J. Phase Equilib. Diffus. (2018) 39:635–649
https://doi.org/10.1007/s11669-018-0654-z

Ocean of Data: Integrating First-Principles Calculations and CALPHAD Modeling with Machine Learning

Zi-Kui Liu[1]

Submitted: 29 May 2018 / in revised form: 13 June 2018 / Published online: 2 July 2018

Abstract Thermodynamics is a science concerning the state of a system, whether it is stable, metastable or unstable, when interacting with the surroundings. Computational thermodynamics enables quantitative calculations of thermodynamic properties as a function of both external conditions and internal configurations, in terms of first and second derivatives of energy with respect to either potentials or molar quantities. Thermodynamic modeling based on the CALPHAD method enables the thermodynamics beyond stable states and is the foundation of Materials Genome and materials design. In last several decades, first-principles calculations based on density functional theory have provided invaluable thermochemical data to improve the robustness of CALPHAD modeling. Today with ever increasing computing power and large amount of data repositories online, it calls for a new paradigm for CALPHAD modeling approach incorporating machine learning to create a sustainable ecosystem for the ocean of data and for emergent behaviors.

Keywords first-principles calculation · CALPHAD · thermodynamics · thermodynamic modeling · thermodynamic stability

This invited article is part of a special issue of the *Journal of Phase Equilibria and Diffusion* in honor of Prof. Zhanpeng Jin's 80th birthday. The special issue was organized by Prof. Ji-Cheng (JC) Zhao, The Ohio State University; Dr. Qing Chen, Thermo-Calc Software AB; and Prof. Yong Du, Central South University.

✉ Zi-Kui Liu
prof.zikui.liu@psu.edu

[1] Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA

1 Introduction

Thermodynamics is a science concerning the state of a system, whether it is stable, metastable or unstable, when interacting with the surroundings. The interactions can involve exchanges of any combinations of heat, work, and mass between the system and the surroundings, dictated by the boundary conditions between the system and the surroundings. The typical work includes contributions from the external mechanic, electric and magnetic fields. The first law of thermodynamics describes those interactions, while the second law of thermodynamics governs the evolution of the state inside the system. Consequently, the combination of the first and second laws of thermodynamics provides an integration of the external and internal parts of a system. It is thus self-evident that thermodynamics in the form of combined law includes both equilibrium and non-equilibrium states of a system.[1,2]

However, in broad scientific communities, thermodynamics is usually divided into equilibrium thermodynamics and irreversible thermodynamics with the former often simply called thermodynamics and the latter largely related to non-equilibrium thermodynamics. This is understandable because Gibbs focused on systems at equilibrium,[3] while the works by Prigogine and co-workers concentrated on systems away from equilibrium.[4,5] On the other hand, thermodynamic modeling based on the CALPHAD approach[2,6–8] establishes the Gibbs energy of individual phases across the complete spaces of temperature and composition of the system covering stable and metastable regions of each phase, and the unstable regions in terms of miscibility gaps. In essence, the CALPHAD approach considers the individual phases as the building blocks of materials so that the competition among phases based on the change of energetics of individual phases with

respect to external conditions dictates phase stability, and the interactions between phases in terms of interfacial/strain/electric/magnetic energies contribute to the morphological evolution of microstructures with multiple phases, resulted in the CALPHAD approach as the foundation of computational materials design.[9–12] It was the success of the CALPHAD approach along with that of Human Genome project[13] that inspired the author to coin the term "Materials Genome®" in 2002[14–16] which was used in the Materials Genome Initiative (MGI) in 2011[17] with consensus.[15,16]

Before 2000, the CALPHAD modeling relied almost exclusively on experimental information and some theoretical estimations, and the integration with results from first-principles calculations based on density functional theory (DFT) was rather limited.[18,19] Some of the early discussions focused on the Gibbs energy differences between stable and non-stable crystal structures of pure elements under ambient pressure under the term of lattice stability.[20–22] The continued development of computation methods and software tools, particularly VASP,[23–25] have fueled the utilization of energetics from DFT-based first-principles calculations in the CALPHAD modeling. In 2009, the author reviewed the first-principles calculations and CALPHAD modeling of thermodynamics.[26] In the present paper, the fundamental thermodynamics will be briefly discussed for the basis of the present review followed by discussion on the development in last decade, focusing on data in line with the MGI goal for data sharing and analysis that aims to provide an enhanced knowledgebase to scientists and engineers in designing new materials and foster the development of computational capabilities, data management, and an integrated approach for materials science and engineering.[17]

2 Fundamentals of Thermodynamics

The combined first and second law of thermodynamics is written in terms of the internal energy change, dU, as follows[2]

$$dU = TdS - V\left(\sum_{i,j,k,l}\sigma_{ij}d\varepsilon_{kl} + \sum_i H_i dB_i + \sum_i E_i dD_i\right) + \sum \mu_i dN_i - Td_{\mathrm{ip}}S \quad \text{(Eq 1)}$$

where T, S, and V are temperature, entropy, and volume; σ_{ij} and ε_{kl} the components of stress and strain; H_i and B_i the components of magnetic field and magnetic induction; E_i and D_i the components of electric field and electric displacement; μ_i and N_i the chemical potential and moles of independent chemical component i, and $d_{ip}S$ is the entropy production due to internal processes (ip). By introducing the driving force for each internal process, the entropy production can be represented by the product of driving force, $D_{ip,i}$, and the change of corresponding internal variable, $d\xi_i$ as follows[2]

$$Td_{ip}S = \sum D_{ip,i}d\xi_i - \frac{1}{2}\sum D_{ip,ij}d\xi_i d\xi_j + \frac{1}{6}\sum D_{ip,ijk}d\xi_i d\xi_j d\xi_k \quad \text{(Eq 2)}$$

where the second and third terms are added for discussion of stability of the system when $D_{ip,i}$ vanishes, the limit of stability when both $D_{ip,i}$ and $D_{ip,ij}$ vanish, and the invariant critical point (ICP) when all $D_{ip,i}$, $D_{ip,ij}$, and $D_{ip,ijk}$ vanish.[27] It should be noted that the entropy change in Eq 1 includes both external and internal contributions, i.e.

$$dS = \frac{dQ}{T} + \sum S_i dN_i + d_{ip}S \quad \text{(Eq 3)}$$

where dQ denotes the amount of heat that the system receives from the surrounding, and S_i the partial entropy of chemical component i. As Gibbs focused on systems at equilibrium, i.e. $d_{ip}S = 0$, all internal variables become dependent variables due to the condition of $D_{ip,i} = 0$, and the combined law of thermodynamics, Eq 1, becomes

$$dU = TdS - V\left(\sum_{i,j,k,l}\sigma_{ij}d\varepsilon_{kl} + \sum_i H_i dB_i + \sum_i E_i dD_i\right) + \sum \mu_i dN_i = \sum Y^a dX^a \quad \text{(Eq 4)}$$

where Y^a represents T, σ_{ij} H_i, E_i and μ_i, called potentials, and X^a denotes S, $V\varepsilon_{kl}$, VB_i, VD_i, and N_i, called molar quantities. Each set of Y^a and X^a forms a pair of conjugate variables. At equilibrium, each potential is homogenous in the system. It is important to realize that the independent variables in Eq 4 are externally controlled, i.e. external independent variables (EIVs).[1,2] Furthermore, by switching EIVs between the conjugate molar quantities and potentials through the Legendre transformation, the combined law for Helmholtz energy, Gibbs energy, and other free energies can be obtained with the general form as follows[2]

$$d\phi = d\left(U - \sum_{c\neq a} Y^c X^c\right) = \sum_{a\neq c} Y^a dX^a - \sum_{c\neq a} X^c dY^c - Td_{ip}S \quad \text{(Eq 5)}$$

where the EIVs are the combination of X^a and Y^c. For example, for systems with temperature, stress, electric and magnetic fields, and the chemical potential of oxygen controlled from the surroundings, the combined law of a new free energy becomes

$$d\psi = d\left[U - TS + V\left(\sum_{i,j,k,l}\sigma_{ij}\varepsilon_{kl} + \sum_i H_i B_i + \sum_i E_i D_i\right) - \mu_O N_O\right]$$
$$= -SdT + V\left(\sum_{i,j,k,l}\varepsilon_{kl}d\sigma_{ij} + \sum_i B_i dH_i + \sum_i D_i dE_i\right) - N_O d\mu_O + \sum_{i\neq O}\mu_i dN_i - Td_{ip}S \quad \text{(Eq 6)}$$

Many physical properties of the system can be calculated through the second derivatives of internal energy to its independent variables in Eq 1 as shown in Table 1,[2] i.e.

$$U_{X^aX^b} = \frac{\partial^2 U}{\partial X^a \partial X^b} = \frac{\partial Y^b}{\partial X^a} = \frac{\partial Y^a}{\partial X^b} \quad \text{(Eq 7)}$$

The diagonal quantities in the table are the second derivatives with respect to the same molar quantity, i.e. $X^a = X^b$ in Eq 7, and the off-diagonal quantities with respect to two different molar quantities, i.e. $X^a \neq X^b$ in Eq 7. The properties in the first four columns/rows of the table are for a closed system with fixed composition, and the properties in the last row/column are related to mass transport under gradient of a potential. There is another set of second derivatives not shown in the table that are with respect to one potential and one molar quantity from Eq 5, i.e.

$$\phi_{X^aY^c} = \frac{\partial^2 \phi}{\partial X^a \partial Y^c} = \frac{\partial Y^a}{\partial Y^c} = \frac{\partial X^c}{\partial X^a} \quad \text{(Eq 8)}$$

Such as the second derivatives of Gibbs energy with respect to N_i and T are related to thermotransport properties[28,29] or with respect to electron (N_e) and T are related to thermoelectric properties,[30,31] as follows,

$$G_{TN_i} = \frac{\partial^2 G}{\partial T \partial N_i} = \frac{\partial \mu_i}{\partial T} = \frac{\partial S}{\partial N_i} = S_i \quad \text{(Eq 9)}$$

$$G_{TN_e} = \frac{\partial^2 G}{\partial T \partial N_e} = \frac{\partial \mu_e}{\partial T} = \frac{\partial S}{\partial N_e} = S_e \quad \text{(Eq 10)}$$

which are partial entropy of the component (S_i) and electron (S_e), respectively, and will be discussed in the future.

