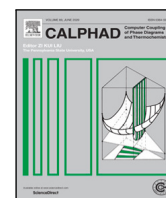




Contents lists available at ScienceDirect

Calphad

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

XTDB, an XML based format for Calphad databases

Bo Sundman^a, Fabio Miani^b, Axel van de Walle^c, Bengt Hallstedt^d, Ursula R. Kattner^e, Florian Tang^f, Taichi Abe^g, Reza Naraghi^h, Erwin Povoden-Karadenizⁱ, Aurelie Jacobⁱ, Shuanglin Chen^j, Richard Otis^k, Kazuhisa Shobu^l, Malin Selleby^m, Alexander Pischⁿ

^a OpenCalphad, Gif sur Yvette, France

^b DPIA, University of Udine, Udine, Italy

^c School of Engineering, Brown University, Providence, RI, USA

^d IWM, RWTH Aachen University, Aachen, Germany

^e MSED National Institute of Standards and Technology, Gaithersburg, MD, USA

^f GTT-Technologies, Herzogenrath, Germany

^g NIMS, Sengen, Tsukuba, Japan

^h Thermo-Calc Software AB, Solna, Sweden

ⁱ TU Wien, Vienna, Austria

^j CompuTherm, Middleton, WI, USA

^k California Institute of Technology, Pasadena, CA, USA

^l RICT, Inc., Saga, Japan

^m MSE, KTH Royal Institute of Technology, Stockholm, Sweden

ⁿ SIMAP, Saint-Martin d'Hères, France

ARTICLE INFO

Keywords:

Computational thermodynamics

Calphad

Models

Model parameters

Databases

XML

ABSTRACT

The calculation of phase diagram (Calphad) method uses models that depend on assessed parameters to describe the thermodynamic properties of materials. These model parameters are assessed by researchers and students using experimental and theoretical data on binary and ternary systems that can be merged to multicomponent databases and used to calculate properties and simulate processes for a wide range of materials.

There are several different software using the Calphad method for calculations and they may use slightly different models and database formats. This paper will provide a short background on the current state of database development and proposes a new format based on the eXtensive Markup Language (XML) as a unified database format. This change is particularly important as several new models for the pure elements are currently being introduced in the Calphad databases.

1. Introduction

The development of software to calculate thermodynamic properties of alloys and other materials has been centered around the calculation of phase diagram (Calphad) organization formed in 1973 [1] and the use of the lattice stabilities for the pure elements proposed by Kaufman [2]. These lattice stabilities made it possible to share and combine the assessments of binary and ternary systems made by researchers all over the world.

The unary database [3] published by the Scientific Group Thermodata Europe (SGTE) in 1991 represented a major improvement as this database included heat capacity data and a separate magnetic model [4,5] using the Curie temperature and the magnetic moment.

This unary database transformed Calphad to a general tool for computational thermodynamics and together with kinetic models and data, it could be applied to all kinds of simulations of phase transformations and processes. The SGTE database proposed a simple database format, called TDB (short for Thermodynamic DataBase), which with slight variations has been adopted by several software developers. During the last 45 years several thousand binary and ternary systems have been assessed based on experimental and theoretical data using the Calphad models.

These assessments have been integrated in several large databases by several different software providers often using TDB formats with slight but significant differences. With such databases the software can perform reliable calculations for multicomponent alloys, nuclear

* Corresponding author.

E-mail address: bo.sundman@gmail.com (B. Sundman).

<https://doi.org/10.1016/j.calphad.2025.102849>

Received 30 March 2025; Received in revised form 31 May 2025; Accepted 18 June 2025

Available online 23 July 2025

0364-5916/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

materials and many other systems, in order to simplify the development of new materials and better processes [6–12]. But even small variations of the TDB format will create problems for the future development, in particular the introduction of more complex models in the new unary database [13], the integration of data from Density Functional Theory (DFT) calculations and adding more materials properties, for example mobilities. The authors of this paper have discussed two very different ways forward: one oriented towards an ontology based schema providing facilities for internal verification of the database [14–16], the other one aims to integrate the simple, compact and human readable notation of the TDB files in a format providing structured tools for extending the database to new properties and models as discussed in [17,18]. This paper follows the second proposal and a way to reconcile the two approaches is discussed in Section 5. The eXtended Markup Language (XML) is proposed as markup language for extending the TDB format (XTDB).

2. The Gibbs energy and the organization of this paper

XML elements are the basic building blocks of an XML document and consist of a start tag (called tag name below), an end tag, and predefined attributes with data in between. These tags define the structure and meaning of the data while attributes of a tag describe the properties of the XML element. For those not familiar with XML an introduction can be found at [19]. The conventions used in this text for the XTDB format are that a tag name is written in **bold** and an attribute is written in *italics* and a short summary of the XTDB format is given in [Appendix A](#). As supplementary material there is a complete database for the Al-C system using new unary models and other examples. There is also a complete **ModelDescriptions** tag with a number of physical models together with proposed composition dependent Model Parameter Identifiers (MPID).

An XTDB database can be written and used by different software packages and the **ModelDescriptions** tag specifies the models that are used in a particular database. This tag is provided by the software generating the database and if another software reads the database it can check if a model extracted from the XTDB file is actually implemented.

2.1. The Gibbs energy expression

Models for the Gibbs energy of materials are important to predict the spontaneity of a process or transformation because a system always tries to reach its equilibrium state and the equilibrium is at a minimum of the Gibbs energy at constant T , P and overall composition, N_i :

$$G = G(T, P, N_i) \quad (1)$$

where T is the absolute temperature, P the pressure and N_i the number of moles of component i .

An essential feature of the Calphad method is that each phase in a system is modeled independently and its Gibbs energy depends on T , P and the fractions of the constituents of the phase. Depending on the software using the database, a user may calculate the set of stable phases and other properties of a system for a very wide range of external conditions.

2.2. The external variables

Sometimes one is more interested in the activity of a component or the heat that is added or removed. All of these properties are available in the Calphad modeling as partial derivatives of the Gibbs energy:

$$dG = \left(\frac{\partial G}{\partial T}\right)_{P, N_i} dT + \left(\frac{\partial G}{\partial P}\right)_{T, N_i} dP + \sum_i \left(\frac{\partial G}{\partial N_i}\right)_{T, P, N_{j \neq i}} dN_i \quad (2)$$

where the variables that must be kept constant for each partial derivative are indicated. With simple mathematical transformations, see for example [20], it is possible to use Eq. (2) to calculate how the local

equilibria in a system respond to changes in the external conditions of a process. In particular the isobaric heat capacity, C_p , and the chemical potential, μ_i , of each component i :

$$C_p = -T \left(\frac{\partial^2 G}{\partial T^2}\right)_{P, N_i} \quad (3)$$

$$\mu_i = \left(\frac{\partial G}{\partial N_i}\right)_{T, P, N_{j \neq i}} \quad (4)$$

are essential for equilibrium calculations and kinetic simulations.

