Viewpoint Paper

CALPHAD, first and second generation – Birth of the materials genome

Larry Kaufman[a] and John Ågren[b,*]

[a]*Brookline, MA, US*
[b]*Materials Science and Engineering, KTH, 100 44 Stockholm, Sweden*

Abstract—The "materials genome" was given a rather vague definition in US President Obama's announcement in 2011. We argue that the materials genome, analogously with biological genomes, should be defined as a set of information (databases) allowing prediction of a material's structure, as well as its response to processing and usage conditions. The materials genome is thus encoded in the language of CALPHAD thermodynamics and kinetics, as such databases are major parts of integrated computational materials engineering.

The aim of this report is to present the early development of CALPHAD (computer coupling of phase diagrams and thermochemistry), which eventually led to the concept of the materials genome, and to set the latter in a proper context. This should lead to a more precise definition of the materials genome, which is needed in order to identify those areas where more research is needed under the banner of the materials genome.

By the end of the last century, it had become evident to a broad group of materials engineers that computational materials science, particularly the CALPHAD method, had reached the stage where it could be used as a powerful engineering tool. The CALPHAD method is unique, in that it is neither empirical nor fundamental. Rather, it is a way to encode experimental and theoretical results in a thermodynamic language to make them applicable to a much wider context than the original experiments or calculations.

The main vision behind CALPHAD is as follows: suppose that the Gibbs energy for a system is known as a function $G(P, T, x_1...., x_n, \xi_1..., \xi_q)$, where P, T, $x_1...., x_n$ are the natural variables of G, i.e. pressure, temperature and the content of the different components, respectively, and $\xi_1..., \xi_q$ are internal variables, which may differ in character from case to case. Then a large number of quantities may be calculated from G. For

example, the state of equilibrium for given conditions, different types of phase diagrams, thermochemical properties such as heat of mixing, chemical activity and vapor pressure, and thermophysical properties such as thermal expansion and bulk modulus. Moreover, by performing calculations for general situations which do not necessarily represent equilibrium, one can calculate driving forces which may be combined with kinetic information to predict the dynamic evolution of a system.

The main challenge, then, is how to obtain the function $G(P, T, x_1...., x_n, \xi_1..., \xi_q)$. Clearly it cannot be mapped by direct experiments. Quantum mechanical abinitio calculations can only be used in extremely simplified situations, and even then the accuracy may not be satisfactory. The idea behind the CALPHAD method is based on the fact that information about quantities that may be extracted from the Gibbs energy function also indirectly tells us something about it. Based on some model, usually stemming from statistical mechanics, a mathematical expression for G that contains a number of unknown parameters is obtained. These parameters may then be adjusted until the G-function corresponding to the chosen model is capable of representing the chosen information well. Consequently, the CALPHAD procedure involves the following steps: (a) selection of the information to be represented; (b) selection of thermodynamic models, one for each phase in the system; (c) fitting the model parameters to achieve the best

* Corresponding author. E-mail: john@mse.kth.se

representation of the selected information; and (d) compilation of the parameters into a database. The whole cycle from (a) to (d) is usually referred to as a CALPHAD assessment. It should be emphasized that the selected information may be experimental measurements, theoretical calculations or estimates made in some other way. The cycle usually has to be performed several times iteratively. For example, it may be necessary to use a more advanced model for at least one of the involved phases, or it may be found that some selected information is be less reliable and should be given a lower weight in the fitting.

The theoretical basis for CALPHAD is thus thermodynamics, the formal rules and concepts of which were rigorously formulated by Gibbs in 1875 [1]. In 1908, van Laar [2] used the concepts of thermodynamic equilibrium and chemical potential, and applied what Hildebrand [3] later coined as regular solution behavior, to demonstrate that phase diagrams could be obtained from thermodynamic data by equilibrium calculations. In other words, we may say that these early workers showed that the structure of a material is encoded in its chemical composition provided that its thermodynamic language is understood. The birth of modern CALPHAD can be set at 1956, when Kaufman and Cohen [4] presented an evaluation of the Fe–Ni system based on the regular solution formalism and a combination of phase diagram information from higher temperatures and information on the martensitic transformation at low temperatures. They also compared their computed heat effects with experimental measurements and found satisfactory agreement.