For a homogeneous system to be stable, the diagonal quantities must be positive. When the homogeneous system approaches its limit of stability, i.e. both $D_{ip,i}$ and $D_{ip,ij}$ in Eq 2 vanish, the inverses of the diagonal quantities approaches zero,[2] and the diagonal quantities diverge to positive infinity. The table is symmetric due to the equality of mixed partial second derivatives, but the signs of the off-diagonal quantities are not directly prescribed by any laws and can be indeed either positive or negative depending on the external conditions such as the negative thermal expansion or thermal contraction.[32,33] Even though

Table 1 Physical quantities related to the second directives of internal energy

	T	σ_{kl}	E_k	H_k	μ_k
S	C/T, heat capacity	α_{kl}, piezocaloric effect	p_k, electrocaloric effect	m_k, magnetocaloric effect	$\frac{\partial \mu_k}{\partial S}$
ε_{ij}	α_{ij}, thermal expansion	s_{ijkl}, elastic compliance	d_{ijk}, converse piezoelectricity	q_{ijk}, piezomagnetic moduli	$\frac{\partial \mu_k}{\partial \varepsilon_{ij}}$
D_i	p_i, pyroelectric coefficients	d_{ikl}, piezoelectric moduli	k_{ik}, permittivities	a_{ik}, magnetoelectric coefficient	$\frac{\partial \mu_k}{\partial D_i}$
B_i	m_i, pyromagnetic coefficient	q_{ikl}, piezomagnetic moduli	a_{ik}, magnetoelectric coefficient	μ_{ik}, permeability	$\frac{\partial \mu_k}{\partial B_i}$
N_i	$\frac{\partial N_i}{\partial T}$	$\frac{\partial N_i}{\partial \sigma_{kl}}$	$\frac{\partial N_i}{\partial E_k}$	$\frac{\partial N_i}{\partial H_k}$	Thermodynamic factor

negative thermal expansion was mentioned in the review paper by the author in 200[9,26] its significance in thermodynamic modeling was realized later[32] and will be discussed in the present paper in terms of in-depth discussion of building blocks for emergent behaviors beyond individual phases in CALPHAD modeling.

For systems not at equilibrium, some of the potentials are not homogeneous, and the difference of a potential gives the driving force for its conjugate molar quantity to change from one state to another state through internal processes until the potential difference is eliminated. The internal variables used to represent the states of the system thus become independent variables of the system. In order to evaluate the driving forces for the changes of internal independent variables (IIV), i.e. the derivatives of the system energy with respect to each IIV, the system energy needs to be modeled as the function of both IIVs in Eq 2 (ξ_i) and EIVs in Eq 5. It is important to realize that the energy landscape of a system can be very complex with many IIVs. When a system is brought into one of unstable regions, some internal processes occur spontaneously and homogeneously without barriers, and the system can decompose into a wide range of self-organized structures depending on the interactions among various internal processes and how the system is constrained from the surroundings. This includes the widely studied spinodal decompositions[34–36] and many dissipative structures and oscillatory behaviors discussed by Kondepudi and Prigogine.[5]

3 First-Principles Calculations

In the 2009 paper by the author,[26] the fundamentals of first-principles calculations approaches were discussed. The central theme of first-principles calculates is to obtain the electronic structures of a state with IIVs either prescribed or from the various degrees of relaxations, from which the total energy and a wide range of properties can be computed. These include their temperature dependences through the quasiharmonic approximation even though DFT-based first-principles calculations are primarily for ground states at zero Kelvin. Therefore, thermodynamic properties of stable and metastable structures can be reasonably well predicted from first-principles calculations with some limitations discussed in the literature,[37] while the thermodynamic properties of unstable states are more challenge and will be discussed in a separate section except the lattice stability for CALPHAD modeling, i.e. the Gibbs energy difference between the stable structure and other structures of pure elements central to the CALPHAD modeling of the Gibbs energy of mixing,[20,38] which is briefly reviewed in this section.

The properties that we have worked on are as follows for stoichiometric compounds and solution phases with the latter mainly based on the special quasirandom structures:[39–41]

- Thermal properties: phonon spectrum,[42,43] heat capacity, enthalpy, entropy, free energy,[44] thermal expansion[45];
- Transport properties: diffusion coefficient,[46,47] Seebeck coefficients[30,31];
- Interfacial properties: stacking fault energy,[48,49] anti-phase boundary energy,[50,51] grain boundary and interfacial energy[52,53];
- Mechanical properties: elastic moduli,[54,55] dislocation mobility,[56] relative creep rate.[57,58]

It is worth noting that we developed a mixed-space approach using the accurate force constants calculated by the direct approach in real space and the dipole–dipole interactions calculated by linear response theory in reciprocal space to predict the phonon spectrum for polar materials, particularly the splitting between longitudinal and transverse optical phonon frequencies in polar materials.[43,59,60] Furthermore, based on the effects of perturbations of an individual atom or a group of atoms on the electron charge density, we showed a first-principles approach to determine unambiguously the amount of electron charge associated with a particular atom or a group of atoms.[61]

One advancement significant to CALPHAD modeling is the prediction of lattice stability. The approach to predict the lattice stability of structures unstable at zero Kelvin based on ab initio molecular dynamics (AIMD) simulations proposed by Ozolins for fcc and bcc W[62] was applied to bcc-Ti with the entropy, enthalpy, and Helmholtz energy differences between bcc- and hcp-Ti shown in Fig. 1,[63] together with the experimental data compiled by Chase.[64] The bcc–hcp transition temperature is predicted to be about 760 K, lower than the experimental result T=1155 K, though much better than that calculated from Debye model. Another significant development in this aspect is to consider the limit of stability as the effective lattice stability used in the CALPHAD method as demonstrated by van de Walle and co-workers.[65,66] As discussed above, the limit of stability is with $D_{ip,ij}$ in Eq 2 being zero, commonly referred as inflection points. This simple and efficient method is grounded in formal statistical mechanics and exhibits the desirable smooth extrapolation behavior between stable, metastable, and unstable regions. Extensive testing on pure elements, binary and ternary systems with three common lattices (fcc, bcc and hcp) confirms the predicted smooth behavior and shows very good agreement

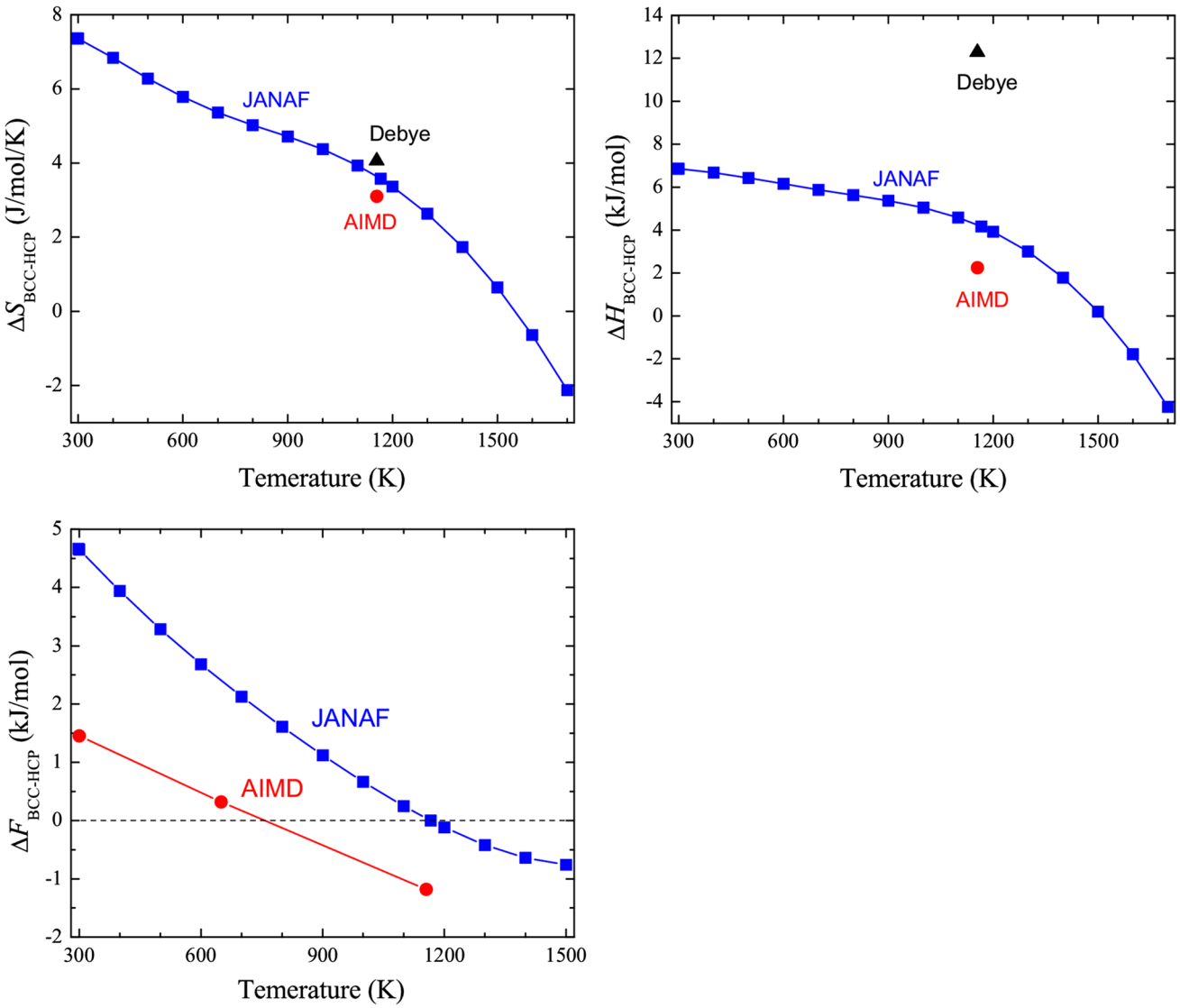

Fig. 1 Entropy, enthalpy and Helmholtz energy differences between BCC and HCP Ti predicted by AIMD simulations with thermodynamic integration and their transition temperature of 760 K[63]

with available experimental and computational estimates.[65,66] This new approach has the potential to be extended to estimate the free energy of end-members of compound energy formalisms[67] commonly used in CALPHAD modeling.

