower state identified by ID 1392319 (which appears in the P-branch of the $v_3 + v_4$ band).

As another example, more than 11,000 of the 193,978 states identified for $\text{H}^{14}\text{N}^{16}\text{O}_3$ are found to fail the test $K_a + K_c = J$ or $J + 1$ for asymmetric top rotational quantum numbers.

It is anticipated that validation of this sort which is facilitated by the relational database structure will lead to improvements in the quality of the data in HITRAN. Some further examples of data validation are given in Appendix B.

6. HITRANonline—an online interface to the HITRAN database

HITRANonline is a web service allowing the HITRAN database to be searched from a web browser or any computer program able to make HTTP requests. The website, at <http://www.hitran.org/> consists of a form in which the criteria on which the database is to be searched can be specified. Various output formats can be selected and a

customised bibliography for the returned data can be returned with that data. This interface provides a user-friendly way of querying the database without having to write SQL (which is performed by the underlying software).

The HITRANonline interface allows users to select which parameters are to be returned from transitions matching a query, and offers several text-based output formats, including the native 160-character HITRAN2004 format. One or more characters separating the data fields can be chosen if required, and default values for missing data can be customised. Therefore, it is hoped that in many cases the output data can be used directly by users' code without further manipulation. An option to return data in the XML-based XSAMS format [11] is also supported. XSAMS has been adopted by the Virtual Atomic and Molecular Data Centre (VAMDC) project [12] as a standard to allow the aggregation and comparison of atomic and molecular data from different sources.

In addition to the line-transition data search interface, the HITRANonline service provides a means to search for absorption cross section data (in both the UV and IR regions) and the recently-introduced collision-induced absorption data [13]. The search may be conducted by molecule (or collision-pair), wavenumber range, temperature and pressure, and the matching files are returned in the native HITRAN format or as a single-column text file with the associated metadata provided in XSAMS format. Similarly, the partition function information [14,15] is stored and referenced as metadata for each isotopologue. The relevant partition function file can be returned with a search for transitions using the standard interface.

Any query that can be specified through the web page interface can also be made directly to the URL <http://www.hitran.org/tap/sync/> by POSTing the query parameters in the VSS2 query language [16] (a subset of SQL). This provides a mechanism for accessing the database directly from code able to make HTTP requests and from command-line tools such as `wget` and `curl`. It also allows HITRANonline to act as a VAMDC 'node': by supporting the VAMDC standard query language and return format, XSAMS, data from HITRAN can be aggregated and compared with that from other databases, for example using the VAMDC portal interface at <http://portal.vamdc.eu/>.

Appendix A. The HITRAN relational database schema

A.1. General notes

The tables in the HITRAN relational database use the UTF-8 character encoding and case-insensitive UTF-8 character set collation (`COLLATE=utf8_unicode_ci` in MySQL). In our implementation, the InnoDB transactional storage engine is used with MySQL Server 5.5. Each transition is stored in an average disk space of 1486 bytes, so that the HITRAN 2012 database of transitions excluding supplementary data occupies 6.4 GB.

A.2. Molecules

The **Molecule** table contains information concerning a molecular entity (which exists as one or more **Isotopologues**).

7. Conclusion

A new relational database model for HITRAN has been developed and implemented, and a web service, HITRANonline, has been deployed to facilitate searching for data and its transformation into different output formats. HITRANonline is available with a full set of the data provided in the 2012 release of HITRAN [17] and will be maintained with subsequent updates.

HITRANonline represents a major change to the way that the line-transition data are organised, and removes significant limitations imposed by the older, fixed-width text-based format. Parsing the data into a relational model has also allowed checks of data consistency and validity that would have been difficult with the previous format.

The online HITRAN interface also acts as a database node conforming to the standards of the Virtual Atomic and Molecular Data Centre (VAMDC) [12,16], allowing HITRAN data to be compared to and aggregated with data from other atomic and molecular databases.

The implementation of HITRANonline will allow the HITRAN database to expand naturally and hence to adopt more complete parameters sets, such as line broadening by a range of different species, and more sophisticated physics, such as line profile models more complex than Voigt. It will also facilitate integration with other databases. For example, the recent edition of the high-temperature spectral database HITEMP [18] was constructed by manually integrating with the contents of the 2008 edition of HITRAN [10]. The new structures allow integration with databases such as HITEMP, or the even more comprehensive high-temperature data provided by the ExoMol project [19], to be accomplished automatically.

