Calphad 61 (2018) 173–178

Contents lists available at ScienceDirect

Calphad

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

The Thermodynamic Database Database

Axel van de Walle[a,*], Chiraag Nataraj[a], Zi-Kui Liu[b]

[a] School of Engineering, Brown University, Providence, RI 02912, USA
[b] Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Keywords:
Database
Materials informatics
Search engine
Thermodynamic data

ABSTRACT

This paper describes an effort to organize the condensed phases thermodynamic data freely available in electronic form within the scientific literature through the design of a special-purpose search engine indexing data electronically available in the "thermodynamic database" (TDB) format. This focus is motivated by the fact that it is widely used and can be readily read or imported into most thermodynamic modeling software. This form of data also provides a rather complete thermodynamic description of a given system and thus enables researchers to generate any phase diagram cross-section of interest, a capability typically not available in traditional phase diagram handbooks. For convenience, users can quickly preview selected cross sections directly online. In designing this system, special emphasis was devoted to ensuring that the bibliographic references of the original data sources are transparently reported and to providing links to the original data sources, rather than the data itself, in order to enforce access rights. This effort was made possible by combining and building upon a number of key components, such as the CALPHAD journal's supplementary information section, the NIMS database, the NIST materials data repository, the Crossref bibliographic service, and various thermodynamic (OpenCalphad, ATAT) and graphical (gnuplot, XTK) software.

1. Introduction

While there exist various comprehensive commercial sources of thermodynamic data (e.g. [1–3]), this paper describes an effort to organize the condensed phases thermodynamic data freely available in electronic form within the scientific literature. This is accomplished via a special-purpose search engine, the Thermodynamic DataBase DataBase (TDBDB) [4], that indexes data that is electronically available in the "thermodynamic database" (TDB) format. This capability aims (i) to facilitate the development of future CALPHAD models, (ii) to enable CALPHAD-based high-throughput computational materials discovery and optimization and (iii) to provide information that is complementary to widely used phase diagram handbooks [3]. Indeed, merely reporting phase diagrams becomes less effective for higher-order systems — researchers rather need the ability to generate specific sections on demand, which demands a comprehensive thermodynamic model. Such models can be readily read or imported into standard thermodynamic software (e.g., [2,5,6,1,7–11]).

In designing this system, emphasis was put on simplicity, ensuring that access rights are followed and transparent attribution of proper credit to the authors of the original data. This effort was made possible by combining and building upon a number of key components, such as the CALPHAD journal's supplementary information section, the NIMS database [12], the NIST materials data repository [13,14], the Crossref bibliographic service [15], and various thermodynamic (OpenCalphad [8,16], ATAT [17,18]) and graphical (gnuplot [19], XTK [20,21]) software.

2. Methodology

The Thermodynamic DataBase DataBase (TDBDB) effort combines a number of existing resources in the form of data and software libraries. This section describes each aspect, summarized in Fig. 1, in more detail.

The raw data was obtained by first scanning the entire publication history of the CALPHAD journal in search of electronic supplementary information in the form of TDB files. This process also incidentally gathered databases containing some kinetic information. This gave a database of thermodynamic data somewhat biased towards the most recently assessed systems (because the systematic inclusion of electronic supplementary information is a relatively new phenomenon). To complete this database with systems that have been well-assessed for a long time, we queried the NIMS database [12], which covers a large number of binary systems as well as a few higher-order multi-component systems. We also searched the NIST materials data repository [13,14] for CALPHAD assessments. Finally, we have manually added a few well-known openly available databases, such as the

* Corresponding author.
E-mail address: avdw@alum.mit.edu (A. van de Walle).

https://doi.org/10.1016/j.calphad.2018.04.003
Received 17 January 2018; Received in revised form 30 March 2018; Accepted 3 April 2018

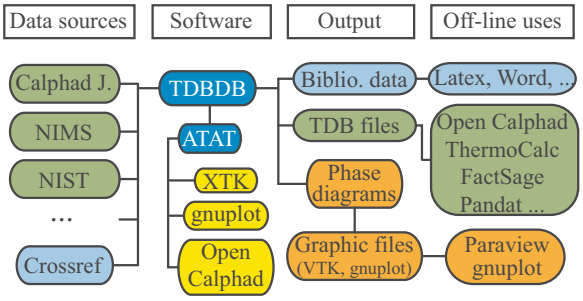

Fig. 1. Overall organization of the Thermodynamic DataBase DataBase effort.

NIST solder database [22], some open databases supplied with MatCalc [11] and some contributed TDB files (e.g. [23]). The end result is (currently) a total of 766 data entries arising from 528 publications. This database will be updated periodically by scanning new issues of the CALPHAD journal and re-scanning the NIMS and NIST databases, among other web crawling efforts.

We then matched each paper to its Digital Object Identifier (DOI), which provides a permanent pointer to the data source. This also enables a simple mechanism to generate appropriate bibliographic citations in any format via the Crossref service [15]. While the DOI assignment for CALPHAD Journal articles and NIST database entries was straightforward, for articles cited in the NIMS database, DOI matching was achieved through queries to the Crossref service. If no correct match is found (which may happen, for instance, for older papers), a standard URL (Uniform Resource Locator) is used instead.

