

Definition of a Format for Transfer, Update and Maintenance of Phase Dependent Databases (PDB) in Materials Science

SGTE
Bo Sundman
bo.sundman@gmail.com

DRAFT

Last update October 11, 2016

This document is based on a previous database format known as TDB proposed by SGTE in 1990 for the interchange of thermodynamic model parameters between different software.

This document defines a format for interchange of thermodynamic model parameters and related data between different software and some advice about structuring, updating and maintenance of a database using this format.

Revision history

Date	Revision	Notes
2016-09-16	First draft	Bo Sundman
October 11, 2016	Second Draft	Bo Sundman's interpretation of some comments.

How to read this draft

Please be aware that many definitions may require a more strict formulation to avoid misinterpretations. Please note of all of these and propose a more strict formulation.

Please also suggest a better order of presentation.

The document is long but the first part contains many explanations that is not necessary in order to use the database format. All essential user information can be found in the appendices. Read those first and go to the main document only when you do not understand or agree.

- The general concept of thermodynamic databases is first described as this document may be of interest to other people than those currently dealing with thermodynamic databases.
- Then some concepts needed for the format are described.
- Then all keywords that is part of the database format are explained. Many of these need additional information and this can be extended whenever there is a new model or parameter that should be included. All such such information are documented in the Appendices to make it easier to modify.
- Finally some advice for the database mangers and discussions of future developments.
- The appendices describe all details for anyone who wants to use this database format.

Test software

There is now a free thermodynamic software, OpenCalphad (OC), that can be used for tests of a new format. This software can read most unencrypted TDB files and a subset of this software will be modified to become a test software for the new format.

I have inserted some explanations, comments and questions in italics in the text. I propose to call this format PDB, Phase oriented DataBase

Proposed changes after 2016.10.10

- Bosse: Moved the description of the T and P dependence to an appendix.
- Bosse: Added the D suffix after the table in Appendix F.
- Bosse: Define errors messages and warnings. There should be an error if the same phase occurs more than once (not in include files) and also if a parameter is duplicated (what about include files?). Constituents that are not entered as species is an error.

Contents

1	General	6
1.1	Phase or Composition Oriented Data Storage	6
1.2	Composition dependence	7
1.3	Temperature and pressure dependence	8
1.4	Helmholtz energy models	8
2	Text Oriented Format	9
2.1	Formatting the text	9
2.1.1	Case insensitivity	9
2.1.2	Line length	9
2.1.3	Super- and subscripts	9
2.2	Terminology: elements, species and phases	10
2.3	Nomenclature	11
2.4	Abbreviations	11
2.5	The model parameters	11
2.6	The SER reference state	12
2.7	Extracting subsets of data	12
3	Syntax of the data interchange format	12
3.1	Reading sequence	13
3.2	Keyword ELEMENT	14
3.3	Keyword SPECIES	14
3.3.1	Ambiguous species names	15
3.4	Keyword TABLE_OF_MODELS	16
3.4.1	The model code	17
3.4.2	The addition code	17
3.4.3	The detail code	18
3.5	Keyword PHASE	18
3.5.1	Phase names	19
3.6	Keyword CONSTITUENT	19
3.6.1	Range of stable phase composition	20
3.7	Keyword TABLE_OF_IDENTIFIERS	20
3.8	Keyword PARAMETER	21
3.8.1	The identifier	21
3.8.2	The component array	21
3.8.3	The sign of the sub-regular parameter	22
3.8.4	Wildcard constituents	22
3.8.5	Parameters for a disordered part	23
3.8.6	The expression	23
3.8.7	Bibliographic reference	23
3.9	Keyword FUNCTION	23
3.10	Keyword DATABASE_INFORMATION	24

3.11	Keyword BIBLIOGRAPHY	25
3.12	Keyword SPECIAL	25
3.13	Keyword VERSION	26
3.14	Keyword CHECKSUM	26
3.15	Keyword DEFAULTS	26
3.16	Keyword ASSESSED_SYSTEMS	26
3.17	Keyword INCLUDE_FILE	26
4	Database manager tasks	27
4.1	The beginning	27
4.1.1	Unary data	27
4.2	Adding an assessment	27
4.2.1	Phase names	27
4.2.2	Model compatibility	28
4.2.3	Simple checks	28
4.2.4	More subtle checks	28
4.3	Estimated range of validity	28
5	Some future features	28
5.1	Databases generated from repositories	28
5.2	Material	30
6	Appendix A: Example of an interchange	32
7	Appendix B. Summary of keywords	33
8	Appendix C. Definition of model codes	34
9	Appendix D. Definition of addition codes	40
10	Appendix E. Definition of detail codes	41
11	Appendix F. Parameter identifiers	42
12	Appendix G. The constituent array for interaction parameters	44
13	Appendix H, The function of T and P	45

1 General

A thermodynamic database for materials science which deals with many different kinds of solid solutions in multicomponent systems contains data for elements and phases which make it possible to calculate the equilibrium state of such a system. Such a database can be used in various software systems and in order to make it easy to interchange data between different software systems this document defines a format using a normal text file with a small number of keywords that can be used to describe the content of the database. In addition to the thermodynamic data other information related to the phase and its composition at various values of T and P can be stored using the same basic format.

Each software which will use a database in this format should have an interface which can read the database written in this format and store the data in the internal format used by the software system. It should also have an interface to transform the internal data format of the software system to a file using this format in order to make it available to another software.

The format defined here can easily handle future extensions of the types of data stored in the database, for example thermo-physical data, mobilities etc. For more information on this see sections 3.7 and 5.2.

1.1 Phase or Composition Oriented Data Storage

In thermodynamic databases for substances it is common to store data according to the composition of a phase. This means that the data for pure Fe with bcc structure (ferrite), with fcc structure (austenite) and liquid are all stored as one dataset for pure Fe. For the given composition a single expression is given for temperatures from 298.15 K and up. This method is not suitable for solution databases for two reasons:

1. The first reason is that for solutions it is necessary to extrapolate the thermodynamic data outside the stable range of the constituents because when other elements dissolve in the austenite, ferrite or liquid the stability range in T changes for the phase. In a substance databases such extrapolations are only made for the gas but in a solution database there must be thermodynamic data for each condensed phase for the whole temperature range. This means also that it is impossible to know the stable phase for a pure element at a given T without calculate and compare the Gibbs energy of different phases for the pure element.
2. The second reason is that interaction parameters are needed to describe the non-ideal solution phases. Such parameters are associated with two or more constituents but only one phase. Thus duplication and ambiguity is avoided if one stores these referred primarily to the phase.

In solution databases one must in addition decide on a model to describe the interaction between the constituents of a phase. Many different models are discussed in detail in [10] and here it is sufficient to note that each phase can have a separate model. However, a specific phase must be described with the same model independent of its constituents because

otherwise it would not be possible to combine components in an arbitrary way. This makes the decision on the model to use for each phase of crucial importance in a solution database.

1.2 Composition dependence

In all solution models of the Gibbs energy the composition dependence is expressed explicitly using a series expansion in the fractions of the constituents of the phase. Some phases can have internal degrees of freedoms, for example a gas phase with many constituents containing the same element, the equilibrium fraction of the constituents must be determined by a minimization of the Gibbs energy. In other cases a phase may have contributions to the Gibbs energy from additions, the value of which depend on other composition dependent properties. One example of this is the Inden model, described for example in Lukas[10], for the magnetic contribution to the Gibbs energy due to ferromagnetic ordering. In this case the composition dependence of the critical temperature for magnetic ordering, and the Bohr magneton number, must be described separately and for a given composition the value of this critical temperature and the Bohr magneton number are needed in order to calculate the magnetic contribution to the Gibbs energy.

The interchange format described here can express the composition dependence of any number of composition dependent quantities, in addition to the Gibbs energy. For each phase the same series expansion in the fraction of the constituents of the phase is used for all quantities. The constituents of a phase can be the components of the phase, molecules in a gas phase or constituents on different sublattice sites in a crystalline phase. The basic expression for this series expansion is:

$$Z_M = \sum_i x_i {}^0Z_i + {}^EZ_M \quad (1)$$

where Z is the quantity which is composition dependent, x_i is the mole fraction of constituent i , 0Z_i is the value of this quantity for the pure constituent i (this value may be a function of temperature and pressure). The summation in eq. 1 represents a linear combination of the values for the constituents. EZ_M is the excess part which depends upon the interaction between the constituents. The subscript M means that the data are for one mole of formula unit of the phase.

Note that eq. 1 does not contain a term for the ideal entropy of mixing, $RT \sum_i x_i \ln(x_i)$, because such a term is only present in the Gibbs energy expression. The ideal entropy does not contain any adjustable parameters and it is not included in the interchange format.

If a phase has sublattices there will be one additional summation in the first term of eq. 1 for each sublattice. On the sublattices one may have the same constituents (in order to describe chemical ordering) or different (for example interstitial solutions or inter-metallic compounds):

$$Z_M = \sum_i y_{i,1} \sum_j y_{j,2} \dots \sum_n y_{n,s} {}^0Z_{ij\dots n} + {}^EZ_M \quad (2)$$

where $y_{n,s}$ is the site fraction of constituent n on sublattice s . For the sublattice model the configurational entropy expression is given by:

$$R \sum_s a_s \sum_i y_{is} \ln(y_{is}) \quad (3)$$

where a_s is the relative number of sites on sublattice s . The values of a_s are given in the definition of the phase.

For a substitutional model the excess part of Z_M must be zero whenever the mole fraction of any constituent is unity. The simplest expression with this property is the regular solution model:

$${}^E Z_M = \sum_i \sum_{j>i} x_i x_j Z_{ij} \quad (4)$$

The value of Z_{ij} is a binary parameter for the system i-j. One may make Z_{ij} temperature, pressure and composition dependent as defined by the model index. If there are sublattices it should be noted that i and j must belong to the same sublattice and there will be one additional summation in eq. 4 for each sublattice. The composition dependence of Z_{ij} is often expressed by a Redlich-Kister polynomial:

$$Z_{ij} = \sum_{\nu} (x_i - x_j)^{\nu} Z_{ij}^{\nu} \quad (5)$$

and one additional subscript, ν , is needed to specify the place of the parameter Z_{ij}^{ν} in the Redlich-Kister expression. This subscript is called the degree of the parameter. There can be other excess models and for these the degree can mean other things. There are no additional complications if one has sublattices also but note that one may have composition dependent excess terms on each sublattice separately.

