npj | Computational Materials

www.nature.com/npjcompumats

ARTICLE OPEN

Combining thermodynamics with tensor completion techniques to enable multicomponent microstructure prediction

Yuri Amorim Coutinho[1], Nico Vervliet[2], Lieven De Lathauwer[2,3] and Nele Moelans[1*]

Multicomponent alloys show intricate microstructure evolution, providing materials engineers with a nearly inexhaustible variety of solutions to enhance material properties. Multicomponent microstructure evolution simulations are indispensable to exploit these opportunities. These simulations, however, require the handling of high-dimensional and prohibitively large data sets of thermodynamic quantities, of which the size grows exponentially with the number of elements in the alloy, making it virtually impossible to handle the effects of four or more elements. In this paper, we introduce the use of tensor completion for high-dimensional data sets in materials science as a general and elegant solution to this problem. We show that we can obtain an accurate representation of the composition dependence of high-dimensional thermodynamic quantities, and that the decomposed tensor representation can be evaluated very efficiently in microstructure simulations. This realization enables true multicomponent thermodynamic and microstructure modeling for alloy design.

npj Computational Materials (2020)6:2 ; https://doi.org/10.1038/s41524-019-0268-y

INTRODUCTION

A number of recent discoveries has largely increased the interest in multicomponent alloy design. High entropy or multi-principle element alloys, for example, can give access to a wide range of properties and combinations of properties that cannot be obtained in alloys based on a single major element.[1] Moreover, the large range over which the diffusion properties of the different elements can vary gives rise to complex precipitation, interdiffusion, and coarsening paths,[2–4] resulting in new ways to stabilize metastable phases or structures for extended times.[5] Adding too many alloying elements may, however, also unnecessarily complicate material recovery and recycling.[6] Hence, in order to exploit in an optimal and responsible way the opportunities multicomponent alloy design can bring, a profound understanding of the effects of compositional variations and heat treatment on microstructure evolution in these alloys is required. Microstructure simulations are indispensable to obtain these insights given the immense parameter space, including not only the current temperature and the amount of each element, but also many parameters describing the history of the material, such as cooling rate and the duration of a heat treatment or mechanical loading.

Phase-field models (PFM) are generally considered as most suitable to study multicomponent microstructure evolution as diffusion and phase morphology often play a crucial role.[7–9] They describe complex morphological changes such as precipitate shape evolution during growth and coarsening and dendrite growth. The effects of various competing thermodynamic driving forces and transport processes can be considered. For multicomponent alloys, the chemical bulk driving force is an important contribution. CALPHAD thermodynamic models are usually adopted to describe the composition and temperature-dependent Gibbs free energy expressions of the different phases. For many technologically relevant alloy systems, CALPHAD models have been assessed to reproduce experimental, theoretical and ab initio data on thermodynamic quantities and phase diagrams.[10–12]

Different approaches have been proposed to include the CALPHAD thermodynamic descriptions in PFM.[13–26] A classification and discussion of the different approaches is given further in the paper. However, to date, only a few results for quaternary or higher-order systems are reported in the literature. The reason is that the proposed methods were either for a specific case and cannot be generalised towards systems containing other types of phases, or become prohibitively complex or computation and data-intensive when the number of elements in the alloy is increased. There is thus no robust simulation approach available facilitating in silico microstructure design for multicomponent alloys.

Here, we introduce the use of tensor completion[27–30] in materials science to enable the representation and use of high-dimensional data sets efficiently. More specifically, we used a canonical polyadic decomposition (CPD) with the factor vectors constrained to polynomial expressions[31] to include high-dimensional thermodynamic data sets obtained from a CALPHAD model in phase-field simulations. Using CPD, the multicomponent composition dependence can be described using a limited number of coefficients. We show that the accuracy obtained with the proposed approach is comparable or better than that obtained with previous approaches, whereas the applicability is general and the computational cost to evaluate data is low. Moreover, the addition of more elements does not lead to a substantial increase of the complexity, and the number of coefficients and the computational cost to evaluate a CPD depend only linearly on the number of components in the system, in contrast to an exponential increase of the amount of thermodynamic data represented. The observed advantages will thus become increasingly prominent when adding more elements. The capabilities of the approach are illustrated for phase-field

[1]Department of Materials Engineering (MTM), KU Leuven, B-3001 Leuven, Belgium. [2]Department of Electrical Engineering (ESAT), KU Leuven, B-3001 Leuven, Belgium. [3]Group Science, Engineering and Technology, KU Leuven Kulak, B-8500 Kortrijk, Belgium. *email: nele.moelans@kuleuven.be

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

npj nature partner journals

simulations of spinodal decomposition and subsequent coarsening, an important phenomenon in multicomponent alloys,[32] for Ag-Cu-Ni-Sn liquid alloys.

RESULTS AND DISCUSSION

Tensor completion of thermodynamic data

In material science, we are familiar with the tensorial character of direction-dependent material properties, like stresses and strains. However, in mathematical engineering, the term tensor has been used much more widely. In general, tensors are higher-order generalisations of vectors (first-order tensors) and matrices (second-order tensors).[33,34] They can be seen as multiway arrays of any order holding numerical values. The term thermodynamic data tensor (TDT) is introduced here to denote a tensor that is obtained by calculating and storing the data set of any thermodynamic quantity of a phase, evaluated from a CALPHAD model. The TDTs typically needed in PFM implementations are the molar Gibbs free energy, $\mathcal{G}$, the diffusion potential of each component n, $\mathcal{P}^{(n)}$, and the derivative of the diffusion potential of the component n with respect to the molar fraction of components m, $\mathcal{D}^{(n,m)}$.

Typically, the molar fractions x_n of the components are used to indicate an alloy composition. For an alloy with C components, the composition dependence of any thermodynamic quantity can be expressed as a function of $C-1$ molar fractions x_n, with $n = 1, \ldots, C-1$ referring to $C-1$ of the C components. The order N of the TDT, this is the number of dimensions of the tensor, then equals $C-1$. A graphical illustration of a third-order TDT, representing composition dependence of a thermodynamic quantity is shown in Fig. 1b. As by definition $x_C + \sum_{n=1}^{N} x_n = 1$ and $0 < x_n < (1 - x_C)$, tensor entries (these are the values in the tensor) for a combination of molar fraction values disobeying this constraint are not physical and cannot be calculated. A TDT is therefore necessarily incomplete. Besides composition dependence, other dependencies of the thermodynamic function, such as temperature, can be considered, increasing the order of the tensor with one for each extra dependency.

If a thermodynamic model is available, TDTs can be precalculated and used in phase-field simulations to get the required thermodynamic quantities. However, the great challenge of dealing with TDTs is that their number of entries increases exponentially with the order of the tensor, as illustrated in Fig. 1a. Therefore, when more components are considered, it quickly becomes impossible to generate, store, or handle the huge data sets. The computational difficulties caused by this exponential dependence are known as the curse of dimensionality.[28]

Moreover, as illustrated in Fig. 1a, b, the step size, δ_{x_n}, with which the molar fraction x_n is sampled, obviously also affects the number of entries in the TDT, as the size of the tensor along each dimension I_n is inversely proportional to the step size. The step size can be chosen differently along each dimension, but in this study we take a same step size for all molar fractions, i.e., $\delta_x \overset{\text{def}}{=} \delta_{x_1} = \ldots = \delta_{x_N}$. If TDTs are used to provide thermodynamic quantities in phase-field simulations, a small step size δ_x is desirable to avoid sudden molar fraction jumps and the need for multi-directional interpolation between entries.

In practice, the use of TDTs in microstructure evolution simulations is thus limited to lower-order systems. The hypothesis motivating our work is that, if a higher-order TDT is too large to be constructed or used, the information present in the TDT may still be accessible if a CPD of the TDT is obtained. See Methods section Tensor Decomposition for a general introduction to this methodology. When applied to a quaternary thermodynamic system ($N = 3$), such as the liquid Ag-Cu-Ni-Sn alloys under study, our technique decomposes the TDT into a sum of R rank-1 terms, of which each term is the outer product of N factor vectors, and with R called the rank of the decomposed tensor model. This is graphically illustrated in Fig. 2a, b. An alternative representation of the same CPD is obtained by collecting all factor vectors of the CPD related to each dimension n into a factor matrix $\mathbf{A}^{(n)}$. In this case, the dimensions n refer to the molar fractions x_n, with $n = 1, 2, 3$, and factor matrices $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, and $\mathbf{A}^{(3)}$ are obtained, as shown in Fig. 2a–c.

A major benefit of this approach is demonstrated in Fig. 2d, where the number of entries in a TDT is shown to grow exponentially with the number of components in the system, whereas the number of parameters in a CPD, representing the same TDT only grows linearly. For alloys with four or more components, the reduction in coefficients when using a CPD representation of a full tensor is gigantic, enabling multicomponent microstructure evolution simulations.

While determining the rank value, R is needed for an exact decomposition of a TDT is difficult, previous applications in other fields have shown that tensors representing physical data can often be well approximated using a low-rank CPD.[35–38] In this work, we applied tensor decomposition to high-dimensional

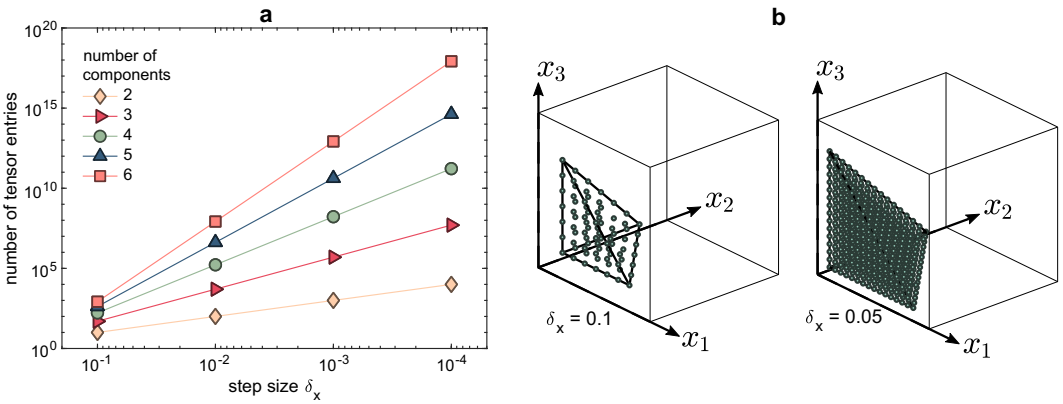


Fig. 1 Characteristics of a thermodynamic data tensor. a The number of entries in a TDT is plotted as a function of the step size δ_x (used for data collection) and the number of components in the system. **b** Visualisation of two third-order incomplete TDTs sampled with different step size. **b** The dashed box represents what would be the shape of a full third-order tensor; however, owing to the constraint $\sum_{i=1}^{C} x_i = 1$ on the molar fractions, the TDT is incomplete and only the entries within the indicated tetrahedron, i.e., the entries satisfying $\sum_{i=1}^{N} x_i \leq 1$ can be evaluated. The step size δ_x determines the density and hence the number of entries in the TDT. On the other hand, the number of TDT entries depends exponentially on the number of components, or, more generally, on the order N of the tensor, making storage, and computation quickly intractable, as can be seen in **a**.

