Hyperfine Interact
DOI 10.1007/s10751-012-0732-2

MossWinn—methodological advances in the field of Mössbauer data analysis

Zoltán Klencsár

Abstract The methodology of Mössbauer data analysis has been advanced via the development of a novel scientific database system concept and its realization in the field of Mössbauer spectroscopy, as well as by the application of parallel computing techniques for the enhancement of the efficiency of various processes encountered in the practice of Mössbauer data handling and analysis. The present article describes the new database system concept along with details of its realization in the form of the MossWinn Internet Database (MIDB), and illustrates the performance advantage that may be realized on multi-core processor systems by the application of parallel algorithms for the implementation of database system functions.

Keywords Internet database · Mössbauer spectroscopy · Spectral analysis · Parallel algorithms

1 Introduction

In the field of natural sciences the enhancing rate of scientific data production witnessed by the 20th century called for an extension and renewal of the traditional ways of disseminating scientific knowledge. Today the challenge of global sharing of the data in an organized, effective and useful manner may be met by the development of database systems exploiting the possibilities offered by the relatively new technology of the internet. A **database system** (**DBS**) can be seen as consisting of a **database** (i.e. collection of data organized according to a given structure) and a corresponding **database management system/software** (**DBMS**) handling the data and providing the functionality of the database system. The database is furthermore usually organized as a collection of implicitly structured **records**.

Zoltán Klencsár (✉)
Pitvar u. 11., Budapest 1141, Hungary
e-mail: z.klencsar@mosswinn.hu

Published online: 07 December 2012

In the field of Mössbauer spectroscopy there already exist several different database systems that utilize the internet in order to provide access to their data via the World Wide Web:

(1) the '**MEDC database**' [1–3] comprises over 50,000 Mössbauer spectroscopy related bibliographical references and a unique collection of over 100,000 sets of data extracted from the corresponding publications;
(2) the '**Mars Mineral Spectroscopy Database**' provides Mössbauer spectra of minerals, collected over a range of temperatures, in order to provide suitable analog spectra for data acquired on remote surfaces such as Mars [4, 5];
(3) the database system '**WWW-Messbauer**' [6] contains ^{57}Fe Mössbauer-spectra of minerals and their crystal-chemical Ge-analogs, along with information on their analysis.

Existing Mössbauer databases reside on (remote) database server computers (**servers**) that can be accessed via the World Wide Web by the means of web browser applications running on the database users' client computers (**clients**). Typically, **database queries** issued by the user are sent to the server and the records matching the query travel back to the client via the internet.

In the present article we introduce the concept and realization of a novel Mössbauer spectroscopy database system (named as the '**MossWinn Internet Database**', **MIDB**) that achieves multiple advances in the field mainly

(1) by delegating the whole database system to the clients, and using remote (internet-based) data storage resources only for the purpose of synchronization of data among the database instances distributed over the client computers, and
(2) by integrating the database management and data analysis functionalities in a single application software, i.e. the MossWinn program [7, 8].

2 Organization of database access

In the database access scheme (Fig. 1) developed for the MIDB the (remote) database server plays the passive role of a data storage medium that hosts a reference version of the database, whereas the client computers all host a standalone copy of the database management software and the database. The database system is operated by the coherent action of the individual copies of the database management software running distributed on the client computers. Communication between the latter and the database server is limited to **synchronization** events during which the clients' local database copy gets compared with the server's reference database, and database records are downloaded or removed from or uploaded to the remote server as required. This ensures that identical copies of the database reside on all the client computers. Database queries and related records, however, do not need to be sent through the internet in this access scheme: by having both the database and the database management software available on the same client computer, the latter can immediately process database queries and return related records on the basis of the local copy of the database, without the need for accessing the database server. Even if the database server cannot be accessed (e.g. due to maintenance or a failure of the server), queries remain processable on the basis of the local copies of the database;