One may have ternary parameters with three subscripts for constituents on the same sublattice and in some cases even higher order excess terms. However, from eqs. 1 to 5 it should be evident that the interchange format need only contain the parameters $Z_{ij\dots}$, where the three dots mean the degree or additional constituent indices, for each phase together with the model index and the structure of the phase in order to reconstruct the full expression for Z_M . The place of a parameter in this expression is determined by the phase, the constituents and the degree of the parameter given as subscripts.

1.3 Temperature and pressure dependence

Temperature, T , and pressure, P , are potentials and most parameters $Z_{ij\dots}$ in eqs. 1 to 5 are functions of temperature and pressure. For T -dependence the function is often described by a power series but special models for the low T dependence can be used. As many model parameters describe metastable states it is important to restrict the T dependence to avoid unphysical features like a solid phase reappearing at very high T .

Modeling the pressure dependence usually require a special model with several parameters that may depend on T , P and composition, for example the thermal expansion and the bulk modulus.

1.4 Helmholtz energy models

This format is intended mainly for models that have explicit P dependence and are thus not well suited for storing models using the Helmholtz energy with explicit volume dependence. The main reason is that all use of the current TDB format is for Gibbs energy models. It

would also be a complication to use a Helmholtz energy model because V is an extensive variable and thus depending on the size of the system whereas P is a potential. But it may be possible to handle such model in the future.

2 Text Oriented Format

This document use a simple text format for the interchange of the thermodynamic data for a system. Inside this text format specific keywords are used for the structured data. The big advantage with this is that the database is readable by a human and can be edited using a normal text editor. In fact the database will normally be updated by editing the database as a text file by a database manager. The format should be designed to to make life easier for him or her, see section 4

The intention is that a freely available software will be developed to check the syntax of a database using the format described here. It is up to the developer of a specific software to write an output form this program to conform with his or her software.

All lines (except blank lines) that are comments should have a dollar sign, “\$” as first non-blank character.

2.1 Formatting the text

There are some basic rules to simplify the reading, editing by a human and also parsing the database by a software.

2.1.1 Case insensitivity

When a database using the interchange format is read by a computer program all lower case letters should be treated as upper case, thus Fe, FE and fe are the same name. To make it easier to read by a human upper and lower case can be mixed when editing the database file.

2.1.2 Line length

Some software have fixed length for reading a text line from a file and a maximum length of a line in the database should not exceed 128 characters.

2.1.3 Super- and subscripts

Super and subscripts are not allowed in this format. Appropriate ways to handle cases where such are used in plain texts will be explained here. A parameter like

$${}^{\circ}G_{\text{Fe}}^{\text{bcc}}$$

will be written as G(BCC,FE;0) (In fact the “;0” is redundant). A Redlich-Kister interaction parameter

$${}^2L_{\text{Cr.Fe}}^{\text{bcc}}$$

is written $G(\text{BCC},\text{CR},\text{FE};2)$, (or $L(\text{BCC},\text{CR},\text{FE};2)$) here the value after the semicolon is important, see section 3.8.2. It is allowed to use L or G for these interaction energy parameters.

To describe the mobility of an element in a phase, for example Fe in bcc, one may use a notation

$$MQ_{\text{Fe}}^{\text{bcc}}$$

and to describe the composition dependence of this property in the Cr-Fe system one will need parameters like $(MQ_{\text{Fe}}^{\text{bcc}})_{\text{Fe}}$ for the self-diffusion of Fe in BCC-Fe and $(MQ_{\text{Fe}}^{\text{bcc}})_{\text{Cr}}$ for the diffusion of Fe at infinite dilution in BCC-Cr. This “double subscript” is handled by incorporating the diffusing element in the parameter identifier like $\text{MQ\&Fe}(\text{BCC},\text{FE})$ and $\text{MQ\&Fe}(\text{BCC},\text{CR})$ which is compatible with the notation used in the current TDB format.

This does not mean all software must use this notation but they can transform this notation to their own.

2.2 Terminology: elements, species and phases

Thermodynamics is used by many different sciences and each has developed its own terminology. Even within materials science there are different use of the same terms and often strong feelings when a term is used differently from what is considered correct. However, the TDB format introduced more than 25 years ago by SGTE has been widely accepted within the CALPHAD community and also for publications so a new format should be based on this.

In the TDB format the ELEMENT is a basic part of a thermodynamic database. In addition to the real elements in the periodic chart one may use fictitious elements. The electron, denoted “/-” and the vacancy, denoted “Va” are also treated as elements but a system with electrons and vacancies have as additional constraints that each phase must be electrically neutral and that the chemical potential of vacancies must be zero at equilibrium. The elements has a unique one or two letter symbol.

The elements can form SPECIES with a fixed stoichiometric ratio of the elements. Electrons and vacancies can be part of a species. This use of species is contested by some chemists who consider that a species cannot exist without a phase designation. For a chemist a species H_2O must either be solid, liquid or gas but in this document a species H_2O is just a stoichiometric formula. This species can be a CONSTITUENT of many phases.

All data are stored in connection with a PHASE. A phase has a name, a model and species as constituents and the same species can be a constituent of many phases. This whole document is devoted to describe how a PARAMETER for the different properties of a phase can be stored depending on its model and constituents, T and P .

In addition to the elements and species the term component is commonly used in thermodynamics but that term not used in the database format because components are related to the way a system is controlled by setting conditions on the amounts or chemical potentials.

As described below this document provides a possibility to invent names for species that is not necessarily the same as the stoichiometric formula because there can be a problem when species names are similar or have some structure information as discussed in section 3.3.1.

For solid phases there is no general standard for names but some recommendations will be given in section 3.5.1.

2.3 Nomenclature

The description of the format used in this paper is simple and all efforts have been used to avoid ambiguities. A string enclosed within brackets “<” “>” describes the type of information that should appear there but other words or symbols used in the description should be used literally. An exception to this is necessary in order to denote that a piece of information may be repeated. Such information is enclosed by slashes and sometimes the number of repetitions will precede the slash. In this case neither brackets nor slashes appear in the actual data string.

All names, symbols, identifiers, codes and keywords defined in this document may consist of the letters A-Z, the numbers 0-9 and the underscore character “_” in any order **but must start with a letter**.

A hyphen “-” can be used instead of underscore and must be treated as identical to an underscore.

In a few cases, explicitly described below, names may contain other characters or have a special form. To make life easier for a database manager a name can be abbreviated if the abbreviation is unique but it is recommended that full length is used. As already stated lower case letters are not distinguished from upper case.

Numeric values that should appear in certain places according to this format but which have no value should be indicated by the word “NONE” rather than “0” (zero). In some cases it is also possible to use the word “UNASSESSED” in as explained in section 3.8.

2.4 Abbreviations

Normally abbreviations of names, symbols, keywords etc should be avoided but as the databases will be edited by humans it is natural that they sometimes use abbreviations when editing the database. As the intention is to develop a software to check if a database conforms to this document document this software will report ambiguities and extend any abbreviations to its full form when requested to write the database on a file.

2.5 The model parameters

The data stored in a thermodynamic database describe the Gibbs energy of each individual phase as a function of the temperature, pressure and the constitution of the phase. The reason for selecting the Gibbs energy rather than any other quantity is discussed in [1].

Already in the TDB format some additional properties that were needed to calculate the Gibbs energy could be stored, for example the critical T for magnetic ordering and the Bohr magneton number. In this document many other properties that are related to the phase can be stored, for example mobilities, individual Bohr magneton numbers and Debye T . See section 3.7.

In the database the elements, species, phases with their constitution create the fundamental structure. All kinds of data are stored as model parameters for each phase for a specified constitution as functions of the potentials T and P .

2.6 The SER reference state

The first Calphad databases 40 years ago usually had one phase with zero for all Gibbs energy parameters for the pure elements and used only lattice stability values for the other phases. With the SGTE unary database in 1991 most databases started to use the Stable Reference State (SER), i.e. the stable state at 298.15 K and 1 bar as reference state. The change made it possible to use the Calphad databases also to calculate heat capacities and heat balance equilibria.

Although it is still possible to have a database with only lattice stabilities in a database following to this document that is discouraged and will not be discussed further in this documentation.

2.7 Extracting subsets of data

The interface software should be able to extract a subset from the database by specifying a set of elements and obtain a complete subset of all species, phases, functions and model parameters. It can be quite complicated to extract a subset of a large database using a text editor.

3 Syntax of the data interchange format

With one exception the interchange format consists of sequences of the following form:

`<keyword> <body of information> !`

The keyword must be the first non-blank part of a line after an LF or CR. It may be preceded by spaces and tab characters only. The exclamation mark, “!” is used to terminate the information that belongs to the keyword and there must not be any exclamation mark inside the body of information part.

Any control character between the keyword and the exclamation mark will be ignored except TAB, FF, CR and LF which will be replaced by a single space.

In the original TDB format all CR and LF were ignored in the body of information (not replaced by space) because the strange behavior of electronic transfer between computers around 1990, often using raw X25. In many cases LF and CR were inserted in arbitrary places, or after each 80 character, in the text that was transferred. Today we can expect that this will not happen

A keyword should be unique with its first 6 characters and its first character must be alphabetic. Note that the keyword itself must be followed by at least one space. The keywords currently defined in this interchange format will be described in detail below and are summarized in Appendix B.

The order of the keywords fairly free, except for a few cases specified below. With one exception, the keywords must not be nested, i.e. after a keyword there must not be any other keywords until after the terminating exclamation mark, “!”. It is possible to add comments after the ! up to the end of the line.

Only one keyword can appear on one line but the body of information may span several lines.