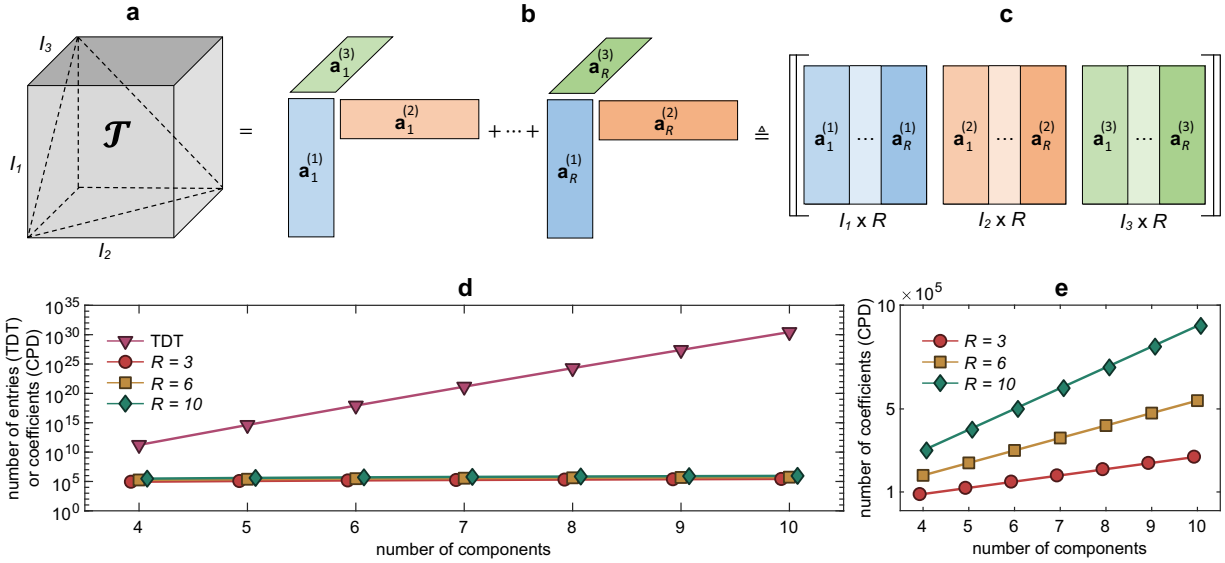


Fig. 2 Canonical polyadic decomposition of a thermodynamic data tensor (TDT). a–c Visualisation of a canonical polyadic decomposition (CPD) when applied to a third-order TDT with dimension $I_1 \times I_2 \times I_3$. The third-order TDT **a** is written as a sum of R rank-1 terms **b**, each of which is represented by the outer product of three nonzero factor vectors $\mathbf{a}_r^{(1)}$, $\mathbf{a}_r^{(2)}$, and $\mathbf{a}_r^{(3)}$. The collection of R factor vectors into factor matrices **c** with dimensions $I_1 \times R, I_2 \times R$, and $I_3 \times R$, provides a convenient representation of the CPD. Each factor matrix contains the coefficients necessary to describe the contribution of the molar fraction of one of the components to the TDT. The entire tensor in **a** corresponds to the dashed box in Fig. 1b; the CPD in **b** will be determined from only the tetrahedron part in Fig. 1b. **d** The number of entries in the TDT and the number of coefficients in the CPDs with rank values $R = 3$, 6, 10 are plotted as a function of the number of components, for step size $\delta_x = 0.0001$. The exponential dependence of the number of entries in the TDT on the number of components is broken when a CPD representation is used instead. Indeed, the CPD needs only $(I_1 + I_2 + I_3)R$ coefficients, which depends linearly on the number of components. **e** The same information as in **d** is plotted, but excluding the TDT size for a better visualisation of the number of coefficients in the TTM as a function of the number of system components.

thermodynamic data sets and we show that a low-rank approximation can be used to obtain an accurate decomposed tensor model describing the dependence of Gibbs energies and derived thermodynamic functions on the molar fractions, representing the alloy composition.

Moreover, when a priori knowledge on solution thermodynamics is taken into account, a CPD can be made even more compact and a continuous description of the composition dependence can be obtained. In this study, for example, we exploited the polynomial dependence of thermodynamic quantities on the molar fractions when the ideal mixing term is subtracted. This approach is reasonable as the contribution from ideal mixing depends only on known information (namely, the molar fractions of all components and the temperature) and can be calculated separately and added to the values evaluated from the CPD in the phase-field simulation (see Methods sections CALPHAD thermodynamic model and Thermodynamic tensor model). A polynomial constraint could therefore be imposed on the factor vectors,[31,39,40] and the CPDs representing the molar Gibbs free energy, ${}^*\mathcal{G}$ (the superscript * denotes that the ideal mixing contribution is subtracted from the thermodynamic quantity), the diffusion potential of component n, ${}^*\mathcal{P}^{(n)}$, (i.e., derivative of the molar Gibbs free energy) and the derivative of the diffusion potential of component n with respect to the molar fractions of the component m, ${}^*\mathcal{D}^{(n,m)}$, (i.e., the second derivatives of the molar Gibbs free energies) were coupled. A single set of factor matrices modeling the composition dependence of all the required thermodynamic quantities at once was then obtained. Such a set containing all factor matrices necessary to completely describe the thermodynamic quantities of the system in a phase-field simulation is further referred as a thermodynamic tensor model of rank R. For a quaternary system, it is represented as

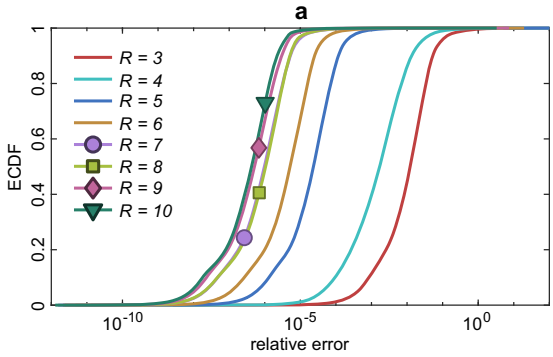
$$\mathrm{TTM}\,R = \left\{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \dot{\mathbf{A}}^{(1)}, \dot{\mathbf{A}}^{(2)}, \dot{\mathbf{A}}^{(3)}, \ddot{\mathbf{A}}^{(1)}, \ddot{\mathbf{A}}^{(2)}, \ddot{\mathbf{A}}^{(3)}\right\}. \quad (1)$$

The factor matrices $\mathbf{A}^{(n)}$ give the contribution related to x_n to the Gibbs free energy TDT. The $\dot{\mathbf{A}}^{(n)}$ model, the contribution related to taking the first partial derivative of the Gibbs energy with respect to x_n, is needed to represent the TDT of the diffusion potential of component n. The $\ddot{\mathbf{A}}^{(n)}$ model contributions related to taking the second partial derivative of the Gibbs energy with respect to x_n. A cross partial derivative with respect to x_n and x_m is obtained using $\dot{\mathbf{A}}^{(n)}$ and $\dot{\mathbf{A}}^{(m)}$.

From these factor matrices, the thermodynamic tensor models representing the required TDTs can be obtained, for instance, for a quaternary system

$$\begin{aligned} {}^*\mathcal{G} &\approx [\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}], \\ {}^*\mathcal{P}^{(1)} &\approx [\dot{\mathbf{A}}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}], \\ {}^*\mathcal{D}^{(1,2)} &\approx [\dot{\mathbf{A}}^{(1)}, \dot{\mathbf{A}}^{(2)}, \mathbf{A}^{(3)}], \\ {}^*\mathcal{D}^{(3,3)} &\approx [\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ddot{\mathbf{A}}^{(3)}], \end{aligned} \quad (2)$$

giving an extremely compact representation of an immense amount of data. Taking, for example, $N = 3$, $\delta_x = 0.0001$, and $R = 7$, a TDT of about $(1/\delta_x)^N/N! \approx 1.6 \times 10^{11}$ entries (the factor $1/N!$ takes into account that only the entries within the tetrahedron in the TDT in Fig. 1a, b are calculated) is represented using factor matrices that contain only $R \times N \times 1/\delta_x \approx 2.1 \times 10^5$ coefficients. The CPD model is thus about a million times more compact than the TDT. Furthermore, the coefficients of the degree d polynomial constraint alone already suffice for the representation of the TTM. The Gibbs free energy and derived TDTs can thus be represented with $R \times N \times (d+1) = 105$ coefficients only. Such impressive compression ratios have recently revolutionised tensor-based scientific computing;[35–38] in materials science applications, the approach is new.



b

Rank	IQR	Q = 0.95	Q = 0.98
3	2.46×10^{-2}	8.70×10^{-2}	3.67×10^{-1}
4	4.92×10^{-3}	2.40×10^{-2}	1.02×10^{-1}
5	5.15×10^{-5}	1.85×10^{-4}	6.72×10^{-4}
6	1.17×10^{-5}	3.94×10^{-5}	1.52×10^{-4}
7	2.32×10^{-6}	8.01×10^{-6}	3.13×10^{-5}
8	2.39×10^{-6}	7.27×10^{-6}	2.93×10^{-5}
9	1.27×10^{-6}	4.40×10^{-6}	1.29×10^{-5}
10	9.87×10^{-7}	3.44×10^{-6}	8.81×10^{-6}

Fig. 3 Thermodynamics tensor model accuracy. a Empirical cumulative distribution function (ECDF) of the relative error of TTMs for rank values $R = 3, \ldots, 10$. **b** Table with the interquartile range (IQR) and quantiles $Q = 0.95$ and $Q = 0.98$ of the relative error of TTMs for rank values $R = 3, \ldots, 10$. In both cases **a**, **b** the relative error is calculated for the Gibbs free energy, diffusion potentials, and derivatives taking data collected using TC-toolbox with COST 531 database as the reference. There is little improvement of the accuracy of the TTM for rank values $R \geq 7$.