2.3. The internal variables

In Calphad the Gibbs energy of a system uses Gibbs energy models for each phase, α , with different sets of constituents, which can be elements or chemical species or ions, and each with its own set of model parameters:

$$G = \sum_{\alpha} \aleph^{\alpha} G_M^{\alpha}(T, P, y_{s,i}^{\alpha}) \quad (5)$$

$$G_M^{\alpha} = \sum_I \prod_{I_s} (y_{s,i}^{\alpha})^{\circ} G_I^{\alpha}(T, P) - T \text{ c}^{\text{fg}} S_M^{\alpha}(y_{s,i}^{\alpha}) + {}^E G_M^{\alpha}(T, P, y_{s,i}^{\alpha}) + {}^{\text{phy}} G_M^{\alpha}(T, P, y_{s,i}^{\alpha}) \quad (6)$$

where \aleph^{α} is the number of moles formula unit of α and G_M^{α} is the Gibbs energy of α per mole formula unit as a function of T , P and $y_{s,i}^{\alpha}$ which denote the fraction of constituent i on sublattice s . The number of constituents in a phase can be much larger than the number of components of the system. In the term $\prod_{I_s} (y_{s,i}^{\alpha})^{\circ} G_I^{\alpha}$ the I is known as an endmember of α and specifies a constituent i in each sublattice s of the phase. The endmember I can be considered as a ‘‘compound’’ and $\prod_{I_s} (y_{s,i}^{\alpha})$ is the probability of this compound and ${}^{\circ} G_I^{\alpha}$ its Gibbs energy. In a solution this is the contribution from this endmember to the total Gibbs energy of the phase. $\text{c}^{\text{fg}} S_M^{\alpha}(y_{s,i}^{\alpha})$ is the configurational entropy and ${}^E G_M^{\alpha}(T, P, y_{s,i}^{\alpha})$ are the contributions from interactions of the constituents mixing on one or more sublattices of the phase. If a single endmember represents the constitution of the α phase then $\text{c}^{\text{fg}} S_M^{\alpha} = {}^E G_M^{\alpha} = 0$.

The term ${}^{\text{phy}} G_M^{\alpha}(T, P, y_{s,i}^{\alpha})$ is the contribution from various physical models, for example magnetism. This contribution can also be split into a sum of the contributions from each endmember and excess terms. Most physical models have parameters that depend on the constitution and sometimes also T and P . They are discussed in Section 3.3. If the phase has charged constituents there is an extra condition for equilibrium that the phase is electrically neutral. From now on we will only deal with expressions for a single phase and the phase superscript is omitted.

Already in the 1991 unary database a separate contribution, ${}^{\text{phy}} G_M$, from magnetic properties was introduced which has been essential for the calculation of steels and other magnetic materials. The introduction of model parameters for physical properties, such as the Curie temperature and magnetic moment, has greatly improved simulations of such materials.

The mixing depends on the model of the phase, which in the simplest case assumes that all constituents mix on a single sublattice. Alternatively, there can be two or more sublattices each with a specific set of constituents, for example interstitial solutions. This affects the configurational entropy as well as all the model parameters used to describe the Gibbs energy and its contributions from various physical models. The sublattices describe Long Range Ordering (LRO) where the constituents mix randomly in each sublattice, but some models take Short Range Ordering (SRO) into account by adding constituents representing clusters to modify the configurational entropy.

2.4. The configurational entropy

The random mixing of constituents $y_{s,i}$ on one or more sublattices s is;

$${}^{\text{cfeg}}S_M = \sum_s a_s \sum_i y_{s,i} \ln(y_{s,i}) \quad (7)$$

$$\sum_i y_{s,i} = 1 \quad (8)$$

$$M = \sum_s a_s \quad (9)$$

where a_s is the number of sites on sublattice s and $y_{s,i}$ is the fraction of constituent i on sublattice s . The sum of the sites is equal to the formula unit, M , of the phase.

The random mixing model is specified using the Compound Energy Formalism (CEF) as value of the *Configuration* attribute together with the sublattices in the **Sublattices** tag, see Section 3.2, Appendix A and Appendix B. The CEF entropy can describe Long Range Ordering (LRO)..

There are several models for Short Range Ordering (SRO), see for example [9,21–24], and some have been implemented in different software. There are no model parameters in the database for the configurational entropy model, the database only provides the set of model clusters as constituents of the phase and possibly their Gibbs energy of formation. The explanation of the model is provided by the software documentation.

3. The content of a Calphad database

Normally, a user selects the elements and phases for a given system from a given database, the data are then extracted by the software with the appropriate models and parameters. However, this can be software and application dependent.

The major part of the current TDB, as well as an XTDB database, are model parameters and their representation in the database as discussed in Section 3.3.

3.1. Elements, species, vacancies and electrons

The elements are the basic part of the XTDB format and they are defined as shown in Appendix A. The vacancy, denoted VA, is included among the chemical elements because the vacancy is needed to model interstitial solutions and defects in different phases use the vacancy. This may seem to violate the Gibbs phase rule:

$$f = n + 2 - p \quad (10)$$

where f is the degrees of freedom in a system with n elements (or components) and p is the number of stable phases. But the vacancy must have its chemical potential equal to zero at equilibrium and thus it will not change this rule.

Several models for the liquid [22,25] also include vacancies, with or without a charge, as an essential model constituent. The electron is never treated as a constituent by itself but a charged vacancy can be used as a free electron or hole to model semiconductors. As each phase must be electrically neutral there is no degree of freedom added by the ions.

In the TDB file the chemical species, including the vacancy and charged species as cations and anions, are introduced directly after the elements as shown in Appendix A. It would be possible to eliminate the **Species** tag in the XTDB file and instead define the constituents separately for each phase. However, the XTDB tags needed for entering a phase, as explained in Section 3.2, and Appendix A.3, is already quite complex and, therefore, in this proposal all constituents of a phase have separate **Species** tags independent of the **Phase** tag.

The sublattices of the thermodynamic model are populated by species and as such also called constituents of the phase. The species can be single atoms, vacancies, molecules/associates and charged ions. A vacancy with a charge can be used as a free electrons or hole. A species has fixed stoichiometry and as constituent of a phase it may have additional properties that must be provided as attributes in the **Species** tag.

3.2. The phase, its constituents and models

The XTDB tags defining a phase are listed in Appendix A. The **Phase** tag itself has only 3 attributes, *Id* is the name of the phase, *Configuration* specifies how the configurational entropy should be calculated and the *State* which is needed if the Equal Entropy Criterion (EEC) model, see Section 4.5, is used for the database.

There are 3 standard values for the *Configuration*:

- CEF, the abbreviation for Compound Energy Formalism, covers gas and all models with random mixing on one or more sublattices [26],
- MQMQA is used for liquid with the Modified Quasichemical Model in the Quadruplet Approximation [22],
- I2SL is used for liquids with the Ionic 2-Sublattice Liquid model [25] with variable site ratios.

At present there are no other established models for the configurational entropy for XTDB databases.

Nested inside the **Phase** tag one must have a **Sublattices** tag even if there is only one (or no) sublattices. In this tag the attributes *NumberOf* specifies the number of sublattices and *Multiplicities* the number of sites on each, i.e., the value of a_s in Eq. (9), which defines the formula unit of the phase. The values of the model parameters must be for one formula unit of the phase.

Within the **Sublattices** tag, there must be a **Constituents** tag for each sublattice. It has two attributes, *Sublattice* which can be omitted if there is only one, and *List* with a list of all constituents (already entered as **Species**), which can be present in the sublattice and separated by a space. Optionally the *WyckoffPosition* can be provided.

If the phase has some physical contributions, ${}^{\text{phy}}G_M$ in Eq. (6), those must be specified using the **AmendPhase** tag. The *Ids* of several models can be included in its *Models* attribute, separated by a space. There can also be a **DisorderedPart** tag if the phase has two separate sets of fraction variables, see Section 3.7.

A solid phase can optionally have a **CrystalStructure** tag, but using Calphad models the crystallographic designation of the phase may change with its constitution. Pure Fe as austenite can have the attribute *Strukturbericht* A1, but the thermodynamic model for the same phase may be ordered with *Strukturbericht* L1₂ in the Fe-Pt system or appear as carbide with the *Strukturbericht* B1 in the Fe-Ti-C system. The software may use the constitution of the phase to indicate this after an equilibrium calculation.