3.1 Reading sequence

The main idea is that it should be possible to read the database file sequentially once to obtain all information. This means there are a few rules of the order of keywords:

- The VERSION keyword must be the first keyword.
- An ELEMENT must be defined before it is used in a species or as constituent.
- A SPECIES must be defined before it is used as a constituent.
- The TABLE_OF_MODELS must appear before any PHASE keyword.
- The CONSTITUENT keyword must follow directly after the PHASE keyword.
- The TABLE_OF_IDENTIFIERS must appear before any PARAMETER keyword.
- A PARAMETER keyword for a phase must be after the keyword for the phase.
- The FUNCTION keyword can appear anywhere.
- The BIBLIOGRAPHY, DATABASE_INFORMATION, ASSESSED_SYSTEMS, can appear anywhere but only once.
- The SPECIAL, INCLUDE_FILE and DEFAULTS keyword can appear anywhere several times.
- The CHECKSUM keyword should be last if used.

The FUNCTION keyword is a bit problematic as it is a tedious work for the database manager to ensure that a function is defined before it is used in another function or parameter. Thus it is recommended that the software interface accepts the use of an undefined FUNCTION when reading a PARAMETER or another FUNCTION. If there are any such undefined functions when the database has been read, the database file can be rewound and searched again for just FUNCTION keywords for missing functions. This may have to be done a few times.

Even if all functions are read it would be necessary to have them in calling order to avoid errors when entering a function calling another function that is not yet entered. So rewinding and just reading FUNCTION keywords for missing functions will ease the work of the database manager.

Some database managers prefer to order all parameters by phase, some prefer to order them by binaries, ternaries etc. This is up to the manager.

3.2 Keyword ELEMENT

Elements are those included in the periodic chart. They are referenced by their chemical symbols, i.e. their names are one or two letters long. It is possible to have hypothetical element names e.g. A or Z. A few extensions to the periodic chart are necessary. Thus vacant sites on a sublattice are treated as a component and the name VA has been adopted. At equilibrium the chemical potential of vacancies must be zero. Electrons are denoted by the combination /- and a positive charge can be denoted by /+ or /- -. At equilibrium the net charge of all stable phases must be zero.

Upper and lower case letters are not distinguished and thus CO, Co, co and cO can all be used to denote cobalt. Note that this does not mean that one cannot use lower case for the second letter in a particular implementation of the database. The interface software can change this when it uploads data from the interchange format.

All elements have a defined state at 298.15 K which will be called Stable Element Reference state (SER). The name of this state (e.g. FCC or H2_GAS etc.), the atomic mass, the entropy at 298.15 K and the difference in enthalpy between 298.15 and 0 K for the element in that same state will be given at the element keyword.

```
ELEMENT <element name> <SER state> <atomic mass>
      <enthalpy at 298.15 K - enthalpy at 0 K>
      <entropy at 298.15 K> !
```

The element name can be one or two characters. Examples:

```
element fe bcc_a2          55.847  4489.0  27.280 !
element am double_hcp      243.06   6407.0  55.396 !
element h 1/2_mole_of_h2_gas 1.0079  4234.0  65.285 !
```

Note that the values of $H_{298} - H_0$ and S_{298} are never used for calculations above 298.15 K.

The current interest to expand the database to 0 K using Debye or Einstein models does not require any changes in this keyword because such data will be given relative to the phase.

3.3 Keyword SPECIES

Species are aggregates of elements with fixed stoichiometry. The elements are the simplest types of species and they do not have to be given as species also. The name of a species is usually its chemical formula but must not exceed 24 characters. The stoichiometric factor 1 can be excluded if there is no possibility of misinterpreting. This means CO1 or CO is cobalt while carbon-monoxide must be written C1O or C1O1 because we does not distinguish between upper and lower case letters and cannot use “Co” for Cobalt and CO for carbon monoxide. In fact all single letter element symbols should be followed by a stoichiometric number except if it is the last element.

Note that in chemistry a stoichiometric formula has almost always a phase designation like g (for gas), l (for liquid) or s (for solid) and many chemists do not like using a stoichiometric formula without such a designation. In this documentation the same species

can be a constituent of many phases and it is convenient to define the stoichiometry of the species separate from the phase. A species which is present in a particular phase is called a constituent of the phase. A phase with fixed composition can have a single species as constituent.

In this document a species can have a name that is different from its stoichiometric formula. This is useful to identify isomers and simplify the name of a species with a complex stoichiometry.

Parenthesis are not allowed either in the name or in the stoichiometric formula. A period “.” is not allowed in the name of a species but can of course be used to specify a non-integer stoichiometry. In the name of a species a hyphen can be used to indicate negative charge and a “+” can be used to indicate a positive charge, although the name of a species does not have to have any indication that it is charged.

SPECIES <species name> /<element name> <stoichiometric factor>/ !

There should be no space between the element name and stoichiometric factor. Examples

```
species h2o h2o1 !
species c1o c1o1 !
species mg2sio4 mg1o2si0.5 ! Can be confusing for the user
species s1o2 s1o2 !
species o-2 o1/-1 !
species fe+2 fe1/--2 !
species s1o4-2 s1o4/-2 !
species c2cl2h4_cis c2cl2h4 ! These two species -
species c2cl2h4_trans c2cl2h4 ! are isomers
species this_is_a_very_long_name fe1 ! Allowed but not recommended.
species feo fe0.9870 !
species ax al2ca4h26o20 ! A very short name
species hole va/+1 !
```

The elements in the stoichiometric formula do not have to be in alphabetical order. The stoichiometric factors can be real numbers. A final stoichiometric value of 1 can be ignored. An ion must have its charge specified in the stoichiometric formula, a divalent positive charge can be written “/-2” or “/+2”. No parenthesis are allowed in species names as well as for the stoichiometry. Vacancies are also species but an electron by itself is not allowed as species but can be entered as a charged vacancy.

3.3.1 Ambiguous species names

Many species names are abbreviations of another species. Thus exact match is required when species are used as constituents in model parameters. A very long stoichiometric formula one can be given a short name.

3.4 Keyword TABLE_OF_MODELS

There are 3 levels of codes that can be provided for a phase, the MODEL, any ADDITION and possibly some DETAIL. At least a MODEL code must be specified for each phase.

All codes used for a phase in the database either MODEL, ADDITION or DETAIL, must be given in this table. The table may contain codes that are not used in the current database.

After this keyword each line must start with MODEL, ADDITION or DETAIL followed by a unique code. The code must be unique among all of the three types.

After the code a free text can be written within single or double quotes. This text is just informative but may be used to display information for a user of the database. A complete list of the codes and detailed explanation are given in the Appendix C, D and E.

TABLE_OF_MODELS

```
model IDEAL  'for example ideal gas'
model SUBRKM 'Substitutional Redlich-Kister_Muggianu'
model SUBPK  'Substitutional polymer_Kohler model'
model I2SL   'ionic_liquid_two-sublattice model'
model KFGL   "Kaphor-Frohberg-Gaye-Lehmann irsid slag model"
model CEFj   'Compound Energy Formalism with j sublattices'
model CEFjD  'CEF with j sublattices and disordered part'
model CEF4F  'CEF with 4 sublattices for FCC ordering and permutations'
model CEF4B  'CEF with 4 sublattices for BCC ordering and permutations'
model CEF5F  'CEF with 4 subl for FCC ordering, permutations and interstitials'
model CEF5BD 'CEF with 4 subl for BCC ordering, permutations, interstitials and disordered par
model GUTS   'FactSage liquid model'
model BOND   'Model based on bond energies between sublattices'
addition IMAGB 'Inden magnetic ordering model for BCC'
addition IMAGF 'Inden magnetic ordering model for FCC and HCP'
addition IWMAGB 'Inden-Wei magnetic ordering model for BCC'
addition IWMAGF 'Inden-Wei magnetic ordering model for FCC and HCP'
addition LOWTD 'Low T Debye model'
addition LOWTE 'Low T Einstein model'
addition MURN  'Murnaghan pressure-volume model'
detail MGAP   'Miscibility gap with major constituents specified'
detail DCON   'Default major constituents specified in all sublattices'
detail STRUCT 'additional information about the structure of the phase'
!
```

This is quite clumsy but normally one does not have to edit this very often. When the table is edited one can use new lines freely as they are ignored when reading the body of information.

When a code is used after a phase additional information can be provided within parenthesis as explained below. The expressions how to calculated the Gibbs energy for the various models are given in Appendix C and additions in Appendix D.

3.4.1 The model code

This code need more information to be understood by the software. Normally the software must have the basic model implemented when reading the database, otherwise a warning or error message must be given that the phase will be ignored.

There is no dilute solution model or model for the activity coefficients but such models can be converted to a model using the integral Gibbs energy using for example the paper by Hillert[3] or Pelton and Bale[4].

For all CEF models there must be a number 1 to 9 directly after CEF specifying the number of sublattices (one may use 1 to have more than one site in the substitutional lattice) and then within parenthesis the number of sites on all sublattices. For example

```
PHASE SIGMA CEF5(2.0 4.0 8.0 8.0 8.0) !
```

The letters F and B after the number of sublattices denote the special case when the software should take care of model parameter permutations. In a 4 sublattice ordered fcc phase a parameter like $^{\circ}G_{\text{Al:Ni:Ni:Ni}}$ has 4 permutations that should have identical values. This is specified by the F. The B means the same for the ordered bcc phase (which has a different set of permutations). The 4 sublattice hcp phase has the same permutations as fcc.

The letter D means that there is a separate set of model parameters for the substitutional disordered state where the 4 ordered sublattices has merged. The model equation for this can be found in Appendix D. If there is an interstitial sublattice that is separate also in the disordered phase.

```
PHASE BCC_4SL CEF5FD(0.25 0.25 0.25 0.25 3) IMAG!
```

A software that does not handle permutations will have to expand these automatically when reading the parameters for this model.

The parameters for the substitutional disordered part can be detected by the number of sublattices in the constituent array of the parameter. But to be on the safe side this format requires that such parameters have a suffix D on the parameter identifier. See section 3.8.5.

For each model identifier a written documentation how to implement the model must be given in Appendix C of this document, preferably with a reference to a published paper.

There are many complicated models that may need extra information. Unsymmetrical models like Toop or power series need additional information and I am not sure how to specify that.

3.4.2 The addition code

Additions are contributions to the Gibbs energy due to some additional physical model. The IMAGB and IMAGF additions are well established. The others are not yet implemented in most software.

