# Information system for molecular spectroscopy.
# 3. Molecular energy levels

## A.V. Kozodoev, A.I. Privezentsev, and A.Z. Fazliev

*Institute of Atmospheric Optics,*
*Siberian Branch of the Russian Academy of Sciences, Tomsk*

A part of the information system for molecular spectroscopy based on problem ontology is described. Two problems (direct and inverse) of determination of molecular energy levels are considered. The database related to these problems is compiled for two triatomic molecules: water vapor and hydrogen sulfide, as well as their isotopic modifications.

## Introduction

Many modern techniques of gas media investigation use the data on molecular spectral parameters. Several research teams, in particular, the team supporting the HITRAN data bank,[1] deal with classification, acquisition, storage, and distribution of such data. Note that the problem of data classification falls in the category of spectroscopic problems, and its solution determines, to a large extent, the organization of acquisition, storage, and distribution of data and knowledge, i.e., the actions, which, in turn, are subjects of study in informatics.

For the nearly four-decade history, the HITRAN team several times modified both the set of physical essences, as well as format of documents and files storing the data. The events occurred in the last decade forced us to consider critically the existing form of acquisition, storage, and distribution of spectral data used by the HITRAN team. First, the recent papers on calculation of the water vapor spectrum by variational methods report hundreds of millions of spectral lines.[2,3] Such a great number of lines make the file structure inconvenient for storage of tens-GB data. Second, there appeared quantum numbers in solution of problems by the variational method, which describe more correctly the water vapor spectrum in the frequency range above $20000$ cm$^{-1}$ [Ref. 3], as well as a technical problem of comparison of large data arrays arose. Third, the Internet made large information volumes more available both as a whole and by parts. All the mentioned features indicate that now it is necessary to pass from operation with data stored in files to information-computation systems (ICSs), providing the user with graphical interfaces for operation with data and information processes characteristic of molecular spectroscopy.

First systems of this kind were developed seven years ago.[4] However, the further intentional development of the data and metadata structures, in particular, compilation of ontologies, took place in designing the "Atmospheric Spectroscopy" ICS.[5–7] This development was based on the semantic grid approach[8] to the designing of modern intellectual ICSs. Within this approach, ICS is represented in the form of three layers: the layer of data and computations, the information layer, and the knowledge layer. Such separation allows simulation areas to be outlined clearly. Simulation in the object domain is connected with the layer of data and computations. Simulation in other two layers belongs to problems of informatics and is connected with information objects and processes of operation with them.

This paper describes the part of the "Atmospheric Spectroscopy" ICS containing the results of solution of the problem of two types: direct and inverse problems of determination of molecular energy levels. The aim of this work is to compile the database and the knowledge base for the part of molecular spectroscopy, which corresponds to molecular energy levels, to testing data for restrictions, characteristic of different types of molecules, to checking the knowledge data for consistency and their belonging to catalogs, to forming metadata, and to the comparing values or quantum numbers of energy levels from different data sources. An extensional part of metadata is formed automatically during the data acquisition, construction of compound data sources, and solution of problems in ICS. The knowledge level contains the notions formalized by semantic web tools.

## 1. Data and metadata on energy levels

Molecular energy levels can be determined from solutions of the direct problem based on quantum mechanics methods and the inverse problem based on the prepared list of transitions (peak list) obtained from processing spectra. The direct problem solution yields information about molecular energy levels, including the values of energy levels and quantum

numbers. The inverse problem solution yields the values of energy levels, quantum numbers, uncertainties in energy levels, and the number of transitions used to calculate an energy level. These data are stored in the database of the information system. Note that energy levels calculated in the direct and inverse problems are of different origins, that is, they are different physical essences.

The information about energy levels along with the listed data of these problems includes metadata. Metadata characterize energy levels with respect, for example, to an article, they were published in, with respect to quantitative characteristics of data obtained from measurements or calculations, with respect to the method used in calculation, with respect to input data of the problem, etc. Figure 1 shows the screenshot demonstrating metadata on energy levels corresponding to the data source formed from the solution of the direct problem. In contrast to quantum numbers and errors, metadata correspond to the whole array of energy levels, being a component of a data source, rather than to a single level. They are necessary for description of declarative knowledge in molecular spectroscopy connected with systematization or, in other words, with construction of taxonomies of information resources in molecular spectroscopy.