3.3. The Id of model parameters

The major part of a Calphad database is the model parameters. In the XTDB file they are written in the same compact way as in the TDB file. The **Parameter** tag for XTDB, see Appendix A, has an *Id* attribute using a simple notation with a Model Parameter Identifier (MPID) followed within parenthesis by the phase *Id*, the *Ids* of the constituents in the sublattices and possibly a degree of a parameter, exactly as in the TDB files. For parameters in the Gibbs energy expression ${}^{\circ}G_T$ a simple G is used as MPID, for parameters used in the ${}^E G_M$ one can use G or L as MPID. For example:

```
<Parameter Id="G(FCC_A1,FE:VA)" Expr="GFCCFE" />
<Parameter Id="L(FCC_A1,FE,CR:VA;0)" Expr="10833-7.477*T" />
```

In the *Id* of the **Parameter** tag the phase constituents are specified by the *Id* of the **Species** included in the *List* attribute for the **Constituents** tag in the order of the **Sublattices** tags used for the **Phase** tag. The constituents in different sublattices are separated by a colon,

“:”. When two or more constituents mix on the same sublattice they are separated by a comma, “,”. For a binary Redlich–Kister parameter the sequence of the mixing constituents may be important, see Section 3.11.

Before the closing parenthesis of the *Id* in the **Parameter** tag there can be a semicolon, “;”, followed by a digit. The meaning of this depends on the model, i.e., the MPID of the parameter.

In the 1991 unary database there were two new MPIDs, the Curie temperature and the magnetic moment for the magnetic model. Since then several models with composition dependent parameters have been introduced in a rather ad hoc way. The XTDB format will formalize all such MPID using the **ModelDescriptions** tag, see Appendix A.7, where the model and all allowed MPIDs for each model are defined. The fact that the MPIDs for the parameters are defined in the **ModelDescription** tags means that different software can use different symbols for the same composition dependent property. This may create some confusion but also adds flexibility when new models are introduced.

Some more examples are:

- G(LIQUID,FE+2:VA;0) is the endmember of pure Fe in the I2SL model.
- TC(FCC_4SL,FE:AL:AL:AL:VA;0) is the Curie temperature of the endmember in a 4 sublattice ordering model of the FCC phase with interstitials, see Appendix A.7. TC is the value of the *MPID2* attribute in the **Magnetic** tag for the magnetic model, see Appendix C.
- BMAGN(BCC_A2,CR,FE:VA;1) is a first order Redlich–Kister term in the Cr-Fe system for the magnetic moment in the BCC phase. The magnetic moment is given by the number of Bohr magnetons as function of the phase constitution. BMAGN is defined in Appendix C. The value after the semicolon indicates the power ν used for parameter in Eq. (20) in Section 3.11.
- G(SIGMA,MO:CR:*:*:*) is an EBEF parameter representing the bond energy parameter for Cr-Mo in a 5 sublattice model of the σ phase, see Appendix B.2. The wildcard character, “*”, is used to indicate that the parameter is independent of the constituent in that sublattice, and is explained in Section 3.8.
- LNTH(AL4C3,AL:C) is the logarithm of the θ parameter in the Einstein model, explained in Section 3.5, for the Al_4C_3 phase, see Appendix B.1.

All the MPIDs used in an XTDB file, except G and L, must be defined in the **ModelDescriptions** tag where the *MPIDx* values are specified for the different models.

As the *Id* of a **Parameter** tag contains the phase *Id* these tags can be arranged in the XTDB file as the database manager prefers (but after the **Phase** tags). Some may prefer to have the parameters for all phases for a binary or ternary system together to simplify updates.

3.4. Contributions from the physical models

In the TDB files the way to specify physical contributions is unstructured. The XTDB format uses **ModelDescriptions** tags, where the physical models used in the database are defined. The current models are listed with very short explanations in Appendix C. Each model has an *Id* attribute and possibly one or more Model Parameter Identifiers (MPID) for parameters describing the composition dependence of the contribution.

For example:

```
<Phase Id="FCC_4SL" . . . >
```

```
. . .
<AmendPhase Models="IHJCX GEIN FCC4PERM" >
  <DisorderedPart Sum="4" Subtract="Y" />
</AmendPhase>
</Phase>
```

the *Models* attribute specifies that the FCC_4SL phase has the Inden–Hillert–Jarl–Chen–Xiong (IHJCX) magnetic model, the Einstein low temperature extrapolation model, explained in Section 3.5, and a permutation of the parameters on the 4 sublattices for ordering explained in Section 3.9. The **DisorderedPart** tag is explained in Section 3.7 with examples in Appendix B. A software package can use the XTDB format even if only a subset of the models described here are implemented.

3.5. The Einstein low T extrapolation model

The Einstein model is used to extrapolate the Gibbs energy of a solid phase of an element down to $T = 0$ K. The integrated Einstein Gibbs energy equation per mole of atoms is:

$$\text{Ein } G_m(\theta) = 1.5R\theta + 3RT \ln(1 - \exp(-\theta/T)) \quad (11)$$

$$\text{Ein } G_M^\alpha = \aleph^\alpha \text{Ein } G_m(\theta) \quad (12)$$

where θ is the assessed Einstein temperature. Eq. (12), including the number of moles of atoms per formula unit, \aleph^α of the α phase, is used in Eq. (6). The value of $\ln(\theta)$ (called LNTH below) is used as a composition dependent variable in the Einstein MPID parameter because it is considered more appropriate physically.

Sometimes the heat capacity of an element is fitted using several Einstein temperatures, each with a different weight factor (the sum of which is unity), but only one of them can be composition dependent using the LNTH parameter. Thus the software must implement Eq. (11) as a function called GEIN with θ as argument, for use in the *Expr* attribute, see Section 4.6. The contributions from all θ are included in the Gibbs energy parameter for the endmember using GEIN functions with their assessed weight factors as shown below for the graphite phase, GRAPHITE, in the assessment of Al-C by [27].

```
<TPfun Id="GEGRACC" Expr="-0.5159523*GEIN(1953.3)+
0.121519*GEIN(448)+0.3496843*GEIN(947)+.0388463*GEIN(192.7)+
0.005840323*GEIN(64.5);" />
<TPfun Id="GTSERCC" Expr="-0.0029531332*T**2-3.3998492E-
16*T**5;" />
<TPfun Id="GHSERCC" Expr="-17752.213+GEGRACC+GTSERCC;" />
<Parameter Id="G(GRAPHITE,C;0)" Expr="GHSERCC;" Bibref="21HE" />
<Parameter Id="LNTH(GRAPH,C;0)" Expr="LN(1953.3);" Bibref="21HE" />
```

where the **TPfun** tag with the *Id* attribute GEGRACC together with the LNTH parameter is the contribution from the Einstein model using 5 different θ . The **TPfuns** GTSERCC and GHSERCC are the contribution at high T and the enthalpy at $T = 0$ K respectively. One of the θ is selected to vary with the composition using the LNTH parameter and its corresponding GEIN function is included in GEGRACC with its factor subtracted by 1.0. The whole Al-C assessment is in Appendix B.1 but in [27] the LNTH parameter is not used.

3.6. The liquid 2-state model

The Gibbs energy of the liquid phase for an element is extrapolated below its melting temperature using the liquid 2 state model, proposed by [28,29]. The Gibbs energy expression for the liquid is:

$$G_M^{\text{Liq}} = G_M^{\text{am}} - RT \ln \left(1 + \exp\left(-\frac{\Delta G_M}{RT}\right) \right) \quad (13)$$

where G_M^{am} describes the metastable amorphous state at low temperature using an Einstein model together with the Gibbs energy parameter, “G”, as a polynomial in T with no heat capacity contribution at $T = 0$ K. The ΔG_M parameter, divided by RT in Eq. (13), describes the transition to the stable liquid and the stable liquid above the melting T .