It is worth mentioning that AIMD simulations combine the strengths of atomic forces from DFT-based first-principles calculations and the extensive statistical sampling from molecular dynamic (MD) simulations.[68] In addition to predict atomic structures,[69–71] diffusion coefficients in both liquid and amorphous,[72,73] and shear viscosity,[74] the AIMD simulations have been used to evaluate the entropy in the liquid phase,[75] melting temperatures,[74,76–78] and the dynamics of atomic structures at high temperatures,[79] which is worth of further discussion. In Ref 79 the atomic structure and polarization evolution of $PbTiO_3$ as a function of temperature and time were investigated by AIMD simulations. Conventionally, the general structural features in MD simulations are characterized through the pair correlation function (PCF) or the triplet correlation function (TCF). However, the PCF and TCF are derived through statistical average over the entire AIMD ensemble and time steps and cannot provide the specific information on instantaneous local atomic distortion and polarization. In the AIMD study of PbTiO[3,79] the atomic configurations in $PbTiO_3$ are characterized as a function of time in terms of Ti-O bond lengths in the nearest-neighboring shell. It was observed that the AIMD predicted lattice parameters are in good agreement with those determined from the XAFS experiments, whereas significantly different from those measured by x-ray or neutron diffraction, especially at the temperatures above ferroelectric–paraelectric (FE-PE) phase transition

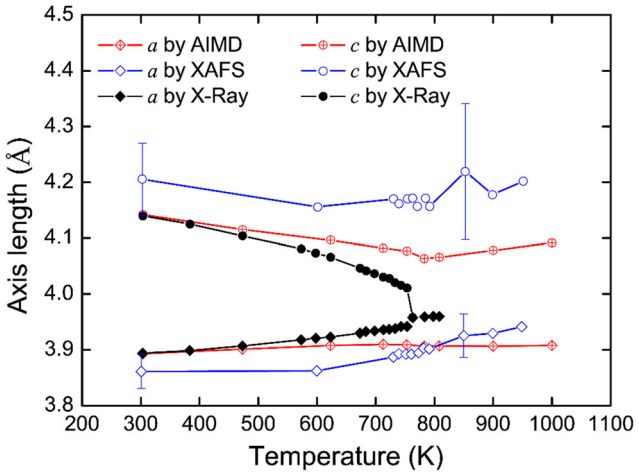

Fig. 2 Temperature dependence of the lattice parameters a, c of $PbTiO_3$ unit cell.[79] The data in crossed symbols are derived from current AIMD simulations, the open symbols are from the EXAFS measurements by Sicron et al.[113] and the closed symbols are from the x-ray diffraction by Shirane et al.[114]

temperature of 763 K as seen in Fig. 2. By following the three lattice parameters of each $PbTiO_3$ unit cell as a function of AIMD simulation time, it is found that the unit cells switch the directions of their longest axis as demonstrated in Fig. 3. During the axis switch, certain amount of unit cells is in the cubic configuration, and this amount increases with temperature due to the increase of the switching frequency (Fig. 3) and becomes saturated at a value slightly higher than 50% above the FE-PE phase transition temperature (Fig. 4). Figure 4 also depicts that the majority of tetragonal unit cells are polarized. Since AIMD simulations automatically take into account the anharmonicity at finite temperatures, it is anticipated that their applications will become more and more significant with continued increase of computing power.

The continued increase of computing power is also reflected by the establishments of a number of large scale online databases from DFT-based first-principles calculations with Materials Project,[80,81] OQMD,[82,83] and AFLOW[84] being the major ones in the United States. These databases cover many known stable and metastable binary and ternary stoichiometric compounds and are widely used by the community. Most energetic data are for values at zero Kelvin with some values at finite temperatures in AFLOW and recently added in Materials Project. Furthermore, more and more functionalities are added such as calculations of phase diagrams and Pourbaix diagrams in Materials Project.[80,81] It is clear that not only these databases will continue to have more data, but also more databases will become available online.

For new phases with unknown crystal structure and atomic positions, we have been testing the two approaches: the "Structure Predictor" code in *pymatgen* (Materials Project)[80,85] based on the data-mined knowledge of experimental crystal data (i.e., the **ICSD** data),[86] and the USPEX (Universal Structure Predictor: Evolutionary Xtallography)[87,88] using evolutionary variable-cell simulations as shown in another publication in this special issue by Jiang and Liu.

4 CALPHAD Modeling

In the 2009 paper by the author,[26] the fundamentals of CALPHAD modeling were discussed. In the last decade, CALPHAD modeling has demonstrated its importance in the Integrated

Computational Materials Engineering (ICME)[89] and MGI.[17] The outcome of CALPHAD modeling is the Gibbs energy functions of individual phases, and their collections are stored in various thermodynamic database (TDB) files. In early publications, those functions were tabulated and often with many typos, which greatly hindered their direct usages by the readers. Since 2009, the CALPHAD journal[90] has had all its publications with the TDB files of binary and ternary systems attached as supplementary document of the publication. More recently, a special-purpose search engine, the Thermodynamic DataBase DataBase (TDBDB), was developed to index TDB data electronically.[91] 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. On the TDBDB web interface users enter the elements of interest, and the system displays a list of databases containing data on all of the requested elements with links to the paper, its bibliographic reference and to the TDB file. It also provides a quick preview of the system's phase diagram using OpenCalphad software.[92,93]

The development of phase-based property data using the CALPHAD method has become a central component of MGI data infrastructure.[94,95] In addition to the expansion of CALPHAD modeling to more properties such as the elastic properties,[96] new models or new experimental and computational information demand the remodeling of low-order (unary, binary, and ternary) systems to develop the new multicomponent descriptions. Therefore, the TDB files alone are not enough, and the all input data used in

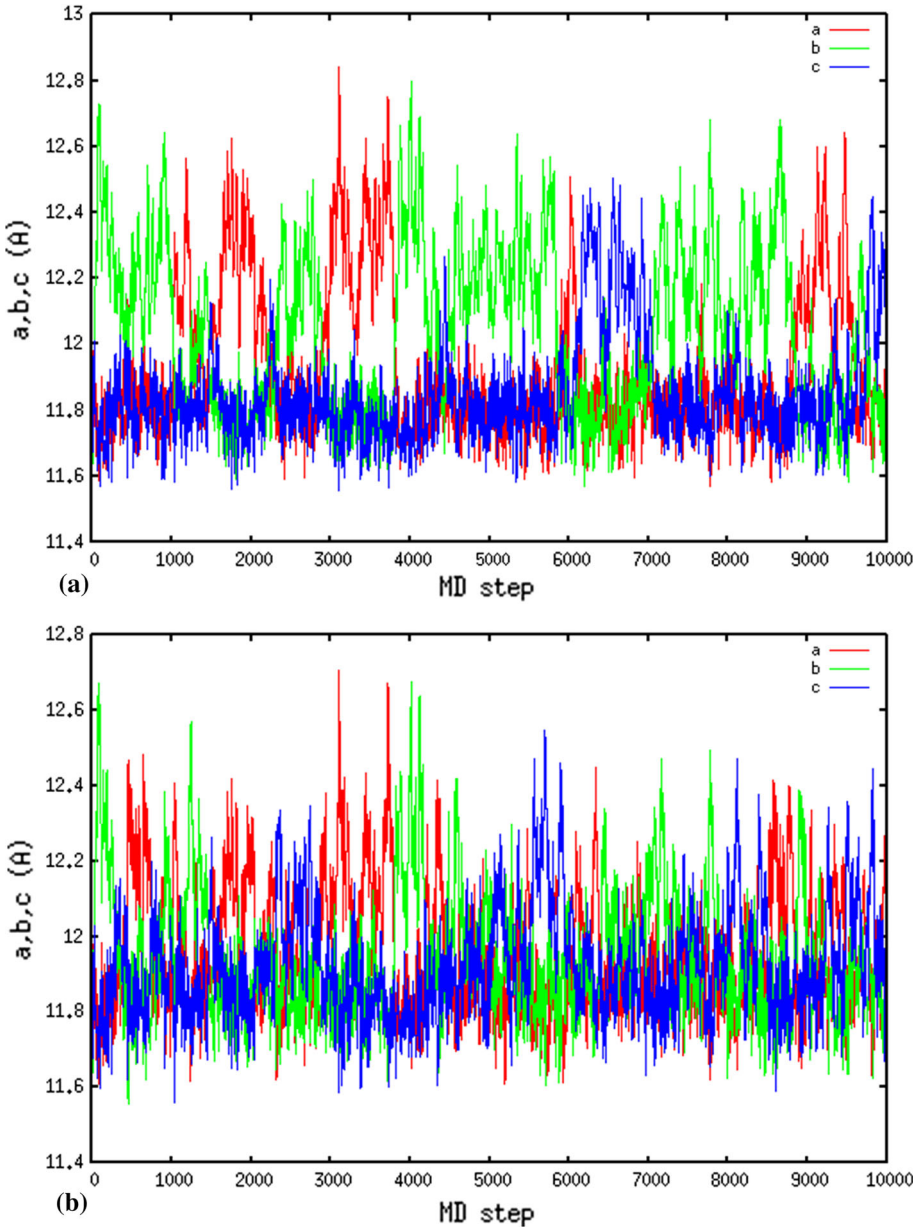

Fig. 3 Instantaneous lattice parameters of $PbTiO_3$ measured in all three directions as a function of AIMD simulation steps, (a) 623 K, and (b) 753 K

CALPHAD modeling are needed. Consequently, ESPEI (extensible, self-optimizing phase equilibrium infrastructure) was developed between 2008 and 2010.[97] ESPEI integrates databases of input data for CALPHAD modeling (crystallographic, phase equilibrium, thermochemical, and modeled Gibbs energy data, etc.) and the development of TDB model parameters (automation of thermodynamic modeling) with GUI (graphical user interface) designed mainly by Microsoft C# and SQL (structured query language). In ESPEI, a two-step automated procedure determines the model parameters of individual phases, first exclusively using thermochemical data from DFT-based first-principles calculations, and then using experimental phase equilibrium data. It used Thermo-Calc[98] for phase equilibrium calculations.

