A proposal of an XML format for TDB files

Bo Sundman et al. September 21, 2023

Background

The current TDB format for Calphad databases has worked well for 40 years but introducing new models and features has revealed some awkward features, in particular the TYPE_DEFINITION keyword. When a change is needed it may be interesting to adopt an accepted markup language as XML or JSON and as it is easy to convert between any generally accepted markup language, the XML has been selected as it is slightly more explicit for a new XTDB format.

As XML is extendable each software may add their own flavour in a controlled way and it will be simpler to use an XTDB file by different softwares as one can indicate what is software specific.

The proposal here is a minimal one intended to:

- Make manually editing an XTDB file as easy as the current TDB format.
- Minimize problems when using software A to read a database file written by software B.
- Provide better facilities to handle new models and features added by the different softwares using Calphad databases. Significant new features should be discussed with SGTE before integrated in the XTDB format.

This XTDB definition may require minor modifications in the XML format already adoped in some software. Commercial companies providing encrypted databases to customers can use whatever format they prefer for such databases. But they should allow users to read and write unencrypted databases in the XTDB format.

How to add comments to this proposal

As we are using different editing facilities I have generated this PDF with limited amount of text on each page and at the end of each page in the attached LaTeX file (.tex) you are invited to add comments and send it back to me. Add a summary of your comments on the last page.

% Add any comments below

%———————————————————————

Please add your comments using your preferred text editor without changing any of the LaTeX formatting.

You are encoraged to ask anyone you think can contribute to defining the new database format to have a look at this and provide their comments.

In the table at the last page you can summarize your comments.

1 XTDB version 0.0.1

The following proposal is for a common definition of an XTDB file format based on XML. It can hopefully become an SGTE standard acceptable by all software which develop or use Calphad databases.

The "note" for the attributes for an XML element in the tables below is "+" for a compulsory attribute, "–" for an optional one or a letter A-Z for special cases explained in section 1.11.

All **Id** attributes are case insenitive as in the TDB file. Except for phase names the value of an **Id** cannot be abbreviated when used in other attributes.

An XML parser is convenient for reading XML files but to simplify the life for the humans reading the XTDB file or software using sequential read operations, it is recommended that the attributes to an XML element are provided in the order listed for each element below. Software specific attributes will come at the end.

The order of XML elements is free but to simplify for a human reading and editing the XTDB file a recommended order is **Defaults, DatabaseInfo, SoftSpecific** followed by **Element, Species, Model** and **Phases** and then **TPfun**s and **Parameter**s for the pure elements. After that **TPfun** and **Parameter** for binary, ternary and higher order systems, normally arranged for each subsystem.

An XML element referenced in the text below is written in **bold**.

1.1 The XTDB file

The XTDB file should be divided into lines (terminated by LF (Line Feed or newline as preferred by UNIX dialects) or CRLF (Carriage Return and Line Feed as preferred by Windows)). A line should not exceed 2000 characters.

Only one XML element per line (to be easily readable by humans) but an XML element may extend over several lines, including nested elements.

Maybe all attributes of an XML element should fit within a line (2000 characters)? Only nested XML elements on following lines.

For an example of a very long attribute see 1.5.

A software reading the XTBD file should ignore XML elements and attributes it does not support, preferably with a warning to the user.

1.2 The system XTDB elements

Element	Attributes	Note	Explanation
XTDB			Containing XML elements for an XTDB database.
	Version	+	Version of XTDB for this file.
	Software	+	Name of software generating the database.
	Date	+	Year/month/day the database was written or edited
	Signature	–	Name/email of person or organisation generating the database.
Defaults			Provides default values of attributes in other XML elements.
	LowT	–	Default value of low T limit.
	HighT	–	Default value of high T limit.
	Refid	–	Default bibliographic reference.
	Elements	–	VA and/or /- (the electron).
	Model	–	Any model applicable for the system, for example EEC
DatabaseInfo			Free text.
	Date	–	Last update.

1.3 The chemical element and species element

Element	Attributes	Note	Explanation
Element			Specifies a chemical element in the database. The vacancy is denoted "VA" and the electron "/-".
	Id	+	Chemical element symbol, for example FE, H
	Refstate	+	Name of a phase, for example GAS
	Mass	+	Mass in g/mol
	H298	–	Enthalpy difference between 0 and 298.15 K in reference state.
	S298	–	Entropy difference between 0 and 298.15 K in reference state.
Species			Specifies a molecular like aggregate used as constituent of phases. The elements, except the electron, are also species.
	Id	A	Species name max 24 letters. For note see 1.11.
	Stoichiometry	B	One or more elements each followed by an unsigned real representing the stoichiometric ratio. For note see 1.11. See also appendix A.1.
	MQMQA	C	For a constituent in the MQMQA model. For note see 1.11.
	UNIQUAC	D	For a constituent in the UNIQUAC model. For note see 1.11.

1.4 The function and temperature range elements

There is at present no way to handle several pressure ranges. The value of a function, as well as its first and second derivative with respect to T, must be continuous across an interval of T range.