There can be several additions for a phase. For each addition identifier a written documentation how to implement the model must be given in this document. The defined additions will be given in Appendix D.

3.4.3 The detail code

The details given for a phase may be ignored by the software reading the database but it can be useful in some cases. One detail can be additional structure information like

```
PHASE HALITE CEF2(1.0 1.0) IMAGF DETAIL(STRUCT(B2, PERICLASE)) !
```

In some cases the details may be useful to generating start constitutions as it is not always possible to calculate a global equilibrium, for example if T is not set as a condition.

For example the MGAP can be used to indicate there can be a miscibility gap in the phase and provide default state values. In the example below it is indicated that the FCC phase can be stable with mainly Ti or V in first sublattice and C and N in second. There may even be a proposal for a modified phase name (with a pre- or suffix), MC_CARB, for this phase included.

```
PHASE FCC_A2 CEF2(1.0 1.0) IMAGF MGAP(MC: TI V : C N:CARB) !
```

Note that a phase may have a miscibility gap even if there is no MGAP specified.

The DCON can also be handled by the % suffix as described in section 3.6. These may be considered to be a way to help the calculation software to find the equilibrium. It can also be used to avoid problems to identify which phase is the carbide and which is the austenite!

There can be several details for a phase. For each detail code a written documentation is given in Appendix E.

3.5 Keyword PHASE

Phases are homogeneous parts of the thermodynamic system with uniform structure and composition. In some cases one may use the word phase also to denote parts which have uniform structure but non-uniform composition although such a system cannot be in equilibrium. For solutions the phase is the central concept as data are stored primarily according to the phase it is associated with.

```
PHASE <phase name> /<model>/ /<addition>/ /<detail>/ !
```

The “model”, “addition” and “detail” are explained below. Some of them have additional information.

Examples, note the text after ! are comments.

```
phase bcc_a2 SUBRKM IMAGB ! Substitutional solution and Inden magnetic bcc model
phase fcc_4SL CEF4F(0.25 0.25 0.25 0.25) IMAGF ! CEF with 4 ordered sublattices
                        and the Inden magnetic fcc model
phase gas IDEAL ! Ideal gas
phase wustite CEF2(1.0 1.0) IMAGF !
phase m23c6 CEF3(2.0 21.0 6.0) ! carbide with 3 sublattices
```

In the previous TDB format the “type definition” keyword was used for many strange things and in this format we have tried to be more explicit about the models and introduced model specifications that should be explained in the database file.

The constituents of a phase are a subset of the species. This subset is given either by a “constituent” keyword.

In the TDB format the CONSTITUENT keyword was optional as the constituents could be deduced from the PARAMETER keywords. However, this is not allowed in this document as it is easy to make spelling mistakes or use wrong abbreviated species names.

3.5.1 Phase names

There is no generally excepted way to name a phase. For an ideal gas the name “gas” is obvious and for a substitutional liquid one may use “liquid” but there are several different models for liquids. For solid stoichiometric phases one may use the chemical formula but if there are several allotrophs it is not so convenient to have suffixes like _S1 and _S2 for that. For crystalline solution phases one may use the lattice type like FCC or BCT but there are many different FCC phases with various interstitial places filled with other atoms or vacancies. The Structur Bericht is useful but it is far from complete. For some phases there are established names like Laves_Phase or MU_phase. In general it is not convenient to use transcriptions of greek letters as ALPHA which can mean different phases in different system. Even if exact match is required to identify a phase one should avoid having a phase name which is an abbreviation of another phase name.

3.6 Keyword CONSTITUENT

The constituent keyword must follow directly after the corresponding PHASE keyword. The syntax is

```
CONSTITUENT <phase name> : <species on first sublattice>
/ : <species on other sublattices in sublattice order>/ : !
```

This keyword must follow directly after the PHASE keyword but anyway the phase must be specified. The number of sublattices with constituents must correspond to the specified for the phase keyword. Note that colon “:” are used between the sublattices and comma “,” or a space between constituents in the same sublattice

Examples

```
phase gase ideal !
constituent gas :c c1h4 c2 c2h2 c2h2o1 c1o c1o2 h h2 h2o h2o2 o o2 o3: !
phase fcc subbrkm !
constituent fcc :fe al si: !
phase wustite cef2(1.0 1.0) !
constituent wustite :fe+2 fe+3 va: o-2 : !
```

3.6.1 Range of stable phase composition

In a general thermodynamic database many phases will dissolve all constituents. When retrieving data from the database for a particular subsystem there will be a large number of phases included that are never stable because the set of stable phases. This is due to the fact that phases may dissolve many constituents but the phase may be stable only for small additions of many of the constituents. The set of stable phases will of course change with T , P and composition of the system and this can only be determined by an equilibrium calculation.

For example in the Cu-Zn system neither Cu nor Zn exist as pure in the bcc phase but nevertheless there are four bcc related phases stable in the middle of this system.

In order to facilitate for the calculation software the CONSTITUENT keyword can be used in order to specify MAJOR constituents on each sublattice of a phase. The major constituents should have a suffix consisting of a per-cent sign “%”. For example:

```
CONSTITUENT FCC_A1 :FE% CR NI% MN% S: VA% C: !
```

Where Fe, Ni and Mn are major in the first sublattice and Va in the second (interstitial) sublattice. There is an alternative way to specify the same information as described in section 3.4.3.

There is no facility to specify limits of the composition range of a phase. This may be a feature tempting facility to implement but would be very complicated to handle in a multi-component database. It is very complicated for a software to handle if the composition of a stable phase cross such a stability limit for a component. Should the calculation terminate with an error message? The calculation may occur during a long time simulation where the phase is transforming.

It is the task of the database manager to verify that an assessed subsystem does not create problems when added to a multicomponent database. Frequently additional model parameters have to be added or even assessed parameters slightly modified. See also section 3.10.

3.7 Keyword TABLE_OF_IDENTIFIERS

A model parameter identifier determine which property the parameter is associated with. In the current TDB files these are normally those listed in section 3.8.1 but in the new format this will be considerably extended.

A parameter identifier describe a property that can be composition dependent using the same model as the Gibbs energy. Some of them can also depend on T , P and be specific for a particular element or constituent.

We should use short identifiers, at present the maximum length is 4 letters and we should try to keep that limit. But it is important to have unique identifiers!

A special keyword should list all the parameter identifiers used in a database before the first parameter in the database. All identifiers that may be used are listed in Appendix F.

```
TABLE_OF_IDENTIFIERS /<identifier> {<explanation>}/ !
```

The table is useful to check for typing errors at least.

It is important to establish a syntax for all identifiers as they may depend also on the constituents of the phase even if several of the identifiers may not have any associated code in a some software. In such a case the software can just ignore the parameters with this identifier.

3.8 Keyword PARAMETER

All temperature and pressure dependent data for a phase is given using the PARAMETER keyword. With this keyword a phase and a composition or composition range must be specified followed by an expression. In the simplest case the parameter gives the Gibbs energy of formation of a phase with a single constituent. In more sophisticated models the parameter may be part of a complex Gibbs energy expression. Note that the Gibbs energy expression of a solution phase may have many coefficients in its composition dependence. Each of these coefficients will appear in a PARAMETER keyword, giving the temperature and pressure dependence of one coefficient. See section 1.2 for more information. The syntax is

```
PARAMETER <identifier> ( <phase name>, <component array> ; <degree> )
    <expression> <bibliographic reference> !
```

The first part of the information after the keyword PARAMETER but before the equality sign is the “name” of the parameter. The <identifier> after the PARAMETER denote the type of composition dependent quantity and it must be standardized. The following identifiers are used currently:

3.8.1 The identifier

The identifier gives the type of parameter. In the TDB files these were limited to a very few:

Symbol	Unit	Notes
G	J/mol formula unit	Energy parameters that are part of the explicit composition dependence of the Gibbs energy
L	J/mol formula unit	Identical to G parameters but mainly used for interaction parameters
TC	Kelvin	Curie temperature parameters
BMAG	dimensionless	Bohr magneton parameters

In the new format this will be extended considerably as described in the section 3.7 and Appendix F.

3.8.2 The component array

The syntax of the <component array> in the PARAMETER information is composed of names of species separated by comma “,” for species in the same sublattice and using colon “:” to separate groups of species that go into different sublattices. A recursive definition is:

<component array> is
 <species name> or
 <species name>,<component array> when there are more species in
 the same sublattice,
 <species name>:<component array> when there are more sublattices.

There must not be any spaces in a component array. Note that the parameter in the Gibbs energy expression should be multiplied with the fraction of the constituents given by the component array.

Whenever the value of a model parameter may depend on the order of the constituents, the alphabetical order will be used unless something else is explicitly stated, see the section 3.15.

The <degree> in a parameter can be void or a value from 0 (zero) to 9. If the parameter is a binary interaction parameter in a Redlich-Kister excess model the degree means the degree as a coefficient in a Redlich-Kister polynomial. For other parameters and models the degree has different meanings. In some cases the parameter depend on the order of the constituents as discussed in Appendix G.

3.8.3 The sign of the sub-regular parameter

In the Redlich-Kister polynomial the sign of the parameter for odd powers depend of the order of the interacting constituents. By default in this format the constituents will always be ordered alphabetically according to the species names.

The database manager can always arrange the sign to be correct but if he or she prefers that the sign should depend on the order the constituents are written in the constituent array.

For example:

PARAMETER L(FCC,FE,CR;1) will by default be multiplied with $(x_{Cr} - x_{Fe})$

If the database manager wants it to be multiplied with $(x_{Fe} - x_{Cr})$

that must be specified in the DEFAULTS keyword. Such a change will apply to the whole database.

3.8.4 Wildcard constituents

It is possible to specify “*” in a sublattice to indicate that the parameter is independent of the constitution on this sublattice.

Parameters with a wildcard in a sublattice will be added together with parameters which also have a specific constituent in that sublattice. Thus both the parameters

PARAMETER L(FCC,FE:*:VA) ...

PARAMETER L(FCC,FE:C,VA) ...

should be included in a calculation of the Gibbs energy of the FCC phase for a system with C and Fe.