Shortly after that, an open source code for CALPHAD modeling, OpenCalphad,[92,93] became available. At the same time, we developed a Python-based open source code called pycalphad[99,100] for designing thermodynamic models, calculating phase diagrams and investigating phase equilibria using the CALPHAD method. Pycalphad provides routines for reading thermodynamic databases and solving the multi-component, multi-phase Gibbs energy minimization problem. The pycalphad software project advances the state of thermodynamic modeling by providing a flexible yet powerful interface for manipulating CALPHAD data and models. One key feature of pycalphad is that the thermodynamic models of individual phases can be programmatically manipulated and overridden at run-time without modifying any internal solver or

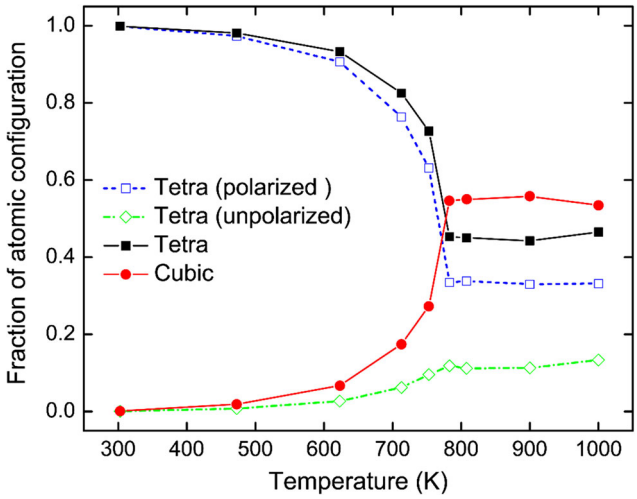

Fig. 4 Fractions of the cubic (closed circles) and tetragonal (closed squares) configurations as a function of temperature, obtained from the AIMD simulations.[79] Among the tetragonal configurations, the fractions of the polarized and unpolarized ones in the <001> direction are shown in the open squares and open diamonds, respectively

calculation code. Because the models are internally decoupled from the equilibrium solver and the models themselves are represented symbolically, pycalphad is an ideal tool for CALPHAD database development and model prototyping and is used to further re-develop ESPEI for high throughput CALPHAD modeling with new capability for uncertainty quantification.[101,102] The quantification of uncertainty reinforces the notion of CALPHAD as the premier method for integrating heterogeneous thermophysical data into self-consistent models as part of an ICME design pipeline.[102]

5 Sustainable Ecosystem of Data

In analogy to the nature, one may consider various databases above as individual ponds, lakes, and reservoirs as a part of a large ecosystem of data. To make this ecosystem of data sustainable, there need to be mechanisms like streams, creeks, and rivers in the nature to overcome the barriers and transport the data in individual data repositories into oceans of data and circulate them back to the individual data repositories. It is evident that there are driving forces for such transportations and circulations because the ocean of data possesses information and properties that the simple summation of individual data repositories do not have and can generate new products that are useful to individual data repositories. The challenges lie in the barriers between them and how to put values back to the individual data repositories. It is our hope that the newly developed open-source ESPEI has the potential to enable the development of pathways with lower barriers between individual data repositories and the ocean of data and add values to data processing in the individual data repositories for their specific applications through communicating with the ocean of data as schematically shown in Fig. 5.

As mentioned above, the re-developed open-source ESPEI[102] uses the open-source pycalphad[100] as its computing engine with the model module decoupled from the solver. This is significant as it enables users to not only investigate various models for one property, but also develop models for properties of their own interests such as those listed in Table 1 and those discussed in section 3 with minimal modifications of the solver. In accordance with the original idea, the high throughput ESPEI approach relies exclusively on the complete prediction of properties of individual phases as a function of EIV and IIV from DFT-based first-principles calculations. It is thus evident that in addition to solid phases, the DFT-based prediction of properties of liquid discussed in section 3 in terms of AIMD simulations is critical or they must be estimated or obtained from experiments.

ESPEI then evaluates the values of model parameters with the number of model parameters determined using the corrected Akaike information criterion (AICc)[103–105] as the scoring method to strike a balance between maximum model predictiveness and minimal model complexity. The statistical distribution of values of model parameters is then investigated in term of a regularization scheme with the prior probability distribution updated to the posterior probability distribution through likelihood (information) and evidence (normalizing quantity) and the posterior probability distribution determined by a stochastic Bayesian optimization approach based on Markov Chain Monte Carlo (MCMC) sampling method.[106] However, classical MCMC requires that model parameters are uncorrelated, while in the CALPHAD method, the model parameters in a phase are correlated. Another challenge is that traditional Metropolis–Hastings MCMC algorithms require determining a prior distribution for each parameter to sample from to construct the posterior. ESPEI solves these potential problems in terms of ensemble samplers, as introduced by Goodman and Weare,[107] that use an ensemble of Markov chains in parallel. These ensemble samplers have the property of affine invariance, where new parameter proposals are scaled by linear transforms depending on the current state of all the parameters in the parallel chains. The Metropolis criteria accept all proposed parameters that improve the likelihood, but sometimes accept proposals that decrease the likelihood. Adding parameters that decrease the likelihood to the Markov chain systematically leads to the uncertainty quantification of each parameter once the Markov chains converge. ESPEI uses an ensemble sampler algorithm in the emcee

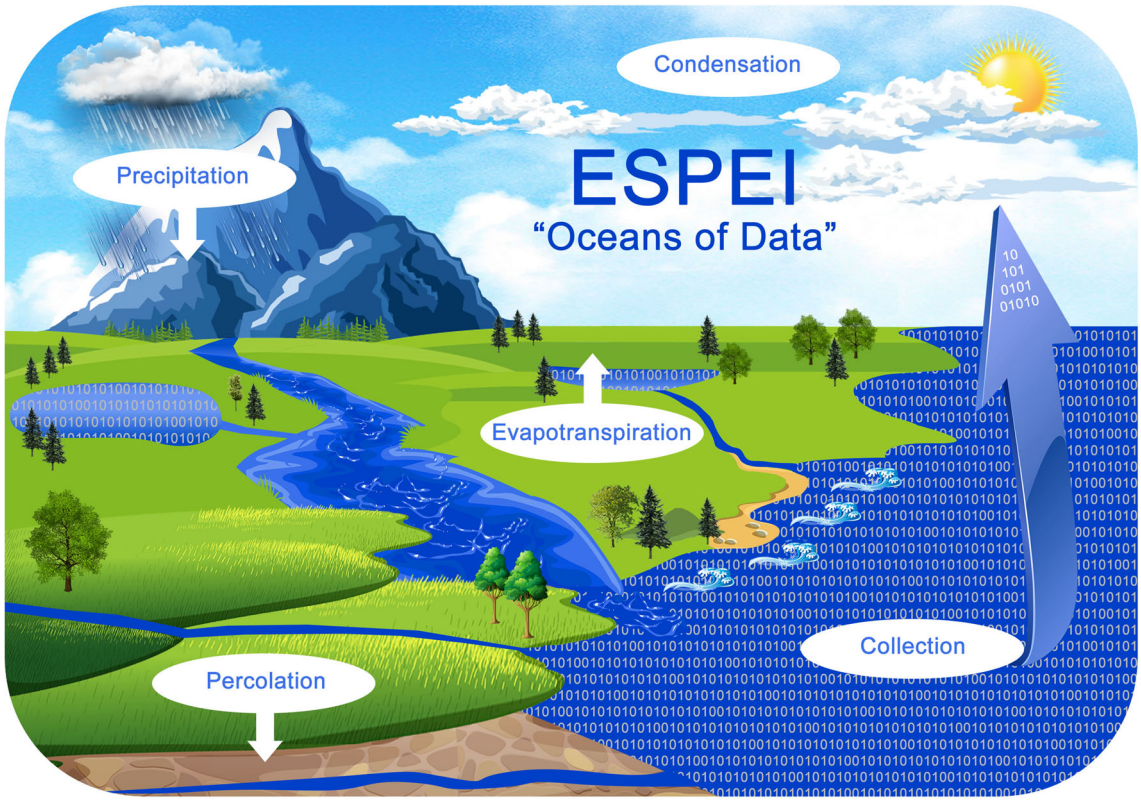

Fig. 5 Schematic diagram of sustainable ecosystem of data showing various data repositories (lakes), interconnections (flows), private data (percolation), processing (evaporation), collection (ocean), and reuse (condensation and precipitation), drawn with clipart from https://www.kisspng.com

package[108] that implements parallelizable ensemble samplers. To use emcee, ESPEI defines the initial ensemble of chains as Gaussian distributions centered on the parameters generated by single phase data in the first step and provides emcee with a function to calculate the total error due to various types of data. This machine learning process automatically determines the optimal number of model parameters, their numerical values, their correlations, and their statistic uncertainties.