Element	Attributes	Note	Explanation
TPfun			Defines a T, P expression to be used in parameters or other functions.
	Id	E	Function name, max 16 characters. The name can be used in the "Expr" attribute of other functions or parameters. For note see 1.11.
	LowT	–	If the default low T limit applies.
	Expr	+	Simple mathematical expression terminated by ;. Use the **Trange** element if several ranges. See section 1.12
	HighT	–	If the default high T limit applies.
Trange			Only inside a **TPfun** or **Parameter** element for an expression with several T ranges.
	Expr	+	Simple mathematical expression terminated by ;. See section 1.12.
	HighT	–	If the default high T limit applies.

1.5 The phase element

Normally the software requires information of sublattices, constituents and models for the phase in order to create the appropriate data structures. The **Parameter** and some models, see section 1.9.3 can appear anywhere in the XTDB file.

The **Splitphase** model (also known as the disordered fraction model) combines the Gibbs energies of an ordered and a disordered phase where the disordered phase has fewer sublattices than the ordered and the constituent fractions on its first sublattice are the average of the fractions summed over several sublattices in the ordered phase as specified by the attribute **Sum**. The **SplitPhase** element must appear within the ordered **phase** element before any other nested XML elements such as **AmendPhase, Parameter**, **KohlerModel** or **ToopModel** elements. Tho **Parameter, ToopModel** and **KohlerModel** elements may appear anywhere in the XTB file.

Some software may not have a separate disordered phase but arrange all **Parameter** elements for the ordered and disordered phase within the same phase.

A phase with very long list of constituents (i.e. GAS) can have several **Constituent** elements for the same sublattice.

Element	Attributes	Note	Explanation
Phase			All data belongs to a phase.
	Id	F	Phase name, see 1.11 note F.
	Configuration	G	Model for the configurational entropy. For note see 1.11.
	State	–	G for gas phase, L for liquid phase. Needed if EEC is used.
Sublattices			Only inside a **Phase** element.
	NumberOf	+	Number of sublattices, an integer value 1-10
	Multiplicities	+	Sites on each sublattice, as many reals as sublattices separated by a space.
Constituents			Only inside a **Phase** element.
	Sublattice	+	Indicates the sublattice for the list of species.
	List	+	Species Id, separated by a space, in the sublattice.
SplitPhase			Only inside the **Phase** element of the ordered phase.
	Refid	+	Where the model is described.
	Disordered	K	Name of disordered phase, for note see 1.11.
	Sum	+	Number of sublattices in the ordered phase to be summed for the disordered phase constitution. All sublattices summed must have the some constituents.
	Subtract	L	Must have value Y if the Gibbs energy of ordered phase as disordered should be subtracted, for note see 1.11.
AmendPhase			Only inside a **Phase** element.
	Model	+	List of models "Id", separated by a space, for this phase. The "Id" is specified in a model element below.

1.6 The simple parameter element

All thermodynamic data are defined by the parameter elements. They can be arranged inside a phase element or separately for each binary, ternary etc. subsystem. See also section 1.7 and appendix A.4.

Element	Attributes	Note	Explanation
Parameter			Specifies the T, P expression of a model parameter for a set of constituents.
	Id	+	As in TDB file, for example G(LIQUID,A,B:VA;2). See also the **Parameter2** element. May include the **Trange** element for parameters with multiple ranges.
	LowT	–	If the default low T limit applies.
	Expr	+	Simple mathematical expression terminated by ;. If several ranges use a **Trange** element. See section 1.12.
	HighT	–	If the default high T limit applies.
	Bibref	+	Bibliographic reference.

The **Id** starts with a model parameter identifier (MPID) followed within parenthesis by a phase name and one or more constituents in each sublattice of the phase. Constituents in the same sublattice (for interaction parameters) are separated by a comma, ",", and a colon, ":", is used to separate constitents in different sublattices.

After the constituents in the last sublattice a semicolon, ";" followed by a single digit, can be used to indicate a degree. This degree can have different meanings in different models but is normally used for the power in a Redlich-Kister series.

A wildcard, "*", can be used as constituent in a sublattice indicating that the parameter is independent of the constituents in that sublattice. See also secition 1.13.

It is not allowed to have a parameter for an interaction between a wildcard and a specific element, for example G(FCC,VA,*:VA) which would represent an interaction $y_{\mathrm{Va}}(1 - y_{\mathrm{Va}})$.

1.7 The elaborate parameter element

An alternative more elaborated XML element can be used for parameters which may be preferred by software. It is straightforward to convert from one to another.

