An artificial neural network model for the unary description of pure substances and its application on the thermodynamic modelling of pure iron

Maximilian Länge¹

Published online: 22 January 2020
© The Author(s) 2020

Abstract

The aim of this work is to introduce a novel approach for the universal description of the thermodynamic functions of pure substances on the basis of artificial neural networks. The proposed approximation method is able to describe the thermodynamic functions ($C_p(T)$, $S(T)$, $H(T) - H(T_{\text{ref}})$, $G(T)$) of the different phases of unary material systems in a wide temperature range (between 0 and 6000 K). Phase transition temperatures and the respective enthalpies of transformation, which are computationally determined by the minimization of the Gibbs free energy, are also approximated. This is achieved by using artificial neural networks as models for the thermodynamic functions of the individual phases and by expressing the thermodynamic quantities in terms of the free network parameters. The resulting expressions are then optimized with machine learning algorithms on the basis of measurement data. A physical basis for the resulting approximation is given by the use of, among others, Planck–Einstein functions as activation function of the neurons of the network. This article provides a description of the method and as an example of a specific application the approximation of the thermodynamic functions of the different phases of pure iron. The article focuses on the problem of the representation of thermodynamic data and their practical application.

Keywords Neural networks · Function approximation · Derivative approximation · Computational thermodynamics

1 Introduction

The approximation of thermodynamic functions is an important task in material science and engineering. A consistent description of thermodynamic data helps to understand, improve and develop materials. With methods and formalisms from the field of computational thermodynamics, e.g. the CALPHAD method (Lukas et al. 2007; Saunders and Miodownik 1998), the numerical calculation of stable phases, phase equilibria, phase transitions, whole phase diagrams and phase diagram-related data is possible. The Gibbs energy G is the central quantity of such calculations, and an exact description of G in terms of the temperature T and

the systems composition is the key for a reliable description of material systems and their properties. G and other thermodynamic quantities are related to one another, and the different quantities of interest can be expressed as sets of partial derivatives of G . Expressions for G are usually derived by approximating the temperature-dependent isobaric heat capacity C_p based on suitable models (Dinsdale 1991; Chase et al. 1995; Chen and Sundman 2001; Roslyakova et al. 2016), where the free model parameters are optimized to fit measurement data. G is then calculated by a subsequent integration of the optimized model for C_p .

Artificial neural networks (ANN) can be used for function regression. With the property of being universal function approximators as described in Hornik (1991) ANNs have the ability to map a set of independent input variables to a set of dependent output variables and thus can detect and model any linear or nonlinear relationship between these. In recent years ANNs are used to solve tasks in science and engineering in many scientific fields and among them in material science. ANNs have been used extensively to model physical properties of materials (Hemmat 2017; Hemmat Esfe

Communicated by V. Loia.

✉ Maximilian Länge
maxoid@gmail.com

¹ Institute for Applied Thermo- and Fluidynamics,
Mannheim – University of Applied Sciences,
Mannheim, Germany

et al. 2016). The latest work of Avrutskiy (2017) delivers the framework for approximating functions and their derivatives simultaneously, and it can therefore be used to solve this specific task. In the present work the question is therefore investigated if thermodynamic functions can be modelled on the basis of ANNs. To answer this question a neural network model for the approximation of thermodynamic functions is introduced. The thermodynamic functions of iron between 0 and 6000 K are approximated and validated as a challenging example.

2 Methods

2.1 Modelling of thermodynamic functions

Unary systems consist of only one compound. The thermodynamic state variables depend therefore only on the pressure p and temperature T . The Gibbs energy $G(T, p)$ is the central quantity when calculating phase diagrams or phase diagram-related data. It is given in Eq. (2.1) by

$$G = H - TS. \quad (2.1)$$

A relationship between G and the entropy S can be established by the derivation of G w.r.t. T at constant pressure as given in Eq. (2.2) by

$$S = - \left. \frac{\partial G}{\partial T} \right|_p. \quad (2.2)$$

A relationship between G and the enthalpy H can also be derived and is known as Gibb–Helmholtz equation as given in Eq. (2.3) by

$$H = -T^2 \left. \frac{\partial}{\partial T} \left(\frac{G}{T} \right) \right|_p. \quad (2.3)$$

The isobaric heat capacity C_p , in the following denoted as C , can also be derived from the Gibbs energy and is given in Eq. (2.4) by

$$C = -T \left. \frac{\partial^2 G}{\partial T^2} \right|_p. \quad (2.4)$$

Equations (2.1)–(2.4) can be used to characterize a material system. In the present work this resulting set of partial derivatives is modelled and solved on the basis of artificial neural networks.