Since the uncertainties of computed energetic data from first-principles calculations are in the order of several kJ/mole-atom,[109] while the accurate predictions of phase relations need an accuracy of 0.1 kJ/mole-atom or better. Therefore, the model parameters evaluated based solely on data from first-principles calculations must be refined using key experimental phase equilibrium data, which is one of the important features of the CALPHAD modeling. In ESPEI, this refinement is also carried out by the MCMC sampling method as the second step of ESPEI automation. As an example the CALPHAD modeling of the Cu-Mg binary system using ESPEI is available online.[105] The CALPHAD modeling of ternary systems will be posted online accordingly. The ESPEI modeling of diffusion coefficients and elastic properties are also being considered. The current uncertainty quantification capability in ESPEI includes the uncertainty of model parameters and model predictions, and future development will be extended to uncertainty of models themselves through machine learning. Furthermore, the inconsistencies between experimental data or between experimental and DFT data is being treated through the K-fold cross validation where one or multiple datasets are excluded from the data used to evaluate the model parameters, the confidence intervals or credible intervals for the frequentist or Bayesian approach are evaluated for all combinations of data to determine outliers by their distance from the confidence/credible interval threshold.

It is our goal that ESPEI will contributes to the establishment of the ocean of data and the development of property databases of multi-component materials with multiple defects on the fly. Its open source architecture can enable the independent creation of models for new properties that have not modeled before. As pointed out by Xiong and Olson and shown in Fig. 6,[110] all current computational materials design approaches reside on the processed data which in turn depend on the proto data with ESPEI playing essential roles in both data, i.e. as if the two sets of data breed all other properties, in line with the

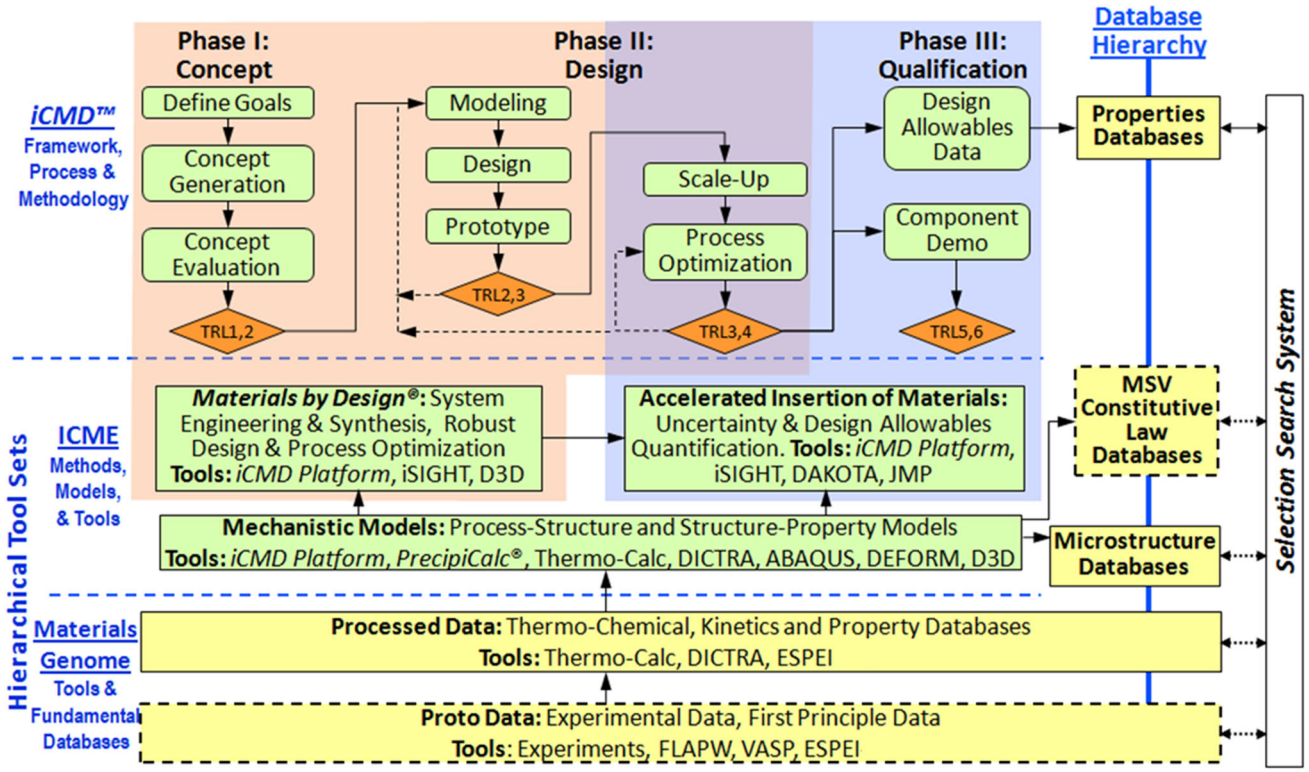

Fig. 6 Overall hierarchical architecture of methods, tools, techniques, and databases for applications of ICME methods based on MGI, showing all of them floating on the "ocean of data" with ESPEI as a key component[110]

concept of "ocean of data" illustrated in Fig. 5. In proto data, ESPEI is linked with the high throughput DFT-based first-principles calculations using the newly revised open-source program atomate,[111] which is also the backbone of Materials Project data repository.[80] In the processed data, ESPEI is used to efficiently and robustly evaluate model parameters and can be extended to other properties needed for computational materials design as discussed above. Furthermore, ESPEI can feed back to various data repositories to guide the needs of new DFT-based first-principles calculations and experimental measurements. In the above, it was mentioned that phase equilibrium data are critical in refining the model parameters. It is anticipated that the high throughput experiments will become more and more desired, such as the diffusion multiple approaches that have been developed by Zhao.[112]

One of our own future modeling efforts is the emergent behaviors of materials, i.e. commonly referred as property anomaly. As discussed in section 2, molar quantities, i.e. the properties of the system, diverge when the system reaches its limit of stability. Usually the state of a system changes from stable to metastable before it reaches the limit of stability to become unstable except at the critical point where stable/metastable/unstable states merge together and all $D_{ip,i}$, $D_{ip,ij}$, $D_{ip,ijk}$, and higher order derivatives in Eq 2 vanish. It is natural to realize that molar quantities would gradually deviate from normal behavior before they diverge as the EIVs change. This is indeed demonstrated in our recent work on thermal expansion anomaly for both Invar (Fe_3Pt) and anti-Invar (Ce) materials.[32] We consider that every system consists of many possible states, and a macroscopically homogeneous single phase is a statistical mixture of all states. When one state is much stable than all other states, the system is in a single-phase state. When several states are in equilibrium with each other, the system is in the state of multiple phase equilibrium. However, when the energy differences between stable and metastable states are comparable with the thermal energy and the penalty for their co-existence due to additional interfacial/strain/electric/magnetic energies is relative small, the metastable states may thus exist statistically due to the thermal fluctuation, similar to the cubic and tetragonal states in $PbTiO_3$ discussed in section 3. It should be noted that this scenario is very different from the state of multiple phase equilibrium mentioned above where each phase exists statically, while in the thermal fluctuation, the metastable states are not in thermodynamic equilibrium with the stable state, and their existence is due to the statistical fluctuations between all states resulted from the thermal energy, therefore statistical rather than static.

This is shown in Fig. 7 where the Helmholtz energies of 12 atom supercells of Fe_3Pt with 512 magnetic spin

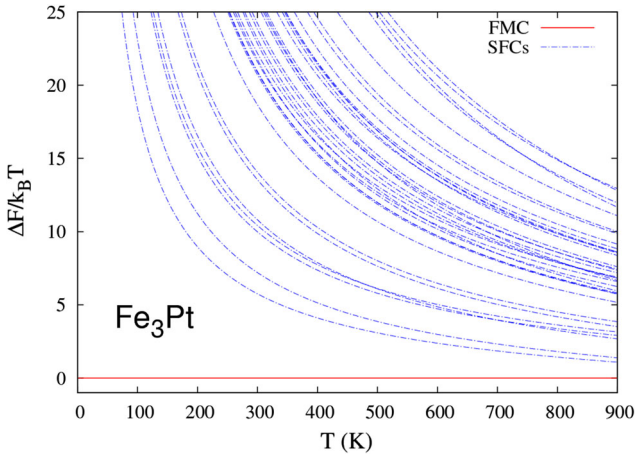

Fig. 7 Helmholtz energies of various spin flipping magnetic states (SPC) of 12 atom supercell Fe_3Pt as a function of temperature with respect to that of ferromagnetic state (FMC)[115]

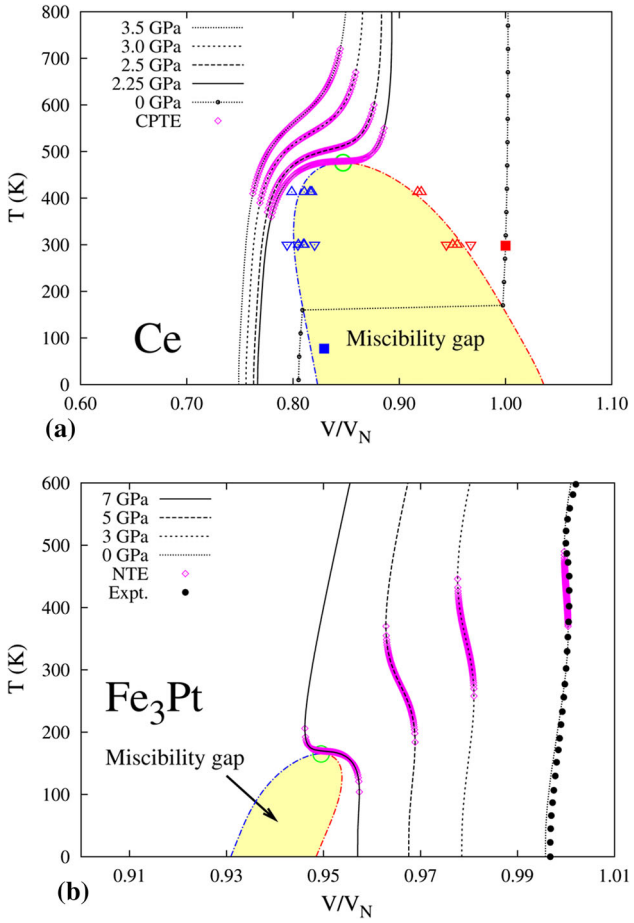

Fig. 8 Predicted temperature–volume phase diagrams with the miscibility gap (yellow shaded area) and critical point (green circle) marked,[32] (a) cerium and (b) Fe3P. Additional isobaric volumes are plotted for several pressures with available experimental data superimposed (Color figure online)

configurations with 37 being unique are plotted. It can be seen that from the energetic point of view, the ferromagnetic state has the lowest energy at all temperatures considered, but the energy differences between the ferromagnetic state and other states decreases with the increase of temperature, meaning that other states have higher entropy than the ferromagnetic state. Consequently, the competition between the states intensifies with the increase of probability of metastable states with temperature, adding additional entropy to the system. This competition reaches the maximum at the critical point, and the changes of all molar quantities of the system with respect to temperature become infinite. It is important to point out that the change of the conjugate molar quantity of temperature, i.e. entropy, becomes positive infinity as discussed in section 2, while the changes of any none conjugate molar quantities, such as volume, can either be positive or negative infinity depending on whether the volumes of the metastable states are larger or smaller than that of the stable state.