Element	Attributes	Note	Explanation
Parameter2			A more detailed parameter element preferred by software.
	Id	–	As in TDB file, for example G(LIQUID,A,B:VA;2). See also the **Parameter** element.
	MPID	H	Model parameter identifier, for example G or TC. For note see 1.11.
	Phase	–	Can be omitted inside a phase element, otherwise the phase name.
	LowT	–	If the default low T limit applies.
	Expr	+	Simple mathematical expression terminated by ;. If several ranges use a **Trange** element. See section 1.12.
	HighT	–	If the default high T limit applies.
	Bibref	+	Bibliographic reference.
ConstArray			Only inside a **parameter2** element.
	Degree	I	Can be omitted if zero, for note see 1.11.
Constituent			Only inside a **ConstArray** element.
	Sublattice	+	Sublattice of constituent.
	Species	+	Name of a constituent.

The **Parameter2** element may be preferred by software but for manual editing **Parameter** is simpler, see the appendix A.4. Both elements can be used in the XTDB file and software can easily convert between them.

If an "Id" is present in a **parameter2** element the software should check that the "Id" is identical to the long form and report an error if not.

The use of wildcards, "*", for species in a parameter must be defined, see section 1.13.

1.8 Bibliography for parameters and models.

For each parameter there must be a bibliographic reference. This can be the same for all parameters from the same source. The models also have a reference which should indicate a paper where the model is explained, see Appendix A.5.

Element	Attributes	Note	Explanation
Bibliography			Contains bibliographic references
Bibitem			Only inside a **Bibliography** element.
	Id	+	Used as bibref attribute for a parameter or model, normally a paper or a comment by the database manager.
	Text	+	Reference to a paper or comment
	Date	–	Date when the bibref element was created
	Sign	–	Signature of the one adding the reference

1.9 XTDB model elements

The XML elements for generally accepted models can be on a separate file but the model "Id" is the important and used in the **AmendPhase** element for each phase which has the model. Most model has one or more model parameter identifiers (MPID) used in the **Parameter** element. See also appendix B. Software specific models should be explained in this XML element.

Element	Attributes	Note	Explanation
Models			Contains model descriptions and identifiers used in "AmendPhase" elements for phases. The models usually specify one or more model parameter identifiers needed by the model.

1.9.1 XTDB some used model elements

The XML elements listed here should be defined inside a library of **Model** elements and their "Id" attribute is used to specify the model parameter identifiers (MPIDs) associated with the model.

Element	Attributes	Note	Explanation
Magnetic			There are several magnetic models.
	Id	+	This Id is used in **AmendPhase** element.
	Refid	+	Where the model is described.
	MPID1	+	Specifies a magnetic model parameter identifier (MPID) for parameters.
	MPID2	+	Specifies a magnetic model parameter identifier (MPID) for parameters.
	MPID3	+	Specifies a magnetic model parameter identifier (MPID) for parameters.
Volume			Specifies the model for volume of a phase.
	Id	+	This Id is used in the **AmendPhase** element.
	Refid	+	Where the model is described.
	MPID1	+	Specifies volume model parameter identifier (MPID) for parameters.
	MPID2	+	Specifies volume model parameter identifier (MPID) for parameters.
	MPID3	+	Specifies volume model parameter identifier (MPID) for parameters.
FCC4Perm			For the FCC 4 sublattice tetrahedron model identical permutations of a parameter are listed only once.
	Id	+	This Id is used in **AmendPhase** element.
	Refid	+	Where the model is described.
BCC4Perm			For the BCC 4 sublattice tetrahedron model identical permutations of a parameter are listed only once.
	Id	+	This Id is used in **AmendPhase** element.
	Refid	+	Where the model is described.

1.9.2 New unary models and other models

There is also a new magnetic model. Other new **Model** elements may be included in the XTDB file in order to specify new MPIDs.

Element	Attributes	Note	Explanation
Einstein			The low T vibrational model.
	Id	+	This Id is used in **AmendPhase** element.
	Refid	+	Where the model is described.
	MPID1	+	Specifies the Einstein model parameter identifier (MPID) for parameters.
Liquid2state			The liquid 2-state model.
	Id	+	This Id is used in **AmendPhase** element.
	Refid	+	Where the model is described.
	MPID1	+	Specifies liquid model parameter identifier (MPID) for parameters.
	MPID2	+	Specifies Einstein model parameter identifier (MPID) for parameters.
EEC			Specifies that the Equi-entropy model applies to current database
	Refid	+	Where the model is described.

1.9.3 New XTDB model elements which are constituent specific

These models are related to specific constituents of a phase and they must be specified explicitly and cannot be included in the **AmendPhase** element for the phase because they include additional information.

The Toop and Kohler ternary extrapolation models are most likely specified together with the parameters for a specific ternary. The Muggianu ternary model is the default extrapolation model.

There are probably many more models to be considered and we must take care how they can be integrated in the XML format.

Element	Attributes	Note	Explanation
ToopModel			Subset of 3 constituents for which the Toop ternary extrapolation model should be used. By default the Muggianu extrapolation is used.
	Refid	+	Where the model is described.
	Phase	+	Can be omitted if used inside a **Phase** element. Otherwise the phase for which the model should be used.
	Constituents	+	Specifies 3 constituents, the Toop constituent first.
KohlerModel			Subset of 3 constituents for which the Kohler ternary extrapolation model should be used. By default the Muggianu extrapolation is used.
	Refid	+	Where the model is described.
	Phase	+	Can be omitted if used inside a **Phase** element. Otherwise the phase for which the model should be used.
	Constituents	+	Specifies 3 constituents in any order.