The structure of data array corresponding to the direct and inverse problems and loaded by a user into ICS must necessarily include energy levels and quantum numbers. It should be noted that three types of quantum number notations for the water vapor molecule are allowed: normal modes, Schwenke notations,[9] and Barber–Tennyson notations (BT2).[3]

The database is now compiled fragmentarily, that is, it includes energy levels only for water vapor and hydrogen sulfide molecules. For the water vapor molecule, the calculated data include values of energy levels up to $30\,000$ cm$^{-1}$, while the data obtained from experiments through solution of the inverse problem include values of energy levels up to $26\,000$ cm$^{-1}$. The upper boundary of values of the total angular momentum $J$ is 50 for the calculated data and 42 for the experimental data. The largest number of calculated energy levels exceeds $220\,000$, and the number of energy levels calculated from experimental data exceeds $20\,000$. The data on energy levels of other isotopic modifications of the water vapor molecule number from several hundreds to several thousands. The information-computation system includes the Polovtseva's data [10] on more than 2200 energy levels of hydrogen sulfide in the region $5800–10950$ cm$^{-1}$ and several hundreds of energy levels for the isotopic modifications $H_2^{34}S$ and $H_2^{33}S$.

---

**Annotation (2007_Zobov_H2_160**　　　　　　　　　　Calculation/Experiment

| Substance | | |
|---|---|---|
| $H_2O$ | | |

| Input data | | |
|---|---|---|

| Method | | |
|---|---|---|
| DVR3D | | |

| Reference | | |
|---|---|---|
| N.F. Zobov, R.I. Ovsyannikov, S.V. Shirin, O.L. Polyansky, The Assignment of Quantum Numbers in the Theoretical Spectra of the $H_2^{16}O$, $H_2^{17}O$ and $H_2^{18}O$ Molecules Calculated by Variational Methods in the Region 0–26 000 cm$^{-1}$. // Optics and Spectroscopy, 2007, v. 102, no. 3, p. 348-353. | | |
| Quantum numbers have been assigned in the theoretical spectra of three isotopologues of the water molecule: $H_2^{16}O$, $H_2^{17}O$, and $H_2^{18}O$. The spectra were calculated by variational methods in the region 0–26 000 cm$^{-1}$ at a temperature of 296 K. For each molecule, the quantum numbers are assigned to more than 28 000 levels. The quantum numbers are assigned to 216 766, 210 679, and 211 073 spectral lines of the $H_2^{16}O$, $H_2^{17}O$, $H_2^{18}O$ molecules, respectively. The theoretical spectra with the assigned quantum numbers are available in the Internet. | | |

| Output data | | |
|---|---|---|
| **Energy levels** | | |
| Unit | cm$^{-1}$ | |
| Minimal value of energy level | 23.794 | |
| Maximal value of energy level | 27888.028 | |
| Number of energy levels | 28232 | [T] |
| **Quantum numbers** | | |
| Quantum numbers type | BT2 | |
| Minimal value of total angular momentum J | 0 | |
| Maximal value of total angular momentum J | 10 | |
| Number of levels with unique quantum numbers | 28232 | [T] |
| Number of levels with nonunique quantum numbers | 0 | [T] |
| **Quantum numbers** | | |
| Quantum numbers type | NormalModes | |
| Minimal value of total angular momentum J | 0 | |
| Maximal value of total angular momentum J | 10 | |
| Number of levels with unique quantum numbers | 28231 | [T] |
| Number of levels with nonunique quantum numbers | 1 | [T] |

**Fig. 1.** Metadata of a data source including energy levels for the main isotopic modification of the water vapor molecule.

The knowledge base includes the description of 70 data sources in the form of statements formed by OWL language tools. Using the free Protégé software,[11] the user, having imported the knowledge base, can perform the semantic search of data sources by metadata values.