Acknowledgments

This work is supported by the VAMDC project and ERC Advanced Investigator Project 267219 (ExoMol). VAMDC is funded under the "Combination of Coordination and Support Actions" Funding Scheme of The Seventh Framework Program. Call topic: INFRA-2012-3.3 Scientific Data Infrastructure. Grant Agreement number: 313284).

A.2.1. Molecule table structure

Molecule	
id	Primary key
molecID_str	Plain text string representation
InChI [20]	InChI identifier
InChIKey	InChIKey identifier
stoichiometric_formula	Plain text stoichiometric formula
ordinary_formula	Plain text conventional chemical formula
ordinary_formula_html	HTML conventional chemical formula
common_name	Plain text common chemical name
cml	Chemical Markup Language representation [21]

A.2.2. Example: the water molecule

id	1
molecID_str	H2O
InChI	InChI=1S/H2O/h1H2
InChIKey	XLYOFNOQVPJJNP-UHFFFAOYSA-N
stoichiometric_formula	H2O
ordinary_formula	H2O
ordinary_formula_html	H ₂ ₀
common_name	Water
cml	<pre> <moleculeid="H2O"> <atomArray> <atomelementType="H id="H1"/> <atomelementType="H id="H2"/> <atomelementType="O id="O1"/> </atomArray> <bondArray> <bond atomRefs2="H1 O1" id="H1_O1" order="s"/> <bond atomRefs2="H2 O1" id="H2_O1" order="s"/> </bondArray> </molecule> </pre>

A.3. Isotopologue

The **Isotopologue** table contains data relating to a particular isotopologue of a molecule.

A.3.1. Isotopologue table structure

Isotopologue	
id	Primary key, "global" isotopologue ID
isoID	HITRAN "per-molecule" isotopologue ID
InChI	InChI identifier
InChIKey	InChIKey identifier
molecule_id	Foreign key to Molecule table
iso_name	Plain text conventional isotopologue formula
iso_name_html	HTML conventional isotopologue formula
abundance	Fractional isotopologue abundance
mass	Isotopologue mass, in atomic units
afgl_code	AFGL identifying code
cml	Chemical Markup Language representation [21]
ZPE_state_id	ID of lowest energy level of this isotopologue

A.3.2. Example: The $H_2^{18}O$ isotopologue of water

id	2
isoID	2
InChI	InChI=1S/H2O/h1H2/i1+2
InChIKey	XLYOFNOQVPJJNP-NJFSPNSNSA-N
molecule_id	1
iso_name	H2 (18O)
iso_name	H ₂ ¹⁸ O
abundance	0.00199983
mass	20.014811
afgl_code	181
cml	<moleculeid="H2O_2"> <atomArray> <atom element="H" id="H1"/> <atom elementType="H2"/> <atom elementType="O" isotopeNumber="18"/> </atomArray> <bondArray> <bondatomRefs2="H1-O1H1_O1" order="S"/> <bondatomRefs2="H2-O1H2_O1" order="S"/> </bondArray> </molecule>
ZPE_state_id	762283

A.4. State

The **State** table contains data relating to each state of every isotopologue in the database

A.4.1. State table structure

State	
id	Primary key
iso_id	Foreign key to Isotopologue table
energy	State energy in cm
g	Total degeneracy [22]
nucspin_label	Character identifying nuclear spin isomer: o, m, or p
s_qns	String representation of state quantum numbers
qns_xml	XSAMS (XML) representation of state quantum numbers [11,16]

A.4.2. Example: The lowest energy level of $H_2^{18}O$

id	762283
iso_id	2
energy	0.
g	1
nucspin_label	p
s_qns	ElecStateLabel=X;v1=0;v2=0;v3=0;J=0;Ka=0;Kc=0
qns_xml	<nltcs : QNs > <nltcs : ElecStateLabel >X</nltcs : ElecStateLabel > <nltcs : v1 >0</nltcs : v1 > <nltcs : v2 >0</nltcs : v2 > <nltcs : v3 >0</nltcs : v3 > <nltcs : J >0</nltcs : J > <nltcs : Ka >0</nltcs : Ka > <nltcs : Kc >0</nltcs : Kc > </nltcs : QNs >