1.10 Software specific XML elements and others

The TYPE_DEFINITION keyword in the TDB file is replaced in this XML format by **AmendPhase, SplitPhase** and some other attributes. But some features of the TYPE_DEFINITION, which do not relate to the thermodynamic data, have to be replaced by software specific XML element.

The ASSESSED_SYSTEM is an nice feature and I propose we keep such a possibility also in the XTDB file.

I think it would also be interesting for users to have an XML element UNASSESSED_BINARY, especially for encrypted databases, in order to indicate composition regions where the user should be careful using calculated results.

The current TYPE_DEFINITION keyword is sometimes used to create composition sets if certain elements have been selected or testing if certain phases should be rejected/restored. We should think about how this can be implemented.

Element	Attributes	Note	Explanation
AssessedSystem			Indication of system with assessed data.
	Elements	+	The elements or constituents assessed.
	TC	-	Software specific character string to calculate the system with TC.
	Pandat	-	Software specific character string to calculate the system with Pandat.
	PyC	-	Software specific character string to calculate the system with PyCalphad.
	MC	-	Software specific character string to calculate the system with MatCalc.
	OC	-	Software specific character string to calculate the system with OC.
UnassessedBinary			Indication of missing binary assessments.
	Elements	+	Two elements with no or incomplete set of assessed parameters.

1.11 Notes

Note that the XML elements and attributes are case sensitive whereas the names of elements, species, functions and other text are case insensitive. For example a chemical element written FE, Fe, fe and fE is the same element. Thus CO is Cobalt and C1O is carbonmonoxide.

As a devoted Fortran programmer I prefer character variables with fixed length and also because it is difficult to format nice output with very long names of functions or phase names. For a human it is also complicated to handle long names, even if they may be abbreviated.

It is important to be strict here, it is easy to misunderstand a definition if one is not very explicit.

A: A species name or **Id** must start with letter A-Z and can contain letters, digits and the special characters "_", "/", "-" and "+". It must not be abbreviated when used as constituent in a **Phase** or **Parameter** element.

Should the characters "-" and "_" be treated as identical(?)

B: The species stoichiometry is a sequence of one or more element names (case insensitive) followed by a real number specifying the stoichiometric ratio. Following the TDB standard an element with a two letter name does not need a stoichiometric ratio equal to unity, for example MGO for MG1O1. A final stoichiometry unity can also be ignored. No parenthesis are allowed. For MQMQA clusters the stoichiometry can be omitted.

C: The MQMQA attribute should contain two or more element names separated a comma or a sublattice separator ":" and followed by equal number of unsigned reals representing the bond fractions, For example:

CS,LA:F 9.0 6.0 4.0

See also appendix A.1. According to the MQMQA model the cluster stoichometry is calculated as 6.0 divided by the bond fractions. The cluster is always electrically neutral. For endmember clusters (with one species in each sublattice) a real representing the SNN/FNN ratio, usually 2.4, must be supplied.

D: For UNIQUAC species. Contains two reals representing volume and area of the species in m^3 and m^2 (?).

E: A **TPfun** name must start with a letter A-Z and may contain letters, digits and the special character "_". It must not be longer than 16 characters. It cannot be abbreviated.

F: A phase name must start with a letter A-Z and have no more than 24 characters. It can contain letters, numbers and the special character "_". In some software"(" and ")" may be used in phase names but they can be converted to "_" in other software.

A phase name may be abbreviated in parameters and some other cases and thus a phase name must be unique and not an abbreviation of another phase.

G: The configuration model, for example CEF, I2SL, MQMQA etc. Maybe also"RKM" for a Redlich-Kister-Muggianu model with a single lattice and one site?

H: An initial set of model parameter identifies (MPID) must be defined, see section 1.14 and appendix C. At present G is the Gibbs energy, TC Curie T, BMAGN the Bohr magneton number etc. The MPID must not be abbreviated.

I: The degree can be omitted if zero.

K: The disordered phase within a **SplitPhase** element can be omitted if the parameters for the disordered phase are provided within the ordered phase. In OC I use a suffix "D" after the MPID name and the parameters has the appropriate reduced number of sublattices.

For example GD(SIGMA,CR) can be used as **Id** in a **Parameter** element for the lattice stability of Cr in SIGMA if the SIGMA phase is modelled with 3 or more sublattices and has the **SplitPhase** element. This avoids the need for a meaningless disordered DIS_SIGMA phase.

L: A BCC or FCC phase with a disordered fraction set (i.e. a **SplitPhase** model) can have the "ordered" part "as disordered" subtracted in order to allow the use of the disordered part separately. If so the value of the **Subtract** attribute should be "Y".