Moreover, the coefficients in the factor vectors can be obtained using only a limited set of entries from the original TDT, called the training set (see Methods section Thermodynamic tensor model training for more details). There is thus no need to compute the full TDT at any point in the derivation.

Another major advantage of the approach is that values approximating individual entries of the TDT can be calculated very efficiently and easily from the TTM. For example, the diffusion potential of the first component $\tilde{\mu}_1$ at a composition given by the molar fraction values x_1, x_2, and x_3, for which the entry in the tensor is indexed by i_1, i_2, and i_3, is calculated as

$$\tilde{\mu}_1(x_1, x_2, x_3) \approx \sum_{r=1}^{R} \dot{a}_{i_1,r}^{(1)} \, a_{i_2,r}^{(2)} \, a_{i_3,r}^{(3)} + \overline{R}\overline{T}(\log(x_1) - \log(1 - x_1 - x_2 - x_3)). \tag{3}$$

Therefore, again we avoid the need to construct explicitly the full TDT ${}^{*}\mathcal{P}^{(1)}$ before data can be extracted. This provides us with an efficient framework to supply PFM with CALPHAD thermodynamic quantities.

First TTMs with rank values $R = 3, \ldots, 10$ are optimised using Tensorlab,[40] modeling the composition dependence of the molar Gibbs energy, the diffusion potentials of the elements Ag, Cu, and Ni, and the partial derivatives of these diffusion potentials with respect to the molar fractions Ag, Cu, and Ni. The data used for training were obtained using Thermo-Calc 2017b with the COST 531 database.[41] Further details are given in Methods section Thermodynamic tensor model training. This particular system was chosen for validation purposes, as we have access to all coefficients of the CALPHAD model for this quaternary system. The rank R of the TTM is an important parameter in this study. A rank as low as possible is preferred, as the size of the factor matrices scales linearly with R (see Fig. 2b, c), and consequently the number of coefficients that has to be optimised and the number of terms that has to be evaluated when extracting data from the tensor model (equation (3)) in the phase-field simulations, increase linearly with R as well.

The dependence of the accuracy of the TTMs on the rank value R is evaluated by comparing thermodynamic quantities approximated with TTMs $R = 3, \ldots, 10$ with the corresponding entries in the validation TDTs (see Methods section Validation). The empirical cumulative density function (ECDF) of the relative errors (RE) on the thermodynamic quantities calculated from the TTMs is plotted for the different rank values in Fig. 3a. The interquartile range (IQR) and $Q = 0.95$ and $Q = 0.98$ quantiles of the relative error are given for each rank value R in the table in Fig. 3b. The plot and table show that up to $R = 7$, the accuracy improves for increasing R, whereas for higher values of the rank further improvement is limited. Using the TTM with rank value $R = 7$, 98% of the data points are represented with more than four digits of accuracy, i.e., $\mathrm{RE} < 10^{-4}$. For a small fraction (<0.1%) of points, the relative error remains large. However, further inspection of these larger RE showed that they are only found when the considered thermodynamic quantity itself has a value close to zero. It is verified, for instance, that TTMs with $R = 7$ or higher all have a maximum absolute error smaller than 32 J/mol, which is very small compared with the range over which the Gibbs free energy data vary, namely between -1.02×10^5 J/mol and 3.82×10^6 J/mol.

Phase-field simulations

Next, the TTMs for different rank values are used to evaluate the diffusion potentials as a function of the local composition in a multicomponent phase-field model simulating spinodal decomposition and further coarsening in Ag-Cu-Ni-Sn liquids, using the model described in Methods section Phase-field microstructure evolution model. The accuracy and advantages of the approach using TTMs are compared with those of the existing approaches for using CALPHAD thermodynamic models in phase-field simulations. Fig. 4 gives an overview of the most common coupling approaches.[13–16,18–20,22–26] The methods using direct implementation of the CALPHAD Gibbs free energy expressions (b) or an interface with a thermodynamic software (d), give an exact and continuous representation of the composition dependence of the CALPHAD Gibbs free energies. They are considered as a reference against which the accuracy of the simulations using TTMs is evaluated. In comparison with the method using lookup tables (e–f), the curse of dimensionality is broken and an arbitrarily fine resolution, and even continuous representation, of the composition dependence can be obtained.

First, 1D simulations were conducted using TTMs and validated against the approach using an interface with thermodynamic software. The conclusions are discussed in the Supplementary Information. Next, 2D simulations were conducted for the six alloy compositions given in Fig. 4i, using TTMs with rank values $R = 3, \ldots, 10$, and validated against the approach where the CALPHAD substitutional model is directly implemented in the PFM. For each of the six alloy compositions an initial condition is created by adding small random noise at each position in the system to the values of the molar fractions as given in Fig. 4i. The initial condition and all phase-field simulation parameters are the same for all simulations. The only difference is the approach used to include the CALPHAD diffusion potentials in the PFM.

A selection of the 2D-simulated microstructures is shown in Fig. 5a, c, e. Results obtained using the direct implementation of the CALPHAD model expressions (CE) are shown in the first column and results obtained using TTMs $R = 4$ and $R = 6$ in the second and third column. Even for $R = 4$ and $R = 6$, the microstructures obtained using the TTMs look very similar to those obtained using

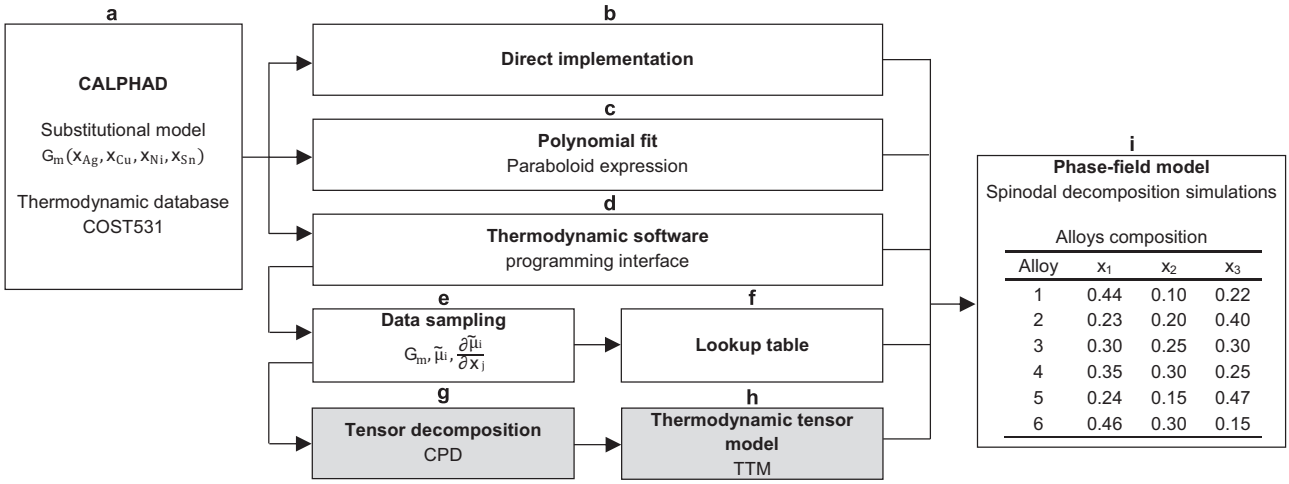


Fig. 4 Schemes coupling CALPHAD and phase-field method. Using the CALPHAD method, an expression for the bulk free energy of the existing phases **a** is coupled with the phase-field model **i** via different approaches. **b** The expression of the CALPHAD Gibbs free energy model and its derivatives with respect to the composition variables are inserted directly in the phase-field model. It is simple if all phases show the behaviour of a substitutional solution. However, most solid phases are described using a sublattice model for which it is complex to relate the phase-field and the CALPHAD variables. **c** A paraboloid expression fitted to data calculated with the CALPHAD method is used to approximate the composition dependence of the Gibbs free energy. However, this approximation can only describe the Gibbs free energy accurately over limited composition ranges. Moreover, for higher-order systems, it becomes hard to prevent molar fractions from taking nonphysical values below 0 or above 100%. **d** External software is used to evaluate thermodynamic quantities as required by the phase-field simulation. The time spent in the communication between software is the main disadvantage of this approach, making it less efficient than **a**. **e**, **f** The sampling of data as a function of composition using a thermodynamic software into lookup tables that are consulted in the phase-field simulation, which is heavily affected by the curse of dimensionality. **g** Proposed methodology in this work: a canonical polyadic decomposition is applied to describe efficiently the thermodynamic quantities, resulting in factor matrices **h** from which the thermodynamic quantities can be evaluated in PFM. The alloy compositions considered for the phase-field simulations are given in **i**.

the direct implementation of the CALPHAD model. More results can be found in the Supplementary Information.

For a more quantitative comparison, the corresponding probability density functions (PDF) of the molar fraction fields are visualised in Fig. 5b, d, f. The distributions are bimodal with two means, m_1 and m_2, corresponding to the different compositions of the co-existing liquid phases. According to the Langer-Bar-On-Miller (LBM) method,[42–44] the microstructure characteristics can be quantified based on the amplitude A, defined as $A = |m_2 - m_1|$ (see Methods subsection Validation). For all six alloys, the bimodal PDFs obtained from the simulations using the TTM with rank value $R = 6$, coincide with those obtained evaluating the full CALPHAD model expressions. When using the TTM $R = 4$, a deviation of the PDF from that obtained for CE is observed in some cases, see Fig. 5d. These deviations are also reflected in the values of the amplitude A. Therefore, the relative error on the values obtained for the amplitude A from those measured from the profiles obtained using the direct implementation of the CALPHAD model (CE) are considered as an appropriate measure to quantify the accuracy of the phase-field simulations using TTMs with a different rank value R.

In Fig. 6a, the CDFs of the RE on the amplitudes of the bimodal distributions of the molar fraction values are plotted for varying rank. Up to rank value $R = 6$, the accuracy improves when the rank value of the TTM is increased. For $R \geq 6$, no further improvement of the accuracy is obtained. The analysis of the IQR given in Fig. 6b indicates a low dispersion in the relative error, even for $R < 6$. The $Q = 0.95$ and $Q = 0.98$ quantile show that even when approaching the upper limit of the distribution, high accuracy is maintained.