3.8.5 Parameters for a disordered part

As explained above in section 3.4.1 there may be parameters with for an ordered phase that belong to a disordered part. These will have a suffix D on the parameter identifier in addition to have fewer sublattices.

```
PARAMETER G(FCC_4SL,AL:NI:NI:NI:VA) ...  
PARAMETER GD(FCC_4SL,NI:C,VA) ....
```

where the first parameter is part of the ordered description of the FCC phase and the second is in the disordered part. (In the TDB format the disordered part is a separate phase that is connected by a TYPE_DEFINITION but that is a software dependent feature.)

3.8.6 The expression

The expression will contain coefficients which should be multiplied with temperature and pressure. The actual form of the expression is described in Appendix H, section 13.

Note that it may be important in some cases to distinguish between parameters that are zero or those which has not been assessed. A parameter that is zero could normally be omitted completely but it is recommended that unary and binary parameters should be included even if they are zero. The reason for this is that all parameters possible according to the model that are not included in the database by default will be considered as un-assessed.

If one particularly wants to stress that a parameter is not assessed the function can have the value UNASSESSED.

3.8.7 Bibliographic reference

The <bibliographic reference> should be found inside the body of the bibliographic keyword and will normally refer to the paper where this parameter was assessed or to the person who entered the parameter, hopefully with an explanation. See further in section 3.11.

3.9 Keyword FUNCTION

This keyword is useful because many thermodynamic parameters are related. For example the Gibbs energy of formation of metastable phases is often based on the Gibbs energy of the stable phase which can be entered as a function and used in many parameters.

```
FUNCTION <name> <expression> !
```

The name of a function can be used as symbol name in expressions. In the current TDB file the length is limited to 8 and it must start with a letter. In OC this is now extended to 16 characters and the proposal is to allow 16 characters in the new PDB format.

NOTE that function names must never be abbreviated!

The expression is described in Appendix H, section 13. In the example below the function GHSECR describes the Gibbs energy of BCC-Cr at 1 bar from 298.15 to 6000 K. The function GCRFCC describes the Gibbs energy for metastable FCC-Cr using the same heat

capacity as for BCC-Cr, there is just a ΔH and ΔS added. These functions are later used when describing the properties of Cr in various phases. Note that the PARAMETER keyword have a bibliographic reference but not the FUNCTION. There is no need for a final hash character, “#” after a function name.

```

FUNCTION GHSECR      2.98150E+02  -8856.94+157.48*T-26.908*T*LN(T)
      +.00189435*T**2-1.47721E-06*T**3+139250*T**(-1);  2.18000E+03  Y
      -34869.344+344.18*T-50*T*LN(T)-2.88526E+32*T**(-9);  6.00000E+03  N !
FUNCTION GCRFCC      2.98150E+02  +7284+.163*T+GHSECR;    6.00000E+03  N !

...

PARAMETER G(BCC_A2,CR:VA;0)  2.98150E+02  +GHSECR;    6.00000E+03  N REF283 !

...

PARAMETER G(FCC_A1,CR:VA;0)  2.98150E+02  +GCRFCC;    6.00000E+03  N REF281 !
PARAMETER TC(FCC_A1,CR:VA;0)  2.98150E+02  -1109;    6.00000E+03  N  REF281 !
PARAMETER BMAG(FCC_A1,CR:VA;0)  2.98150E+02  -2.46;    6.00000E+03  N REF281 !

...

PARAMETER G(SIGMA,FE:CR:CR;0)  2.98150E+02  +8*GFEFCC+22*GHSECR+92300
      -95.96*T;    6.00000E+03  N REF107 !

```

It is illegal for a function to refer to itself in the expression part. Circular references are not allowed either.

In order to represent \sqrt{T} , i.e the square root of T , one must use two functions

```

FUNCTION F1      1 0.5*LN(T) 6000 N; !
FUNCTION SQRTT 1 EXP(F1); 6000 N !

```

Then the function SQRTT can be used wherever \sqrt{T} is needed. The low T limit must not be less than 1 as LOG of zero is -Infinity.

Some software may not use functions but it is always possible add up the coefficients with the same T dependence, possibly multiplied with some coefficient, in order to obtain the expression for a parameter.

3.10 Keyword DATABASE_INFORMATION

After this keyword follows a free text (without any !) where the database manager can inform a user about the database. Typically it should inform about the purpose of the database, the name of the database manager(s), the elements in the database, last update, what type of data it contains and preferably an estimate of the range of validity. See also section 3.16.

`DATABASE_INFORMATION <free text without any !> !`

There should be some way to format this text so it is not written as a very long line. Suggestions are welcome.

3.11 Keyword BIBLIOGRAPHY

Most parameters in the database should originate from some assessment and it is vital to give a reference to the paper or report where the parameter was determined. The reference must be given for each parameter as it is not unusual that some parameters from a published assessment are later modified to give a better extrapolation to a higher order system. If each parameter has a bibliographic reference those changed have their reference modified at the same time. It is thus not sufficient to have a single bibliographic reference for a binary system, it must be possible to have individual references for all parameters.

Parameters that are modified manually should have as bibliographic reference the date of the change, the name of the responsible and the reason. Please think of the database manager that will replace you, without such references he or she will be completely lost.

It is important to keep the bibliographic information in the same file as the parameters and the database manager never has time to update two or more files simultaneously.

The syntax of the bibliographic keyword is

`BIBLIOGRAPHY /<bibliographic reference> '<text>'/ !`

The `<bibliographic reference>` is the same as used in the `PARAMETER` keyword. The text should give the publication or where to find the report or simply '2016.09.10 Bo Sundman to make HCP less stable'. The text must be enclosed within single or double quotes because all bibliographic references are defined in the same keyword.

A bibliographic reference may start with a number, for example "91Din" but one should use only letters and digits. It should not be case sensitive as nothing else is.

Maybe it is better to enclose the text by curly braces as a missing ' (or added) may upset reading of the references. But we use curly braces in references to papers?

3.12 Keyword SPECIAL

This keyword is intended for parts of a database that include data which is associated with a particular software and does not fit the current format. The syntax of this is a bit special as it may include keyword sequences that should be ignored by other software.

Thus the `SPECIAL` keyword must be terminated by a double exclamation mark, "!!", two ! following directly after each other. It is thus allowed to have single exclamation marks inside its body.

`SPECIAL <software id> <body of text not including double !> !!`

Of course it would be rather meaningless if the whole database was enclosed by a `SPECIAL` keyword. All cases of `SPECIAL` should eventually be replaced by accepted keywords.

3.13 Keyword VERSION

If this format will be used it will gradually include new features and some may become obsolete or changed. When reading a database it is thus important to know for which version it was created. At the same time one can specify some default values for the high and low T limits for the data.

VERSION <date> <default low T limit> <default high T limit> !

3.14 Keyword CHECKSUM

The checksum was never been implemented in TDB, I think it would be a good idea to have one if anyone can come up with a simple algorithm. The checksum should be calculated for the part between keyword and ! after converting the text to upper case and removing the control characters as specified above.

3.15 Keyword DEFAULTS

At present there is only one use for this keyword:

By default the odd parameters in a Redlich-Kister polynomial are multiplied with the difference of the constituents in alphabetical order. By specifying:

DEFAULT RKORDER=NOTALPHABETIC !

after this change the order of the elements when calculating the difference will be that specified in the constituent array. This will apply to all parameters read after this keyword.

The alphabetical order can be reset by:

DEFAULT RKORDER=ALPHABETIC !

3.16 Keyword ASSESSED_SYSTEMS

The database manager can provide information about the assessed systems in the database. Possibly also some information how to calculate them with his or her preferred software within (). The elements should be given in alphabetical order.

ASSESSED_SYSTEMS AL-FE AL-NI AL-SI FE-MN AL-MN AL-FE-MN !

3.17 Keyword INCLUDE_FILE

This keyword can be used anywhere and means that the file will be opened and read as a database file. Redundant keywords will be ignored but new elements, species and phases will be added. New constituents for an already entered phase will be added provided the models are the same. New functions and parameters will be added.

Question: should existing functions and parameters be overwritten or ignored or give error message?

4 Database manager tasks

A large solution database is created by adding individual assessments of the pure elements, binary, ternary and higher order systems. The assessment procedure is not part of this document, read for example [10]

4.1 The beginning

Normally a database starts from a few binary systems with the same models for the phases and where extrapolations can be easily checked but when it starts to grow by adding new elements and assessments the task to verify the database becomes cumbersome.

4.1.1 Unary data

Until today the unary data has been very easy using the SGTE unary database[5]. As there is considerable efforts today to model the Gibbs energy down to 0 K and eliminate the unphysical breakpoints at the melting T of the elements this will maybe be more difficult in the future.

4.2 Adding an assessment

When adding an an assessment the manager must unify the names of phases and their models with those used in the database.

4.2.1 Phase names

When adding an assessment one should take care that phases that can form a solution have the same name as well as model. The current databases use a mix of phase names which may seem strange to a beginner but which is usually easy to understand with some experience. Thus FCC_A1 (where A1 is the structur bericht) means a phase with a substitutional FCC lattice but which can have interstitials, like C or N, in the octahedral sites, although this should actually have structur bericht B1 which is represented for example by the TiC cubic carbide. But the same lattice with ionic constituents, like MgO, is called HALITE as that is the name of the NaCl structure. However, in the C-O-Ti system there is a complete solubility between the TiC phase called FCC_A1 and the high temperature form of TiO which is called HALITE.

There is also a phase CaF_2 where Ca occupies an FCC lattice and F tetrahedral interstitial sites. This has structurbericht C1 and UO_2 has the same structure but there is no solubility between CaF_2 and UO_2 and they should be given different names.

So the the field is open for anyone to argue about the best name for a phase. What is important is that phases in different systems that have complete solubility with phases in other systems should be treated as the same phase.

4.2.2 Model compatibility

Many inter-metallic phases have slightly different models in different systems, for example the ordered B1, the μ phase and the σ phases. As it is important not to have several μ phases it can be rather challenging to select model and modify the parameters in those systems that have another model.

This will probably become more and more challenging as new assessments using ordered descriptions of FCC and BCC phases are published.