The *MPID1* attribute in the **Liquid2State** tag is for the ΔG_M parameter, see Appendix C, and the “G” parameter describes the low T amorphous state. The ΔG_M polynomial used in Eq. (13) must not have any heat capacity contribution at $T = 0$ K, see [27,29].

3.7. Using two Gibbs energy functions for a phase

For phases with many sublattices it is possible to use two separate sets of parameters for the Gibbs energy. This may significantly reduce the number of model parameters, see Sections 3.8 and 3.10. It can be used for phases such as face centered cubic (FCC), body centered cubic (BCC) and hexagonal close packed (HCP), which can be both ordered and disordered, and for phases which are always ordered, such as σ , μ , etc. It is indicated by a **DisorderedPart** tag inside the **AmendPhase** tag of the ordered **Phase** tag.

The **DisorderedPart** tag has 3 attributes, the *Sum* which specifies the number of sublattices, n , in the ordered phase to be summed to obtain the fractions of the disordered phase as:

$$x_i^{\text{dis}} = \frac{\sum_{s=1}^n a_s \sum_i y_{s,i}}{\sum_{s=1}^n a_s} \quad (14)$$

where s represent the sublattices to be summed and $y_{s,i}$ are the site fraction of i on the sublattice s . All ordered sublattices should have the same constituents but there can be a final interstitial sublattice in both the ordered and disordered part. It is not necessary that the sublattices which are summed are crystallographically identical, a σ phase can also have a disordered part, in particular when using Effective Bond Energy Formalism (EBEF) discussed in Section 3.10. The software is responsible to ensure the number of atoms in the ordered and disordered parts are the same.

In the disordered part the Gibbs energies of the endmembers, ${}^\circ G_i$, provide a surface of reference relative to the stable state of the constituent i . Thus all parameters in the ordered part become a kind of excess Gibbs energy parameter relative to this surface. For a phase with the **DisorderedPart** tag the Gibbs energy is calculated by one of these 2 equations:

$$G_M = {}^{\text{dis}}G_M(x) + {}^{\text{ord}}G_M(y) \quad (15)$$

$$G_M = {}^{\text{dis}}G_M(x) + {}^{\text{ord}}G_M(y) - {}^{\text{ord}}G_M(y=x) \quad (16)$$

where x are averaged values of y according to Eq. (14). The configurational Gibbs energy is calculated for the ordered part only. Eq. (15) is mainly used for phases with many sublattices which never disorder, for example topologically close-packed (TCP) phases.

In order to use Eq. (16) one must specify the second attribute *Subtract*="Y" in the **DisorderedPart** tag and it is used for phases with order/disorder transitions such as FCC, BCC and HCP when the model parameters for the disordered phase, including disordered excess parameters, have been assessed separately. Subtracting ${}^{\text{ord}}G_M(y=x)$ means that the model parameters of the ordered phase will not affect the disordered state. But there are many relations between the parameters to be considered using this, for details see [17,30].

The third attribute is *Disordered*, which can indicate the *Id* of the disordered phase if those parameters are entered in a separate phase. The parameters for the disordered phase can also be part of the ordered one, using parameters with fewer sublattices, see Appendix B.4.

3.8. The use of wildcards as constituents

It is possible to use parameters that do not specify the constituent in one or more of the sublattices. An extreme case is to use a parameter $G(\text{C14_Laves},*)$ which will add the value of the parameter to the Gibbs energy of the C14_Laves phase at all compositions. In a simulation, a positive value of such a parameter can be considered as a nucleation threshold.

In some compounds that are stable in a limited composition range, but may dissolve small amounts of many elements, an excess parameter $G(\text{C14_Laves},\text{A},\text{B}:*)$ can be a useful approximation of the interaction between A and B independent of the constituent in the second sublattice.

The implementation of the wildcard in a parameter identifier has sometimes been misunderstood. It is wrong if the software replaces the wildcard in the parameter with the actual constituents in the sublattice for which the wildcard is used. Instead any parameter with an explicit constituent in the sublattice for which the wildcard is used in another parameter should be added to this parameter. For example:

```
<Parameter Id="L(C14,A,B:C)" Expr="-7000" />
<Parameter Id="L(C14,A,B:*)" Expr="5000" />
```

should have their contribution to the Gibbs energy calculated as:

$$\Delta G = y_{1,A}y_{1,B}y_{2,C}L_{\text{A},\text{B}:C} + y_{1,A}y_{1,B}L_{\text{A},\text{B}:*} \quad (17)$$

where the two parameters are totally independent and can be modified separately. Even if C is the only constituent in the 2nd sublattice, both parameters should be extracted from the database.

3.9. Permutations of parameters in some ordered phases

When modeling ordering in phases with crystallographically identical sublattices such as L1_2 and L1_0 ordering in FCC, the endmember $G(\text{FCC},\text{A}:\text{B}:\text{B})$ has 4 permutations of the constituent A on the 4 sublattices that must have identical Gibbs energies. Using the **DisorderedPart** tag and the endmembers in the disordered part to represent the Gibbs energy relative to the stable state of the constituents, the endmembers in the ordered part are related just to the bonds between the constituents. If the AB bond, u_{AB} is independent of composition one has:

$$\begin{aligned} G(\text{FCC}, \text{A} : \text{A} : \text{A} : \text{B}) &= G(\text{FCC}, \text{A} : \text{A} : \text{B} : \text{A}) = \dots = 3u_{\text{AB}} \\ G(\text{FCC}, \text{A} : \text{A} : \text{B} : \text{B}) &= G(\text{FCC}, \text{A} : \text{B} : \text{A} : \text{B}) = \dots = 4u_{\text{AB}} \quad (18) \\ G(\text{FCC}, \text{A} : \text{B} : \text{B} : \text{B}) &= G(\text{FCC}, \text{B} : \text{A} : \text{B} : \text{B}) = \dots = 3u_{\text{AB}} \end{aligned}$$

By specifying FCC4PERM in the *Models* attribute of the **AmendPhase** tag a parameter $G(\text{FCC},\text{A}:\text{A}:\text{A}:\text{B})$ needs to be included only once in the XTDB file. In a multicomponent system a parameter $G(\text{FCC},\text{A}:\text{B}:\text{C}:\text{D})$ has 24 permutations. It is the software that must take care of calculating these parameters with the different sets of sublattice fractions, either by an internal loop or by storing the parameter at several appropriate positions in its data structure, see also Appendix B.4.

Excess parameters, such as $G(\text{FCC},\text{A},\text{B}:*:*:*)$, must also be permuted by the software for all identical sublattices when read from the database. Such a parameter indicates that the interaction between A and B in a sublattice is assumed to be independent of the constituents on the other 3 sublattices, which is reasonable even if the wildcard represents a third element. The reciprocal parameter, $G(\text{FCC},\text{A},\text{B}:\text{A},\text{B}:*:*:*)$, with 6 permutations is of particular interest because it can approximate the SRO contribution to the phase, see [31,32].

How the permutations are implemented in the software is not discussed here. Either the software stores the parameter explicitly for all possible permutations or stores the parameter only once and uses an internal loop to multiply the parameter with all relevant constituent fractions in the sublattices. The latter is recommended as it makes it easier to change the expression of the parameter. The BCC4PERM permutation is more complicated for the software as the tetrahedron is not symmetrical.