As an additional comparison, the volume fraction of one of the phases is measured as a function of time from the simulations. In Fig. 6c, d, results from simulations using the full CALPHAD expression (CE) are plotted along results from simulations using TTMs with rank values $R = 3, \ldots, 6$. The volume fractions obtained using a TTM with rank value $R = 6$ are almost identical to those obtained with the direct implementation of the CALPHAD expression in the PFM. However, if the main interest is in the volume fraction measurements, a TTM with rank value $R = 5$ or even $R = 4$ may already give sufficient accuracy (see Fig. 5c, d). Depending on the application and the desired accuracy, a lower rank value R can thus be chosen to limit the number of coefficients in the TTM and the computation time needed to extract data from the tensor by evaluation of equation (3).

In conclusion, exploiting the low-rank tensor character of thermodynamic properties for higher-order multicomponent systems, a highly accurate and compact representation of their composition dependence can be obtained, which can be evaluated easily and efficiently in microstructure evolution simulations. Although tensor decomposition and tensor completion can be applied as a purely data-driven method, we illustrated that there is the flexibility to also account for a priori knowledge on the models or physical phenomena behind the data by applying functional constraints to the factor vectors. It is possible to extend the proposed method to other types of thermodynamic solution models, to include more elements or the dependence on other variables than composition, or to represent different types of composition-dependent phase properties, such as diffusion mobilities, other kinetic coefficients, interface properties, and mechanical properties. The rank value may be slightly higher for more complex dependencies. It is also possible to adjust the degree of the polynomial constraint or even use different types of constraints to model thermodynamic functions with certain features. The statement that the number of coefficients in the CPD depends only linearly on the number of components in the system is an intrinsic property of the CPD and will thus remain valid. The important advantage is that the efficiency and lower complexity is not compromised when adding more elements or variables, which opens up the possibility for detailed higher-order

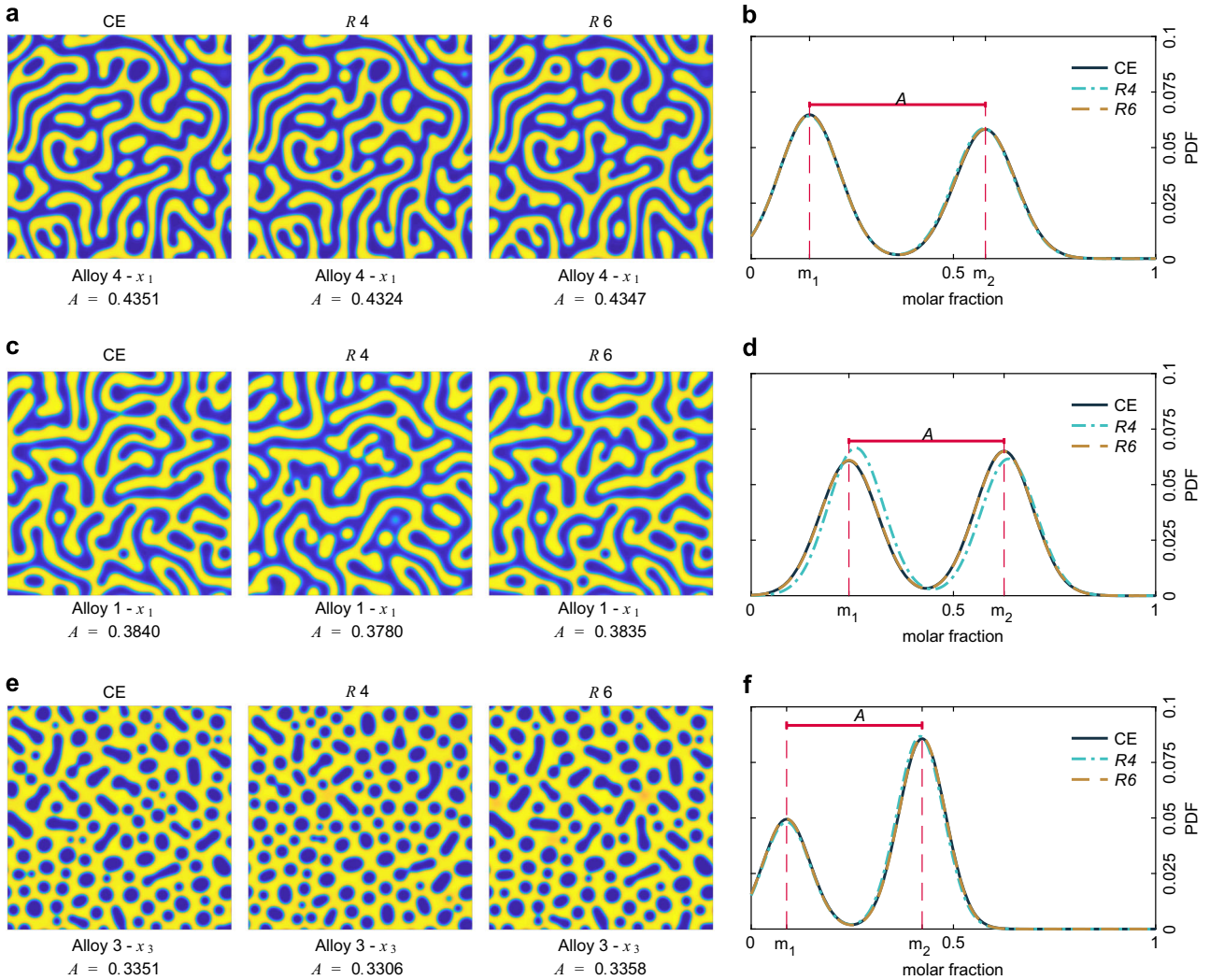


Fig. 5 2D microstructures and probability distribution function of the molar fraction profiles. a, **c**, **e** show a selection of microstructures resulting from the 2D simulations of three different alloys for visual comparison. The first column shows the microstructures obtained with direct use of the CALPHAD full expressions in the phase-field model. The second and third column show the results obtained using the coupling through decomposed TTMs for rank values $R = 4$ and $R = 6$. Each row represents one alloy and the molar fraction of one component as described by the label. **a**, **c**, **e** show the PDF of the molar fraction values in the different grid cells of the numerical domain. The two means m_1 and m_2 of the bimodal distribution are indicated. The amplitude A of the molar fraction variation in the microstructure is defined as the difference between m_1 and m_2.

multicomponent microstructure evolution studies, essential for future alloy development.

METHODS

CALPHAD thermodynamic model

Ag-Cu-Ni-Sn liquid alloys are considered in this study. According to the CALPHAD method,[10–12,45] the required thermodynamic quantities are described assuming a random substitutional solution model. They are a function of $C = 4$ molar fractions x_i, with $i = 1, \ldots, C$, and the temperature T. The molar Gibbs free energy consists of three parts

$$G_{\mathrm{m}} = G^{\mathrm{o}} + G^{\mathrm{id}}_{\mathrm{mix}} + G^{\mathrm{xs}}_{\mathrm{mix}}, \tag{4}$$

in which G^{o} refers to the Gibbs free energy of a reference surface, $G^{\mathrm{id}}_{\mathrm{mix}}$ to the contribution from ideal random mixing (i.e., owing to the configurational entropy), and $G^{\mathrm{xs}}_{\mathrm{mix}}$ to the excess energy owing to deviations from ideal behavior. The term G^{o} is given by

$$G^{\mathrm{o}} = \sum_{i=1}^{C} x_i G^{\mathrm{o}}_i, \tag{5}$$

where G^{o}_i is the temperature-dependent molar Gibbs free energy of pure component i with respect to the SGTE[46] reference state. $G^{\mathrm{id}}_{\mathrm{mix}}$ is given by

$$G^{\mathrm{id}}_{\mathrm{mix}} = \overline{R}T \sum_{i=1}^{C} x_i \log x_i, \tag{6}$$

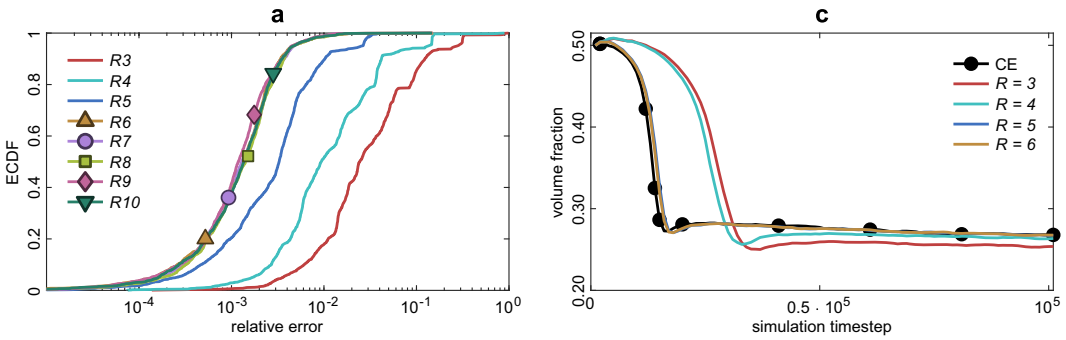
in which $\overline{R}$ is the ideal gas constant. $G^{\mathrm{xs}}_{\mathrm{mix}}$ is expressed using a Redlich–Kister–Muggianu polynomial[47]

$$G^{\mathrm{xs}}_{\mathrm{mix}} = \sum_{i=1}^{C}\sum_{j>i}^{C} x_i x_j \sum_{v} L^{(v)}_{ij}(x_i - x_j)^v + \sum_{i=1}^{C}\sum_{j>i}^{C}\sum_{k>j}^{C} x_i x_j x_k \left(x_i L^{(0)}_{ijk} + x_j L^{(1)}_{ijk} + x_k L^{(2)}_{ijk}\right), \tag{7}$$

in which $L^{(v)}_{ij}$ are binary interaction parameters and $L^{(0,1,2)}_{ijk}$ ternary interaction parameters.

As the molar fractions should always sum to one, i.e., $\sum_{i=1}^{C} x_i = 1$, one molar fraction, e.g., the Cth, is dependent on the other fractions, i.e., $x_C = 1 - \sum_{n=1}^{N} x_n$. Therefore, the molar Gibbs free energy can be expressed as a function of $N = C - 1$ molar fractions x_n, with $n = 1, \ldots, N$, and the temperature T.