A.5. Quantum numbers

The **Quantum Number** table describes every quantum number, label or symmetry for each state in the database

A.5.1. Quantum number table structure

Quantum number	
id	Primary key
state_id	Foreign key to State table
qn_name	Name of the quantum number, symmetry or label
qn_val	Value of the quantum number, symmetry or label
qn_attr	Additional contextual information for this quantum number
xml	XML representation of this quantum number, for its XSAMS description [11,16]

A.5.2. Example: F quantum number of a state of $^{16}\text{O}^{12}\text{C}^{17}\text{O}$

id	6033763
state_id	801367
qn_name	F
qn_val	2.5
qn_attr	nuclearSpinRef:01
xml	< ltcs : F nuclearSpinRef = "01" > 2.5 < /ltcs : F >

A.6. Transitions

The **Transition** table describes each transition for every isotopologue in the database. The fields `nu`, `sw`, `a`, and `elower` duplicate the respective parameters stored in other tables and are indexed so that searching on these attributes is fast and efficient.

A.6.1. Transition table structure

Transition	
id	Primary key
iso_id	Foreign key to Isotopologue table
statep_id	Foreign key to State table identifying upper state
statepp_id	Foreign key to State table identifying upper state
nu	Transition wavenumber in cm^{-1}
sw	Transition intensity, weighted by isotopologue abundance, in $\text{cm}^{-1}/(\text{molec cm}^{-2})$ at 296 K.
a	Einstein A-factor in s^{-1}
multipole	Transition multipole (E1 for electric dipole, E2 for electric quadrupole, M1 for magnetic dipole)
elower	Lower state energy in cm^{-1}
gp	Upper state total degeneracy
gpp	Lower state total degeneracy
valid_from	Date that data for this transition entered the database
valid_to	Date that data for this transition expires
band	Plain text representation of the vibrational band to which this transition belongs

A.6.2. Example: The P(24) transition of the $2\nu_1-\nu_1$ band of $^{12}\text{C}^{16}\text{O}$

id	2501262
iso_id	26
statep_id	1127932
statepp_id	1127933
nu	2016.0272

sw	4.87e-25
a	29.51
multipole	E1
elower	3284.0845
gp	47
gpp	49
valid_from	2009-04-15
valid_to	2009-04-15
band	2v1-v1

A.7. Parameters

The **Parameter** tables describe the parameters associated with the transitions in the database. Each type of parameter is stored in its own table.

A.7.1. Parameter table structure

Parameter	
id	Primary key
trans_id	Foreign key to Transition table
val	Parameter value
err	Absolute uncertainty in the parameter value
ierr	HITRAN error code [2] .
source_id	Foreign key to the Source table

A.7.2. Example: The γ_{air} line-broadening parameter of a transition

prm_gamma_air	
id	10002
trans_id	1668605
val	0.0233
err	0.0012
ierr	6
source_id	52

A.8. Sources

The **Source** table describes every source (reference) to the original data in the database. In many cases, this will be a published article, but other sources such as private communications, notes and other database compilations are supported. Cross-referencing is achieved by referring to a Source's primary key (*id* field).

A.8.1. Source table structure

Source	
id	Primary key
sourcetype_id	Integer identifying the type of source (journal publication, private communication, etc.)
authors	Authors list
title	Plain text title
title_html	HTML title
title_latex	LaTeX version of the title, used to create BibTeX output
journal	Journal title
volume	Volume
page_start	First page number of article
page_end	Last page number of article
year	Year of publication

institution	Institution associated with this reference
note	Plain text version of a reference taking the form of a note
note_html	HTML version of a reference taking the form of a note
note_latex	LaTeX version of a reference taking the form of a note
doi	Document Object Identifier for published articles
url	A URL providing online access to the resource referenced, if available
article_number	Article number, used instead of page numbers by some journals