4.2.3 Simple checks

This may require adding parameters for un-assessed binaries and for phases which dissolve all elements in the new assessment but is not stable inside the assessed system to avoid that this phase becomes stable where it should not be, for example adding a positive interaction to HCP in an assessment of the Fe-Cu system, where the HCP is not stable.

For binary systems the Gibbs energy curves at various T can reveal strange things.

4.2.4 More subtle checks

The story of the interaction parameter for FCC in the Cr-Mo system.

A useful method to find strange parameters is to calculate binary or ternary diagrams suspending all but one or a few phases.

Calculating phase diagrams at several temperatures extrapolating the new assessed systems with the other elements in the database, one at a time may reveal strange things.

4.3 Estimated range of validity

There is no way a large solution database can be maintained without a skilled manager. This manager should of course have many computational tools as help but it is his or her skills that determine how easy and accurate the database is when used by industry or for academic research.

One task of the manager is to estimate limits for the composition for which he or she can believe the database will give reasonable results. Feedback from users is very important for this as it is impossible to check things without experimental data.

5 Some future features

This section is intended for collection of ideas that is not part of the formal definition but may guide future extensions.

5.1 Databases generated from repositories

In the future there may exist large general databases with materials data from which data can be extracted for a particular system. In most cases the extracted data will not be model parameters but original experimental data or data generated from DFT calculations. Such

data can be used for testing the database and also for assessments and maybe a simple way to define such informations should be included here.

The experimental data could be described like POP files used in Thermo-Calc or the “enter many_equilibria” command available in OC. This means to set the conditions for which the experimental data has been generated and then recalculate the same information using the software and available database. Possibly the model parameters for the phases could be adjusted to obtain a better fit.

Examples how one could describe conditions and experimental data for some different kinds of equilibria:

- An experimental tie-line in a binary two-phase region. After the enter command a name is given and the Y means the following commands refer to the new equilibrium.

```
ENTER EQUILIBRIUM AB1 Y
SET CONDITION T=1273 P=1E5
SET STATUS PHASE *=SUS
SET STATUS PHASE FCC BCC=FIX 0
EXPERIMENT X(FCC,A)=0.2:.01 x(BCC,A)=0.3:0.01
```

- The experimental value of the congruent melting of UO_2 . The condition $x(\text{liq},\text{O}) - x(\text{C1_mo2})=0$ means both phases should have the same mole fraction of O.

```
ENTER EQUILIBRIUM CONGRUENT Y
SET CONDITION P=1E5 X(LIQ,0)-X(C1_M02,0)=0
SET STATUS PHASE *=SUS
SET STATUS PHASE LIQ C1_M02=FIX 0
EXPERIMENT T=3340:25, X(O)=.66:.01
```

- Tables with experimental data like enthalpies of mixing can be entered like below for a binary liquid. The “enter table” command has first a “head” part where the conditions and phase status are specified (all phases suspended by default) in a simplified way (no SET needed). The values that vary in the table are indicated with a symbol @j which means the value is in column j. Note the leftmost column is a “name” of the equilibrium which is considered as column zero!

```
ENTER TABLE_OF_EQUIL
ENTERED 1 LIQUID
CONDITION P=1E5 N=1 T=1100 X(B)=@1
REFERENCE_STATE A LIQ * 1E5
REFERENCE_STATE B LIQ * 1E5
EXPERIMENT HM(LIQ)=@2:5%
TABLE_START
LIQH1 0.1 -3700
LIQH2 0.2 -6000
```

```
LIQH3 0.5 -7800
LIQH4 0.8 -5000
LIQH5 0.9 -2800
TABLE_END
```

As liquid is the only entered phase it is not really necessary to specify it in the experiment.

Extending to these kinds of data we must also define how to specify various kinds of state variables.

5.2 Material

In some cases data is characterized for a material which may consist of several phases. In order to accommodate such data, particularly for thermo-physical use, one may use the keyword MATERIAL. The data for this keyword is

```
MATERIAL <name> <composition> '<comments>' !
```

The name is 1 to 24 alphanumerical characters or underscore, it must start with a letter. The composition is specified in mass percent of the components, normally the elements but species may be used. The chemical symbol of the component is followed by the mass percent, the major component is specified by giving the value as *. The comments may be any text.

Examples

```
MATERIAL PIG_IRON C 4 SI 2.5 MN 0.4 FE * !
MATERIAL LIMESTONE CAC03 3 CAC103 * 'Almost pure' !
```

Such materials definitions may be helpful for specifying the limits of accuracy for the database.

References

- [1] Reasons to model the Gibbs energy.
- [2] M. Hillert, CALPHAD, **4** (1980) 1–12
- [3] M. Hillert, Met Trans A **17A** (1986) 1878–1879
- [4] A.D. Pelton and C.W. Bale, Met Trans A **17A** (1986) 1211–1215
- [5] A.T. Dinsdale, CALPHAD, **15** (1991) 317–425
- [6] J.-O. Andersson et al. CALPHAD, **26** (2002) 273–312
- [7] C.W. Bale et al. CALPHAD, **26** (2002) 189–228

- [8] B. Sundman et al. *Intr Mat and Manu Innov* **4:1** (2015)
- [9] A.D. Pelton, *CALPHAD*, **25** (2001) 319–328
- [10] H.L. Lukas et al. *Computational Thermodynamics, the CALPHAD method*, Camb. Univ. Press (2007)
- [11] Wei Xiong, Thesis on magnetic model in Cr-Fe paper

6 Appendix A: Example of an interchange

On the following pages an example of the use of the interchange format is given in order to describe a database for ??.

new example needed

to be generated ...

7 Appendix B. Summary of keywords

VERSION <date> <lower T limit> <upper T limit>!

ELEMENT <element name> <SER state> <atomic mass in g/mol>
<enthalpy at 298.15 K - enthalpy at 0 K>
<entropy at 298.15 K> !

SPECIES <species name> /<element name> <stoichiometric factor>/ !

TABLE_OF_MODELS /<code> '<explanation>' /!

PHASE <phase name> <model code> /<addition code>/ /<detail code>/ !

CONSTITUENT <phase name> : <species on first sublattice>
/ : <species on other sublattices in sublattice order>/ : !

TABLE_OF_IDENTIFIERS /<identifier> '<explanation>' /!

PARAMETER <identifier> (<phase name>, <component array> ; <degree>)
<expression> <bibliographic id> !

FUNCTION <name> <expression> !

DATABASE_INFORMATION free text !

BIBLIOGRAPHY /<bibliographic identifier> '<text>' / !

SPECIAL <software> <any text and keywords including single !> !!

DEFAULTS /option=value/ !

ASSESSED_SYSTEMS assessed systems with a hyphen between the elements in
alphabetical order. !

INCLUDE_FILE filename !

CHECKSUM <value> !

8 Appendix C. Definition of model codes

All model codes in the version (October 11, 2016) of the PDB format:

Model code	Meaning
IDEAL	no sublattices and no excess parameters
SUBRKM	Substitutional model with Redlich-Kister Mugianu (RKM) excess model
SUBPK	Substitutional model with polynom model and Kohler extrapolation
SUBMIX	Substitutional model with mixed excess model
CEFj	Sublattice model with j sublattices, $1 \leq j \leq 9$, RKM excess model
CEFjD	as CEFj but also a disordered substitutional model added, RKM excess
CEF4F	A 4 sublattice model for ordered fcc or hcp with permuted parameters only once, RKM excess
CEF4B	A 4 sublattice model for ordered bcc with permuted parameters only once, RKM excess
CEF5F	A 4 sublattice model for ordered fcc or hcp with extra interstitial sublattice and permuted parameters only once, RKM excess
CEF5B	A 4 sublattice model for ordered bcc with extra interstitial sublattice and permuted parameters only once, RKM excess
CEF5FD	A 4 sublattice model for ordered fcc or hcp with extra interstitial sublattice and permuted parameters only once and a disordered substitutional model for the ordered sublattices, RKM excess
CEF5BD	A 4 sublattice model for ordered bcc with extra interstitial sublattice and permuted parameters only once and a disordered substitutional model for the ordered sublattices, RKM excess
I2SL	The 2 sublattice ionic liquid model.

The Gibbs energy expression for these models are found below. The symbols used in the Gibbs energy expressions are:

Symbol	Explanation
R	gas constant
T	absolute temperature
G_M^α	is the Gibbs energy of α phase for one formula unit. The formula unit is important because phases can have vacancies as constituent and thus the total number of moles of atoms may vary. In fact the normal model for an ideal gas is per mole formula units, not per mole of atoms.
$^\circ G_i^\alpha$	is the Gibbs energy of formation of species i in the α phase from the reference states of the elements.
$^\circ G_I^\alpha$	is the Gibbs energy of formation of the endmember I in the α phase from the reference states of the elements.
y_i or y_{is}	the constituent fraction of i (on sublattice s).

- model code IDEAL

The constituents can be any kind of species, including vacancies and ions.

$$G_M = \sum_i y_i {}^\circ G_i + RT y_i \ln(y_i) \quad (6)$$

- model code SUBRKM

The constituents can be any kind of species, including vacancies and ions.

The substitutional model using Redlich-Kister Muggianu extrapolations.

$$G_M = \sum_i y_i {}^\circ G_i + RT y_i \ln(y_i) + {}^E G_M \quad (7)$$

$${}^E G_M = \sum_i \sum_{j>i} y_i y_j (L_{ij} + \sum_k y_k (L_{ijk} + \dots)) \quad (8)$$

The binary interaction parameters are described with Redlich-Kister polynom.

$$L_{ij} = \sum_\nu (y_i - y_j)^\nu {}^\nu L_{ij} \quad (9)$$

The ternary parameters can be composition dependent according to Hillert's proposal[2]:

$$L_{ijk} = \sum_\nu z_i {}^\nu L_{ijk} + z_j {}^\nu L_{ijk} + z_k {}^\nu L_{ijk} \quad (10)$$

$$\begin{aligned} z_i &= y_i + (1 - y_i - y_j - y_k)/3 \\ z_j &= y_j + (1 - y_i - y_j - y_k)/3 \\ z_k &= y_k + (1 - y_i - y_j - y_k)/3 \end{aligned} \quad (11)$$

The use of z_i instead of y_i makes this formulation symmetric also in multicomponent systems.