The diffusion potential of the nth component, needed in multicomponent phase-field simulations, is given by the first partial derivative



b

Rank	IQR	Q = 0.95	Q = 0.98
3	4.32×10^{-2}	2.76×10^{-1}	3.17×10^{-1}
4	2.09×10^{-2}	1.34×10^{-1}	1.48×10^{-1}
5	3.93×10^{-3}	2.59×10^{-2}	2.90×10^{-2}
6	1.58×10^{-3}	4.66×10^{-3}	7.93×10^{-3}
7	1.65×10^{-3}	4.80×10^{-3}	7.70×10^{-3}
8	1.64×10^{-3}	4.65×10^{-3}	7.93×10^{-3}
9	1.38×10^{-3}	4.48×10^{-3}	7.44×10^{-3}
10	1.64×10^{-3}	4.69×10^{-3}	7.78×10^{-3}

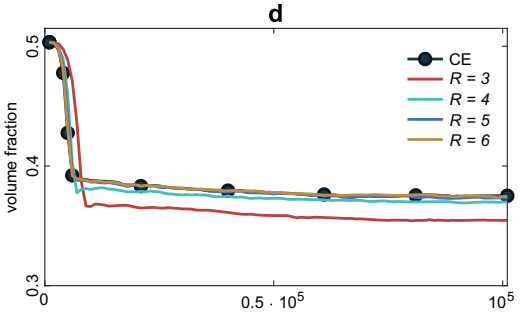


Fig. 6 Rank dependence of the accuracy of the 2D simulations using thermodynamic tensor models. a The ECDF of the relative errors on the amplitude of the molar fraction variation plotted in different line colours for TTM rank values $R = 3, \ldots, 10$. For $R = 6$ and above, markers are also used for better identification of overlapping lines. **b** Table showing the interquartile region (IQR) and quantiles $Q = 0.95$ and $Q = 0.98$ of the relative errors for rank values $R = 3, \ldots, 10$. **c**, **d** Evolution of the volume fraction of the liquid phases as a function of time steps obtained from simulations using directly the CALPHAD model expressions (CE) and using TTMs with rank values $R = 3, \ldots, 6$. No further improvement in the accuracy of the simulations using TTMs is obtained when using a rank value higher than $R = 6$. It should be noticed, however, that $R = 5$ already gives high accuracy. Moreover, for certain applications, lower rank values might already give sufficient accuracy.

of the molar Gibbs free energy with respect to x_n,

$$\tilde{\mu}_n = \frac{\partial G_\mathrm{m}}{\partial x_n}, \tag{8}$$

which is related to the chemical potentials μ_n of component n and μ_C of component C, as defined in standard solution thermodynamics, as[14]

$$\tilde{\mu}_n = \mu_n - \mu_C. \tag{9}$$

The partial derivatives of the diffusion potential of the nth component with respect to x_m,

$$\frac{\partial \tilde{\mu}_n}{\partial x_m} = \frac{\partial \mu_n}{\partial x_m} - \frac{\partial \mu_C}{\partial x_m} = \frac{\partial^2 G_\mathrm{m}}{\partial x_n \partial x_m}, \tag{10}$$

are often necessary to numerically solve phase-field equations.

Inspection of the different terms in equation (4) shows that only the contributions G^o and $G^\mathrm{xs}_\mathrm{mix}$ contain model parameters from the thermodynamic database, namely G^o_i, $L^{(v)}_{ij}$, and $L^{(0,1,2)}_{ijk}$. As these parameters may not be accessible by the user or the PFM, the terms G^o and $G^\mathrm{xs}_\mathrm{mix}$ must be modelled. The contribution $G^\mathrm{id}_\mathrm{mix}$ given by equation (6), however, is a function of known variables only, namely the independent molar fractions x_n and the temperature T. Therefore, the contribution of $G^\mathrm{id}_\mathrm{mix}$ to equation (4) can be calculated at any moment and for any entry of the TDT without consulting the thermodynamic database. As the logarithmic dependence on the molar fractions of this term complicates the formulation of an accurate TTM, we define a starred version of the Gibbs free energy, diffusion potentials, and their partial derivatives with respect to the molar fractions in which the contribution of $G^\mathrm{id}_\mathrm{mix}$ is not included:

$${}^*G_\mathrm{m} = G_\mathrm{m} - \overline{R}T\left[\left(\sum_{n=1}^{N} x_n \log x_n\right) + x_C \log x_C\right], \tag{11}$$

$${}^*\tilde{\mu}_n = \tilde{\mu}_n - \overline{R}T(\log x_n - \log x_C), \tag{12}$$

$$\frac{\partial\, {}^*\tilde{\mu}_n}{\partial x_m} = \begin{cases} \dfrac{\partial \tilde{\mu}_n}{\partial x_m} - \overline{R}T\left(\dfrac{1}{x_n} + \dfrac{1}{x_C}\right), & n = m \\ \dfrac{\partial \tilde{\mu}_n}{\partial x_m} - \overline{R}T\left(\dfrac{1}{x_C}\right), & n \neq m. \end{cases} \tag{13}$$

Equations (5) and (7) show that ${}^*G_\mathrm{m}$, ${}^*\tilde{\mu}_n$, and $\frac{\partial\, {}^*\tilde{\mu}_n}{\partial x_m}$ depend polynomially on the molar fractions of the different components. As discussed in Methods section Thermodynamic tensor model, only these polynomial contributions to the TTMs are modelled, knowing that equation (7) for $G^\mathrm{id}_\mathrm{mix}$ and its derivatives can be evaluated and added during the phase-field simulation.

For the simulations, temperature-dependent expressions for all binary and ternary interaction parameters $L^{(v)}_{ij}$, $L^{(0,1,2)}_{ijk}$, and the Gibbs free energies G^o_i of the pure elements for the liquid phase are obtained from the COST 531 thermodynamic database.[41]

Tensor decomposition

A CPD writes a Nth-order TDT $\mathcal{T}$ as a sum of R rank-1 terms,[28,33,34] each of which is the outer product, denoted by $\otimes$, of N nonzero factor vectors $\mathbf{a}^{(n)}_r$, with $r = 1, \ldots, R$ and $n = 1, \ldots, N$:

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}^{(1)}_r \otimes \cdots \otimes \mathbf{a}^{(N)}_r. \tag{14}$$

Elementwise, this means that each entry indexed with i_n, for $n = 1, \ldots, N$, of $\mathcal{T}$ is given by

$$\mathcal{T}(i_1, \ldots, i_N) = \sum_{r=1}^{R} a^{(1)}_{i_1,r} \ldots a^{(N)}_{i_N,r}. \tag{15}$$

Collecting all factor vectors $\mathbf{a}^{(n)}_r$ into factor matrices $\mathbf{A}^{(n)} = \left[\mathbf{a}^{(n)}_1, \ldots, \mathbf{a}^{(n)}_R\right]$, the CPD can be compactly written as

$$\mathcal{T} = [\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]. \tag{16}$$

There exist various techniques to determine the coefficients in the factor matrices based on a limited set of data entries of the original tensor.[28] Moreover, when available, a priori knowledge on the model or physical process behind the data can be taken into account, resulting in an even more compact and continuous tensor model of the data. In this work, for example, we exploited the polynomial dependence of ${}^*G_\mathrm{m}$ on the molar fractions, and the fact that the ${}^*\tilde{\mu}_n$ and $\frac{\partial\, {}^*\tilde{\mu}_n}{\partial x_m}$ are the first and second partial derivatives of ${}^*G_\mathrm{m}$. A polynomial constraint could therefore be applied on the factor vectors and the CPDs for ${}^*G_\mathrm{m}$ and its first and second partial

derivatives ${}^*\tilde{\mu}_n$ and $\frac{\partial\,{}^*\tilde{\mu}_n}{\partial x_m}$ could be coupled within a single tensor model, as discussed in Methods Thermodynamic tensor model.

Thermodynamic tensor model

We first show how a TDT can be written as a CPD in which each factor vector $\mathbf{a}_r^{(n)}$ is a discretized smooth function in a single variable x_n, with $n = 1, \ldots, N$. Moreover, these functions are linear combinations of a set of basis functions common to each factor vector. This way, the problem of modeling all TDTs is recast into finding N coefficient matrices. The method is built upon two mild assumptions: separability and a linear combination of basis functions.

The first assumption is that the underlying function $f(x_1, \ldots, x_N)$ is a sum of separable functions, i.e., the terms can be written as the product of N functions in a single variable:

$$f(x_1, \ldots, x_N) = \sum_{r=1}^{R} f_r^{(1)}(x_1) \cdots f_r^{(N)}(x_N), \tag{17}$$

in which R is preferably low. This is exactly the continuous formulation of equation (15). For example, $G_\mathrm{m}(x_1, \ldots, x_N)$ is not separable with low R because of the term $G_\mathrm{mix}^\mathrm{id}$ (equation (6)), which contains $x_C \log x_C$ with $x_C = 1 - \sum_{n=1}^{N} x_n$, but ${}^*G_\mathrm{m}(x_1, \ldots, x_N)$ is, as it is the sum of products of polynomials in x_n. As $G_\mathrm{mix}^\mathrm{id}$ (equation (6)) contains only known constants, we can subtract this term and work with ${}^*G_\mathrm{m}(x_1, \ldots, x_N)$ (equation (11)) without loss of generality. If the function f, and therefore also each $f_r^{(n)}(x_n)$, is evaluated on (a subset of points of) an N-dimensional grid, an (incomplete) tensor is obtained. For example, in this paper, an equidistant grid is defined by $\mathbf{x}_1 = \cdots = \mathbf{x}_N = [\,\delta_x \quad 2\delta_x \quad \ldots \quad (1 - N\delta_x)\,]^\top$ for some δ_x and the TDT ${}^*\mathcal{G}$ is ${}^*G_\mathrm{m}(x_1, \ldots, x_N)$ evaluated on this grid. If we collect the functions $f_r^{(n)}(x_n)$ evaluated in $\mathbf{x}_n$ as columns of $\mathbf{A}^{(n)}$, $n = 1, \ldots, N$, we obtain the CPD:

$${}^*\mathcal{G} = [\![\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)}]\!], \tag{18}$$

where ${}^*\mathcal{G}$ is the tensor containing the values ${}^*G_\mathrm{m}(x_1, \ldots, x_N)$. Similarly, the TDTs for ${}^*\tilde{\mu}_n$ and $\frac{\partial\,{}^*\tilde{\mu}_n}{\partial x_m}$ can be written as a CPD. Note that the grid points do not need to be chosen in the same way for different components n, nor do they have to be equidistant.