$G(T, p)$ as a thermodynamic potential can be used to calculate the stable phase for a chosen state, phase equilibria and

whole phase diagrams. A detailed description of the underlying physical principles is above the scope of this work.

2.2 Neural networks for the approximation of functions and their derivatives

Artificial neural networks (ANNs) consist of a large number of elementary processing units, the so-called neurons. The different neurons are interconnected to each other, and the connections are weighted individually. Neural networks can be trained to a desired behaviour by trial-and-error procedures based on training data. The purpose of the training is to adjust the individual weights of the network, so the network exhibits a desired behaviour. The neural network model presented in this work is a feedforward network and is trained with the resilient propagation (rProp) learning algorithm. Expressions for the derivatives of the networks output w.r.t. its inputs, the calculation of the different thermodynamic quantities as well as the training procedure itself will be discussed in the following section.

2.2.1 Notation

The proposed model can be used with an arbitrary number of layers but will be fixed to three consisting of one input layer (1), one hidden layer (2) and one output layer (3). Neurons of the different layers of a network will be denoted by the Greek letters α , for layer 1 and β for layer 2. The general structure of an arbitrary neuron β is shown in Fig. 1. The number of neurons for the output layer will be fixed to 1. Network parameters are numerated due to their multiple use. The weight matrices connecting the neurons of different layers will be in this sense referred to as $W_{21,\beta\alpha}$ that connects a neuron α from layer 1 with a neuron β from layer 2 and $W_{32,1\beta}$, respectively. Vector-valued network parameters apply only on one layer and are addressed by the number of the layer and one Greek letter for the neuron, like $b_{2,\beta}$. The value a neuron receives before applying its activation function will be denoted as *net input* s . The net input of each layer is organized as vector $s_{2,\beta}^t$ where the superscript t runs through the number of training examples in general. The absolute number of iterable objects will be denoted by amount lines, like $|\gamma|$ or $|t|$. The activation function of the input and output layer is linear. The activation function of the hidden layer will be denoted as f for the general description and will be specified later. The activation functions are applied on every neuron of a layer. Expressions for the neural network representations of the different thermodynamic quantities are in the following denoted by the subscript N . The overall input pattern is denoted as x_α^t and the overall output pattern as y^t .

2.2.2 Representation of thermodynamic functions with neural network variables

Feedforward networks are suitable networks for solving regression tasks (Hornik 1991). Every layer is fully connected with its adjacent layer, and the information flows monodirectionally from the input through the hidden layer to the output of the network. In the following section the net input of an arbitrary layer, e.g. $s_{2,\beta}^t$, is defined as the weighted sum of all input values and an additional threshold value $b_{2,\beta}$ as given in Eq. (2.5) by

$$s_{2,\beta}^t = \sum_{\alpha=1}^{|\alpha|} W_{21,\beta\alpha} \cdot x_{\alpha}^t + b_{2,\beta} \quad (2.5)$$

Using this definition the overall output of the network y^t as function of its input x^t for any input pattern t is calculated as given in Eq. (2.6) by

$$\begin{aligned} y^t(x) &= W_{2,1\beta} \cdot f \left(\sum_{\alpha=1}^{|\alpha|} W_{1,\beta\alpha} \cdot x_{\alpha}^t + b_{2,\beta} \right) + b_{3,1} \\ &= W_{2,1\beta} \cdot f \left(s_{\beta}^t \right) + b_{3,1}. \end{aligned} \quad (2.6)$$