## 2. Acquisition and storage of data on energy levels

The ICS user has a possibility to load private data on energy levels into the database and the knowledge base in order to compare them with the data recommended by experts. For this purpose, the software was developed, which controls for the correspondence of input data to typical restrictions characteristic of data on energy levels based on XML schemes. To classify the data of molecular spectroscopy, we used the model, in which the object domain is represented in the form of hierarchy of problems.[12]

The *data source* is the key concept for ICS formation based on data division into problems of molecular spectroscopy. Data sources can be elementary and compound. A source containing the data on energy levels, published in one paper and corresponding to a molecule of one type, is an elementary source. All other data sources (for example, HITRAN and Geisa) are compound.

For an elementary or compound source, four levels of description are used.

1. *Description of data source attributes*: data source name, reference, comments, type of the problem, and the name of the substance, whose energy levels are presented in the data source.

2. *Description of intensional of data presented in the source*: energy levels, quantum numbers, errors, and the number of transitions used to obtain the energy level.

3. *Metadata on the problem related to the data source* are defined in ontologies of spectroscopic problems and contain the information about input data and the method of solution of the problem. In particular, for the direct problem of determination of water vapor energy levels, such data are represented by URLs of information resources, such as the potential function and the basis wave functions.

4. *Metadata on energy levels* contain the information about minimal and maximal values of energy levels, the number of energy levels, the type of quantum numbers, the maximal value of the total angular momentum, the number of unique and non-unique values of quantum numbers assigned to energy levels.

For acquisition of data arrays from elementary data sources, a loading system with a web interface has been developed. The loading of energy levels obtained from solution of both direct and inverse problems consists of three steps:

1. Selection of a data source, specification of a molecule and the number of physical essences in the loaded data array.

2. Formation of the ordered list of physical essences in the order determined by the order of columns in the loaded file (Fig. 2).

3. Setting the data file address.

The ICS user can form private data sources. To relate a source to a bibliographic reference, the database includes hundreds of bibliographic references in the field of molecular spectroscopy. If the needed reference is absent, the user can form private references to information resources. Other users will have no access to the private bibliographic references.

**Data File Schema**

| Line number | Physical Entity | Beginning Position of the Column | End Position of the Column |
|---|---|---|---|
| 1 | Quantum Numbers. Normal Modes ⌄ | 1 | 24 |
| 2 | Energy level (cm^-1) ⌄ | 26 | 36 |
| 3 | Quantum Numbers. BT2 ⌄ | 38 | 50 |
| 4 | Quantum numbers. Schwenke ⌄ | 52 | 76 |

**Fig. 2.** Description of columns of the file containing energy levels and quantum numbers.

The system determines the following errors in formation of private files: inconsistent types of data; incorrect structure of physical essences described by the user when preparing the loading (step 2); inconsistence to allowable ranges of values for physical parameters, etc.

Now the information system includes the data from 70 elementary data sources on energy levels of the molecules: $H_2O$ ($H_2^{17}O$, $H_2^{18}O$, HDO, $HD^{18}O$, $D_2O$) and $H_2S$ ($H_2^{34}S$, $H_2^{33}S$). Most data sources deal with solution of inverse problems. Some data sources are divided into parts containing energy levels for *orto-* and *para-*water. The data on energy levels of the water vapor molecule are represented most thoroughly. It should be noted that the data on energy levels of $SO_2$ and $O_3$, having the same set of quantum numbers as the water vapor molecule, can also be loaded in ICS.

Figure 3 shows a simplified scheme of the database, demonstrating the relations between tables storing energy levels. The simplification concerns the list of quantum number notations and the relation to the bibliographic database.
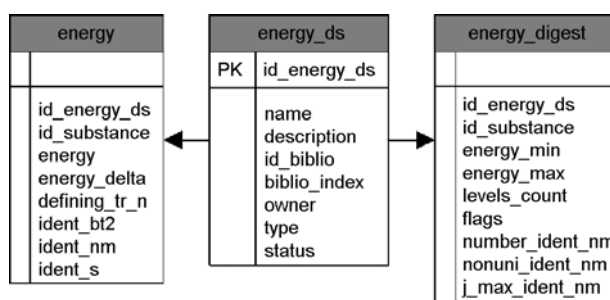


**Fig. 3.** A simplified structure of database of energy levels.

# 3. Search and comparison of energy levels from different data sources

Energy levels connected with data sources and corresponding to the direct or inverse problems can be compared both in values and in quantum numbers. This comparison is important in analysis of experimental data, whose number is an order of magnitude smaller than that of calculated data, and it becomes particularly significant for energy levels above $20\,000$ cm$^{-1}$.