Second, we assume that each $f_r^{(n)}$ can be expressed as a linear combination of K basis functions $b_k^{(n)}(x_n)$, $k = 1, \ldots, K$:

$$\begin{aligned} f_r^{(n)}(x_n) &= u_{1,r}^{(n)} b_1^{(n)}(x_n) + \cdots + u_{K,r}^{(n)} b_K^{(n)}(x_n) \\ &= \left[\, b_1^{(n)}(x_n) \ldots\, b_K^{(n)}(x_n) \right] \left[\, u_{1,r}^{(n)} \ldots\, u_{K,r}^{(n)} \right]^\top. \end{aligned} \tag{19}$$

By evaluating $f_r^{(n)}(x_n)$ on the grid, we obtain

$$\mathbf{A}^{(n)} = \left[\, \mathbf{b}_1^{(n)} \ldots\, \mathbf{b}_K^{(n)} \right] \mathbf{U}^{(n)} = \mathbf{B}^{(n)} \mathbf{U}^{(n)}, \tag{20}$$

in which $\mathbf{B}^{(n)}$ collects the basis functions evaluated in $\mathbf{x}_n$ and $\mathbf{U}^{(n)}$ the coefficients. As the basis functions are fixed, only the coefficients $\mathbf{U}^{(n)}$ need to be determined. In this paper, we choose a basis of monomial functions evaluated in the same points in all dimensions, i.e., $b_k^{(n)}(x_n) = b_k(x) = x^{(k-1)}$ for $k = 1, \ldots, d+1$, and $n = 1, \ldots, N$, which is reasonable given the polynomial nature of the theoretical models; see equations (11), (12), and (13). This way, each TDT can be written as a CPD in which each factor matrix $\mathbf{A}^{(n)}$ is subject to a linear constraint:

$${}^*\mathcal{G} = [\![\mathbf{B}\mathbf{U}^{(1)}, \ldots, \mathbf{B}\mathbf{U}^{(N)}]\!] \quad \text{with} \quad \mathbf{B} = \left[\, \mathbf{1} \quad \mathbf{x} \quad \mathbf{x}^2 \quad \ldots \quad \mathbf{x}^d \,\right] \tag{21}$$

in which the powers are elementwise. Note that the coefficient matrices $\mathbf{U}^{(n)}$ do not necessarily contain the coefficients G_i^o, $L_{ij}^{(v)}$, and $L_{ijk}^{(0,1,2)}$ as their entries.

A key property following from the definition of the TDTs and the CPD formulation with linear constraints is that computing derivatives only requires the computation of derivatives of the known basis functions $b_k(x_n)$:

$$\frac{\mathrm{d}f_r^{(n)}(x_n)}{\mathrm{d}x_n} = u_{1,r}^{(n)} \frac{\mathrm{d}b_1(x_n)}{\mathrm{d}x_n} + \cdots + u_{K,r}^{(n)} \frac{\mathrm{d}b_K(x_n)}{\mathrm{d}x_n}. \tag{22}$$

For example, as ${}^*\tilde{\mu}_1 = \frac{\partial\,{}^*G_\mathrm{m}}{\partial x_1}$, the corresponding TDT ${}^*\mathcal{P}^{(1)}$ can be represented as

$${}^*\mathcal{P}^{(1)} = [\![\dot{\mathbf{A}}^{(1)}, \mathbf{B}\mathbf{U}^{(2)}, \ldots, \mathbf{B}\mathbf{U}^{(N)}]\!] \quad \text{with} \quad \dot{\mathbf{A}}^{(1)} = \dot{\mathbf{B}}\mathbf{U}^{(1)}, \tag{23}$$

as only $\mathbf{A}^{(1)}$ depends on x_1. $\dot{\mathbf{B}}$ contains the evaluated derived functions $\dot{b}_k(x) = (k-1)x^{k-1}$:

$$\dot{\mathbf{B}} = \left[\, \mathbf{0} \quad \mathbf{1} \quad 2\mathbf{x} \quad \ldots \quad d\mathbf{x}^{d-1} \,\right]. \tag{24}$$

Similarly, the derivatives of the potentials depend on $\dot{\mathbf{B}}$ and $\ddot{\mathbf{B}}$, in which the latter matrix contains the second-order derivatives:

$$\ddot{\mathbf{B}} = \left[\, \mathbf{0} \quad \mathbf{0} \quad 2\times\mathbf{1} \quad 6\mathbf{x} \quad \ldots \quad d(d-1)\mathbf{x}^{d-2} \,\right]. \tag{25}$$

For example, the CPDs of the TDTs corresponding to $\frac{\partial\,{}^*\tilde{\mu}_1}{\partial x_1}$ and $\frac{\partial\,{}^*\tilde{\mu}_1}{\partial x_2}$ are given by

$${}^*\mathcal{D}^{(1,1)} = [\![\ddot{\mathbf{B}}\mathbf{U}^{(1)}, \mathbf{B}\mathbf{U}^{(2)}, \ldots, \mathbf{B}\mathbf{U}^{(N)}]\!], \tag{26}$$

$${}^*\mathcal{D}^{(1,2)} = [\![\dot{\mathbf{B}}\mathbf{U}^{(1)}, \dot{\mathbf{B}}\mathbf{U}^{(2)}, \ldots, \mathbf{B}\mathbf{U}^{(N)}]\!], \tag{27}$$

respectively. Hence, once the coefficient matrices $\mathbf{U}^{(n)}$, $n = 1, \ldots, N$, are determined, all TDTs are modeled. Moreover, as the coefficients are independent of the grid, different grids can be used in the training phase, i.e., to find $\mathbf{U}^{(n)}$ and when actually using the model, i.e., during the phase-field simulations.

Thermodynamic tensor model training

In the optimisation routine for modeling the TDTs of a quaternary system ($N = 3$), we want to find the coefficients in the matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$ that minimise the least squares cost function

$$\begin{aligned} \min_{\mathbf{U}^{(1)},\mathbf{U}^{(2)},\mathbf{U}^{(3)}} \quad & \tfrac{1}{2} \left\| \mathcal{S} * \left({}^*\mathcal{P}_\mathrm{T}^{(1)} - [\![\dot{\mathbf{B}}\mathbf{U}^{(1)}, \mathbf{B}\mathbf{U}^{(2)}, \mathbf{B}\mathbf{U}^{(3)}]\!] \right) \right\|^2 \\ & + \tfrac{1}{2} \left\| \mathcal{S} * \left({}^*\mathcal{P}_\mathrm{T}^{(2)} - [\![\mathbf{B}\mathbf{U}^{(1)}, \dot{\mathbf{B}}\mathbf{U}^{(2)}, \mathbf{B}\mathbf{U}^{(3)}]\!] \right) \right\|^2 \\ & + \tfrac{1}{2} \left\| \mathcal{S} * \left({}^*\mathcal{P}_\mathrm{T}^{(3)} - [\![\mathbf{B}\mathbf{U}^{(1)}, \mathbf{B}\mathbf{U}^{(2)}, \dot{\mathbf{B}}\mathbf{U}^{(3)}]\!] \right) \right\|^2. \end{aligned} \tag{28}$$

${}^*\mathcal{P}_\mathrm{T}^{(1)}$, ${}^*\mathcal{P}_\mathrm{T}^{(2)}$, and ${}^*\mathcal{P}_\mathrm{T}^{(3)}$ are training TDTs, which are computed with a step size of $\delta_x = 0.025$ using the TC-Toolbox (Thermo-Calc 2017b) with the COST 531 thermodynamic database[41] at 1400 K, each with 11,482 entries. The residual between the tensor and its low-rank, polynomial-constrained CPD model, e.g., ${}^*\mathcal{P}_\mathrm{T}^{(1)} - [\![\dot{\mathbf{B}}\mathbf{U}^{(1)}, \mathbf{B}\mathbf{U}^{(2)}, \mathbf{B}\mathbf{U}^{(3)}]\!]$, is multiplied elementwise (denoted by $*$) with a binary tensor $\mathcal{S}$ which is one for each computed training entry and zero otherwise, such that only the limited set of training entries contribute to the objective function. $\mathcal{S}$ can be different for each term of the objective problem during the optimisation, but here we choose to keep it constant. Only the diffusion potentials are used in the optimisation routine since other approaches, e.g., using ${}^*\mathcal{G}$ or a combination of all TDTs, have shown to be less accurate. Therefore, as only derivatives are used, the Gibbs free energy is determined up to a constant a, which can be estimated as the mean of the residual and must be added to the computed CPD for the Gibbs free energy.

In this work, a conservative, i.e., small enough, value for δ_x is chosen when selecting the training data to guarantee a highly accurate model. No further investigations are conducted on the influence of the step size used for the training data on the accuracy of the tensor model or phase-field simulations. Various other aspects can affect the accuracy of the tensor model, such as the degree of the polynomial constraint, the rank value of the CPD and the distribution of the training data over the system. However, the grid with which data are sampled is not very relevant, as we are interested in the optimisation of the coefficient in the $\mathbf{U}^{(n)}$ matrices and not the factors in the $\mathbf{A}^{(n)}$ matrices. The factors in the latter are only computed after the optimisation routine with equation (20) and the grid spacing can be chosen according to the necessity. This procedure also assures that the thermodynamic quantities are accurately modelled over the entire composition domain (including the dilute solution regions, where at least one of the molar fractions takes a very low value).

The least squares cost function is optimised in Tensorlab[40] using the data independent algorithm for the CPD of incomplete tensors subject to linear constraints (CPDLI DI).[31] A few decompositions are computed starting from random initial guesses for $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$, given R and d, and the best result is retained. CPDs for rank values $R = 3, \ldots, 10$ are computed. (Models with maximal degrees per variable $d = 3, 4, 5$ have been computed as well to validate that the correct degree $d = 4$ can indeed be recovered. This degree $d = 4$ can be obtained from equation (4) and the parameters obtained from the thermodynamic database COST 531.[41])

Once the matrices with the unknown coefficients $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$ are found, the factor matrices $\mathbf{A}^{(n)}$, $\dot{\mathbf{A}}^{(n)}$ and $\ddot{\mathbf{A}}^{(n)}$, for $n = 1, 2, 3$, are

constructed using equation (20) and basis matrices $\mathbf{B}$, $\dot{\mathbf{B}}$, $\ddot{\mathbf{B}}$, which are evaluated on a grid with $\delta_x = 0.0001$.