For phases which never disorder, for example the sigma phase, one can also have a "disordered" part with a single sublattice containing the pure elements as endmembers (and possibly some long range interaction parameters). In Thermo-Calc terminology this is the "NEVER" model. In order not to subtract the ordered part as disordered in XTDB the **Subtract** attribute can be omitted or given any other value than "Y".

1.12 The mathematical expression used in Expr

The mathematical expressions for T and P used in **TPfun, Trange** and **Parameter** are the same as in TDB files. It is very limited because some software must calculate first and and second derivatives with respect to T, P (and constitution). A more extended mathematical expression could be allowed for expressions which are not used for database parameters. In OC it is allowed to enter more complex expressions for post processing in batch/macro files.

In the 1990 definition of the TDB file the type of expression allowed consists of "simple terms" such as:

[signed real number] * [**TPfun Id**] ** [power] *T** [power] *P**[power]

where [power] is an integer (a negative power must be within parenthesis). No spaces allowed in a simple term. A "complex term" is a simple term multiplied with a math function of a simple term, such as:

[simple term] *LN([simple term])

An **Expr** attribute in **TPfun, Trange** or **Parameter** consists one or more terms. A positive sign of the first term can be omitted.

The following general math functions are allowed in OC:
$\exp(), \ln(), \log()$, erf(). Note that log() and ln() is the same and erf() is the error function. The number of math functions can be extended.

The following math function is needed for the unary project:
GEIN(θ) to calculate $1.5R\theta + 3RT\ln(1 - \exp(-\theta/T))$;

It is not allowed to group several terms together using parenthesis.

A square root of T is entered as two **TPfun**://

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

1.13 The use of wildcards for constituents in parameters

In some parameters a wildcard, "*", is used to indicate the parameter is independent of the constituent in this sublattice. For example:

A parameter G(sigma,GA:GA:*;0)=100000; means that the Gibbs energy of formation of "phase" with GA in first and second sublattice is independent of the constituent in the third. If one later finds that for example the Gibbs energy for sigma with TA in the third sublattice should be just +70000 then, in OpenCalphad, one must set G(sigma,GA:GA:TA)=-30000; because OC stores the two parameters separately and will add them when calculating.

This may be handled differently in other software and requires attention.

1.13.1 The EBEF model use wildcards

This model use the same notation for parameters as in CEF, and can replace a large number of endmembers parameters by fewer **Parameter** elements, representing bond energies between pairs of elements. The bond energy parameters specify constituents in only 2 (or 3) sublattices in phases with 3 or more sublattices and use wildcards in remaining sublattices.

A parameter G(SIGMA,CR:FE:*:*:*) represent the bond energy between CR in first and FE in second sublattice, independent on the constituents in the other sublattices (note that G(SIGMA,FE:CR:*:*:*) is not the same value!).

Using such parameters (fitted for example to DFT calculated endmember energies) the number of parameters can be reduced by more than an order of magnitudes and, in addition, extrapolations can be improved!

1.13.2 An alternative to wildcards

To avoid using explicit wildcards one could indicate the sublattice after the constituent, i.e. to use

G(SIGMA,CR@1:FE@2) instead of G(SIGMA(CR:FE:*:*:*), where the number after the character "@" indicates the sublattice.

1.13.3 A missing wildcard in the I2SL model parameters

An I2SL model may exist with only neutrals in the anion sublattice. For example the elements C and S can be neutrals in the I2SL model and their parameters are in Thermo-Calc are written as G(I2SL,C), omitting the cation sublattice. In OpenCalphad it is written G(I2SL,*:C) which may be a bit more consistent as it indicates that the constituent on the cation sublattice is irrelevant.

Alternatively the parameter could be written G(I2SL,C@2) if the proposal above is accepted.

Note that an interaction between neutrals with a single specific cation is forbidden in the I2SL model.

1.14 Model parameter identifiers, MPID

An MPID must start with a letter A-Z and contain letters and digits and not exceed 8? characters. It cannot be abbreviated. There are already some defined but there should be an extensive list of future MPIDs to avoid that different software use the same for (slightly) different things. See appendix C.

The letter "&", frequently used for mobilities, is forbidden in XML but it can be replaced by some other character, for example "@". Using "&" seems unnecessary and quite clumsy.

1.15 CVM and the cluster site model

These should eventually be included also.

2 Points of discussions

1. Should mathematical expressions be terminated by a ";"? It is not necessary because there is a final double quote of the expression but it may be nice.
2. The GEIN function is used in the **Parameter** for a pure element modeled with several Einstein θ. Only one of these is selected to vary with composition using the LNTH model parameter identifier (MPID).
3. Should the sign of Redlich-Kister terms depend on the alphabetrical order of the constituents or as the order the constituents are written in the parameter **Id**? Currently the TDB file use alphabetical.

3 Summary

There are certainly many more things to take care of but I think it is more important to agree on a minimum common XML format which can make the thermodynamic databases more useful both for experimentalists, assessments, database maintenance and thermodynamic software in particular for the development of new models and applications. We have to take one step at a time.