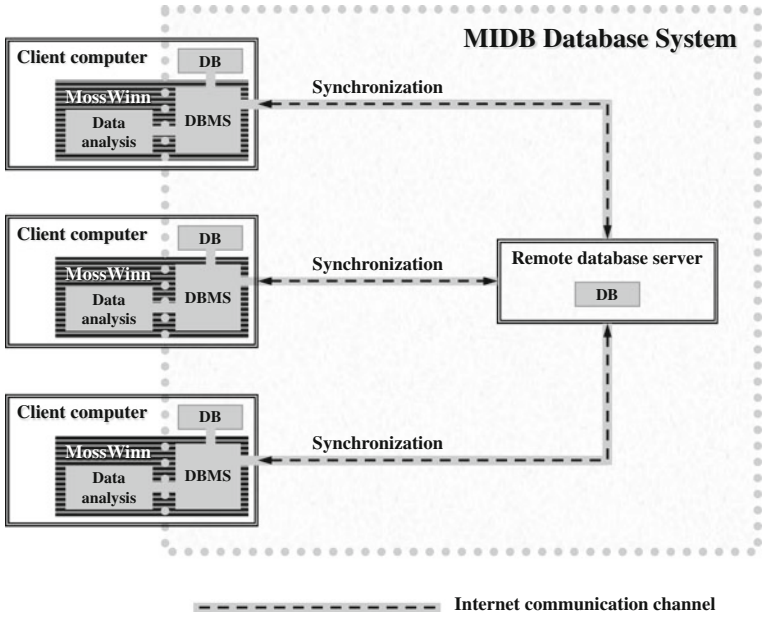

Fig. 1 Schematic structure and access scheme of the MIDB database system. The MIDB is operated via the coherent action of identical copies of the database management system (DBMS, integrated into the MossWinn program) distributed over the client computers. The DBMS copies handle queries by turning exclusively to the corresponding local copy of the database (DB), and connect to the remote database server via the internet only for the purpose of synchronization of data between the local and remote versions of the database. The DBMS also provides fast access to the database for the data analysis subsystem of MossWinn, thereby allowing the program to utilize database data for data analysis functions

only the synchronization events need to be postponed until the database server is available again.

3 The structure of database records

The schematic structure of the MIDB records is depicted in Fig. 2. Each record is associated with a single Mössbauer measurement, concerning which it can include the following information (r stands for 'required', o for 'optional'): the source nuclide (r), the stoichiometry of the measured compound (r), total surface density of the measured sample (o), temperature of the measurement (r), flux density of the externally applied magnetic field (if any) (r), orientation of the external magnetic field with respect to the gamma ray (restricted to *parallel* or *perpendicular*, o), isomer shift of the applied isomer shift reference material given relative to a specific standard [9] (r), source matrix (or absorber material in case of emission-type Mössbauer measurements) (o), publication year of the associated paper or the year of measurement in the absence of publication (o), internet link to the associated publication's official site (o), additional textual information or internet link to a website containing supplementary information concerning the measurement