In designing this system, we wanted to ensure that copyrights or the terms of service of the data sources were respected. For these reasons, the TDBDB website does not actually provide the content of the TDB files or the papers. It acts as a search engine and yield links to the requested resource. It is up to the users to ensure they have the credential to access the data. For data hosted in the NIMS database this simply involves creating an account on the NIMS site (which is free). For data or articles hosted in scientific journals, one needs an electronic subscription to the journal (which most researchers have via their respective institutions). As a result, the TDBDB service will most likely increase the visibility and access statistics of the data providers rather than reduce them.

The user interface is deliberately minimalist, as this provides crisper user feedback, eases code development and maintenance and makes it easier to access the website on devices with smaller screens (e.g. smart phones). Users simply enter the elements of interest and the system displays a list of databases containing data on all of the requested elements. Along with each database entry are reported links to the paper, its bibliographic reference and to the TDB file.

The TDBDB web interface also provides a quick preview of the system's phase diagram (based on the database selected). This feature is made possible by the OpenCalphad [8,16] software. As the plotting capabilities of OpenCalphad currently require some user guidance, we opted for a simple robust brute-force method to generate the phase diagram preview. A large number of randomly chosen temperature-composition coordinates are generated and the corresponding phase equilibria are calculated with OpenCalphad. The points yielding single-phase equilibria are discarded and the remaining multiphase equilibria are sorted by phases and the temperature composition data points belonging to the same multiphase equilibrium are re-connected to form the phase boundaries. For 2D sections, this re-connection algorithm reduces to a simple nearest neighbor search. In higher dimension, this amounts to a type of Delaunay triangulation [26]. The advantages of this scheme are fourfold. First, it is easily scalable from a fast low-precision representation to a slower high-precision one by merely adjusting the number of randomly chosen points. Second, it can easily capture phase fields located anywhere in the chosen section (without preference for those near axes or close to a previously known equilibrium). Third, the probability of properly "capturing" a phase is proportional to the size of the set of points in temperature-composition space that involve an equilibrium with that phase, i.e. if the algorithm "misses" a phase, it is likely to have been small anyway. Finally, this approach can be easily parallelized and the current implementation exploits that fact to deliver a faster feedback. This method does have some drawbacks. It does not currently apply to the calculation of sections that involve composition constraints (but such capabilities will be included in the future). This method is also less efficient than boundary-following techniques because each equilibrium is re-computed "from scratch" without using information from nearby equilibria, although this too could be improved in the future, since OpenCalphad has the ability to re-use previously calculated equilibria to speed up the calculations. Nevertheless, the robustness and scalability of the current method seem ideal to deliver phase diagram previews with minimal user input.

This data is then plotted and posted on the website in one of a number of ways. For binary systems, the phase diagram data is plotted using gnuplot [19]. It is rendered online using gnuplot's "canvas" terminal (see Fig. 2). One can also download the gnuplot script to generate a plot in any graphic format off-line. Ternary isothermal sections can similarly be plotted within a standard Gibbs triangle by specifying identical minimum and maximum temperatures (see Fig. 3). For ternary systems, a 3D plot (including temperature as the third axis) can be generated. The website displays a static 3D view but, by downloading the gnuplot script and running it on a local copy of gnuplot, one can interactively rotate the phase diagram off-line or generate a plot in any graphic format. For quaternary systems, only isothermal sections are implemented and are shown within a 3D Gibbs tetrahedron, again using gnuplot. While the gnuplot-generated 3D plots show the phase boundaries as discrete points (with tie lines connecting them), another option is available that provides higher-quality and interactive 3D graphics. This feature makes use of the XTK package [20,21], which enables 3D graphics via WebGL that can be interactively rotated. Phase boundaries are shown as partially transparent triangulated surfaces with tie-lines connecting them (see Fig. 4). Both Gibbs triangle and Gibbs tetrahedron (see Fig. 5) representations are available. The underlying graphic files can also be download in standard VTK

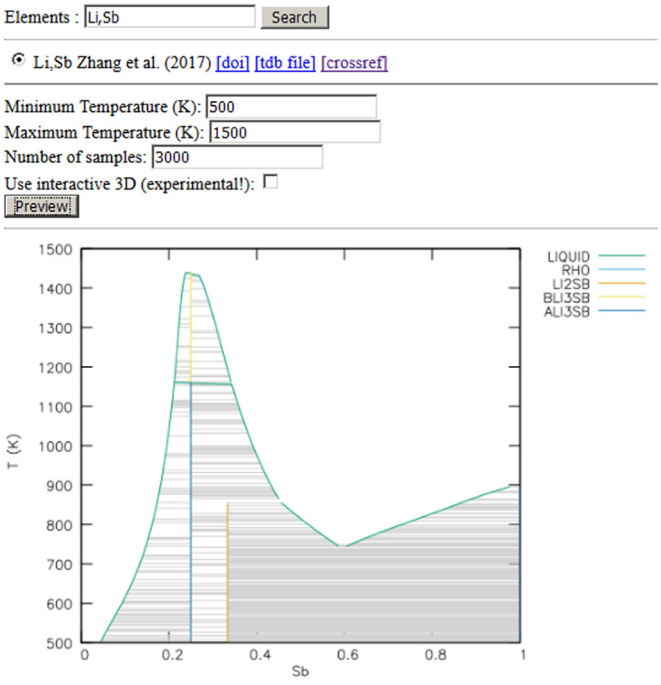

Fig. 2. Example of binary phase diagram using the data from Reference [24].