Appendix A Appendix A: Some examples

It may not be completely consistent with the definitions above as I have taken them from my previous conversion software.

A.1 Elements and species.

The way to define MQMQA constituents and stoichiometry is tentative.

```
  <Element Id="/-" Refstate="ELECTRON_GAS" Mass="  0.000000E+00" />
  <Element Id="VA" Refstate="VACUUM" Mass="  0.000000E+00" />
  <Element Id="AL" Refstate="FCC_A1" Mass="  2.698200E+01" H298="  4.577300E+03" S298="  2.832200E+01" />
  <Element Id="FE" Refstate="BCC_A2" Mass="  5.584700E+01" H298="  4.489000E+03" S298="  2.728000E+01" />
...
  <Species Id="VA" Stoichiometry="VA" />
  <Species Id="AL" Stoichiometry="AL" />
  <Species Id="FE" Stoichiometry="FE" />
  <Species Id="ALFE" Stoichiometry="AL0.5FE0.5" />
  ...
  <Species Id="LA/F" MQMQA="LA:F 6.0 2.0 2.4" />
  <Species Id="LACS/F" MQMQA="LA,CS:F 9.0 6.0 4.0" />
```

The stoichiometry of the MQMQA species is calculated from the bond ratios according to the model. The **MQMQA** attribute for an "endmember" also include a factor needed in the configurational entropy expression.

Note that species cannot be abbreviated as constituents in phases or parameters and thus one can have species which are abbreviations of another species. For example "O" and "O2".

A.2 Defaults, TPfun and Trange

Using default T limits the function are not much more complex than in the TDB file.

```
  <Defaults LowT="298.15" HighT="6000" Refid="U.N.Known" Elements="VA /-" />
...
  <TPfun Id="GHSERAL"  >
    <Trange HighT="700" Expr=" -7976.15+137.093038*T-24.3671976*T*LN(T)-.001884662*T**2-8.77664E-07*T**3+74092*T**(-1);"
    <Trange HighT="933.47" Expr=" -11276.24+223.048446*T-38.5844296*T*LN(T)+.018531982*T**2  -5.764227E-06*T**3+74092*T**
    <Trange HighT="2900" Expr=" -11278.378+188.684153*T-31.748192*T*LN(T)-1.230524E+28*T**(-9);" />
  </TPfun>
...
  <TPfun Id="LFALFE0"  Expr="-104700+30.65*T;" />
  <TPfun Id="LFALFE1"  Expr="+30000-7*T;" />
  <TPfun Id="LFALFE2"  Expr="+32200-17*T;" />
  <TPfun Id="UFALFE"  Expr="-4000+T;" />
  <TPfun Id="GAL3FE"  Expr="+3*UFALFE+9000;" />
  <TPfun Id="GAL2FE2"  Expr="+4*UFALFE;" />
```

A.3 Phase

Entering phases in the XTDB file is a bit more complex but we have get rid of the TYPE_DEFINITION. It is not so nice to use CEF for the liquid, that is why I think RKM might be a nice option for a liquid model without sublattices and maybe **Sublattices** can be omitted?

```
 <Phase Id="LIQUID" Configuration="RKM" state="L" >
    <Sublattices Number_of="1"  Ratios="1" >
      <Constituents sublattice="1" list="AL C" />
    </Sublattices>
    <AmendPhase model="LIQ2STATE" />
  </Phase>
...
  <Phase id="A2_BCC" Configuration="CEF" state="S" >
    <Sublattices Number_of="2"  Ratios="1  3" >
      <Constituents sublattice="1" list="AL FE" />
      <Constituents sublattice="2" list="VA" />
    </Sublattices>
    <AmendPhase model="IHJBCC" />
  </Phase>
...
  <Phase id="AL8FE5_D82" Configuration="CEF" state="S" >
    <Sublattices Number_of="2"  Ratios="8  5" >
      <Constituents sublattice="1" list="AL FE" />
      <Constituents sublattice="2" list="AL FE" />
    </Sublattices>
  </Phase>
...
  <Phase id="BCC_4SL" Configuration="CEF" state="S" >
    <Sublattices Number_of="5"  Ratios="0.25  0.25  0.25  0.25  3" >
      <Constituents Sublattice="1" List="AL FE" />
      <Constituents Sublattice="2" List="AL FE" />
      <Constituents Sublattice="3" List="AL FE" />
      <Constituents Sublattice="4" List="AL FE" />
      <Constituents Sublattice="5" List="VA" />
    </Sublattices>
    <SplitPhase Disordered="A2_BCC" Sum="4" Subtract="Y" Refid="09Sun" />
    <AmendPhase model="IHJBCC BCC4Perm" />
  </Phase>
```

For phases one can abbreviate its name and thus one cannot have a phase with a name which is an abbreviation of another phase name. Phase names can be abbreviated for each part separated by an underscore, "_". A phase name "AL_X" is thus an abbreviation of "AL2_X".