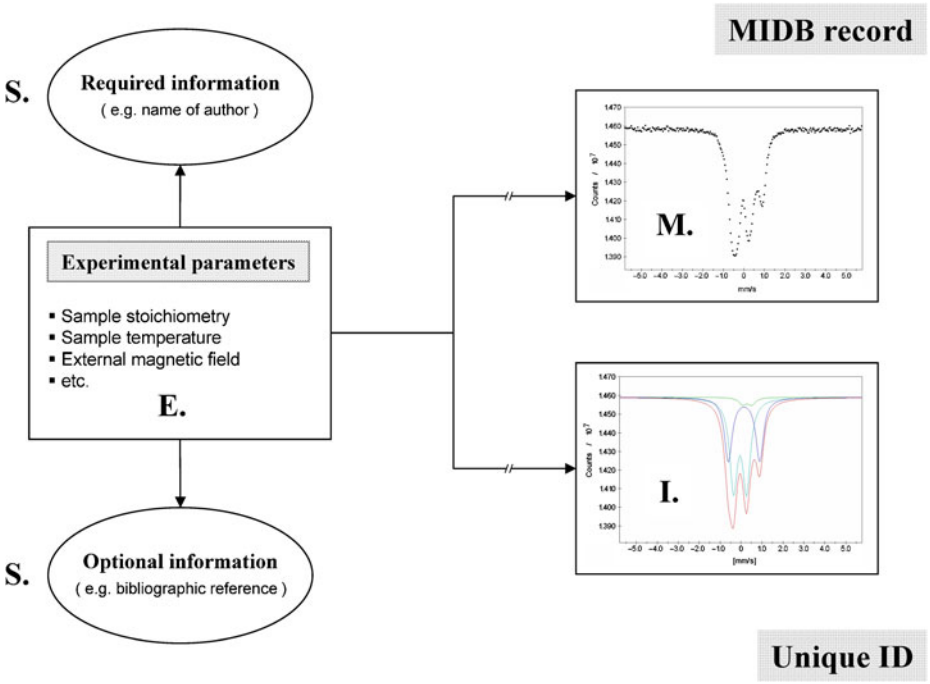

Fig. 2 Schematic structure of a single record of the MIDB database. The experimental parameters (E.) are the main parameters that influence the outcome of the measurement (M.) on which the researcher's interpretation (I.) is based. The experimental parameters are supplemented (S.) by required and optional information. The line breaks indicate that removal of either the measured spectrum (M.) or the corresponding interpretation (I.) shall be possible without rendering the record unusable. The label 'Unique ID' stands for a unique identification code associated with the record

(o), arbitrary keywords and further attributes associated with the measurement (o), name (r), title and E-mail (o) of the researcher/user publishing the record, name of the first author of the associated publication (r), and finally either a downsampled version of the measured spectrum or the interpretation (fit model) associated with it, or both. Experimental parameters not explicitly considered (e.g. effective thickness of the absorber, orientation of the sample with respect to the γ-ray direction, etc.) that may nevertheless influence the outcome of a measurement may be optionally given in the field of keywords.

4 The source of the data

The main difficulty concerning the establishment of a general purpose wide-scope database with the given structure (Fig. 2) lies in the fact that the corresponding data are accessible only to their respective owners (i.e. researchers, research institutions) with the required detail. It follows that the records of a database with the envisaged structure (Fig. 2) can originate only directly from the researchers who have measured, analyzed and presumably archived the Mössbauer spectra. It is thus plausible to operate the database system such that it relies on records submitted

to the database directly by its users, i.e. the researchers themselves. The MIDB was developed along this idea, and consequently in its case it is up to the users to decide which of their own works should be included in the database. The MIDB furthermore provides the users with the possibility to edit or withdraw the records they submitted to the database. In this sense, users themselves act as editors of the database.

Though the complete realization of the record structure shown in Fig. 2 is throughout possible, there is a reason not to include the measured spectra with full resolution into the database records. Namely, disclosure of the original numerical data of a measurement, be that in the form of a database record or otherwise, may compromise the ability of the data owner to demonstrate ownership over the corresponding data. For this reason, in the case of the MIDB the record structure displayed in Fig. 2 is realized with the restriction that instead of the original measurement only a limited precision, downsampled version of it is included in the record, which beside providing a solution for the aforementioned problem, also turns out to be adequate for the purposes of the database.

5 The functions of the database system

In practice the most important function of a database system is the one that allows the user to query the database for records satisfying certain search criteria. By applying a general purpose (lossless) compression method, it turns out to be possible to store a single MIDB record on around 2,000 bytes only. Thus, a database consisting of, e.g., 10^5 such records can be stored on $\sim 2 \times 10^8$ bytes, i.e. on ~200 MB. Not only is this kind of storage space readily available on clients and servers, but the corresponding data amount can also be loaded in the memory of typical client computers. Consequently, the speed of execution of MIDB database queries is limited neither by internet communication nor by hard disk access: in the execution of queries only memory operations are involved.

As in the case of the MIDB emphasis is put on the individual measurements and the associated experimental conditions, it is mainly the corresponding set of experimental parameters that is offered for the purpose of setting search criteria for queries. Figure 3 shows the page of the MIDB's database browser form on which the criteria can be set. There are two kinds of criteria: one that excludes records not satisfying the corresponding criterion (*exclusive criteria*), and one that influences the ranking of records (*ranking criteria*) not excluded by other criteria. Whereas the stoichiometry of the sample is treated as a ranking criterion, the source nuclide, 'Required elements', 'Excluded elements' and keywords (Fig. 3) act as exclusive criteria. The criteria concerning the sample temperature, the magnitude of the applied magnetic field and the year of publication are treated as exclusive criteria if given as a range of values, but serve as ranking criteria if given by a single value only. The exclusive criteria of 'TMS' and 'RMS' (Fig. 3) when set make the query to consider only records referring to transmission and reflection geometry measurements, respectively. The latter criteria are also mutually exclusive, similarly to the criteria of parallel and perpendicular orientation of the external magnetic field with respect to the gamma ray direction. Records can furthermore be filtered on the basis of the surname of the first author associated with the work included in the