3.10. The EBEF model

The wildcard feature has been further explored by Dupin et al. [18] improving the modeling of intermetallic phases with many sublattices. Using the **DisorderedPart** feature described in Section 3.7, the Gibbs energy of the ordered part of such phases is independent of the reference states of the elements. In a σ phase with 5 sublattices an endmember parameter $G(\text{SIGMA},\text{A}:\text{B}:\text{A}:\text{A}:\text{A})$ thus represents the bond energy between constituent A in sublattice 1 and constituent B in sublattice 2, independent of the constituents in the other sublattices. As the sublattices are not crystallographically identical (as in an ordered FCC phase) an endmember $G(\text{SIGMA},\text{B}:\text{A}:\text{A}:\text{A}:\text{A})$ will not have the same value. Using the notation $E_{A,B}^{s,t}$ for a parameter with constituent A in sublattice s and constituent B in sublattice t , we can describe the energy for exchanging of A and B in any pair of sublattices as:

$$\Delta G_{A,B}^{\sigma} = \sum_{s=1}^4 \sum_{t=s+1}^5 (y_{s,A} y_{t,B} E_{A,B}^{s,t} + y_{s,B} y_{t,A} E_{B,A}^{s,t}) \quad (19)$$

where the $E_{A,B}^{s,t}$ can be fitted to DFT calculations for all combinations of endmembers in all 5 sublattices. This set of $E_{A,B}^{s,t}$ is similar to an excess parameter $L_{A,B}$ in a phase with a single sublattice and very good results have been found when extrapolating such parameters, evaluated in binary systems, to higher order systems. An example can be found in Appendix B.2.

Using model parameters such as $G(\text{SIGMA},\text{A}:\text{B}:\text{A}:\text{A}:\text{A})$ it is essential that the software treats them as Eq. (17), i.e., independent of the fractions in the sublattices with wildcards.

3.11. Composition dependence of excess parameters

This section applies to all kinds of excess parameters for the different physical models, not just the Gibbs energy parameters. In a binary system the composition dependence of the constituent fractions can be expressed in different ways. Most frequently used is the so called Redlich-Kister (RK) model, which for a substitutional model is:

$$\text{bin}G_{AB} = x_A x_B \sum_{v=0}^n (x_A - x_B)^v \cdot {}^v L_{AB} \quad (20)$$

where ${}^v L_{AB}$ can be a function of T . There are also other series expansions but they can always be transformed to a RK polynomial. Thus the XTDB file will only support the RK binary excess expression. However, some software packages calculate the difference $x_A - x_B$ in the alphabetical order and some in the order the constituents are listed in the parameter. This may require the software to change the sign of the parameters with odd powers. The composition dependent RK parameters have the degree specified after a semicolon in the *Id* attribute in the **Parameter** tag, see Section 3.3.

3.11.1. Ternary excess parameters

Parameters for ternary excess terms can be composition dependent according to the suggestion in [33]. The ternary composition dependence for a parameter $L_{1,2,3;*}$ is implemented in TDB files as:

$$\text{ter}G_{1,2,3} = y_1 y_2 y_3 (v_1 \cdot {}^0 L_{1,2,3;*} + v_2 \cdot {}^1 L_{1,2,3;*} + v_3 \cdot {}^2 L_{1,2,3;*}) \quad (21)$$

where 1, 2, 3 are constituents in alphabetical order. The values of v_i are:

$$v_1 = (1 + 2y_1 - y_2 - y_3)/3$$

$$v_2 = (1 + 2y_2 - y_3 - y_1)/3 \quad (22)$$

$$v_3 = (1 + 2y_3 - y_1 - y_2)/3$$

where $v_1 + v_2 + v_3 = 1$ ensures that the contribution is symmetrical in higher order systems. When implemented in the TDB format 40 years ago there was an inconsistency with the degrees of the parameter. A single ${}^0 L_{1,2,3;*}$ was assumed to be composition independent and in order to have a composition dependent ternary parameter one must specify all 3 parameters with indices 0, 1 and 2 even if one of the parameters is zero.

This will be modified in the XTDB by including an ${}^0 L_{1,2,3;*}$ which is composition independent and using 1, 2, 3 for the composition dependent parameters, i.e.,

$$\text{ter}G_{1,2,3} = y_1 y_2 y_3 ({}^0 L_{1,2,3;*} + v_1 \cdot {}^1 L_{1,2,3;*} + v_2 \cdot {}^2 L_{1,2,3;*} + v_3 \cdot {}^3 L_{1,2,3;*}). \quad (23)$$

It is not possible to make such a change reading a TDB files because there are so many old TDB files that cannot be changed consistently. But in the XTDB files this change can be made automatically by any software converting a TDB file to XTDB. There is no need for any special information in the XTDB file.

3.11.2. Reciprocal excess parameters

The reciprocal parameters with simultaneous interaction in two sublattices are important for an approximate SRO contribution as shown in [31] but they rarely have a higher order composition dependence. For a reciprocal parameter there is a composition dependence using 0 as index for the 0 degree term and the indices 1 and 2 for the 1st degree terms:

$$\Delta G = y_{1,A} y_{1,B} y_{2,C} y_{2,D} ({}^0 L_{A,B:C,D} + (y_{1,A} - y_{1,B}) \cdot {}^1 L_{A,B:C,D} + (y_{2,C} - y_{2,D}) \cdot {}^2 L_{A,B:C,D}) \quad (24)$$

with the constituents in alphabetical order.

3.12. Ternary extrapolation methods

In ternary systems the composition dependence of the binary excess parameters can be extrapolated in different ways. Three methods are used and explained in [33,34] and there is the **TernaryXpol** tag that can specify the extrapolation method for each case. For example:

```
<TernaryXpol Phase="FCC" Constituents="Fe Cr Si" Xpol="KT3T3" />
```

the Xpol attribute has one letter for the extrapolation method for each binary, K and M mean Kohler or Mugganu, T means Toop and one must also indicate the Toop constituent: 1, 2 or 3. The binaries are in the order of the constituents listed in the *Constituent* attribute. In the example above this means that the binaries Fe-Cr is Kohler, Fe-Si is Toop with Si as Toop element and Cr-Si is Toop also with Si as Toop element.

4. Some explanations related to Appendix A

The summary of XTDB in Appendix A is short and some additional information is provided here.

4.1. Names, texts and upper and lower case

The XML itself is case sensitive. The tags and attributes in XTDB must be written exactly as defined in [Appendix A](#). New tags and attributes may be added when needed to handle new features.

For the data provided in the attributes the thermodynamic software is free in how it handles names and other texts. In the TDB format upper and lower case was treated as equivalent and XTDB will follow the same rule. Except for the cases listed below, any *Id* must start with a letter A-Z and contain only letters, numbers 0-9 and the underscore character, “_”.

The length of *Ids* for species, phases, functions, etc. in the TDB files has been quite restricted. In XTDB there will be no restriction but it is recommended not to exceed 24 characters for any *Id*, except for the **Parameter** tag. However, a software reading an XTDB file is allowed to modify any *Id* according to its own rules and such changes of the *Id* relative to the XTDB file should be displayed to the user when reading the database.

4.2. The MQMQA and UNIQUAC constituents

The constituents needed to calculate the configurational entropy and Gibbs energy for the MQMQA model are provided as in the FactSage DAT file and must be specified in the *MQMQA* attribute of the **Species** tag. Also for the UNIQUAC model [35] for organic liquids each **Species** has an attribute *UNIQUAC* which defines the area and volume needed for its configurational entropy expression. Information for this is provided in the full XTDB documentation available at [36].

4.3. AppendXTDB files

The XTDB file for a large system can be very large. The set of model parameters for a σ phase with 5 sublattices and 10 constituents in each is 10^5 . There are models, such as the EBEF, explained in [Section 3.10](#), which can reduce this significantly but a single file for a database may still be difficult to manage. A way to handle this is to use the **AppendXTDB** tag. Such a tag is allowed only in the primary XTDB file and the primary XTDB file should also include all **Defaults**, **DatabaseInfo**, **AppendXTDB**, **Element**, **Species**, **Phase** tags and their subtags. But all or parts of the **Tpfun**, **Trange**, **Parameter**, **ModelDescriptions** and **Bibliography** tags can be in separate **AppendXTDB** files.