A.8.2. Example: The γ_{air} line-broadening parameter of a transition

id	50
sourcetype_id	1
authors	M.-F. Mérianne, A. Jenouvrier, C. Hermans, A.C. Vandaele, M. Carleer, P.-F. Coheur, R. Colin, S. Fally, M. Bach
title	Water vapor line parameters in the 13 000–9250 cm ⁻¹ region
title_html	Water vapor line parameters in the 13 000 – 9250 cm ^{<sup>-1</sup>} region
title_latex	Water vapor line parameters in the 13 000–9250 cm^{-1} region
journal	Journal of Quantitative Spectroscopy and Radiative Transfer
volume	82
page_start	99
page_end	117
year+	2003
institution+	-
note	-
note_html	-
note_latex	-
doi	10.1016/S0022-4073(03)00148-1
url	http://www.sciencedirect.com/scie...5b66
article_number	-

Appendix B. SQL queries for data validation

Structured Query Language (SQL) is a well-established, special-purpose declarative programming language for managing data in relational database management systems. The relational database implementation of HITRAN described in this paper uses SQL to update, manage, validate and search the database. Some example SQL queries are given below.

B.1. Inconsistent state energies

This command lists the IDs, energies and quantum number assignments of all the states of $^{14}\text{NH}_3$ ($\text{iso_id}=45$) which have the same assignments (s_qns) but energies differing by more than 10 cm^{-1} .

```
SELECT s1.id, s1.energy, s2.id, s2.energy, s1.s_qns
FROM state s1, state s2
WHERE s1.iso_id=45
AND s1.iso_id=s2.iso_id
AND s1.s_qns=s2.s_qns
AND & s1.id > s2.id
AND ABS(s1.energy-s2.energy) > 10.;
```

B.2. Quantum number validity

This command lists the state IDs and F and J quantum numbers of states of $^{16}\text{O}^{17}\text{O}$ ($\text{iso_id}=38$) which do not obey the condition $|J - \frac{5}{2}| \leq F \leq J + \frac{5}{2}$.

```
SELECT s.state_id, ss.qn_val, s.qn_val FROM
qns s, qns ss, state st
WHERE s.state_id=st.id
AND st.iso_id=38
AND s.state_id=ss.state_id
AND s.qn_name='F'
AND ss.qn_name='J'
AND & (s.qn_val > ss.qn_val + 2.5
OR s.qn_val < ABS(ss.qn_val2.5));
```

B.3. Selection rules

This command lists the IDs of the transitions of HOCl (*iso_id*=67, 68) which have $\Delta J > 1$.

```
SELECT t.id from trans t, state sp, state spp, qns qJp,
qns qJpp
WHERE t.iso_id > 66 AND t.iso_id < 69
AND t.statep_id=sp.id AND t.statepp_id=spp.id
AND qJp.state_id=sp.id AND qJp.qn_name='J'

AND qJpp.state_id=spp.id AND qJpp.qn_name='J'
AND (qJp.qn_val-qJpp.qn_val) > 1;
```