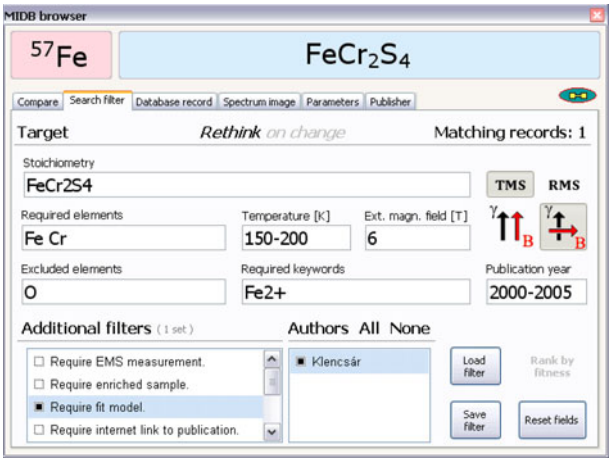

Fig. 3 The 'Search filter' page as a typical example of the database browser form on which one can specify MIDB database query criteria. The functions of the pages accessible via the tabs are described in detail in Table 1

Table 1 Functions of the tabbed pages available via the database browser form (Fig. 3) of the MIDB database system

Caption of page tab	Functions of the page
Compare	Enables one to inspect the fitness of the fit model included in the selected record with respect to one's own spectrum under analysis, and provides the possibility to apply the fit model in question in the fit menu system of MossWinn.
Search filter	Allows one to specify MIDB database query criteria.
Database record	Displays information about experimental conditions and supplementary information associated with the selected record.
Spectrum image	Displays the downsampled measurement (M. in Fig. 2) and the fitting curve corresponding to the interpretation part (I. in Fig. 2) of the selected record.
Parameters	Displays a detailed textual fit report associated with the interpretation part of the selected record.
Publisher	Displays information (name, E-mail) about the publisher of the record, and provides access to the tools that enable the filtering of records on the basis of their publisher.

record, on the basis of additional exclusive search criteria to be selected from a list (Fig. 3), and on the basis of the publisher of the record (see Table 1).

Due to the fact that in the case of the MIDB the execution of a query involves only memory operations, the execution time of the type of queries treated up to this point is very short, the associated highest-ranked records being returned instantaneously.

At the same time, by having the whole database ready for access in the RAM, computationally more intensive queries can be implemented as well. It is, for example, often a reasonable aim to find the records whose associated measurements match well one's own. The execution of the corresponding query has to involve the ranking of the records according to the fitness of their measurement part with respect to one's own spectrum under study. The evaluation of the fitness is complicated by the facts (1) that the data points of the downsampled spectrum of a record (referred to as M_{record} in the followings) are in general given for abscissa (velocity) values

different from those for which one's own spectrum (M_{own}) was measured, and (2) that the absolute amplitudes of the absorption peaks and that of the baseline counts in M_{record} may strongly differ from the corresponding parameters of M_{own}, even if the fitness of the two measurements should be evaluated as to be excellent.

Denoting the range of abscissa values of M_{record} and M_{own} by D_{record} and D_{own}, respectively, the comparison of M_{record} and M_{own} can rely only on those data points of M_{record} whose abscissa is in the range of $D_{\cap} = D_{record} \cap D_{own}$. An approximated squared deviation between M_{record} and M_{own}, denoted by Δ, can then be calculated as

$$\Delta = \sum_{\substack{\text{Over } (x_i, y_i) \text{ points of} \\ M_{record} \text{ for which } x_i \in D_{\cap}.}} \left(y_i - Y_j\right)^2. \tag{1}$$