4.4. The phase and species Id

As explained in [Section 4.1](#) upper and lower case are the same in phase *Id*. In a species *Id* the characters “/”, “+” and “-” are also allowed.

Any parenthesis (“(” and “)”), comma “,”, colon “:” and semicolon “;” are forbidden as they are used to separate the various parts of the model parameter *Id* as explained in [Section 3.3](#).

4.5. The equal entropy criterion, EEC

The equal entropy criterion [37] was proposed as a method to avoid artifact breakpoints in the extrapolated Gibbs energy of solid phases. It means that the software, while calculating an equilibrium, should prevent any solid phase from becoming stable if it has higher entropy than a liquid phase at the same *T*, irrespective of the composition of the phases. When EEC is applicable, the database must indicate if a phase represents a liquid or a solid phase in the *State* attribute.

4.6. The Expr attribute in Tpfun and parameter tags

Many endmember parameters include an expression depending on the data for the pure elements and to avoid repetitions, one can use the **Tpfun** tag for an expression including the *T* and *P* as variables, and use the *Id* of these **Tpfuns** in the *Expr* attribute of other **Tpfun** or **Parameter** tags.

The *Expr* attribute in **Tpfun**, **Trange** and **Parameters** should be kept simple. Below are some rules:

A “simple term” in *Exp* is a signed real or integer value possibly multiplied with an integer power of *T* and *P*. A negative power must be enclosed by parenthesis. Examples:

$$12000 \quad -5*T**2 \quad +3.1416E12*T**(-9) \quad -4*T*P$$

A simple term can be used as argument of EXP, LN or LOG function (both LN and LOG are the natural logarithm) or it can be multiplied with the *Id* of a **Tpfun**. There is also the special GEIN function, defined by [Eq. \(11\)](#) in [Section 3.3](#), which is needed when a phase has been fitted with several Einstein θ , only one of which can be used as composition dependent in the LNTH parameter, see [Appendix B.1](#).

Multiplying several simple terms enclosed by parenthesis with another term is not allowed in *Expr* and neither is division. To handle such cases and for example \sqrt{T} , one can combine **Tpfuns**:

```
<Tpfun Id="LNT" Expr="LN(T) ;" />
<Tpfun Id="SQRT" Expr="EXP(0.5*LNT) ;" />
```

but *Expr*= ‘ ‘EXP(0.5*LN(T)) ; ’ ’ is not allowed.

If the **Trange** tag is used, the expression, as well as its first and second derivatives with respect to *T* and *P*, must be continuous across the breakpoint; otherwise the breakpoint represents a phase transition.

5. Summary

The XTDB format proposed here is a strict definition based on the TDB format that has been used for more than 30 years to develop large multicomponent databases for Calphad applications. The TDB format has been sufficiently flexible to handle many new models and may even have inspired some of them. But different software packages have implemented slightly different variants of the models and maybe not all their features. Thus it is important to establish a common format such as XTDB for future collaboration, especially considering the new unary database. The XTDB format described in this paper will be continuously updated where the current version is hosted at [36].

XTDB has not any facilities for internal verification of a database, that would require a much more complex splitting of the data, in particular the parameter tag with the MPID. The commercial databases that have such verification software for maintaining their databases can easily implement a conversion between their internal format and the XTDB format. The XTDB format is intended to be read and easily understood by researchers and scientists, not necessarily experts on thermodynamic modeling, but who develop and use the models in small databases as part of an assessment or an application. The XTDB format can easily be integrated in other databases for materials science using simple tags and attributes.

A future project is the development of a Markup Language schema for experimental and DFT data for use in assessments for Calphad databases, for this the use of an ontology and FAIR principles [14,15] is

important. The FAIR principles are a set of guidelines for making data Findable, Accessible, Interoperable, and Reusable.

CRedit authorship contribution statement

Bo Sundman: Writing – original draft, Validation, Software, Methodology, Conceptualization. **Fabio Miani:** Methodology, Project administration. **Axel van de Walle:** Writing – review & editing, Methodology, Conceptualization. **Bengt Hallstedt:** Writing – review & editing, Validation, Formal analysis, Conceptualization. **Ursula R. Kattner:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Florian Tang:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Taichi Abe:** Methodology, Formal analysis. **Reza Naraghi:** Methodology, Formal analysis, Conceptualization. **Erwin Povoden-Karadeniz:** Conceptualization. **Aurelie Jacob:** Writing – review & editing, Formal analysis, Conceptualization. **Shuanglin Chen:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Richard Otis:** Methodology. **Kazuhisa Shobu:** Methodology. **Malin Selleby:** Resources, Methodology. **Alexander Pisch:** Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful for creative discussion with Nathalie Dupin. Economic support from Stiftelsen för Tillämpad Termodynamik (STT), Sweden and the Scientific Group Thermodata Europe (SGTE), France are gratefully acknowledged. AvdW acknowledges support from the US National Science Foundation under grant DMR-2209027. URK thanks B. Long for constructive comments and L.M. Hale and C.E. Campbell for testing the XTDB files.

Appendix A. A summary of the XML tags and attributes

See Tables A.1–A.9. A complete version is available at [36]. Attributes which are mandatory are indicated by a !M to the left.

Table A.1
Global tags.

Tag	Attributes	Note
XTDB		Containing XML tags for an XTDB database.
!M	Version	Version of XTDB used for this database.
!M	Software	Name of software generating the database.
!M	Date	Year/month/day the database was written or last edited.
!M	Signature	Person or organization generating the database.
Defaults		Optional tag to provide default values of attributes in different XML tags and some other things.
	LowT	Default value of low T limit.
	HighT	Default value of high T limit.
	Elements	For example “VA” (vacancy) and “/–” (the electron).
	GlobalModel	Any model applicable to the whole database.
DatabaseInfo		Optional tag with information about the database.
	Info	Free text (excluding the characters <> &).
	Date	Last update of the database information.
AppendXTDB		Optional tag with additional files for the XTDB database. It should contain XTBD tags see Section 4.3.
	Models	The ModelDescriptions tag, see Section 3.4.
	Parameters	With mainly Parameter tags.
	TPfun	With some or all TPfun tags.
	Bibliography	With bibliographic tags.
	Misc	Whatever the database manager needs.

Table A.2
The element and species tags.

Tag	Attributes	Explanation
Element		Specifies a chemical element in the database. In addition the vacancy, denoted “VA”, and the electron, denoted “/–”, are included to handle defects and ions.
!M	Id	Chemical element symbol, one or two letters, for example FE, H. The Id is case insensitive, see Section 4.1.
	Refstate	Name of the reference phase for the element.
!M	Mass	Mass in g/mol.
	H298	Enthalpy difference between 0 and 298.15 K in the reference state.
	S298	Entropy difference between 0 and 298.15 K in the reference state.
Species		Specifies a chemical species used as a constituent of phases. The elements, except the electron, are also species.
!M	Id	Species name, see Section 4.1.
!M	Stoichiometry	One or more element Id each followed by an unsigned real or two integers separated by a “/” representing the stoichiometric ratio. A “/–” or “/+” followed by a digit means a negative or positive electric charge, if no digit 1 is assumed. See Section 4.1.
	MQMQA	For a constituent in the MQMQA model, see Section 4.2.
	UNIQAC	For a constituent in the UNIQAC model, see Section 4.2.

Table A.3
The phase tags.