The predicted temperature-volume phase diagrams for Ce and Fe_3Pt are plotted in Fig. 8, respectively. In Ce, the stable, ground state is non-magnetic with smaller volume, and the metastable states are magnetic with larger volumes, resulting in infinite volume increase with respect to temperature at the critical point. While in Fe_3Pt, the stable state is ferromagnetic with the largest volume, and all other spin flipping magnetic states have smaller volumes, resulting in infinite volume decrease with respect to temperature at the critical point. It is noted that the anomalies, either increase or decrease, occurs before the critical point is reached, and happens at different temperature ranges under different pressures away from the pressure at the critical point, i.e. higher than 2.25 GPa for Ce and lower than 7 GPa for Fe_3Pt. Outside the anomaly temperature range, the competition between states decays, and the behavior of molar quantity returns back to normal even the system contains both stable and metastable states. It is evident that this emergent behavior of the system cannot be predicted by the properties of individual states alone, and can only be predicted and controlled by understanding the properties of individual states and their relative dependences on EIVs and IIVs. In the cases of Ce and Fe_3Pt, the main contribution to the emergent behavior is due to the additional entropy originated from the probabilities of metastable states, and the interfacial energy and strain energy are taken into account by the magnetic domain walls inside the supercells of individual states and their volume constrains during energy minimization. It is the author's perspective that future CALPHAD databases may be built upon the individual states beyond the current individual phases, and in this aspect the DFT-based first-principles calculations will play an even more significant role in building up the ocean of data.[15,16]

6 Summary

CALPHAD modeling of properties of individual phases as a function of external and internal variables of phases builds the foundation of computational materials design. Over the last two decades, the energetics from DFT-based first-principles calculations have contributed significantly to the enhancement of CALPHAD modeling in terms of both quality and quantity. The continued increase of computing power, the number of data repositories, and the amount and type of data in each data repository calls for the development of a data infrastructure that can not only generate the proto data, but can also model the proto data to produce the processed data for the downstream applications, both efficiently and accurately. The recent re-developed ESPEI open-source code is briefly presented, which makes use of the modern data analytics in terms of Markov Chain Monte Carlo simulations with a corrected Akaike information criterion and ensemble samplers of Markov chains. It seems that ESPEI can help to build an efficient ecosystem integrating the proto and processed data by creating the ocean of data. Furthermore, it is pointed out that CALPHAD modeling needs to go beyond individual phases if the emergent behaviors of materials properties to be modeled and predicted to include the states that are not observable individually under experimental conditions and using statistical mechanics.

Acknowledgments The research results reported in the paper are accumulated from the author' publications over the years, as listed in the cited references with financial supports acknowledged in them, including the National Science Foundation (NSF), the Department of Energy, Army Research Lab, Office of Naval Research, Wright Patterson AirForce Base, NASA Jet Propulsion Laboratory, and the National Institute of Standards and Technology, in addition to a range of national laboratories and companies that supported the NSF Center for Computational Materials Design, and the LION clusters at the Pennsylvania State University, the resources of NERSC supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and the resources of XSEDE supported by NSF with Grant ACI-1053575. Particularly, pycalphad and ESPEI have been supported by a NASA Space Technology Research Fellowship under Grant No. NNX14AL43H and NSF Research Traineeship Program (CoMET: Computational Materials Education) under Grant No. 1449785. The author would like to thanks Dr. Richard Otis and Mr. Brandon Bocklund for comments of the manuscript and Patricia Lee Craig at Penn State for drawing Fig. 5.

References

1. M. Hillert, *Phase Equilibria, Phase Diagrams and Phase Transformations: Their Thermodynamic Basis*, Cambridge University Press, Cambridge, 2008
2. Z.K. Liu and Y. Wang, *Computational Thermodynamics of Materials*, Cambridge University Press, Cambridge, 2016
3. J.W. Gibbs, *The Collected Works of J. Willard Gibbs*, Yale University Press, New Haven, 1948
4. I. Prigogine, *Thermodynamics of Irreversible Processes*, Interscience, New York, 1961
5. D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures*, Wiley, Hoboken, 1998
6. L. Kaufman and H. Bernstein, *Computer Calculation of Phase Diagrams. With Special Reference to Refractory Metals*, Academic Press, London, 1970
7. N. Saunders and A.P. Miodownik, *CALPHAD (Calculation of Phase Diagrams): A Comprehensive Guide*, Vol 1, Pergamon, New York, 1998
8. H.L. Lukas, S.G. Fries, and B. Sundman, *Computational Thermodynamics: The CALPHAD Method*, Vol 131, Cambridge University Press, Cambridge, 2007
9. G.B. Olson, Computational Design of Hierarchically Structured Materials, *Science*, 1997, **277**, p 1237–1242
10. L. Kaufman, Computational Thermodynamics and Materials Design, *CALPHAD*, 2001, **25**, p 141–161
11. G.B. Olson and C.J. Kuehmann, Materials Genomics: From CALPHAD to Flight, *Scr. Mater.*, 2014, **70**, p 25–30
12. L. Kaufman and J. Agren, CALPHAD, First and Second Generation: Birth of the Materials Genome, *Scr. Mater.*, 2014, **70**, p 3–6
13. Human Genome Project (2012), Available at https://www.genome.gov/10001772/all-about-the–human-genome-project-hgp/. Accessed 26 June 2018
14. The Term of Materials Genome Coined in 2002 and Trademarked in 2004/2012 (78512752/85271561) by MaterialsGenome, Inc. http://www.materialsgenome.com, Available at http://tess2.uspto.gov. Accessed 26 June 2018
15. 刘梓葵. 关于材料基因组的基本观点及展望. *Chin. Sci. Bull.*, 2013, **58**, p 3618–3622
16. Z.K. Liu, Perspective on Materials Genome®, *Chin. Sci. Bull.*, 2014, **59**, p 1619–1623
17. National Science and Technology Council. Materials Genome Initiative for Global Competitiveness (2011), Available at https://www.mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf. Accessed 22 Mar 2017
18. P.J. Spencer, Computational Thermochemistry: From its Early CALPHAD Days to a Cost-Effective Role in Materials Development and Processing, *CALPHAD*, 2001, **25**, p 163
19. P.J. Spencer, A brief history of CALPHAD, *CALPHAD*, 2008, **32**, p 1–8
20. L. Kaufman, The Lattice Stability of the Transition Metal, in *Phase Stability in Metals and Alloys,* eds. P.S. Rudman, J.S. Stringer, R.I. Jaffee) (McGraw-Hill, New York, 1967), pp. 125–150
21. H.L. Skriver, Crystal-Structure from One-Electron Theory, *Phys. Rev. B*, 1985, **31**, p 1909–1923
22. G. Grimvall, Reconciling Ab Initio and Semiempirical Approaches to Lattice Stabilities, *Phys. Chem. Chem. Phys.*, 1998, **102**, p 1083–1087
23. J. Hafner and F. Sommer, Ab-Initio Pseudopotential Calculations of Structure and Stability of Binary-Alloys and Intermetallic Compounds, *CALPHAD*, 1977, **1**, p 325–340
24. G. Kresse and J. Hafner, Ab-Initio Molecular-Dynamics for Open-Shell Transition-Metals, *Phys. Rev. B*, 1993, **48**, p 13115–13118
25. J. Hafner, Ab-Initio Simulations of Materials Using VASP: Density-Functional Theory and Beyond, *J. Comput. Chem.*, 2008, **29**, p 2044–2078
26. Z.K. Liu, First-Principles Calculations and CALPHAD Modeling of Thermodynamics, *J. Phase Equilib. Diffus.*, 2009, **30**, p 517–534
27. Z.K. Liu, X.Y. Li, and Q.M. Zhang, Maximizing the Number of Coexisting Phases Near Invariant Critical Points for Giant