where Y_j is the ordinate (counts) value of the data point in M_{own} whose abscissa value is closest to x_i. In order to be able to compare the values of the deviation between M_{own} and M_{record} observed for different records, Δ has to be normalized for the $n_{\cap}$ number of points included in the sum of Eq. (1), as well as for the $w_{\cap}$ width of the range $D_{\cap}$:

$$d\left(\Delta\right) = \frac{\Delta}{n_{\cap} w_{\cap}}, \tag{2}$$

where the normalized quantity of d (measured in inverse units of velocity) is a proper measure of distance between M_{record} and M_{own} if complication (2) mentioned above can be disregarded. One can overcome the latter by considering that deviations between M_{record} and M_{own} that can be sufficiently accounted for by a linear transformation are usually caused by differences in the corresponding experimental conditions that are mostly unessential in practice. The calculation of the distance between any M_{record} and M_{own} should therefore start with the determination of the coefficients m and b characterizing the linear transformation that minimizes the $\Delta_T(m,b)$ approximated squared deviation between M_{own} and the corresponding transformed version of M_{record}:

$$\Delta_T\left(m, b\right) = \sum_{\substack{\text{Over } (x_i, y_i) \text{ points of} \\ M_{record} \text{ for which } x_i \in D_{\cap}.}} \left(m y_i + b - Y_j\right)^2, \tag{3}$$

where the meaning of Y_j is the same as before. The measure of distance between M_{record} and M_{own} is then given by $d_T = d(\Delta_T(m,b))$, and $M_{record} = M_{closest}$ providing the lowest d_T value can be considered as the best match to one's own measurement.

6 Database-assisted Mössbauer data analysis

As in its interpretation part (Fig. 2) an MIDB record may include sufficient information to reconstruct the corresponding fit model on the client computers, in practice the MIDB database also straightforwardly doubles as a library for Mössbauer fit models. The combination of the database management functionalities with the data analysis functionalities of MossWinn enables the visualization of the corresponding fit model curves (1) together with the records' measurement part, as well as (2) in comparison with one's own spectrum under study. The latter facilitates the prompt

visual inspection of the selected record's fit model concerning its suitability to fit one's own spectrum. The fit model included in the selected record may then be conveniently applied as a starting point of Mössbauer data analysis.

The database queries described in the preceding section can also serve the purpose of a search for fit models associated with specific compounds and measuring conditions represented by the database records. Alternatively, an automatic fit model selection may also be realized on the basis of the reasonable assumption that the fit model associated with $M_{closest}$ (Section 5) is likely to provide a sufficient starting point for the fit of M_{own}. For the case when the fit model associated with $M_{closest}$ is not a physically reasonable model to fit M_{own}, the program also offers a list of records ordered in ascending order of distance between M_{record} and M_{own}, such that to promote the manual selection of a physically reasonable model from the list of those that are capable to provide a suitable mathematical fit.

Via the above functionalities of the MIDB one can effectively realize a database-assisted Mössbauer data analysis, where spectrum fitting starts immediately with the selection of a complete, suitable fit model as opposed to the usual practice of the model's gradual, stepwise construction and refinement. The success/efficiency of the database-assisted fit procedure will naturally depend on the scope and the number of records of the database. Provided that the database is characterized by a large number of records and a corresponding, appropriately wide scope, database-assisted Mössbauer data analysis may well form the basis of the automation of Mössbauer spectral analysis including automatic fit model selection and fitting.

The mathematical procedure (Section 5) used for the selection of the record with $M_{record} = M_{closest}$ may also be seen as a tool for the (not necessarily unequivocal) identification of unknown substances on the basis of their Mössbauer spectra. Namely, the substance whose Mössbauer measurement resulted in $M_{closest}$ may also be treated as the most reasonable origin of M_{own} on the basis of the actual content of the database. The success/accuracy of such an identification process will again critically depend on the scope and number of available database records, as well as on the additional query criteria (related, e.g., to the experimental conditions characteristic to the measurement leading to M_{own}) set out in order to minimize the probability of erroneous identification.