Tag	Attributes	Explanation
Phase		All thermodynamic data is part of a phase.
!M	Id	Phase name, see sections 4.1 and 4.4.
!M	Configuration State	Model for the configurational entropy, see Section 2.4. See Section 4.5.
CrystalStructure	Prototype Strukturbericht PearsonSymbol SpaceGroup	Optional inside a Phase tag. Prototype phase. For example A3, B2, C14, D0_3, etc. For example hR21 cI2. For example 127, 166.
Sublattices		Only once inside a Phase tag.
!M	NumberOf	Number of sublattices, an integer value > 0.
!M	Multiplicities	One real value > 0 for each sublattice, see Appendix B.1
Constituents	Sublattice WyckoffPosition	Only inside the Sublattices tag. Can be omitted if only one sublattice. Optional specification.
!M	List	Species <i>Id</i> separated by a space, see Section 3.2.
AmendPhase	Models	Optional tag inside a Phase tag to specify a contribution from a physical model, see Section 3.4, Appendix C and Appendix B.1. One or more model <i>Ids</i> , separated by a space, for this phase. There can also be a DisorderedPart tag inside this tag.

Table A.4
The function tag. There is no provision for breakpoints in pressure. For high P special models are needed.

Tag	Attributes	Explanation
TPfun		Defines a <i>T, P</i> expression to be used in parameters or other functions.
!M	Id	The <i>Id</i> can be used in the <i>Expr</i> attribute of other functions or parameters, see Section 4.1.
!M	LowT Expr	Can be omitted if the default low <i>T</i> limit applies. Simple mathematical expression terminated by “;”, see Section 4.6. Use the Trange tag if several ranges.
!M	HighT	Can be omitted if the default high <i>T</i> limit applies.
Trange		Only inside a TPfun or Parameter tag for an expression with several <i>T</i> ranges. The function and its first and second derivative must be continuous.
!M	Expr	Simple mathematical expression terminated by “;”, see Section 4.6.
!M	HighT	Can be omitted if the default high <i>T</i> limit applies.

Table A.5
The parameter tag. The very compact form used in the TDB files is retained.

Tag	Attributes	Explanation
Parameter		Specifies the <i>T, P</i> expression of a model parameter for a set of constituents.
!M	Id	As in a TDB file, see Section 3.3
!M	LowT Expr	Can be omitted if the default low <i>T</i> limit applies. Simple mathematical expression terminated by “;”. If several ranges use a Trange tag, see Section 4.6.
!M	HighT	Can be omitted if the default high <i>T</i> limit applies.
!M	Bibref	Bibliographic reference.

Table A.6
The bibliography for parameters.

Tag	Attributes	Explanation
Bibliography		Contains bibliographic references. There are no attributes.
Bibitem		Only inside a Bibliography tag.
!M	Id	Used as value in the <i>bibref</i> attribute for a parameter or model, normally a paper or a comment by the database manager.
!M	Text DOI	Reference to a paper or comment. DOI of paper where the parameter was assessed.

Table A.7
The tags for current models.

Tag	Attributes	Explanation
ModelDescriptions		Contains model tags usually with an <i>Id</i> attribute used in AmendPhase tags inside a Phase tag. Most models have one or more model parameter identifiers (MPID).
!M	Software	Name of software using these models.
Magnetic		There are several magnetic models, see Appendix C.
!M	Id	This is used in <i>Models</i> attribute of the AmendPhase tag. There are 3 variants: IHJBCC, IHJREST and IHJCX.
	Aff	The antiferromagnetic factor (-1, -3 or 0).
!M	MPID1	Specifies the Bohr magneton number MPID.
!M	MPID2	Specifies a Curie or combined Curie/Néel temperature MPID.
!M	MPID3	Specifies a Néel temperature MPID for IHJCX.
!M	Bibref	Where the model is described.
Permutations		For FCC, HCP and BCC lattices a 4 sublattice tetrahedron model, see Section 3.9.
!M	Id	This is used in the <i>Models</i> attribute of the AmendPhase tag. Its can be either FCC4PERM or BCC4PERM.
!M	Bibref	Where the model is explained.
DisorderedPart		Optional tag inside the AmendPhase tag of an ordered phase with or without order/disorder, see Section 3.7.
	Disordered	Optional attribute with the <i>Id</i> of the disordered phase.
!M	Sum	Number of sublattices in the ordered phase in Eq. (14).
	Subtract	Must be specified if Eq. (16) in Section 3.7 should be used.
!M	Bibref	Where the model is described.

Table A.8
The tags for the new unary database.

Tag	Attributes	Explanation
Einstein		The low <i>T</i> vibrational model, see Section 3.5.
!M	Id	This <i>Id</i> is used in AmendPhase tag.
!M	MPID1	Specifies the MPID for the Einstein θ .
!M	Bibref	Where the model is described.
Liquid2state		The liquid 2-state model, see Section 3.6.
!M	Id	This <i>Id</i> is used in AmendPhase tag.
!M	MPID1	Specifies the MPID for the 2-state transition energy.
!M	MPID2	Specifies the MPID for the Einstein θ for the low <i>T</i> extrapolation.
!M	Bibref	Where the model is described.
EEC		Specifies that the Equal entropy model applies to the whole database, see Section 4.5. The liquid Phase tag must also have the <i>State</i> attribute equal to L.
!M	Id	This <i>Id</i> has the value EEC.
!M	Bibref	Where the model is described.

Table A.9
Miscellaneous tags.

Tag	Attributes	Explanation
TernaryXpol		The extrapolation method for a ternary, see Section 3.12.
!M	Phase	The <i>Id</i> of a phase.
!M	Constituents	The <i>Ids</i> of 3 Species that are constituents of the phase.
!M	Xpol	The type of extrapolations, for example KKK if the Kohler method is used for all 3 binaries.
!M	Bibref	Where the model is described.
BinarySystem		Optional tag for a database manager to surround a set of model parameters for a binary system. It can be used to list which assessed systems are present in the database.
!M	System	The <i>Ids</i> of the two elements inside the tag.
	CalcDia	Software dependent way to calculate the binary system.

Data availability

No data was used for the research described in the article.

Supplementary material

Appendix B and C referenced in the text are available as supplementary material at <https://doi.org/10.1016/j.calphad.2025.102849> or at <https://github.com/sundmanbo/XTDB>.

References

- [1] P.J. Spencer, A brief history of CALPHAD, *Calphad* 32 (2008) <http://dx.doi.org/10.1016/j.calphad.2007.10.001>.