- model code SUBPK

The constituents can be any kind of species, including vacancies and ions.

The substitutional model polymer for binary excess with Kohler extrapolations.

$$G_M = \sum_i y_i {}^\circ G_i + RT y_i \ln(y_i) + {}^E G_M \quad (12)$$

$${}^E G_M = \sum_i \sum_{j>i} z_i z_j (L_{ij} + \sum_k z_k (L_{ijk} + \dots)) \quad (13)$$

where z_i, z_j are extrapolated from ternary to "binary" compositions using a Kohler model.

$$L_{ij} = ?? \quad (14)$$

I am not sure how to specify the independent binary variable and what fractions should be used for any ternary term

- model code SUBMIX

The constituents can be any kind of species, including vacancies and ions.

The substitutional model using mixed ternary extrapolations (Toop, Koohler, Redlich-Kister etc). Using the elegant method proposed by Pelton[9] one can define for each

ternary subsystem which extrapolation method should be used. However this means one must for each ternary combination of constituents in a phase give which extrapolation model that should be used. I am not sure how this can be done most simply.

$$G_M = \sum_i y_i {}^\circ G_i + RT y_i \ln(y_i) + {}^E G_M \quad (15)$$

$${}^E G_M = \sum_i \sum_{j>i} z_i z_j (L_{ij} + \sum_k z_k (L_{ijk} + \dots)) \quad (16)$$

where z_i, z_j are extrapolated from ternary to “binary” compositions using a Kohler model.

I am not sure what fraction should be used for the ternary term

For a quaternary system A-B-C-D where A is Toop element in the A-B-C system and B a Toop element in B-C-D and for the rest one use Kohler extrapolations I could imagine this specified as:

PHASE BCC SUBMIX(TK(A B C) TK(B C D) KK()) !*

where TK(A B C) means A is the Toop element and for B and C use Kohler. KK() means use Kohler model for all other ternaries. But I am not sure if this is the sufficient or the simplest way.*

I am not sure if anyone used Toop or Kohler combined with sublattices and how one should handle that.

- model code I2SL (Partially Ionic 2 sublattice Liquid model)

The constituents must be cations on the first sublattice and there must not be any cations on the second sublattice.

This is a 2 sublattice model with cations on the first sublattice and anions, vacancies and neutrals on the second sublattice. The site ratios, P and Q , depend on the average charge on the opposite sublattice.

$$G_M = \sum_i \sum_j y_i y_j {}^\circ G_{ij} + Q(y_{Va} \sum_i y_i {}^\circ G_i + \sum_k y_k {}^\circ G_k) + RTP \sum_i y_i \ln(y_i) + RTQ(\sum_j y_j \ln(y_j) + y_{Va} \ln(y_{Va}) + \sum_k y_k \ln(y_k)) + {}^E G_M \quad (17)$$

$$P = \sum_j \nu_j y_j + Q y_{Va} \quad (18)$$

$$Q = \sum_i \nu_i y_i \quad (19)$$

$${}^E G_M = \sum_{i_1} \sum_{i_2} \sum_j y_{i_1} y_{i_2} y_j L_{i_1, i_2, j} + \sum_i \sum_{j_1} \sum_{j_2} y_i y_{j_1} y_{j_2} L_{i, j_1, j_2} + \sum_i \sum_j y_i y_j y_{Va} L_{i, j, Va} + \sum_i \sum_j \sum_k y_i y_j y_k L_{i, j, k} +$$

$$\begin{aligned}
& Q \sum_{i_1} \sum_{i_2} y_{i_1} y_{i_2} y_{Va}^2 L_{i,i_2:Va} + Q \sum_i \sum_k y_i y_k y_{Va} L_{i:k,Va} + \\
& Q \sum_{k_1} \sum_{k_2} y_{k_1} y_{k_2} L_{k_1,k_2}
\end{aligned} \tag{20}$$

where i denote a cation with charge $+\nu_i$, j an anion with charge $-\nu_j$, Va vacancies and k neutrals.

For a discussion of the excess term and other details see Lukas [10]. Note eq. 20 is wrong in Lukas [10] as the factors Q have are missing.

Eq. 20 has been corrcted.

- model code KFGL

The constituents are “cells” like A-O-A and A-O-B and a special ordering of the constituents is needed.

The cell model proposed by Kaphor-Frohberg and extended by Guy and Lehmann

$$G_M = \text{complicated} \tag{21}$$

- model code GUTS

The constituents can be any kind of species, including vacancies and ions??

Used by FactSage

$$G_M = \text{I do not know} \tag{22}$$

- model code CEFj

The constituents can be any kind of species, including vacancies and ions.

This is the normal Compound Energy Formalism (CEF) and the number of sublattices is given as j , $1 \leq j \leq 9$. The value $j=1$ can be used when the formula unit should have more than one site. The site ratios are specified within parenthesis like

PHASE SIGMA CEF5(2.0 4.0 8.0 8.0 8.0) !

$$G_M = \sum_I \Pi(y_{i \in I}) {}^\circ G_I + RT \sum_s a_s \sum_i y_{is} \ln(y_{is}) + {}^E G_M \tag{23}$$

where $\Pi(y_{i \in I})$ is the product of the constituent fractions of the endmember I , ${}^\circ G_I$ is the Gibbs energy of formation of the endmember (compound) I from the reference state of the elements. a_s are the site ratio for sublattice s .

For details especially about the excess model see Lukas [10].

As far as I know there has never been any assessment using CEF with any other excess models than Redlich-Kister-Muggianu.

- model code CEFjD

The constituents can be any kind of species, including vacancies and ions.

This adds a disordered substitutional part to a phase modeled with j sublattices, $1 \leq j \leq 9$, for example σ or Laves_C14. The configurational entropy of the substitutional part is ignored.

$$G_M = G_M^{\text{ord}}(y) + G_M^{\text{dis}}(x) - RT \sum_i x_i \ln(x_i) \quad (24)$$

Example:

PHASE SIGMA CEF5D(2.0 4.0 8.0 8.0 8.0) !

- model code CEF4F and CEF4B

The constituents can be any kind of species, including vacancies and ions. The set of constituents must be identical on all sublattices.

This is a special case of CEF describing ordering on FCC, HCP or BCC using 4 sublattices (tetrahedron). The number of sites on all 4 sublattices are equal. There can be several ordered forms with different fractions of constituents on the sublattices. There is also a totally disordered state with same fractions on all sublattices. The latter is equivalent to a normal disorderd FCC, HCP or BCC.

PHASE FCC_4SL CEF4F(0.25 0.25 0.25 0.25) !

The structure imposes several restrictions on the model parameters.

$$G_M = \sum_I \Pi(y_{i \in I}) \circ G_I + RT \sum_s a_s \sum_i y_{is} \ln(y_{is}) + {}^E G_M \quad (25)$$

The constituents can be any kind of species, including vacancies and ions.

For model codes CEF4F, CEF4B, CEF5F and CEF5B the model parameters have several permutations, for example

$${}^\circ G_{A:A:A:B} = {}^\circ G_{A:A:B:A} = {}^\circ G_{A:B:A:A} = {}^\circ G_{B:A:A:A} \quad (26)$$

and such parameters need only be given once in the database, the software will take care of the permutations.

NOTE permutations applies also to interaction parameters! So it can be quite complicated to implement.

- model code CEF5F and CEF5B

The constituents can be any kind of species, including vacancies and ions. The set of constituents must be identical on the first 4 sublattices.

This is the same as CEF4F and CEF4B but has an extra interstitial sublattice.

PHASE BCC_4SL CEF5B(0.25 0.25 0.25 0.25 3) !

$$G_M = \sum_I \Pi(y_{i \in I}) \circ G_I + RT \sum_s a_s \sum_i y_{is} \ln(y_{is}) + {}^E G_M \quad (27)$$

- model code CEF4FD, CEF4BD, CEF5FD and CEF5BD

The constituents can be any kind of species, including vacancies and ions. The set of constituents must be identical on the first 4 sublattices.

The final D means that the model parameters describing the disordered state has been extracted to a single sublattice model. The parameters for the ordered 4 sublattice model describe only the contribution due to ordering.

$$G_M = G_M^{\text{dis}}(x) + \Delta G_M^{\text{ord}}(y) \quad (28)$$

$$\Delta G_M^{\text{ord}}(y) = G_M^{\text{4sl}}(y) - G_M^{\text{4sl}}(y = x) \quad (29)$$

where the mole fractions x are calculated from the constitution, y . The last equation above means that the Gibbs energy for the ordered part is calculated twice, once with the original constitution, once with the constitution equal to the disordered state. When the phase is disordered the contribution from ΔG_M^{ord} is thus zero.

In OC I use a D after the normal parameter identifier to specify that the parameter belong to the disordered set, for example TCD(FCC,FE) is the magnetic ordering T for disordered FCC-Fe. Parameters for the disordered set could also be detected by the number of sublattices in the constituent array of the parameter but I think it is useful to have a final D also for the disordered parameters as that may avoid a lot of errors.

- model code BONDj

The constituents can be any kind of species, including vacancies and ions.

This is a special variant of CEF with j sublattices, $2 \leq j \leq 9$, using bond energies between constituents on pairs of sublattices.

PHASE FCC BOND4(1 1 1 1) !

$$G_M = \sum_i x_i {}^\circ G_i + \sum_{s=1}^n \sum_{r=2}^{n-1} (n-1) y_{is} y_{jr} G_{ij} + RT \sum_s \sum_i y_{is} \ln(y_{is}) \quad (30)$$

where ${}^\circ G_{ij}$ is a Gibbs energy for the bond between constituent i in sublattice s and constituent j in sublattice r . Such a parameter will have a wildcard, “*”, for the constituent in all other sublattices, like G(BOND,*:FE:NI:*) to give the bond energy between Fe in sublattice 2 and Ni in sublattice 3.

9 Appendix D. Definition of addition codes

These define models of additions which at present mainly is for the ferromagnetic transition but may be expanded to include many new things like heat capacity models at low T , volumes, strain-stress relations etc. All of these may include parameters with specific identifiers that depend on T , P and constitution of the phase.