The field of CALPHAD really took off with the book by Kaufman and Bernstein [5]. They used regular-solution-type models to parametrize thermodynamic information for a large number of binary and ternary alloy systems in a rather simple form. In particular, the concept of lattice stability was invented and allowed a self-consistent representation of data far beyond the dilute-solution formalism that was common in those days. All the parameters used to calculate the phase diagrams, given as functions of temperature, and all computer codes were given in appendices. However, as pointed out by Saunders and Miodownik [6] in their overview, the book by Kaufman and Bernstein was not very well received. It was criticized by both experimentalists and theoreticians. The former thought it was based on questionable data and the latter that the theoretical analysis was too crude. At that time, other methods were available to predict some aspects of phase stability from the composition of a material. Some were completely empirical, e.g. the Schaeffler diagram [7] for the structure of stainless steel weldments. In that method, the alloy elements were divided into austenite stabilizers and ferrite stabilizers, respectively, the equivalent Ni and Cr contents were determined and the structure could finally be read from a diagram. Other methods were completely theoretical and based on quantum-mechanical principles, e.g. PHACOMP [8] for the tendency to form unwanted topologically close-packed phases in Ni-base alloys. Even though the Schaeffler diagram and PHACOMP were both reasonably successful at handling the issues they were developed for and are still used

today, in their original form or after modification, they did not present the grand view given by CALPHAD and certainly lack its general applicability. In fact, the acronym CALPHAD (CALculation of PHAse Diagrams) was invented to express an alternative to PHACOMP.

Despite its poor initial reception, the book by Kaufman and Bernstein became a source of inspiration for many groups, and the annual meetings (later called CALPHAD conferences) initiated by Kaufman some years later were instrumental in the rapid expansion of the CALPHAD technique. In the early 1970s, several compilations of thermodynamic data were available, e.g. Kubaschewski et al. [9], Janaf [10], Barin and Knacke [11] and Hultgren et al. [12] tables for pure elements and binary systems. Thermodynamic measurements were also available in numerous articles. Using the CALPHAD method, consistent thermodynamic descriptions of multicomponent systems could be put together. Soon, several computer codes became generally available within the scientific community and for the general public which could be combined with the emerging CALPHAD databases to perform calculations in complex materials such as steel and semiconductors. Some of these codes were freeware, e.g. Eriksson's Solgasmix [13] and Lukas et al.'s code [14], whereas others were distributed on a commercial basis, e.g. Aspen (initially freeware for chemical simulations) [15] and FACT [16]. We refer to the situation in the 1970s as the first generation of CALPHAD, which grew up to reach some level of maturity.

The second generation of CALPHAD becomes evident in the late 1980s, with the appearance and rapid acceptance of more advanced models accounting for various ordering tendencies. These models stem from the 1970 paper by Hillert and Staffansson [17], which gave a two-sublattice version of the regular solution model. The model could be applied to ionic mixtures of two cations and two anions, or ternary systems with one element being interstitial. At the time, the model was considered as a purely formal extension of the regular solution formalism. In 1981 it was generalized to any number of components and sublattices by Sundman and Ågren [18], which was also seen as a formal excercise. However, it was later realized that the model was well suited for treating phases with complex crystallography and for various ordering phenomena (see e.g. [19]). New computer codes, such as Thermo-Calc [20] and MTDATA [21], and somewhat later PANDAT [22] – all distributed on a commercial basis – could tackle more complex problems. There were also more ambitious representations of experimental data and an increased use of quantum mechanical calculations. From a principal point of view, a very important development was the introduction of non-equilibrium phenomena to the CALPHAD scheme, ranging from the Scheil-type of solidification simulations, where only thermodynamic data are needed, to multicomponent diffusion kinetics [23]. In addition to thermodynamic data, the latter also needs diffusional mobilities, i.e. quantities that give the diffusional flux of a species under the influence of a chemical potential gradient. The commercial code DICTRA [24] became more widespread. Andersson and Ågren [23] suggested that the diffusional