A.4 Parameters

```
    <Parameter Id="G(A2_BCC,FE:VA;0)"   Expr="+GHSERFE;" Bibref="91Din" />
    <Parameter Id="TC(A2_BCC,FE:VA;0)"   Expr="1043;" Bibref="91Din" />
    <Parameter Id="BMAGN(A2_BCC,FE:VA;0)"   Expr="2.22;" Bibref="91Din" />
    <Parameter Id="G(AL8FE5_D82,AL:AL;0)"   Expr="+13*GALBCC;" Bibref="08Sun" />
...
    <Parameter Id="G(BCC_4SL,AL:AL:FE:FE:VA;0)"   Expr="+GB2ALFE;" Bibref="08Sun" />
    <Parameter id="G(BCC_4SL,AL,FE:*:*:*:VA;1)"   Expr="-634+0.68*T;" Bibref="08Sun" />
```

In the last parameter above the "wildcard" or asterisk, "*", is used for species in three of the sublattices and it means that the parameter is independent of the constituents in these sublattices. See the discussion in section 1.13 how this is treated.

Using the **Parameter2** element the last two parameters above are:

```
   <Parameter2 Id="G(BCC_4SL,AL:AL:FE:FE:VA;0)" Expr="+GB2ALFE;" Bibref="08Sun" MPID="G" Phase="BCC_4SL" >
      <ConstArray Degree="0">
          <Constituent Sublattice="1" Species="AL" />
          <Constituent Sublattice="2" Species="AL" />
          <Constituent Sublattice="3" Species="FE" />
          <Constituent Sublattice="4" Species="FE" />
          <Constituent Sublattice="5" Species="VA" />
      </ConstArray>
   /Parameter2>
...
   <Parameter2 id="G(BCC_4SL,AL,FE:*:*:*:VA;1)" Expr="-634+0.68*T;" Bibref="08Sun" MPID="G" Phase="BCC_4SL" >
      <ConstArray Degree="1">
          <Constituent Sublattice="1" Species="AL" />
          <Constituent Sublattice="1" Species="FE" />
          <Constituent Sublattice="2" Species="*" />
          <Constituent Sublattice="3" Species="*" />
          <Constituent Sublattice="4" Species="*" />
          <Constituent Sublattice="5" Species="VA" />
      </ConstArray>
   /Parameter2>
```

A.5 Models

An extended form of the **Models** should be defined a separate global library file with the **Id** and full description but the models used in an XTDB file should appear in a short form as below.

```
<Models>
  <Magnetic Id="IHJBCC"  MPID1="TC" MPID2="BMAGN" Bibref="82Her" />
  <Magnetic Id="IHJREST"  MPID1="TC" MPID2="BMAGN" Bibref="82Her" />
  <Einstein Id="GLOWTEIN" MPID1="LNTH" bibref="01Qing" />
  <Liquid2state Id="LIQ2STATE" MPID1="G2"  MPID2="LNTH" bibref="14Becker" >
     Unified model for the liquid and the amorphous state treated as an Einstein solid
  </Liquid2state>
  <Volume Id="VOLOWP" MPID1="V0"  MPID2="VA" MPID3="VB" bibref="05Lu" />
...
</Models>
...
<ToopModel Phase="Liquid" Constituents="C Fe Cr" />
```

The **ToopModel** or **KohlerModel** elements can appear anywhere in the XTDB file, even within the **Phase** element or, more likely together with all the **Parameter** elements from an assessment of the ternary C-Fe-Cr system in order to simplify editing a large XTDB file. Either way may be a bit complicated to handle by the calculating software reading the XTDB file.

The **SplitPhase** in example A.3 must appear within a **Phase** element for the ordered phase.

Appendix B Proposal of Id for models.

The agreed models need only to added by their idenfiers to the phases.