Validation

The TC-toolbox (Thermo-Calc 2017b) and MATLAB are used to generate validation TDTs (indicated with subscript V) of the Gibbs free energy $\mathcal{G}_V$, diffusion potentials $\mathcal{P}_V^{(1)}$, $\mathcal{P}_V^{(2)}$, and $\mathcal{P}_V^{(3)}$ and of the derivatives of the diffusion potentials $\mathcal{D}_V^{(1,1)}$, $\mathcal{D}_V^{(2,2)}$, $\mathcal{D}_V^{(3,3)}$, $\mathcal{D}_V^{(1,2)}$, $\mathcal{D}_V^{(2,3)}$, and $\mathcal{D}_V^{(1,3)}$ of the liquid phase at 1400 K, with a step size of $\delta_x = 0.0001$ and using the COST 531 thermodynamic database.[41]

The accuracy of the TTMs with rank value $R = 3, \ldots, 10$ is verified based on the RE on the entries of the validation TDTs. For example, for the Gibbs free energy tensor, the relative error $E_R^{(\mathcal{G})}$ given by

$$E_R^{(\mathcal{G})} = \left| \frac{\mathcal{G}_V(i_1,i_2,i_3) - \left[\sum_{r=1}^{R} a_{i_1,r}^{(1)} a_{i_2,r}^{(2)} a_{i_3,r}^{(3)} + \overline{R}T\left(\sum_{i=1}^{N} x_i \log x_i + x_C \log x_C\right) + a\right]}{\mathcal{G}_V(i_1,i_2,i_3)} \right|. \tag{29}$$

The contribution from the ideal mixing term $G_{\text{mix}}^{\text{id}}$ is calculated separately and added to the values calculated from the TTMs, which only include the ${}^{*}G_{\text{m}}$ in equation (11) (see Methods section CALPHAD thermodynamic model). As discussed in Methods Thermodynamic tensor model training, a constant value a has to be added to the Gibbs energy values computed from the TTMs. Analogously, the relative error $E_R^{(\mathcal{P})}$ of the diffusion potentials and $E_R^{(\mathcal{D})}$ of the partial derivatives of the diffusion potentials with respect to the molar fractions are calculated including their ideal mixing contributions: see equations (12) and (13). As a is a constant, it does not appear in the expressions for the diffusion potentials and their partial derivatives with respect to the molar fractions.

An ECDF is then determined for each rank value R considering all the values obtained for $E_R^{(\mathcal{G})}$, $E_R^{(\mathcal{P})}$, and $E_R^{(\mathcal{D})}$ together. The MATLAB fuction `ecdf` is used for this. The ECDFs are plotted in Fig. 3a; the IQR, and quantiles $Q = 0.95$ and $Q = 0.98$ are presented in Fig. 3b. Additional quantile values can be found in the Supplementary Information.

The validation of the 2D simulations using TTMs is conducted based on a comparison with simulations performed using the direct implementation of the CALPHAD model (approach b in Fig. 4). The reason for this choice is the fact that most of the alternatives, presented in Fig. 4, cannot be applied to a quaternary system owing to certain limitations. For instance, in the approach that uses polynomial fitting (Fig. 4c), only a narrow region of the thermodynamic function can be accurately described and since diffusion paths can become complex for multicomponent systems, it is difficult to identify which composition region will be most relevant for a simulation. Moreover, polynomial fits cannot prevent molar fractions from taking an unphysical value below zero or above one. Although this can be avoided in binary system, by designing the simulation in such a way that the molar fraction does not take a negative value, when considering four or more elements, there is always a time or location in the simulation where one of the molar fraction fields goes below zero, making the remainder of the simulation unstable. The approach that uses lookup tables (Fig. 4f) is also unfeasible in this case owing to the size of the sampled data sets given the required high resolution, as seen in (Fig. 1a). These approaches, however, are suitable and straightforward to be applied to binary or ternary systems, whereas the proposed use of tensor completion requires at least three independent variables in the thermodynamic descriptions.

The approach that uses the direct implementation (Fig. 4b) of the CALPHAD model expression and the one that uses an interface with a thermodynamic software (Fig. 4d) can be applied for quaternary and higher-order systems. With our implementation of the method that uses an interface with a thermodynamic software we could only run 1D simulations in a reasonable amount of time (these are presented in the Supplementary Information). It should be mentioned that optimisations (mainly based on the idea of interpolation between previously calculated data) of phase-field codes using an interface with a thermodynamic software have been reported in the literature;[18,19] however, to the best of our knowledge, only for binary and ternary systems.

For the method using the direct implementation of the CALPHAD model expression (CE), we found that (for our implementation) the efficiency is similar to that using the TTM for the quaternary system. However, such a comparison is often not relevant. The direct implementation of the CALPHAD model expression requires access to an open thermodynamic database (while today, the most advanced thermodynamic databases are commercial). Moreover, if the phase under study is modelled using sublattices, it is mostly impossible (except for some special cases) to relate the phase-field molar fractions with the site fractions in the CALPHAD expression. Hence, in these cases, the direct implementation of the CALPHAD model expression is not even possible, whereas the tensor model can still be formulated. We also observed that the current implementation of the TTM in the phase-field model is memory bound, which can be addressed with further optimisation studies. Theoretically, the TTM method should be cheaper to evaluate than CE.

For the validation of the 2D simulation, the LBM method is used to estimate the amplitudes of the molar fraction distributions of the simulated microstructures during spinodal decomposition and consequent coarsening.[42–44] For each simulation, the microstructures obtained at every thousandth time step are considered. For each microstructure, a Gaussian mixture model is used to fit the distributions of the molar fraction values for each of the three independent components with the MATLAB function `gmdistribution`. A bimodal distribution is obtained with two means m_1 and m_2, corresponding to the compositions of the two liquid phases, as shown in Fig. 5b–d. The amplitude A is then calculated as the difference between these two means, i.e., $A = |m_1 - m_2|$. The amplitude A^R from the simulation using the TTM with rank value R is compared with the amplitude A obtained from the simulation using the direct implementation of the CALPHAD model (approach b in Fig. 4). The relative error E_R^{2D} is defined by

$$E_R^{2D} = \left| \frac{A - A^R}{A} \right|, \tag{30}$$

and is determined for each molar fraction field and each alloy composition and for rank values $R = 3, \ldots, 10$. Then, the ECDF is calculated for each rank value R, considering all the RE determined for the different alloy compositions and the three independent molar fraction fields, obtained for the given rank value R, together. The ECDFs are plotted in Fig. 6a and the information of the IQR, $Q = 0.95$ and $Q = 0.98$ are presented in Fig. 6b.

In addition, the volume fractions of the two liquid phases are measured at every thousandth time step as follows. First, the mean of one of the molar fraction distributions is computed. Then, the grid cells of the simulation domain are divided into two groups, one containing the grid cells with a molar fraction value greater than the mean and the other with the grid cells with a molar fraction value smaller than the mean. The volume fractions of the two phases are then obtained by dividing the number of grid cells in each group by the total number of cells in the simulation domain. This information is presented in Fig. 6c, d.

Phase-field microstructure evolution model

Assuming a volume-fixed frame of reference [48], the temporal and spatial evolution of the non-equilibrium microstructure of the quaternary Ag-Cu-Ni-Sn system ($C = 4$), is described with $N = C - 1 = 3$ conserved field variables, namely the molar fractions $x_1(\mathbf{r}, t)$, $x_2(\mathbf{r}, t)$, and $x_3(\mathbf{r}, t)$ of silver, copper, and nickel, respectively. The molar fraction of tin is then obtained as $x_C = 1 - x_1 - x_2 - x_3$.

Following the Cahn-Hilliard[49] formalism, the non-uniform free energy functional is defined as

$$F(x_1, x_2, x_3) = \int_V \left[f_0(x_1, x_2, x_3) + \sum_n^N \frac{k_n}{2} (\nabla x_n)^2 \right] \mathrm{d}\mathbf{r}, \tag{31}$$

where $f_0(x_1, x_2, x_3)$ is the homogeneous free energy density and k_n are positive gradient energy coefficients for $n = 1, 2, 3$. The variational derivative of F with respect to x_n is given by

$$\frac{\delta F}{\delta x_n} = \frac{\partial f_0(x_1, x_2, x_3)}{\partial x_n} - k_n \nabla^2 x_n. \tag{32}$$

The partial derivative of the homogeneous free energy is obtained given the thermodynamic relation[14] in equation (9) as

$$\frac{\partial f_0}{\partial x_n} = \frac{1}{V_{\text{m}}} (\mu_n - \mu_C) = \frac{\tilde{\mu}_n}{V_m}, \tag{33}$$

with $\tilde{\mu}_n$ the diffusion potential and μ_n the chemical potential of component n ($n = 1, 2, 3$) and μ_C the chemical potential of Sn.

The diffusion equation describes the evolution of the conserved field variables for the multicomponent system[7] and is given by

$$\left(\frac{1}{V_{\text{m}}}\right) \frac{\partial x_n(\mathbf{r}, t)}{\partial t} = -\nabla \times \mathbf{J}_n, \tag{34}$$

in which V_{m} is a constant coefficient approximating the molar volume of

the alloy. $\mathbf{J}_n$ is the diffusion flux of the nth component, which is defined with the Onsager reciprocal relations[50] as

$$\mathbf{J}_n = \sum_m^N -M_{nm}\nabla\frac{\delta F}{\delta x_m}, \tag{35}$$

in which M_{nm} is the mobility of component n taken with component m as reference. As information on the diffusion coefficients for quaternary systems are scarce, equal atomic mobilities β for all components are adopted, giving[48]

$$M_{nm} = \begin{cases} \frac{1}{V_\mathrm{m}}\beta x_n(1-x_m), & n=m \\ -\frac{1}{V_\mathrm{m}}\beta x_n x_m, & n\neq m. \end{cases} \tag{36}$$

Equation (34) is implemented in MATLAB for 1D and 2D simulations using finite difference discretization with explicit Euler time stepping and a central three-point (1D) or five-point (2D) scheme for the spatial coordinates. The parameters $\beta = 10^{-12}\,\mathrm{m}^5/(\mathrm{Js})$, $V\mathrm{m} = 10^{-5}\,\mathrm{m}^3/\mathrm{mol}$, $k_\mathrm{n} = 5\times10^{-11}\,\mathrm{J/m}$ are used. A domain size of $5\times10^{-8}\,\mathrm{m}$ for the 1D and $25\times10^{-16}\,\mathrm{m}^2$ for the 2D simulations is considered with a grid spacing of $2\times10^{-10}\,\mathrm{m}$. A time step of $10^{-13}\,\mathrm{s}$ is used. For each simulation, an evolution time of 10^5 simulation time steps, equivalent to$10^{-8}\,\mathrm{s}$, is considered.