- [2] L. Kaufman, H. Bernstein, *Computer Calculation of Phase Diagrams. with Special Reference To Refractory Metals*, Academic Press NY, 1970.
- [3] A. Dinsdale, SGTE data for pure elements, *Calphad* 15 (1991) [http://dx.doi.org/10.1016/0364-5916\(91\)90030-N](http://dx.doi.org/10.1016/0364-5916(91)90030-N).
- [4] G. Inden, The role of magnetism in the calculation of phase diagrams, *Physica* 103B, [http://dx.doi.org/10.1016/0378-4363\(81\)91004-4](http://dx.doi.org/10.1016/0378-4363(81)91004-4).
- [5] S. Hertzman, B. Sundman, A thermodynamic analysis of the fe-cr system, 6, 1982, [http://dx.doi.org/10.1016/0364-5916\(82\)90018-9](http://dx.doi.org/10.1016/0364-5916(82)90018-9).
- [6] M. Hillert, *Phase Equilibria, Phase Diagrams and Phase Transitions*, 2nd Ed., Cambridge Univ Press, 2008, <http://dx.doi.org/10.1017/CBO9780511812781>.
- [7] H.L. Lukas, S.G. Fries, B. Sundman, *Computational Thermodynamics, the Calphad Method*, Cambridge Univ Press, 2007, <http://dx.doi.org/10.1017/CBO9780511804137>.
- [8] P.E.A. Turchi, I. Abrikosov, B. Burton, S.G. Fries, G. Grimvall, L. Kaufman, P. Korzhavyi, V.R. Manga, M. Ohno, A. Pisch, A. Scott, W. Zhang, Interface between quantum-mechanical-based approaches, experiments and calphad methodology, *Calphad* 31 (2007) <http://dx.doi.org/10.1016/j.calphad.2006.02.009>.
- [9] A.D. Pelton, *Phase Diagrams and Thermodynamic Modeling of Solutions*, Elsevier, 2019, <http://dx.doi.org/10.1016/C2013-0-19504-9>.
- [10] R. Schmid-Fetzer, Phase diagrams, the beginning of wisdom, *J. Phase Equilib. Diffus.* 35 (2014) <http://dx.doi.org/10.1007/s11669-014-0343-5>.
- [11] A. van de Walle, M. Asta, High-throughput calculations in the context of alloy design *MRS bulletin*, 44, 2019, <http://dx.doi.org/10.1557/mrs.2019.71>.
- [12] C. Guéneau, M. Asta, B. Sundman, 2nd Ed, *Comprehensive Nuclear Materials*, Vol. 1, 2020, <http://dx.doi.org/10.1016/B978-0-12-803581-8.12054-5>.
- [13] Z. He, Q. Chen, M. Selleby, Third generation calphad for key elements, *J Phase Equilib Diffus.* 45 (2024) <http://dx.doi.org/10.1007/s11669-024-01166-z>.
- [14] X. Zhang, C. Zhao, X. Wang, A Survey on Knowledge Representation in Materials Science and Engineering: An Ontological Perspective, *Comp in Industry*, 2015, <http://dx.doi.org/10.1016/j.compind.2015.07.005>.
- [15] D. Riccardi, Z. Trautt, A. Bazyleva, E. Paulechka, V. Dikym, J.W. Magee, A.F. Kazakov, S.A. Townsend, C.D. Muzny, Towards improved FAIRness of the thermoML archive, *Comp. Chem.* (2022) <http://dx.doi.org/10.1002/jcc.26842>.
- [16] B. Bocklund, R. Otis, A. Egorov, A. Obaied, I. Roslyakova, Z.-K. Liu, ESPEI for efficient thermodynamic database development, modification and uncertainty quantification: application to Cu-Mg, *MRS Commun* 9 (2019) <http://dx.doi.org/10.1557/mrt.2019.59>.
- [17] B. Hallstedt, N. Dupin, M. Hillert, L. Höglund, H.L. Lukas, J. Schuster, N. Solak, Thermodynamic models for crystalline phases. Composition dependent models for volume, bulk modulus and thermal expansion, *Calphad* 31 (2007) 28–37, <http://dx.doi.org/10.1016/j.calphad.2006.02.008>.
- [18] N. Dupin, U.R. Kattner, B. Sundman, M. Palumbo, S.G. Fries, S. G, Implementation of an effective bond energy formalism in the multicomponent calphad approach, *J Res. NIST* 123 (2018) <http://dx.doi.org/10.6028/jres.123.020>.
- [19] <https://www.w3schools.com/xml/>.
- [20] B. Sundman, X.-G. Lu, H. Ohtani, The implementation of an algorithm to calculate thermodynamic equilibria for multicomponent systems with non-ideal phases in a free software, *Calphad* 101 (2015) <http://dx.doi.org/10.1016/j.commatsci.2015.01.029>, 0927-0256.
- [21] R. Kikuchi, A theory of cooperative phenomena, *Phys. Rev.* 81 (1951) pp 988–1003, <http://dx.doi.org/10.1103/PhysRev.81.988>.
- [22] A. Pelton, P. Chartr, G. Eriksson, The modified quasi-chemical model: Part IV. Two-sublattice quadruplet approximation, *Met. Trans A* 32A (2001) <http://dx.doi.org/10.1007/s11661-001-0230-7>.
- [23] E. Kremer, Associated solution model rebuilt, *Calphad* 77 (2022) 102408, <http://dx.doi.org/10.1016/j.calphad.2022.102408>.
- [24] C.-L. Fu, R.P. Gorrey, B.-C. Zhou, A cluster-based computational thermodynamics framework with intrinsic chemical short-range order, *Acta Mater.* 277 (2024) 1201388, <http://dx.doi.org/10.1016/j.actamat.2024.120138>.
- [25] M. Hillert, B. Jansson, B. Sundman, J. Ågren, A two-sublattice model for molten solutions with different tendency for ionization, *Metall. Trans. A* 16A (1985) 261–266, <http://dx.doi.org/10.1007/BF02816052>.
- [26] M. Hillert, The compound energy formalism, *J. Alloys Compd.* 320 (2001) 161–176, [http://dx.doi.org/10.1016/S0925-8388\(00\)01481-X](http://dx.doi.org/10.1016/S0925-8388(00)01481-X).
- [27] Z. He, B. Kaplan, H. Mao, M. Selleby, The third generation calphad description of Al-C including revision of pure Al and C, *Calphad* 72 (2021) <http://dx.doi.org/10.1016/j.calphad.2021.102250>.
- [28] J. Ågren, Thermodynamics of supercooled liquids and their glass transition, *Phys. Chem. Liq.* 18 (1988) <http://dx.doi.org/10.1080/00319108808078586>.
- [29] C.A. Becker, J. Ågren, M. Baricco, Q. Chen, S.A. Decterov, U.R. Kattner, J.H. Perepezko, G.R. Pottlacher, M. Selleby, Thermodynamic modelling of liquids, *Phys Stat Sol B* (2013) <http://dx.doi.org/10.1002/pssb.201350149>.
- [30] I. Ansara, T.G. Chart, A. Fernández Guillermet, P.C. Hayes, U.R. Kattner, D.G. Pettifor, N. Saunders, K.Zeng, Thermodynamic modelling of selected topologically close-packed intermetallic compounds, *Calphad* 21 (1997) [http://dx.doi.org/10.1016/S0364-5916\(97\)00021-7](http://dx.doi.org/10.1016/S0364-5916(97)00021-7).
- [31] B. Sundman, S.G. Fries, W.A. Oates, A thermodynamic assessment of the Au–Cu system, *Calphad* 22 (1998) 335–354, [http://dx.doi.org/10.1016/S0364-5916\(98\)00034-0](http://dx.doi.org/10.1016/S0364-5916(98)00034-0).
- [32] B. Sundman, N. Dupin, M.H.F. Sluiter, S.G. Fries, C. Guéneau, B. Hallstedt, U.R. Kattner, M. Selleby, The legacy of the regular solution model for stoichiometric phases and ionic melts, *J. Phase Equilib. Diffus.* 45, <http://dx.doi.org/10.1007/s11669-024-01163-2>.
- [33] M. Hillert, Empirical methods of predicting and representing thermodynamic properties of ternary solution phases, *Calphad* 4, pp 1–12, [http://dx.doi.org/10.1016/0364-5916\(80\)90016-4](http://dx.doi.org/10.1016/0364-5916(80)90016-4).
- [34] A. Pelton, A general geometric thermodynamic model for multicomponent solutions, *Calphad* 25 (2001) pp 319–328, [http://dx.doi.org/10.1016/S0364-5916\(01\)00052-9](http://dx.doi.org/10.1016/S0364-5916(01)00052-9).
- [35] J. Li, B. Sundman, J.G.M. Winkelman, A.I. Vakis, F. Picchioni, Implementation of the UNIQUAC model in the OpenCalphad software, *Fluid Phase Eq.* 507 (2020) <http://dx.doi.org/10.1016/j.fluid.2019.112398>.
- [36] <https://github.com/sundmanbo/XTDB/>.
- [37] B. Sundman, U.R. Kattner, M. Hillert, M. Selleby, J. Ågren, S. Bigdeli, Q. Chen, A. Dinsdale, B. Hallstedt, A. Khvan, H. Mao, R. Otis, A method for handling the extrapolation of solid crystalline phases to temperatures far above their melting point, *Calphad* 68 (2020) 101737, <http://dx.doi.org/10.1016/j.calphad.2020.101737>.