Expressions for the first and the second derivative of y^t w.r.t. x^t are derived by the application of the chain rule as given in Eqs. (2.7) and (2.8) by

$$\frac{\partial y^t}{\partial x_{\alpha}} = W_{2,1\beta} \cdot f' \left(\sum_{\alpha=1}^{|\alpha|} W_{1,\beta\alpha} \cdot x_{\alpha}^t + b_{2,\beta} \right) \cdot W_{1,\beta\alpha} \quad (2.7)$$

and, respectively,

$$\begin{aligned} \frac{\partial^2 y^t}{\partial x_{\alpha} \partial x_{\alpha}} &= W_{2,1\beta} \cdot f'' \left(\sum_{\alpha=1}^{|\alpha|} W_{1,\beta\alpha} \cdot x_{\alpha}^t + b_{2,\beta} \right) \cdot W_{1,\beta\alpha} \\ &\quad \cdot W_{1,\beta\alpha}. \end{aligned} \quad (2.8)$$

Using Eqs. (2.6)–(2.8) one can express the Gibbs energy and the related quantities given by Eqs. (2.1)–(2.4) solely by neural network variables. In the proposed model, the input to the network is directly given by the temperature T and the output of the network y^t is directly connected to the Gibbs energy G . Therefore, the neural network representation G_N^t of the Gibbs energy G^t at a given temperature T is calculated as given in Eq. (2.9) by

$$G_N^t = y^t = W_{2,1\beta} \cdot f \left(s_{\beta}^t \right) + b_{3,1} \quad (2.9)$$

The entropy as defined in Eq. (2.2) is calculated as the first derivative of G w.r.t. T . The neural network representation

of the entropy S_N is therefore given as in Eq. (2.10) by

$$\begin{aligned} S_N^t &= - \frac{dy^t}{dx^t} \\ &= -W_{2,1\beta} \cdot f' \left(s_{\beta}^t \right) \cdot W_{1,\beta 1} \end{aligned} \quad (2.10)$$

The neural network representation of the enthalpy from Eq. (2.3) is given as in Eq. (2.11) by

$$\begin{aligned} H_N^t &= G_N^t + x^t S_N^t \\ &= W_{2,1\beta} \cdot f \left(s_{\beta}^t \right) + b_{3,1} - x^t \left(W_{2,1\beta} \cdot f' \left(s_{\beta}^t \right) \cdot W_{1,\beta 1} \right) \end{aligned} \quad (2.11)$$

And finally the neural network representation of the isobaric heat capacity from Eq. (2.4) is given as in Eq. 2.12 by

$$\begin{aligned} C_N^t &= -x^t \frac{d^2 y^t}{d(x^t)^2} \\ &= -x^t \cdot W_{2,1\beta} \cdot f'' \left(s_{\beta}^t \right) \cdot (W_{1,\beta 1})^2 \end{aligned} \quad (2.12)$$

Equations (2.9)–(2.12) consist now solely of neural network parameters and reflect at the same time the physical relations of the different thermodynamic quantities. As described in the work of Avrutskiy (2017) the expressions for the derivatives of a neural network outputs w.r.t. its inputs can be considered as standalone neural networks and could be trained individually. Due to the fact that each of the expressions above is formed from the same network parameters the calculation of G_N , S_N and H_N is also possible when the training data only contain values for C . This results in two major advantages: firstly, all available measurement data can be used for the modelling process and, secondly, the

Fig. 1 Structure of an arbitrary artificial neuron β

self-consistency of resulting network for the different thermodynamic quantities is always fulfilled.