The user can compare energy levels with a preset accuracy from 2 to 7 significant digits (in cm$^{-1}$). The selection of restrictions on the values of the total angular quantum number $J$ allows the user to separate arrays of energy levels with $J$ in the defined interval in the compared data sources. Thus, it becomes possible, for example, to cut off calculated energy levels, which still cannot be reconstructed from the experiment. The comparison can be performed both in tabulated and graphical forms. The screenshot of the tabulated comparison is shown in Fig. 4.

The comparison of energy levels of the water vapor molecule in quantum numbers (normal modes) allows the user to separate the energy levels with coinciding quantum numbers and to obtain the difference in the corresponding energy levels. For the water vapor molecule, it is possible to compare data sources containing the sets of quantum numbers in different notations. At such comparison, the quantum numbers ($v_1$, $v_2$, $v_3$, $J$, $K_a$, $K_c$) are transformed into the pair of quantum numbers ($J$, $s$), in which the quantum number $s$ describes the symmetry.[3] Although such comparison does not provide the one-to-one relation between levels, it is useful for comparison of energy levels obtained in the direct and inverse problems.

The user can export the data on molecular energy levels when browsing or comparing data sources, using the tabulated representation. The data can be received through the Internet in the text format or in the MS Excel format.

## Conclusions

The principal idea of the new approach to the acquisition, storage, and classification of data on energy levels of water vapor, hydrogen sulfide, and some other molecules is that energy levels and accompanying data loaded in the system are connected with the data sources representing them in ICS. Data sources are associated with typical problems of molecular spectroscopy.[12] The compilation of the database with calculated and experimental data allows it to be considered now as a prototype of the information system oriented at representation of the data on energy levels of atmospheric molecules.

| Information source (Left) Calculations/Experiments | Information source (Right) Calculations/Experiments |
|---|---|
| 1997JMS_Polyansky | 2007_Zobov_H2_16O |
| O.L.Polyansky, N.F.Zobov, S.Vitia, J.Tennyson, P.F.Bernath and L.Wallace, High-Temperature Rotational Transitions of Water in Sunspot and Laboratory Spectra. // Journal of Molecular Spectroscopy, 1997, v. 186, no. 2, p. 422-447. | N.F. Zobov, R.I. Ovsyannikov, S.V. Shirin, O.L. Polyansky, The Assignment of Quantum Numbers in the Theoretical Spectra of the $H_2^{16}O$, $H_2^{17}O$ and $H_2^{18}O$ Molecules Calculated by Variational Methods in the Region 0–26 000 cm$^{-1}$. // Optics and Spectroscopy, 2007, v. 102, no. 3, p. 348-353. |

The number of coincided energy levels values is 10. Round-off 3 digits.

Show [12] rows starting from [12]   In all rows 765   << < > >>   Control

| $v_1^{NM}$ | $v_2^{NM}$ | $v_3^{NM}$ | $J^{NM}$ | $K_a^{NM}$ | $K_c^{NM}$ | Energy Levels (E (cm$^{-1}$)) | $v_1^{NM}$ | $v_2^{NM}$ | $v_3^{NM}$ | $J^{NM}$ | $K_a^{NM}$ | $K_c^{NM}$ | $J^{BT2}$ | $s^{BT2}$ | $n^{BT2}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 212.156 | 0 | 0 | 0 | 3 | 2 | 1 | 3 | 3 | 2 |
| 0 | 0 | 0 | 4 | 0 | 4 | 222.053 | 0 | 0 | 0 | 4 | 0 | 4 | 4 | 1 | 1 |
| | | | | | | 224.839 | 0 | 0 | 0 | 4 | 1 | 4 | 4 | 4 | 1 |
| | | | | | | 275.497 | 0 | 0 | 0 | 4 | 1 | 3 | 4 | 2 | 1 |
| | | | | | | 285.218 | 0 | 0 | 0 | 3 | 3 | 1 | 3 | 2 | 2 |
| | | | | | | 285.417 | 0 | 0 | 0 | 3 | 3 | 0 | 3 | 4 | 2 |
| 0 | 0 | 0 | 4 | 2 | 3 | 300.362 | 0 | 0 | 0 | 4 | 2 | 3 | 4 | 3 | 1 |
| 0 | 0 | 0 | 4 | 2 | 2 | 315.779 | 0 | 0 | 0 | 4 | 2 | 2 | 4 | 1 | 2 |
| 0 | 0 | 0 | 5 | 0 | 5 | 325.348 | | | | | | | | | |
| | | | | | | 325.349 | 0 | 0 | 0 | 5 | 0 | 5 | 5 | 3 | 1 |
| 0 | 0 | 0 | 5 | 1 | 5 | 326.625 | | | | | | | | | |
| | | | | | | 326.626 | 0 | 0 | 0 | 5 | 1 | 5 | 5 | 2 | 1 |