For models that are not agreed a full description of the model and its model parameter identifiers (MPIDs) should be provided in the XTDB file. Note the different software may use different model parameter identifiers for the same property...

```
<Models>
  <Magnetic id="IHJBCC"  MPID1="TC" MPID2="BMAGN" anti-ferromagnetic_factor=" -1.00"
     f_below_TC=" +1-0.905299383*TAO**(-1)-0.153008346*TAO**3-.00680037095*TAO**9-.00153008346*TAO**15;"
     f_above_TC=" -.0641731208*TAO**(-5)-.00203724193*TAO**(-15)-.000427820805*TAO**(-25);" bibref="82Her" > in G=f(TAO)*LN(BETA+1) where TAO=T/TC
  </Magnetic>
  <Magnetic id="IHJREST"  MPID1="TC" MPID2="BMAGN" anti-ferromagnetic_factor=" -3.00"
     f_below_TC=" +1-0.860338755*TAO**(-1)-0.17449124*TAO**3-.00775516624*TAO**9-.0017449124*TAO**15;"
     f_above_TC=" -.0426902268*TAO**(-5)-.0013552453*TAO**(-15)-.000284601512*TAO**(-25);" bibref="82Her" > in G=f(TAO)*LN(BETA+1) where TAO=T/TC
  </Magnetic>
  <Magnetic id="IHJQX"  MPID1="CT" MPID2="NT" MPID3="BMAGN" anti-ferromagnetic_factor="  0.00"
     f_below_TC=" +1-0.842849633*TAO**(-1)-0.174242226*TAO**3-.00774409892*TAO**9-.00174242226*TAO**15-.000646538871*TAO**21;"
     f_above_TC=" -.0261039233*TAO**(-7)-.000870130777*TAO**(-21)-.00018426298*TAO**(-35)-6.65916411E-05*TAO**(-49);" bibref="12Xiong" > in G=f(TAO)*LN(BETA+1) where TAO=T/CT or T/NT
  </Magnetic>
  <Einstein id="GLOWTEIN" MPID1="LNTH" bibref="01Qing" >
     Gibbs energy due to the Einstein low T vibrational entropy model, G=1.5*R*THETA+3*R*T*LN(1-EXP(-THETA/T)).
  </Einstein>
  <Liquid2state id="LIQ2STATE" MPID1="G2"  MPID2="LNTH" bibref="14Becker" >
     Unified model for the liquid and the amorphous state treated as an Einstein solid
  </Liquid2state>
  <Volume id="VOLOWP" MPID1="VO"  MPID2="VA" MPID3="VB" bibref="05Lu" >
     The volume of a phase is described as function of T, P and its constitution.
  </Volume>
  <SplitPhase Disordered=" " Sum=" " Subtract=" " Refid="09Sun" >
     The disordered fractions are summed over the ordered sublattices indicated by Sum.  The Gibbs energy calculated 2 or 3 times as indicated by the CEF_appendix but the Subtract but the configurational entropy o
  </SplitPhase>
  <FCC-permutations id="FCC4Perm" bibref="09Sun" >
     Permutations of ordered FCC parameters with the same set of elements are listed only once.
  </FCC-permutations>
  <BCC-permutations id="BCC4Perm" bibref="09Sun" >
     Permutations of ordered BCC parameters with the same set of elements are listed only once.
  </BCC-permutations>
  <EEC id="EEC" bibref="20Sun" >
     Equi-Entropy Criterion means that solid phases with higher entropy that the liquid phase must not be stable.
  </EEC>
</Models>
```

Appendix C Proposal of initial set of Model Parameter Identifiers, MPID

The 32 model parameter identifiers, MPID, defined in OC as shown in Table 1. In OC a parameter for a disordred part of a **SplitPhase** use the same phase name but the MPID has a suffix "D".

Table 1: Current set of model parameter identifiers in OC. For each parameter it is indicated if it can depend on T, P or have an extra constituent specification. Most of them have no associated code.

Indx	Ident	T	P	Specification	Status	Note
1	G	T	P		0	Gibbs Energy
2	TC	-	P		2	Combined Curie/Neel T
3	BMAG	-	-		1	Average Bohr magneton number
4	CTA	-	P		2	Curie temperature
5	NTA	-	P		2	Neel temperature
6	IBM	-	P	&<constituent#sublattice>;	12	Individual Bohr magneton number
7	LNTH	-	P		2	Einstein temperature
8	V0	-	-		1	Volume at T0, P0
9	VA	T	-		4	Thermal expansion
10	VB	T	P		0	Bulk modulus
11	VC	T	P		0	Alternative volume parameter
12	VS	T	P		0	Diffusion volume parameter
13	MQ	T	P	&<constituent#sublattice>;	10	Mobility activation energy
14	MF	T	P	&<constituent#sublattice>;	10	RT*ln(mobility freq.fact.)
15	MG	T	P	&<constituent#sublattice>;	10	Magnetic mobility factor
16	G2	T	P		0	Liquid two state parameter
17	THT2	-	P		2	Smooth step function T
18	DCP2	-	P		2	Smooth step function value
19	LPX	T	P		0	Lattice param X axis
20	LPY	T	P		0	Lattice param Y axis
21	LPZ	T	P		0	Lattice param Z axis
22	LPTH	T	P		0	Lattice angle TH
23	EC11	T	P		0	Elastic const C11
24	EC12	T	P		0	Elastic const C12
25	EC44	T	P		0	Elastic const C44
26	UQT	T	P	&<constituent#sublattice>;	10	UNIQUAC residual parameter (OC)
27	RHO	T	P		0	Electric resistivity
28	VISC	T	P		0	Viscosity
29	LAMB	T	P		0	Thermal conductivity
30	HMVA	T	P		0	Enthalpy of vacancy formation (MatCalc)
31	TSCH	-	P		2	Schottky anomaly T (OC)
32	CSCH	-	P		2	Schottky anomaly Cp/R. (OC)
33	NONE	T	P		0	Unused

Appendix D Replies

If you have added notes on any pages please indicate who you are and make a summary here.

Name	Reply	Name	Reply	Name	Reply	Name	Reply
Abe		Alex		Bengt		Bosse	
Erwin		Fabio		Fan		Lina	
Malin		Nath		Reza		Richard	
Shuanglin		Ursula					