Activation function

Activation functions within neural network regression predefines the shape of the function approximation. In general any shape can be approximated by a neural network using the standard sigmoid activation function and a sufficient number of layers and neurons as proven in Hornik (1991). A major disadvantage is the extrapolation ability of a neural network. The only information about the function to be approximated is provided by the training data. Without further assumptions and constrains a reliable prediction of function values outside the borders of the training data is not possible. When approximating thermodynamic functions and especially when calculating stable phases, phase transitions or phase equilibria a model must deliver reasonable values for the thermodynamic functions of the different phases even outside their stable temperature regimes. The extrapolation abilities of the proposed model are provided by the use of two different activation functions f_a and f_b as also shown in Fig. 2:

$$f_a(s) = E_0 + \frac{3}{2}R\Theta_E + 3Rs\log(1 - e^{-\Theta_E/s}) - \frac{1}{2}as^2 - \frac{1}{6}bs^3 \quad (2.13)$$

$$f_b(s) = \log(e^s + 1) \quad (2.14)$$

The expression from Eq. (2.13) is derived by the integration of the Chen and Sundman model for the isobaric heat capacity, where each term describes a different physical contribution. For a detailed description of this model the author refers to Chen and Sundman (2001). By using Eq. (2.13) as an activation function a physical basis for the approxima-

tion of thermodynamic functions is provided which describes the general shape of the approximated thermodynamic functions. In Eq. (2.13) R stands for the universal gas constant. The parameters E_0 , Θ_E , a and b are treated as additional network parameters and are optimized during the learning process. In many material systems, such as iron, which will be approximated in this work, effects, for example magnetic ordering effects, can occur locally and affect the thermodynamic functions. The local deviations are approximated by a second network with its activation function f_b as given in Eq. (2.14). f_b is based on the so-called *softplus* activation function (Nair and Hinton 2010). The reason to use Eq. 2.14 lies in the bell shape of the second derivative of f_b w.r.t. s that allows to model local and peak-shaped effects in the heat capacity curve.

Training neural networks and derivatives with resilient propagation

The training data consist of sets of pairs $\{(x^t, G^t)\}$ defining a desired network output G^t at given x^t . When dealing with partial derivatives there need to be additional data provided for every unknown quantity of the system. The sets $\{(x^u, S^u)\}$, $\{(x^v, H^v)\}$ and $\{(x^w, C^w)\}$ provide these additional trainings data for each of the derived thermodynamic quantities approximated in this work. The different indices t , u , v and w stand for the fact that the absolute number of trainings examples and the locations x^t , x^u , x^v and x^w do not necessarily need to be the same. To measure the error between the resulting output from the network and the desired output defined by the training data, a cost function $E = e(p^\alpha)$ with $p^\alpha = [W^{\gamma\beta}, W^{\delta\gamma}, b^\gamma, \dots]^T$ that depends on all free network parameters is defined. In the present work the least squares sum is chosen as cost function as given in Eq. (2.15) by

Fig. 2 Normalized prototype functions for the different thermodynamic quantities for the ANN approximation **a** $g_a(x) = f_a(s_a(x))$ with $s_a(x) = x$ and **b** $g_b(x) = f_b(s_b(x))$ with $s_b(x) = x - 5$

Fig. 3 Structure of the ANN model for the approximation of thermodynamic functions

$$\begin{aligned}
 E &= \frac{q_G}{|t|} \sum_t E_G^t + \frac{q_S}{|u|} \sum_u E_S^u + \frac{q_H}{|v|} \sum_v E_H^v + \frac{q_C}{|w|} \sum_w E_C^w \\
 &= \frac{q_G}{|t|} \sum_t (G_N^t(x^t) - G^t)^2 \\
 &\quad + \dots \frac{q_S}{|u|} \sum_u (S_N^u(x^u) - S^u)^2 \\
 &\quad + \dots \frac{q_H}{|v|} \sum_v (H_N^v(x^v) - H^v)^2 \\
 &\quad + \dots \frac{q_C}{|w|} \sum_w (C_N^w(x^w) - C^w)^2
 \end{aligned} \quad (2.15)$$

The main goal of the training process is now to find an optimal set of parameters so that the error defined in Eq. (2.15) reaches a minimum. By repeated application of the chain rule $\partial E / \partial p^\alpha$ can be calculated. A step in the opposite gradient direction is then performed to move towards the minimum of the error surface. The factors q_G , q_S , q_H and q_C are introduced to equalize the contributions of the different thermodynamic quantities, which differ in orders of magnitude, to the error expression.