**Fig. 4.** Comparison of energy levels of water vapor accurate to the third decimal place (for coinciding values, quantum numbers are presented in the same row).

## Acknowledgements

## References

1. L.S. Rothman, D. Jacquemart, A. Barbe, D.C. Benner, M. Birk, L.R. Brown, M.R. Carleer, C. Chakkerian, Jr., K. Chance, L.H. Coudert, V. Dana, V.M. Devi, J.-M. Flaud, R.R. Gamache, A. Goldman, J.-M. Hartmann, K.W. Jucks, A.G. Maki, J.-Y. Mandin, S.T. Massie, J. Orphal, A. Perrin, C.P. Rinsland, M.A.H. Smith, J. Tennyson, R.N. Tolchenov, R.A. Toth, J. Vander Auwera, P. Varanasi, and G. Wagner, J. Quant. Spectrosc. Radiat. Transfer **96**, No. 2, 139–204 (2005).
2. Vl.G. Tyuterev, Atmos. Oceanic Opt. **16**, No. 3, 220–231 (2003).
3. R.J. Barber, J. Tennyson, G.J. Harris, and R.N. Tolchenov, Mon. Notic. Roy. Astron. Soc. **368**, 1087–1094 (2006).
4. Yu.L. Babikov, A. Barbe, V.F. Golovko, S.N. Mikhailenko, and Vl.G. Tyuterev, in: *Proc. of the Third Russian Conf. "Electronic Libraries: Promising Methods and Technologies"* (Petrozavodsk, 2001), pp. 183–187; Spectroscopy of Atmospheric Gases, http://spectra.iao.ru
5. A.D. Bykov, B.A. Voronin, A.V. Kozodoev, N.A. Lavrent'ev, O.B. Rodimova, and A.Z. Fazliev, Atmos. Oceanic Opt. **17**, No. 11, 816–820 (2004).
6. A.V. Kozodoev and A.Z. Fazliev, Atmos. Oceanic Opt. **18**, No. 9, 680–684 (2005).
7. A.V. Kozodoev, A.I. Privezentsev, and A.Z. Fazliev, Vychislit. Tekhnol. **10**, Spec. Issue, 82–91 (2005).
8. D. De Roure, N. Jennings, and N. Shadbolt, *A Future e-Science Infrastructure, Report commissioned for EPSRC/DTI Core e-Sci.* Program (2001), 78 pp.
9. H. Partridge and D.W. Schwenke, J. Chem. Phys. **106**, No. 11, 4618–4639 (1997).
10. E.R. Polovtseva, *"Analysis of Rotational-Vibrational Spectrum of Hydrogen Sulfide in the Region from 4500 $cm^{-1}$ to 11000 $cm^{-1}$,"* Author's Abstract of Cand. Phys.-Math. Sci. Dissert. (Tomsk, 2006), 22 pp.
11. Protégé-ontology editor and knowledge-base framework, http://protege.stanford.edu/
12. A.D. Bykov, A.V. Kozodoev, A.I. Privezentsev, L.N. Sinitsa, M.V. Tonkov, N.N. Filippov, A.Z. Fazliev, and M.Yu. Tretyakov, Proc. SPIE **6580**, 65800W (2006).