Electrocaloric and Electromechanical Responses in Ferroelectrics, *Appl. Phys. Lett.*, 2012, **101**, p 82904
28. A.V. Evteev, E.V. Levchenko, I.V. Belova, R. Kozubski, Z.K. Liu, and G.E. Murch, Thermotransport in Binary System: Case Study on $Ni_{50}Al_{50}$ Melt, *Philos. Mag.*, 2014, **94**, p 3574–3602
29. T. Ahmed, E.V. Levchenko, A.V. Evteev, Z.K. Liu, W.Y. Wang, R. Kozubski, I.V. Belova, and G. Murch, Molecular Dynamics Prediction of the Influence of Composition on Thermotransport in Ni-Al Melts, *Diffus. Found.*, 2017, **12**, p 93–110
30. Y. Wang, Y.-J. Hu, S.-L. Shang, S.A. Firdosy, K.E. Star, J.-P. Fleurial, V.A. Ravi, L.-Q. Chen, and Z.-K Liu, A First-Principles Approach to Predict Seebeck Coefficients: Application to La3-xTe4 (2017). http://arxiv.org/abs/1708.00949
31. Y. Wang, Y.-J. Hu, S.-L. Shang, B.-C. Zhou, Z.-K. Liu, and L.-Q. Chen, First-Principles Thermodynamic Theory of Seebeck Coefficients (2017). http://arxiv.org/abs/1708.01591
32. Z.K. Liu, Y. Wang, and S.L. Shang, Thermal Expansion Anomaly Regulated by Entropy, *Sci. Rep.*, 2014, **4**, p 7043
33. Z.K. Liu, S.L. Shang, and Y. Wang, Fundamentals of Thermal Expansion and Thermal Contraction, *Materials*, 2017, **10**, p 410
34. M. Hillert, A Solid-Solution Model for Inhomogeneous Systems, *Acta Metall.*, 1961, **9**, p 525–535
35. J.W. Cahn, On Spinodal Decomposition, *Acta Metall.*, 1961, **9**, p 795–801
36. J. Zhou, J. Zhong, L. Chen, L. Zhang, Y. Du, Z.-K. Liu, and P.H. Mayrhofer, Phase Equilibria, Thermodynamics and Microstructure Simulation of Metastable Spinodal Decomposition in c-$Ti_{1-x}Al_xN$ Coatings, *CALPHAD*, 2017, **56**, p 92–101
37. M.G. Medvedev, I.S. Bushmarinov, J. Sun, J.P. Perdew, and K. A. Lyssenko, Density Functional Theory is Straying from the Path Toward the Exact Functional, *Science*, 2017, **355**, p 49–52
38. Y. Wang, S. Curtarolo, C. Jiang, R. Arroyave, T. Wang, G. Ceder, L.Q. Chen, and Z.K. Liu, Ab Initio Lattice Stability in Comparison with CALPHAD Lattice Stability, *CALPHAD*, 2004, **28**, p 79–90
39. C. Jiang, C. Wolverton, J. Sofo, L.Q. Chen, and Z.K. Liu, First-Principles Study of Binary bcc Alloys Using Special Quasirandom Structures, *Phys. Rev. B*, 2004, **69**, p 214202
40. A. van de Walle, P. Tiwary, M. de Jong, D.L. Olmsted, M. Asta, A. Dick, D. Shin, Y. Wang, L.-Q. Chen, and Z.-K. Liu, Efficient Stochastic Generation of Special Quasirandom Structures, *CALPHAD*, 2013, **42**, p 13–18
41. A.J. Wang, L.C. Zhou, Y. Kong, Y. Du, Z.K. Liu, S.L. Shang, Y. F. Ouyang, J. Wang, L.J. Zhang, and J.C. Wang, First-Principles Study of Binary Special Quasirandom Structures for the Al-Cu, Al-Si, Cu-Si, and Mg-Si Systems, *CALPHAD*, 2009, **33**, p 769–773
42. Y. Wang, Z.K. Liu, and L.Q. Chen, Thermodynamic Properties of Al, Ni, NiAl, and Ni_3Al from First-Principles Calculations, *Acta Mater.*, 2004, **52**, p 2665–2671
43. Y. Wang, S.L. Shang, H. Fang, Z.K. Liu, and L.Q. Chen, First-Principles Calculations of Lattice Dynamics and Thermal Properties of Polar Solids, *NPJ Comput. Mater.*, 2016, **2**, p 16006
44. S.L. Shang, Y. Wang, D. Kim, and Z.K. Liu, First-Principles Thermodynamics from Phonon and Debye mOdel: Application to Ni and Ni_3Al, *Comput. Mater. Sci.*, 2010, **47**, p 1040–1048
45. D. Kim, S.L. Shang, and Z.K. Liu, Effects of Alloying Elements on Thermal Expansions of Gamma-Ni and Gamma′-Ni_3Al by First-Principles Calculations, *Acta Mater.*, 2012, **60**, p 1846–1856
46. M. Mantina, in *A First-Principles Methodology for Diffusion Coefficients in Metals and Dilute Alloys*. Department of Materials Science and Engineering (The Pennsylvania State University, 2008).
47. B.-C. Zhou, S.-L. Shang, Y. Wang, and Z.-K. Liu, Diffusion Coefficients of Alloying Elements in Dilute Mg Alloys: A Comprehensive First-Principles Study, *Acta Mater.*, 2016, **103**, p 573–586
48. Y. Wang, L.Q. Chen, Z.K. Liu, and S.N. Mathaudhu, First-Principles Calculations of Twin-Boundary and Stacking-Fault Energies in Magnesium, *Scr. Mater.*, 2010, **62**, p 646–649
49. S.L. Shang, W.Y. Wang, Y. Wang, Y. Du, J.X. Zhang, A.D. Patel, and Z.K. Liu, Temperature-Dependent Ideal Strength and Stacking Fault Energy of fcc Ni: A First-Principles Study of Shear Deformation, *J. Phys.: Condens. Matter*, 2012, **24**, p 155402
50. V.R. Manga, J.E. Saal, Y. Wang, V.H. Crespi, and Z.K. Liu, Magnetic Perturbation and Associated Energies of the Antiphase Boundaries in Ordered Ni_3Al, *J. Appl. Phys.*, 2010, **108**, p 103509
51. W.Y. Wang, F. Xue, Y. Zhang, S.-L. Shang, Y. Wang, K.A. Darling, L.J. Kecskes, J. Li, X. Hui, Q. Feng, and Z.-K. Liu, Atomic and Electronic Basis for Solutes Strengthened (010) Anti-Phase Boundary of $L_{12}Co_3$(Al, TM): A Comprehensive First-Principles Study, *Acta Mater.*, 2018, **145**, p 1008–1014
52. H.Z. Fang, Y. Wang, S.L. Shang, P.D. Jablonski, and Z.K. Liu, First-Principles Calculations of Interfacial and Segregation Energies in Alpha-Cr_2O_3, *J. Phys.: Condens. Matter*, 2012, **24**, p 225001
53. W.Y. Wang, S.L. Shang, Y. Wang, Y.J. Hu, K.A. Darling, L.J. Kecskes, S.N. Mathaudhu, X.D. Hui, and Z.K. Liu, Lattice Distortion Induced Anomalous Ferromagnetism and Electronic Structure in FCC Fe and Fe-TM (TM=Cr, Ni, Ta and Zr) Alloys, *Mater. Chem. Phys.*, 2015, **162**, p 748–756
54. S.L. Shang, Y. Wang, and Z.K. Liu, First-Principles Elastic Constants of Alpha- and Theta-Al_2O_3, *Appl. Phys. Lett.*, 2007, **90**, p 101909
55. S.-L.L. Shang, H. Zhang, Y. Wang, and Z.-K.K. Liu, Temperature-Dependent Elastic Stiffness Constants of Alpha- and Theta-Al_2O_3 from First-Principles Calculations, *J. Phys.: Condens. Matter*, 2010, **22**, p 375403
56. Y.-J. Hu, M.R. Fellinger, B.G. Bulter, Y. Wang, K.A. Darling, L.J. Kecskes, D.R. Trinkle, and Z.K. Liu, Solute-Induced Solid-Solution Softening and Hardening in bcc Tungsten, *Acta Mater.*, 2017, **141**, p 304–316
57. C.L. Zacherl, Ph.D. Dissertation, *A Computational Investigation of the Effect of Alloying Elements on the Thermodynamic and Diffusion Properties of fcc Ni Alloys, with Application to the Creep Rate of Dilute Ni-X Alloys*. (The Pennsylvania State University, 2012)
58. C.Z. Hargather, S.L. Shang, and Z.K. Liu, A Comprehensive First-Principles Study of Solute Elements in Dilute Ni Alloys: Diffusion Coefficients and Their Implications to Tailor Creep Rate. *Acta Mater.* (2018) (under revi)
59. Y. Wang, J.J. Wang, W.Y. Wang, Z.G. Mei, S.L. Shang, L.Q. Chen, and Z.K. Liu, A Mixed-Space Approach to First-Principles Calculations of Phonon Frequencies for Polar Materials, *J. Phys.: Condens. Matter*, 2010, **22**, p 202201
60. Y. Wang, S.L. Shang, Z.K. Liu, and L.Q. Chen, Mixed-Space Approach for Calculation of Vibration-Induced Dipole–Dipole Interactions, *Phys. Rev. B*, 2012, **85**, p 224303
61. Y. Wang, W.Y. Wang, L.-Q. Chen, and Z.-K. Liu, Bonding Charge Density from Atomic Perturbations, *J. Comput. Chem.*, 2015, **36**, p 1008–1014
62. V. Ozolins, First-Principles Calculations of Free Energies of Unstable Phases: The Case of fcc W, *Phys. Rev. Lett.*, 2009, **102**, p 65702
63. Z.G. Mei, in *First-Principles Thermodynamics of Phase Transition: From Metal to Oxide*. Ph.D. (The Pennsylvania State University, 2011)