Usually the magnitude and the sign of the different partial derivatives are used for the optimization of the network parameters. This can be a problem when the error surface described by E is too rugged and nonlinear. Under such conditions the magnitude of $\partial E / \partial p^\alpha$ can vary by orders of magnitude from one iteration to the next resulting in a unstable learning process and even in the failing

Fig. 4 **a** The neural network representation of the isobaric heat capacity of FeBCC and the underlying trainings data, **b** relative differences $\Delta C/C = (C_N - C)/C_N$ of the trainings data from the neural network approximation for FeBCC

of the optimization routine. The resilient propagation algorithm (rProp) (Riedmiller and Braun 1993) was developed to become independent from the magnitude of $\partial E / \partial p^\alpha$. rProp adapts local stepsizes based only on the sign of $\partial E / \partial p^\alpha$ of the current iteration k and of the previous iteration $k - 1$.

The rProp learning algorithm introduces an individual update value η^α for each of the p^α network parameters. The individual η^α are changed during the learning process, and its

Fig. 5 **a** The neural network representation of the enthalpy of Fe_{BCC} and Fe_{FCC} together with their underlying trainings data for $300 \text{ K} < T < 1600 \text{ K}$, **b** relative differences $\Delta H/H = (H_N - H)/H_N$ of the trainings data from the neural network approximation for Fe_{BCC} and Fe_{FCC}

evolution is influenced only by the sign of $\partial E/\partial p^\alpha$. According to Riedmiller and Braun (1993) the change of the different η^α is given by

Fig. 6 **a** The neural network representation of the isobaric heat capacity of Fe_{FCC} and the underlying training data, **b** relative differences $\Delta C/C = (C_N - C)/C_N$ of the trainings data from the neural network approximation for Fe_{FCC}

$$\eta_k^\alpha = \begin{cases} \eta^+ \cdot \eta_{k-1}^\alpha, & \text{if } \left(\frac{\partial E}{\partial p^\alpha}\right)_{k-1} \cdot \left(\frac{\partial E}{\partial p^\alpha}\right)_k > 0 \\ \eta^- \cdot \eta_{k-1}^\alpha, & \text{if } \left(\frac{\partial E}{\partial p^\alpha}\right)_{k-1} \cdot \left(\frac{\partial E}{\partial p^\alpha}\right)_k < 0 \\ \eta_{k-1}^\alpha, & \text{else.} \end{cases} \quad (2.16)$$

Fig. 7 **a** The neural network representation of the enthalpy of Fe_{BCC} and Fe_{LIQ} together with their underlying trainings data for $1600 \text{ K} < T < 2200 \text{ K}$, **b** relative differences $\Delta H/H = (H_N - H)/H_N$ of the trainings data from the neural network approximation for Fe_{BCC} and Fe_{LIQ}

After the update values the individual p^α are updated as given in Eq. (2.17) by

$$\Delta p_k^\alpha = \begin{cases} -\eta_k^\alpha, & \text{if } \left(\frac{\partial E}{\partial p^\alpha} \right)_k > 0 \\ \eta_k^\alpha, & \text{if } \left(\frac{\partial E}{\partial p^\alpha} \right)_k < 0 \\ 0, & \text{else.} \end{cases} \quad (2.17)$$

As in regular gradient descent optimization procedures the update is made in opposite gradient direction. The updates are applied as given in Eq. (2.18) by

$$p_k^\alpha = p_{k-1}^\alpha + \Delta p_k^\alpha \quad (2.18)$$

Learning algorithms can be used in online or batch modes. During online learning only one training example per iteration is used at a time for updating the free network parameters, while batch learning uses a mean error over more than one training example per iteration. One cycle through all the available training data is called epoch. It is worth to mention that the rProp algorithm does only work in batch mode and with a large batch size.