7 Figures of performance

In order to test the feasibility of the described database concept and that of its realization to handle a reasonably large database, the MIDB was loaded with 50,000 artificially created database records, and the performance of the database system was tested on different client PCs running MS Windows (32 bit) and having respectively (1) a single-core processor (AMD Athlon 64) running at 2 GHz, (2) a double-core processor (Intel Core 2) running at 1.6 GHz, (3) a triple-core and (4) a hexa-core processor (each AMD Phenom II) running at 2.6 GHz and 3.2 GHz, respectively. The processors (1)–(4) are characterized by their respective 'effective processor speed' calculated as the product of their clock rate and the number of their cores.

For testing purposes the parallel algorithm developed in connection with the computationally most intensive ranking process—that ranks database records according to the fitness of their measurement part with respect to one's own spectrum

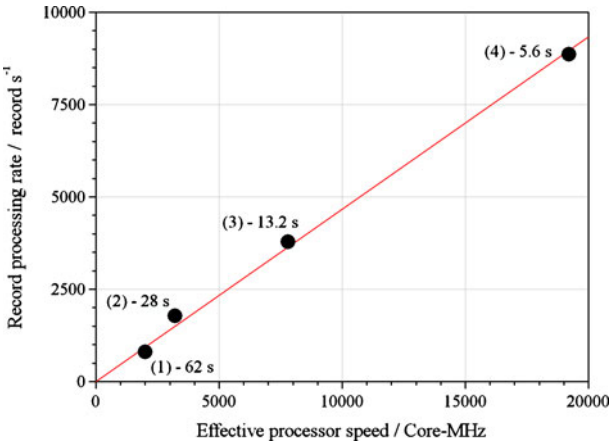

Fig. 4 Average record processing rate while ranking the records of the database according to the fitness of their measurement part with respect to one's own spectrum, as a function of effective processor speed. The actual time needed to complete the ranking for 50,000 records are shown beside the points. The straight line is a guide to the eye

(Section 5)—was executed on the different processors. The successful parallelization of the algorithm is confirmed by the approximately linear scaling of the average record processing rate of the ranking process with the effective processor speed (Fig. 4). Processor (4) completes the ranking of the 50,000 records in only 5.6 s, which is quite suitable in practice.

8 Conclusions

A new Mössbauer spectroscopy database system has been developed according to a scheme that relies on the coherent action of distributed database management softwares operating on local copies of the whole database stored on the client computers, and interfacing the remote database server via the internet only for the sake of synchronization of database records between the server and the clients. Combined with the integration of the database management and data analysis functionalities in the same application software, in the field of Mössbauer spectroscopy this approach has been realized in the form of the MossWinn Internet Database that offers novel functions and multiple potential advantages over previous database solutions. These include particularly the possibilities for

- building the database directly via the collaboration among its users,
- immediate exchange of database records among database users via the database server,
- editing and withdrawal of database records,
- database queries whose execution time is not limited by the available internet bandwidth,
- database queries carried out directly on the basis of measured spectrum data counts, and
- the seamless utilization of database records for the purposes of database-assisted Mössbauer data analysis.

References

1. Stevens, J.G.: Comput. Phys. Commun. **33**, 105 (1984)
2. Mössbauer Effect Data Center, at the University of North Carolina at Asheville (1969–2010), at the Dalian Institute of Chemical Physics, Chinese Academy of Science. http://www.medc.dicp.ac.cn/ (2010)
3. Wang, J., Jin, C.Z., Liu, X., Liu, D.R., Sun, H., Wei, F.F., Zhang, T., Stevens, J.G., Khasanov, A., Khasanova, I.: Hyperfine Interact. **204**, 111 (2012)
4. Mars Mineral Spectroscopy Database, at Mount Holyoke College. http://www.mtholyoke.edu/courses/mdyar/marsmins/ (2003)
5. Dyar, M.D., Schaefer, M.W.: Earth Planet. Sci. Lett. **218**, 243 (2004)
6. Novikov, G.V., Varlamov, D.A., Novikov, A.G.: WWW-Messbauer, Messbauer Spectral Database for Minerals and Analogues, at the Institute of Experimental Mineralogy RAS. http://messbauer.iem.ac.ru (2004)
7. Klencsár, Z., Kuzmann, E., Vértes, A.: J. Radioanal. Nucl. Chem. **210**, 105 (1996)
8. http://www.mosswinn.hu
9. Klencsár, Z.: MossWinn Internet Database Manual. http://www.mosswinn.hu/midbmanual.pdf (2011)