References

- [1] Rothman LS. The evolution and impact of the HITRAN molecular spectroscopic database. *J Quant Spectrosc Radiat Transfer* 2010;111:1565–7.
- [2] Rothman LS, Jacquemart D, Barbe A, Benner DC, Birk M, Brown LR, et al. The HITRAN 2004 molecular spectroscopic database. *J Quant Spectrosc Radiat Transfer* 2005;96:139–204.
- [3] McClatchey RA, Benedict WS, Clough SA, Burch DE, Calfee RF, Fox K, et al. AFCRL atmospheric absorption line parameters compilation. AFCRL-Technical Report-0096, Technical Report, Air Force Cambridge Research Laboratory; 1973.
- [4] Wallace CJ, Jeon C, Anderson CN. H₂O broadening of a CO₂ line and its nearest neighbors near 6360 cm⁻¹. *J Phys Chem A* 2011;115:13804–10.
- [5] Boone CD, Walker KA, Bernath PF. Speed-dependent Voigt profile for water vapor in infrared remote sensing applications. *J Quant Spectrosc Radiat Transfer* 2007;105:525–32.
- [6] Hodges JT, Lisak D, Lavrentieva N, Bykov A, Sinitsa L, Tennyson J, et al. Comparison between theoretical calculations and high-resolution measurements of pressure-broadening for water vapor spectra near 935 nm. *J Mol Spectrosc* 2008;249:86–94.
- [7] Lamouroux J, Tran H, Laraia AL, Gamache RR, Rothman LS, Gordon IE, et al. Updated database plus software for line-mixing in CO₂ infrared spectra and their test using laboratory spectra in the 1.5–2.3 um region. *J Quant Spectrosc Radiat Transfer* 2010;111:2321–31.
- [8] Rosenkranz P. Shape of the 5 mm oxygen band in the atmosphere. *IEEE Trans Antennas Propag* 1975;AP-23:498–506.
- [9] Devi VM, Chris Benner D, Smith MAH, Mantz AW, Sung K, Brown LR, et al. Spectral line parameters including temperature dependencies of self- and air-broadening in the 20 band of CO at 2.3 m. *J Quant Spectrosc Radiat Transfer* 2012;113:1013–33.
- [10] Rothman LS, Gordon IE, Barbe A, Benner PE, Chris an d Bernath D, Birk M, et al. The HITRAN 2008 molecular spectroscopic database. *J Quant Spectrosc Radiat Transfer* 2009;110:533–72.
- [11] Dubernet ML, Humbert D, Ralchenko Y, Roueff E, Schultz DR, Chung H-K, et al. XSAMS: XML schema for atomic, molecular and solid data, version 0.1.1, International Atomic Energy Authority, <http://www-amdis.iaea.org/xsams/documents/>, 2011.
- [12] Dubernet ML, Boudon V, Culhane JL, Dimitrijevic MS, Fazliev AZ, Joblin C, et al. Virtual atomic and molecular data centre. *J Quant Spectrosc Radiat Transfer* 2010;111:2151–9.
- [13] Richard C, Gordon IE, Rothman LS, Abel M, Frommhold L, Gustafsson M, et al. New section of the HITRAN database: Collision-induced absorption (CIA). *J Quant Spectrosc Radiat Transfer* 2012;113:1276–85.
- [14] Fischer J, Gamache RR, Goldman A, Rothman LS, Perrin A. Total internal partition sums for molecular species in the 2000 edition of the HITRAN database. *J Quant Spectrosc Radiat Transfer* 2003;82:401–12. [http://dx.doi.org/10.1016/S0022-4073\(03\)00166-3](http://dx.doi.org/10.1016/S0022-4073(03)00166-3).
- [15] Laraia AL, Gamache RR, Lamouroux J, Gordon IE, Rothman LS. Total internal partition sums to support planetary remote sensing. *Icarus* 2011;215:391–400.
- [16] Dubernet ML, Doronin M, Endres C, Hill C, Marquart T, Nenadovic L, et al. VAMDC Standards Documentation and Reference Guides, version r12.07, Virtual Atomic and Molecular Data Centre, (<http://www.vamdc.eu/standards/>) (2012).
- [17] Rothman LS, Gordon IE, Babikov Y, Barbe A, Benner DC, Bernath PF, et al. The HITRAN2012 molecular spectroscopic database, *J Quant Spectrosc Radiat Transfer*. In this issue.
- [18] Rothman LS, Gordon IE, Barber RJ, Dothe H, Gamache RR, Goldman A, et al. HITEMP, the high-temperature molecular spectroscopic database. *J Quant Spectrosc Radiat Transfer* 2010;111:2139–50.
- [19] Tennyson J, Yurchenko SN. ExoMol: molecular line lists for exoplanet and other atmospheres. *Mon Not R Astr Soc* 2012;425:21–33.
- [20] Heller S, McNaught A. The status of the InChI project and the InChI trust. *J Cheminformatics* 2010;2:1–2.
- [21] Murray-Rust P, Rzepa HS. Chemical basic principles markup, XML, and the worldwide web. 1. *J Chem Info Comput Sci* 1999;39:928–42.
- [22] Šimečková M, Jacquemart D, Rothman L, Gamache R, Goldman A. Einstein A-coefficients and statistical weights for molecular absorption transitions in the HITRAN database. *J Quant Spectrosc Radiat Transfer* 2006;98:130–55.