3 Results

3.1 A neural network for the approximation of thermodynamic functions

The proposed model for the approximation of thermodynamic functions consists of two interconnected subnetworks as shown in Fig. 3. The first subnetwork consists of a $1-1-1$ ANN and has f_a (Eq. 2.13) as its activation function for the hidden neurons. As described in the previous section this first subnetwork provides a base level for the approximation. The second subnetwork has a $1-N-1$ structure and uses f_b (Eq. 2.14) as activation function for the hidden neurons. The number of hidden neurons is depending on the case.

The proposed method is implemented in Python and uses *Theano* (Theano Development Team 2016) as tensor library as well as Theanos built-in automated differentiation algorithm for the derivation of the gradients needed for the calculation of the derived thermodynamic quantities and during the learning procedure. rProp is used in batch mode with a batch size equal to the number of available measurement data. The different phases of a chemical element are approximated each with a separate neural network. The optimization of the network parameters is carried out at the same time for all phases of the system. This is achieved by formulating the error for every phase separately and adding up the different phase-wise error expressions to an overall system error. The reason for this approach lies in the calculation thermodynamic quantities like the Gibbs energy $G_{a \rightarrow b}(T_{\text{tr}})$ or the enthalpy change $\Delta H_{a \rightarrow b}(T_{\text{tr}})$ at the transitions from an arbitrary phase a to a phase b at the transition temperature T_{tr} . $\Delta H_{a \rightarrow b}(T_{\text{tr}}) = H_b(T_{\text{tr}}) - H_a(T_{\text{tr}})$ for example depends on the values for $H_a(T_{\text{tr}})$ and $H_b(T_{\text{tr}})$. The optimization of the neural network representation is based on measurements for $H_a(T_{\text{tr}})$, $H_b(T_{\text{tr}})$ and $\Delta H_{a \rightarrow b}(T_{\text{tr}})$, and it is more than likely that measurements from different sources are not fully consistent. The optimization has the goal to minimize the error

Fig. 8 The thermodynamic functions of Fe_{GAS} between $3000 \text{ K} < T < 6000 \text{ K}$ together with the training data from Chase (1998). **a** $G(T)$, **b** $H(T)$, **c** $S(T)$, **d** $C(T)$

between the network output and its underlying training data leading to an error $E > 0$ even for a fully optimized network. Optimizing $H_a(T_{\text{tr}})$ in a first step and $H_b(T_{\text{tr}})$ and $\Delta H_{a \rightarrow b}(T_{\text{tr}})$ in the second step would distribute the error $E > 0$ due to inconsistencies of the measurement data only on the second phase and would lead to a reduced quality of its approximation. By optimizing all the different phases of a system simultaneously on the basis of a overall system error the error due to inconsistencies of the measurement data is distributed evenly among the different phases of the system.

3.2 Approximation of the thermodynamic functions of pure iron

The thermodynamic functions of Fe were considered to evaluate the performance of the proposed neural network model. The thermodynamic functions of each phase of Fe are approximated for the temperature range between $0 < T < 6000 \text{ K}$ in its stable and metastable regime. Between 0 and 1184 K Fe has a BCC crystal structure and is denoted $\alpha\text{-Fe}$ in the literature. At 1184 K there is a phase transition $\text{Fe}_{\text{BCC}} \rightarrow \text{Fe}_{\text{FCC}}$. The FCC phase is denoted $\gamma\text{-Fe}$ and is

stable between 1184 and 1665 K. At 1665 K a second phase transition occurs in the solid state $\text{Fe}_{\text{FCC}} \rightarrow \text{Fe}_{\text{BCC}}$. The second BCC phase is referred to as δ -Fe, but since the crystal structure is the same as for α -Fe, the two BCC phases are modelled by a single ANN. Fe has its melting point at 1809 K and the transition to the gaseous phase at 3134 K. The approximation of the thermodynamic functions of Fe is challenging due to the strong magnetic peak in the C_p curve of BCC iron at 1042 K and to the four phase transitions between 0 and 6000 K. There exist reviews of the available measurements for Fe, among which Desai (1986) and Chen and Sundman (2001) are used in this work to gather the training data for the neural network model. The used training data, if not mentioned explicitly, consist solely of raw measurement data. Published values from other already optimized models are not taken as training data but are used to compare the obtained results to. The calculation of phase transitions is based on a bisection method which is implemented in python.