64. M.W. Chase, NIST-JANAF Thermochemical Tables 2 Volume-Set (Journal of Physical and Chemical Reference Data Monographs). (American Chemical Society Washington, DC, 1998)
65. A. van de Walle, Q. Hong, S. Kadkhodaei, and R. Sun, The Free Energy of Mechanically Unstable Phases, *Nat. Commun.*, 2015, **6**, p 8559
66. A. van de Walle, Invited Paper: Reconciling SGTE and Ab Initio Enthalpies of the Elements, *CALPHAD*, 2018, **60**, p 1–6
67. M. Hillert, The Compound Energy Formalism, *J. Alloys Compd.*, 2001, **320**, p 161–176
68. R. Car and M. Parrinello, Unified Approach for Molecular-Dynamics and Density-Functional Theory, *Phys. Rev. Lett.*, 1985, **55**, p 2471–2474
69. X. Hui, H.Z. Fang, G.L. Chen, S.L. Shang, Y. Wang, and Z.K. Liu, Icosahedral Ordering in $Zr_{41}Ti_{14}Cu_{12.5}Ni_{10}Be_{22.5}$ Bulk Metallic Glass, *Appl. Phys. Lett.*, 2008, **92**, p 201913
70. X. Hui, H.Z. Fang, G.L. Chen, S.L. Shang, Y. Wang, J.Y. Qin, and Z.K. Liu, Atomic Structure of $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10}Be_{22.5}$ Bulk Metallic Glass Alloy, *Acta Mater.*, 2009, **57**, p 376–391
71. W.Y. Wang, S.L. Shang, Y. Wang, F. Han, K.A. Darling, Y. Wu, X. Xie, O.N. Senkov, J. Li, X.D. Hui, K.A. Dahmen, P.K. Liaw, L.J. Kecskes, and Z.K. Liu, Atomic and Electronic Basis for the Serrations of Refractory High-Entropy Alloys, *NPJ Comput. Mater.*, 2017, **3**, p 23
72. Y. Wang, H.Z. Fang, C.L. Zacherl, Z.G. Mei, S.L. Shang, L.Q. Chen, P.D. Jablonski, and Z.K. Liu, First-Principles Lattice Dynamics, Thermodynamics, and Elasticity of Cr_2O_3, *Surf. Sci.*, 2012, **606**, p 1422–1425
73. Z.F. Zheng, H.Z. Fang, F. Yang, Z.K. Liu, and Y. Wang, Amorphous $LiLaTiO_3$ as Solid Electrolyte Material, *J. Electrochem. Soc.*, 2014, **161**, p A473–A479
74. H. Zhang, S.L. Shang, W.Y. Wang, Y. Wang, X.D. Hui, L.Q. Chen, and Z.K. Liu, Structure and Energetics of Ni from Ab Initio Molecular Dynamics Calculations, *Comput. Mater. Sci.*, 2014, **89**, p 242–246
75. J. Han, W.Y. Wang, C. Wang, Y. Wang, X. Liu, and Z.-K. Liu, Accurate Determination of Thermodynamic Properties For Liquid Alloys Based on Ab Initio molecular Dynamics Simulation, *Fluid Phase Equilib.*, 2013, **360**, p 44–53
76. L.G. Wang, A. van de Walle, and D. Alfè, Melting Temperature of Tungsten from Two Ab Initio Approaches, *Phys. Rev. B*, 2011, **84**, p 92102
77. A. van de Walle, Melting Mechanisms. Simulations Provide a Rare Look at Real Melting, *Science*, 2014, **346**, p 704–705
78. Q.-J. Hong and A. van de Walle, A User Guide for SLUSCHI: Solid and Liquid in Ultra Small Coexistence with Hovering Interfaces, *CALPHAD*, 2016, **52**, p 88–97
79. H.Z. Fang, Y. Wang, S.L. Shang, and Z.K. Liu, Nature of Ferroelectric–Paraelectric Phase Transition and Origin of Negative Thermal Expansion in $PbTiO_3$, *Phys. Rev. B*, 2015, **91**, p 24104
80. A. Jain, O. Shyue Ping, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K.A. Persson, Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation, *Appl. Mater.*, 2013, **1**, p 11002
81. Materials Project, Available at http://materialsproject.org/
82. J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD), *JOM*, 2013, **65**, p 1501–1509
83. OQMD: An Open Quantum Materials Database. Available at http://oqmd.org. Accessed 26 June 2018
84. AFLOW: Automatic Flow for Materials Discovery, http://www.aflowlib.org, Available at http://www.aflowlib.org. Accessed 26 June 2018
85. S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, and G. Ceder, Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis, *Comput. Mater. Sci.*, 2013, **68**, p 314–319
86. G. Bergerhoff, R. Hundt, R. Sievers, and I.D. Brown, The Inorganic Crystal Structure Data Base, *J. Chem. Inf. Model.*, 1983, **23**, p 66–69
87. A.O. Lyakhov, A.R. Oganov, H.T. Stokes, and Q. Zhu, New Developments in Evolutionary Structure Prediction Algorithm USPEX, *Comput. Phys. Commun.*, 2013, **184**, p 1172–1182
88. C.W. Glass, A.R. Oganov, and N. Hansen, USPEX: Evolutionary Crystal Structure Prediction, *Comput. Phys. Commun.*, 2006, **175**, p 713–720
89. National Research Council, *Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security* (2008)
90. CALPHAD: Computer Coupling of Phase Diagrams and Thermochemistry. Available at https://www.sciencedirect.com/journal/calphad. Accessed 26 June 2018
91. A. van de Walle, C. Nataraj, and Z.-K. Liu, The Thermodynamic Database Database, *CALPHAD*, 2018, **61**, p 173–178
92. B. Sundman, U.R. Kattner, M. Palumbo, and S.G. Fries, OpenCalphad-A Free Thermodynamic Software, *Integr. Mater. Manuf. Innov.*, 2015, **4**, p 1
93. B. Sundman, Open CALPHAD, Available at http://www.opencalphad.com. Accessed 26 June 2018
94. C.E. Campbell, U.R. Kattner, and Z.-K.K. Liu, The Development of Phase-Based Property Data Using the CALPHAD Method and Infrastructure Needs, *Integr. Mater. Manuf. Innov.*, 2014, **3**, p 12
95. C.E. Campbell, U.R. Kattner, and Z.K. Liu, File and Data Repositories for Next Generation CALPHAD, *Scr. Mater.*, 2014, **70**, p 7–11
96. Z.K. Liu, H. Zhang, S. Ganeshan, Y. Wang, and S.N. Mathaudhu, Computational Modeling of Effects of Alloying Elements on Elastic Coefficients, *Scr. Mater.*, 2010, **63**, p 686–691
97. S.L. Shang, Y. Wang, and Z.K. Liu, ESPEI: Extensible, Self-optimizing Phase Equilibrium Infrastructure for Magnesium Alloys, in *Magnesium Technology 2010*, (eds. S.R. Agnew, N.R. Neelameggham, E.A. Nyberg, W.H. Sillekens) (The Minerals, Metals and Materials Society (TMS), Pittsburgh, PA, 2010), pp. 617–622
98. J.O. Andersson, T. Helander, L. Höglund, P.F. Shi, B. Sundman, L.H. Hoglund, P.F. Shi, B. Sundman, L. Höglund, P.F. Shi, and B. Sundman, THERMO-CALC & DICTRA, Computational Tools for Materials Science, *CALPHAD*, 2002, **26**, p 273–312
99. R.A. Otis and Z.K. Liu, Pycalphad: CALPHAD-Based Computational Thermodynamics in Python, *J. Open Res. Softw.*, 2017, **5**, p 1
100. R.A. Otis and Z.K. Liu, Pycalphad. Available at http://pycalphad.org. Accessed 12 Apr 2018
101. R.A. Otis and Z.-K. Liu, High-Throughput Thermodynamic Modeling and Uncertainty Quantification for ICME, *JOM*, 2017, **69**, p 886–892
102. Z.K. Liu, R.A. Otis, and B. Bocklund, ESPEI: Extensible Self-optimizing Phase Equilibria Infrastructure. Available at http://espei.org. Accessed 11 Apr 2018
103. H. Bozdogan, Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions, *Psychometrika*, 1987, **52**, p 345–370
104. H. Akaike, Selected Papers of Hirotugu Akaike. (Springer New York, 1998). https://doi.org/10.1007/978-1-4612-1694-0_15. Accessed 26 June 2018
105. ESPEI: Cu-Mg Example. Available at https://github.com/PhasesResearchLab/ESPEI-datasets/tree/master/CU-MG. Accessed 26 June 2018

106. W.K. Hastings, Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, 1970, **57**, p 97–109
107. J. Goodman and J. Weare, Ensemble Samplers with Affine Invariance, *Commun. Appl. Math. Comput. Sci.*, 2010, **5**, p 65–80
108. D. Foreman-Mackey, D.W. Hogg, D. Lang, and J. Goodman, Emcee: The MCMC Hammer, *Publ. Astron. Soc. Pac.*, 2013, **125**, p 306–312
109. S. Curtarolo, D. Morgan, and G. Ceder, Accuracy of Ab Initio Methods in Predicting the Crystal Structures of Metals: A Review of 80 Binary Alloys, *CALPHAD*, 2005, **29**, p 163–211
110. W. Xiong and G.B. Olson, Cybermaterials: Materials by Design and Accelerated Insertion of Materials, *NPJ Comput. Mater.*, 2016, **2**, p 15009
111. K. Mathew, J. Montoya, A. Faghnaninia, S. Dwarakanath, M. Aykol, H. Tang, I. Chu, T. Smidt, B. Bocklund, Z.-K. Liu, J. Neaton, S.P. Ong, K. Persson, and J. Anubhav, Atomate: A High-Level Interface to Generate, Execute, and Analyze Computational Materials Science Workfows, *Comput. Mater. Sci.*, 2017, **139**, p 140–152
112. J.C. Zhao, Combinatorial Approaches as Effective Tools in the Study of Phase Diagrams and Composition–Structure–Property Relationships, *Prog. Mater Sci.*, 2006, **51**, p 557–631
113. N. Sicron, B. Ravel, Y. Yacoby, E.A. Stern, F. Dogan, and J.J. Rehr, Nature of the Ferroelectric Phase-Transition in $PbTiO_3$, *Phys. Rev. B*, 1994, **50**, p 13168–13180
114. G. Shirane and S. Hoshino, On the Phase Transition in Lead Titanate, *J. Phys. Soc. Jpn.*, 1951, **6**, p 265
115. Z.-K. Liu, Y. Wang, and S.-L. Shang, Origin of Negative Thermal Expansion Phenomenon in Solids, *Scr. Mater.*, 2011, **66**, p 130