DATA AVAILABILITY

All data are kept available at KU Leuven and can be accessed upon request. Alternatively, they can be generated in limited time using the provided codes in combination with Tensorlab 3.0, MATLAB and TC-toolbox (Thermo-Calc 2017b).

CODE AVAILABILITY

All codes used to generate the data for this project are publicly available.[51]

Received: 26 August 2019; Accepted: 5 December 2019;
Published online: 07 January 2020

REFERENCES

1. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
2. Clouet, E. et al. Complex precipitation pathways in multicomponent alloys. *Nat. Mater.* **5**, 482–488 (2006).
3. Voorhees, P. W. Scandium overtakes zirconium. *Nat. Mater.* **5**, 435–436 (2006).
4. Orthacker, A. et al. Diffusion-defining atomic-scale spinodal decomposition within nanoprecipitates. *Nat. Mater.* **17**, 1101–1107 (2018).
5. Raabe, D., Li, Z. & Ponge, D. Metastability alloy design. *MRS Bull.* **44**, 266–272 (2019).
6. King, A. H. Our elemental footprint. *Nat. Mater.* **18**, 408–409 (2019).
7. Moelans, N., Blanpain, B. & Wollants, P. An introduction to phase-field modeling of microstructure evolution. *CALPHAD* **32**, 268–294 (2008).
8. Provatas, N. & Elder, K. *Phase-field methods in materials science and engineering* (Wiley online library, 2011).
9. Heulens, J., Blanpain, B. & Moelans, N. Phase-field analysis of a ternary two-phase diffusion couple with multiple analytical solutions. *Acta Mater.* **59**, 3946–3954 (2011).
10. Lukas, H.L., Fries, S.G. & Sundman, B. *Computational thermodynamics: The CALPHAD method* (Cambridge University Press, 2007).
11. Liu, Z.-K. & Wang, Y. *Computational thermodynamics of materials* (Cambridge University Press, 2016).
12. Kroupa, A. Modelling of phase diagrams and thermodynamic properties using CALPHAD method - development of thermodynamic databases. *Comp. Mater. Sci.* **66**, 3–13 (2013).
13. Kitashima, T. Coupling of the phase-field and CALPHAD methods for predicting multicomponent, solid-state phase transformations. *Philos. Mag.* **88**, 1615–1637 (2008).
14. Eiken, J., Böttger, B. & Steinbach, I. Multiphase-field approach for multicomponent alloys with extrapolation scheme for numerical application. *Phys. Rev. E* **73**, 066122 (2006).
15. Grönhagen, K., Ågren, J. & Odén, M. Phase-field modelling of spinodal decomposition in TiAlN including the effect of metal vacancies. *Scripta Mater.* **95**, 42–45 (2015).
16. Koyama, T. & Onodera, H. Computer simulation of phase decomposition in Fe-Cu-Mn-Ni quaternary alloy based on the phase-field method. *Mater. Trans.* **46**, 1187–1192 (2005).
17. Chen, Q., Ma, N., Wu, K. & Wang, Y. Quantitative phase field modeling of diffusion-controlled precipitate growth and dissolution in Ti-Al-V. *Scripta Mater.* **50**, 471–476 (2004).
18. Grafe, U., Böttger, B., Tiaden, J. & Fries, S. G. Coupling of multicomponent thermodynamic databases to a phase field model: Application to solidification and solid state transformations of superalloys. *Scripta Mater.* **42**, 1179–1186 (2000).
19. Cardon, C., LeTellier, R. & Plapp, M. Modelling of liquid phase segregation in the Uranium-Oxygen binary system. *CALPHAD* **52**, 47–56 (2016).
20. Zhu, J. Z., Liu, Z. K., Vaithyanathan, V. & Chen, L. Q. Linking phase-field model to CALPHAD: application to precipitate shape evolution in Ni-base alloys. *Scripta Mater.* **46**, 401–406 (2002).
21. Liu, H. et al. A simulation study of the shape of β' precipitates in Mg-Y and Mg-Gd alloys. *Acta Mater* **61**, 453–466 (2013).
22. Jokisaari, A. & Thornton, K. General method for incorporating CALPHAD free energies of mixing into phase field models: application to the α-zirconium/α-hydride system. *CALPHAD* **51**, 334–343 (2015).
23. Guan, Y. & Moelans, N. Influence of the solubility range of intermetallic compounds on their growth behavior in hetero-junctions. *J. Alloys Compd.* **635**, 289–299 (2015).
24. Heulens, J., Blanpain, B. & Moelans, N. A phase field model for isothermal crystallization of oxide melts. *Acta Mater.* **59**, 2156–2165 (2011).
25. Böttger, B., Eiken, J. & Apel, M. Multi-ternary extrapolation scheme for efficient coupling of thermodynamic data to a multi-phase-field model. *Comp. Mater. Sci.* **108**, 283–292 (2015).
26. Larsson, H. & Höglund, L. A scheme for more efficient usage of CALPHAD data in simulations. *CALPHAD* **50**, 1–5 (2015).
27. Tomasi, G. & Bro, R. PARAFAC and missing values. *Chemom. Intell. Lab. Syst.* **75**, 163–180 (2005).
28. Vervliet, N., Debals, O., Sorber, L. & De Lathauwer, L. Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis. *IEEE Signal Proc. Mag.* **31**, 71–79 (2014).
29. Smilde, A., Bro, R., Geladi, P. & Wiley, J. *Multi-way analysis with applications in the chemical sciences*. (Wiley Chichester, UK, 2004).
30. Liu, J., Musialski, P., Wonka, P. & Ye, J. Tensor completion for estimating missing values in visual data. *IEEE T. Pattern Anal.* **35**, 208–220 (2013).
31. Vervliet, N., Debals, O. & De Lathauwer, L. Canonical polyadic decomposition of incomplete tensors with linearly constrained factors. *ESAT-STADIUS KU Leuven, Technical Report* **16–172** (2016). ftp://ftp.esat.kuleuven.be/pub/stadius/nvervliet/vervliet2017cpdli.pdf.
32. Morral, J. E. & Chen, S. L. High entropy alloys, miscibility gaps and the rose geometry. *J. Phase Equilib. Diff.* **38**, 319–331 (2017).
33. Cichocki, A. et al. Tensor decompositions for signal processing applications: from two-way to multiway component analysis. *IEEE Signal Proc. Mag.* **32**, 145–163 (2015).
34. Sidiropoulos, N. D. et al. Tensor decomposition for signal processing and machine learning. *IEEE T. Signal Proces.* **65**, 3551–3582 (2017).
35. Khoromskij, B. N. *Tensor numerical methods in scientific computing*. (De Gruyter, Berlin, Boston, 2018).
36. Grasedyck, L., Kressner, D. & Tobler, C. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen* **36**, 53–78 (2013).
37. Khoromskij, B. Tensors-structured numerical methods in scientific computing: Survey on recent advances. *Chemometr. Intell. Lab.* **110**, 1–19 (2012).
38. Hackbusch, W. *Tensor spaces and numerical tensor calculus*, vol. 42 of *Springer series in computational mathematics* (Springer, Heidelberg, 2012).
39. Vervliet, N., Debals, O. & De Lathauwer, L. Tensorlab 3.0 âĂŤ Numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, 1733–1738 (2016).
40. Vervliet, N., Debals, O., Sorber, L., Van Barel, M. & De Lathauwer, L. *Tensorlab 3.0* (2016). https://www.tensorlab.net.
41. Kroupa, A. et al. The development of the COST 531 lead-free solders thermodynamic database. *JOM* **59**, 20–25 (2007).
42. Langer, J., Bar-On, M. & Miller, H. D. New computational method in the theory of spinodal decomposition. *Phys. Rev. A* **11**, 1417 (1975).
43. Barkar, T., Höglund, L., Odqvist, J. & Ågren, J. Effect of concentration dependent gradient energy coefficient on spinodal decomposition in the Fe-Cr system. *Comp. Mater. Sci.* **143**, 446–453 (2018).
44. Barkar, T. *Modelling phase separation in Fe-Cr alloys: A continuum approach*. Ph.D. thesis, KTH Royal Institute of Technology (2018). http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Akth%3Adiva-235394.

45. Saunders, N. & Miodownik, A.P. *CALPHAD (calculation of phase diagrams): a comprehensive guide*. No. v. 1 in Pergamon materials series (Pergamon, Oxford; New York, 1998).
46. Dinsdale, A. SGTE data for pure elements. *CALPHAD* **15**, 317–425 (1991).
47. Muggianu, Y. M., Gambino, M. & Bros, J. Enthalpies of formation of liquid alloys bismuth-gallium-tin at 723K - choice of an analytical representation of integral and partial thermodynamic functions of mixing for this ternary-system. *J. Chim. Phys. PCB* **72**, 83–88 (1975).
48. Andersson, J. & Ågren, J. Models for numerical treatment of multicomponent diffusion in simple phases. *J. Appl. Phys.* **72**, 1350–1355 (1992).
49. Cahn, J. W. & Hilliard, J. E. Free energy of a nonuniform system. I. Interfacial free energy. *J. Chem. Phys.* **28**, 258–267 (1958).
50. Onsager, L. Reciprocal relations in irreversible processes. I. *Phys. Rev.* **37**, 405 (1931).
51. Moelans, N., Coutinho, Y.A., Vervliet, N. & De Lathauwer, L. *Combining thermodynamics with tensor completion techniques to enable multicomponent microstructure prediction —codes* (2019). https://doi.org/10.17632/tjm3yjs3pc.1.

ACKNOWLEDGEMENTS

This work is supported by: (1) European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (INTERDIFFUSION, grant agreement no 714754). (2) Fonds de la Recherche Scientifique—FNRS and the Fonds Wetenschappelijk Onderzoek—Vlaanderen under EOS Project no 30468160 (SeLMA). (3) European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Advanced Grant: BIOTENSORS (no. 339804). (4) KU Leuven Internal Funds (C16/15/059). (5) KU Leuven Internal Funds (PDM/18/146).

AUTHOR CONTRIBUTIONS

Aspects related to tensor decompositions by N.V. and L.D.L. Aspects related to the thermodynamic and phase-field computations by Y.A.C. and N.M. Data-analysis by Y.A.C. and N.V.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41524-019-0268-y.

Correspondence and requests for materials should be addressed to N.M.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020