- Addition code IMAGB and IMAGF

Inden magnetic model for BCC and FCC. The expression for this can be found in almost any Calphad paper or in Lukas[10].

The value of the magnetic contribution will depend on parameters with identifiers TC and BMAG.

- Addition code IWMAGB and IWMAGF

Inden magnetic model as modified by Qing-Wei with separate Curie and Neel temperatures and individual Bohr magneton number.[11]

The value of the magnetic contribution will depend on parameters with identifiers CTA, NTA and IBM&A.

- Addition DEBYE and EINSTEIN

Whatever is proposed for extrapolating C_P and G down to 0 K.

The value of the Debye/Einstein T contribution will depend on parameters with identifier THET

- Addition MURN

The Murnaghan model for pressure dependence.

There are no identifiers for this model as yet.

10 Appendix E. Definition of detail codes

These are for things related to the phase but maybe more informative or software specific and can possibly be ignored.

- Detail code STRUCT

After this and within parenthesis one can give free text like:

```
PHASE HALITE CEF2(1.0 1.0) IMAGF DETAIL(STRUCT(B2, PERICLASE)) !
```

- Detail code MGAP

specifies there is a miscibility gap in a phase. It also gives the major constituents in all sublattices and possibly a pre and suffix to the phase name.

```
DETAIL(prefix : /major constituent in sublattice :/ suffix)
```

```
PHASE FCC_A1 CEF2( 1.0 1.0) MAGIF MGAP(MC :TI V:C N : CARB) !
```

Here the phase has a CEF model with 2 sublattices. It has the Inden magnetic addition and there is a detail that there can be a miscibility gap with major constituents Ti and V in the first sublattice and C and N in the second. The text CC (meaning Cubic Carbide) is the prefix and CARB is a suffix that could be added to the phase name for this composition set. If no pre- or suffix needed just start with a : and give nothing after the last :.

Note that a phase can have miscibility gaps even if there is no MGAP keyword! Of course miscibility gaps are conveniently detected by global minimization but sometimes one may want to know which composition set has a specific major constitution.

There are also cases when a calculation cannot rely on global minimization, for example if there is no condition on T .

- Detail code DCON

This can be used to specify default constituents (used for the first composition set). For example

```
PHASE FCC_A1 CEF2( 1.0 1.0) MAGIF MGAP(CC :Ti V:C N:) DCON(FE NI:VA) !
```

This may be used instead of the % used for major constituents in the current TDB format, see section 3.6.1.

11 Appendix F. Parameter identifiers

A parameter identifier should not be longer than 4 letters (excluding possible element or constituent specification). In the table below a parameter identifier may depend on T or P and that is indicated by a *. If it is specific for an element that is marked by an &X and if specific for a constituent &X#s.

Identifier	Spec	Unit	T	P	Description
G	-	J/FU	*	*	The Gibbs energy per formula unit
L	-	J/FU	*	*	Identical to G but mainly used for interactions
TC	-	K	-	*	The Critical T for magnetic ordering
BMAG	-	B.mag	-	*	The average Bohr magneton number
CTA	-	K	-	*	The Curie T for ferro-magnetic ordering
NTA	-	K	-	*	The Neel T for anti-ferromagnetic ordering
IBM	&A#s	B.mag	-	*	The Bohr magneton for constituent A in sublattice s
THET	-	K	-	*	Debye or Einstein T
MQ	&X	?	*	*	LN mobility of X
RHO	-	?	*	*	Electric resistivity
MAGS	-	?	*	*	Magnetic susceptibility
GTT	-	K	-	*	Glass transition temp
VISC	-	?	*	*	Viscosity
LPX	-	m	*	*	Lattice parameter in X direction
LPY	-	m	*	*	Lattice parameter in Y direction
LPZ	-	m	*	*	Lattice parameter in Z direction
LAA	-	degree	*	*	Lattice angle α
LAB	-	degree	*	*	Lattice angle β
LAG	-	degree	*	*	Lattice angle γ
EC11	-	?	*	*	Elastic constant C11
EC12	-	?	*	*	Elastic constant C12
EC44	-	?	*	*	Elastic constant C44
V0	-	m3	-	-	Molar volume at 298.15 K and 1 bar
VA	-	m/K	*	-	Thermal expansion at 1 bar
VB	-	m2/N	*	*	Bulk modulus

A parameter can have a suffix D if it belongs to a disordered part. Such a parameter will also have fewer sublattices in the constituent array.

The parameter identifiers are used to describe how a property varies with composition and possibly T and P . Most properties are for the whole phase but some, like the mobility, describe how an element or constituent specific property varies with T , P and composition of the phase.

Many parameters may not influence the Gibbs energy but are used for simulations and it is convenient to store them together with the thermodynamic data, as long as they are phase specific. Some properties may overlap, like the volume and lattice parameter. But hopefully things like this will be sorted out eventually.

An element specific property like $\ln(\text{mobility})$ of Fe in a phase is given with the identifier MQ&FE and this can vary with the composition and T and P . For example $\ln(\text{mobility})$ of

Fe in the BCC phase in the Cr-Fe system can have the following parameters:

MQ&FE(BCC,FE) ln(mobility) of Fe in pure BCC-Fe
MQ&FE(BCC,CR) ln(mobility) of an Fe atom in pure BCC-Cr
MQ&FE(BCC,CR,FE) describe composition dependence

These parameters can depend on T and P . The mobility is calculated as:

$$MQ\&Fe(BCC) = x_{Fe}MQ\&Fe(BCC, FE) + x_{Cr}MQ\&Fe(BCC, CR) + x_{Fe}x_{Cr}MQ\&Fe(BCC, CR, FE) \quad (31)$$

where MQ&FE(BCC) is the value of ln(mobility) of FE in BCC for the current values of T , P and composition.

The Bohr magneton number of Fe^{+3} in sublattice 3 of magnetite as a function of composition is described by the constituent specific property IBM&Fe+3#3.

We are creating a database with a large number properties where we need model parameters when the phase is not stable!

12 Appendix G. The constituent array for interaction parameters

I have not thought much about these

Sign depend on alphabetical order of constituents for odd Redlich-Kister terms

Toop model

Polynomial models for constituents ...

13 Appendix H, The function of T and P

I think there is no need for an EXPTYPE keyword as we probably do not need any other type than the old 5. Dealing with solution phases it seems not very meaningful to have C_P expressions and with fixed T coefficients they can always be integrated to a Gibbs energy.

The rather lengthy description below is due to the fact that in 1990 all software except Thermo-Calc stored the coefficients for these expressions in tables, usually with a fixed number of coefficients.

In the TDB files the keyword EXPTYPE could be used to specify the way functions of T and P were entered. But only exptype 5 are used in current TDB files so this keyword is no longer needed.

As some software calculate first and second derivatives of T and P using these functions there should be some restrictions on the complexity of the expressions that are allowed. The proposed syntax is that of the old exptype 5:

`<low temperature limit> / <function> ; <high temperature limit> <Y or N> /`

The syntax of the function part is given below. A semicolon must be use to terminate the function and it is followed by a space and the high temperature limit. Finally a “Y” or an “N” must be given. N means that there are no more temperature ranges and Y that there will be another function above the current high temperature limit. The information between the slashes is repeated until it is terminated by an N.

In the TDB format a “,” (space followed by a comma) was allowed for the low and hight T limit. The , meant that the default value was taken, for the low limit 298.15 and for the high limit 6000 K. There is no reason not to allow this also in the new format as it simplifies a little for the manager and the default low and high limits can be set with the VERSION keyword.

The function is written similar to a Fortran like statement but without the use of parenthesis for grouping terms. The basic entity of the function is called a “simple term”. A simple term is:

`<real number> * <symbol name> **<power> *T** <power> *P** <power>`

The text between `<` and `>` describe the item which should appear there. The other items must be given literally and have their usual meaning i.e. `*` is multiplication and `**` is exponentiation, `T` is the temperature and `P` is the pressure. A symbol can be void or another expression, see the FUNCTION keyword below.

If the simple term does not depend on any symbol or T or P that part can be omitted. The power can only be an integer but non-integer powers can be handled with a LOG and EXP pair as described below. Negative powers must be surrounded by parenthesis. Redundant parts of a simple term can be omitted and if the power is unity the exponentiation can be omitted altogether. Note that the real number must be the first item and any symbol must precede the T and P . Examples of simple terms are:

1.15*T -V1 1E-12*P**2 -456754.65*T**(-1)*P 10*R*T

In order to include the logarithm and exponential in these functions it is allowed to multiply a simple term with the logarithm or exponential of another simple term. This more generalized entity is called a term and it is defined

```
<term> = <simple term> or
<simple term> * <symbol> or
<simple term> * LN( <simple term> ) or
<simple term> * LOG( <simple term> ) or
<simple term> * EXP( <simple term> ) or
<simple term> * ERF( <simple term> )
```

A term can be equal to one of the lines above. It is illegal to have both exponentials and logarithms in the same term. It would not be difficult to add mode unary functions.

Examples of terms are:

```
+1.15*T*LN(T)          +1E-6*LOG(-32000*T**(-1))    -5*V3*T*EXP(V4*P)
```

A function of this type is thus a number of terms written after each other in order to form an expression. All terms except the first must be preceded by a sign. Examples of expressions are:

```
-10000 +10*T+1.15*T*LN(T) +7.5E-5*P +134567*T**(-1) -1.13E-12*T*P
```

```
F1+2.5*R*T*LN(T)+R*T*LN(P)
```

where F1 is another FUNCTION. Note that there must be a semicolon to separate the function from the following high temperature limit. Note also that the symbols can denote a numeric value or another function. See the FUNCTION keyword.

Some of the restrictions on a function are:

- The order of the factors in a simple term must be followed,
- no parenthesis allowed except for exponentials, logarithms or negative powers,
- no division allowed,
- no spaces allowed between a sign and a numeric value,
- only one symbol in each simple term,
- only integers as powers.

Some of these restrictions are due to simplify the parsing of the functions and some are due to the requirement that it must be possible to calculate quickly the value of the function as well as its first and second derivatives with respect to T and P .