3.2.1 BCC

Figure 4a shows the ANN representation of the specific heat C_N of Fe_{BCC} iron together with the training data used for the approximation and the results from the *FactSage 7.0* FactPS database (Bale et al. 2016). Figure 4b shows the respective relative differences between C_N and the training data. The calculated thermodynamic function is over wide ranges in good agreement with the measurement data. The overall standard deviation of the neural network approximation to the training data is $\sigma_{\text{BCC}} = 2.9 \text{ J/mol K}$. Between 500 and 700 K the calculated curve tends to predict slightly lower heat capacity values than the pure training data would suggest. This behaviour can be explained by the nature of the neural network regression itself. The error expression combines the error of different quantities of the different phases. The overall system error has a minimum, but can still be seen as the best compromise between the individual errors the system error consists of.

The calculated heat capacity function from this work is in good agreement with curve from the *FactSage 7.0* FactPS database (Bale et al. 2016). Both approximations show deviations from the measurement data between 500 and 700 K. The values near the magnetic peak at 1042 K are better represented by the neural network representation than by the *FactSage 7.0* FactPS database (Bale et al. 2016). Furthermore, the *FactSage 7.0* FactPS database (Bale et al. 2016) representation has a jump at the transition between Fe_{BCC} and Fe_{LIQ} at 1809 K. This jump is avoided in the results obtained in this work, and the transition into the metastable regime is smooth and continuous. In addition, Fig. 5a shows the results for the ANN representation of the enthalpy H_N together with the training data and the representation from the *FactSage 7.0* FactPS database (Bale et al. 2016). Figure 5b

shows in addition the relative difference between H_N and the training data. The calculated values of H_N are in good agreement with the measurement data and the results from Bale et al. (2016). Another interesting aspect is the model's ability to approximate the different thermodynamic quantities, especially C_N , between 0 and 298.15 K, which is a problem for the polynomial-based models as reported in Roslyakova et al. (2016).

3.2.2 FCC

Figure 6a shows the ANN representation of the specific heat C_N of Fe_{FCC} iron together with the underlying training data and the results obtained from the *FactSage 7.0* FactPS database (Bale et al. 2016). The relative differences between C_N and the training data for Fe_{FCC} iron are shown in figure (6b). The available measurement data are scattered. For example the difference between $C_p(1600 \text{ K})$ from Bendick and Pepperhoff (1982) and $C_p(1600 \text{ K})$ from Rogez and Le Coze (1980) is $\Delta C_p = 11.44 \text{ J/(mol K)}$. The neural network approximation lies in good agreement with the measurement data in view of their strong scattering. The overall standard deviation of the neural network approximation to the training data is $\sigma_{\text{FCC}} = 2.1 \text{ J/(mol K)}$. The slope of the neural network approximation of the heat capacity in its stable regime between 1184 and 1665 K is lower than the representation from the *FactSage 7.0* FactPS database (Bale et al. 2016). Like in the Fe_{BCC} phase the FactPS database approximation has a jump above at the melting temperature. This

Table 1 Comparison between calculated and experimental temperatures of phase transformations for iron

	T_{tr} in K	Source
$\text{Fe}_{\text{BCC}} \rightarrow \text{FCC}$	1184	Bendick and Pepperhoff (1982)
	1184	Tsuchiya et al. (1971)
	1184	Anderson and Hultgren (1962)
	1184	Vollmer et al. (1966)
	1184.08	This work
$\text{Fe}_{\text{FCC}} \rightarrow \text{BCC}$	1665	Bendick and Pepperhoff (1982)
	1665	Tsuchiya et al. (1971)
	1665	Anderson and Hultgren (1962)
	1665	Vollmer et al. (1966)
	1665.09	This work
$\text{Fe}_{\text{BCC}} \rightarrow \text{LIQ}$	1811	Desai (1986)
	1809	Anderson and Hultgren (1962)
	1809	Vollmer et al. (1966)
	1808.99	This work
$\text{Fe}_{\text{LIQ}} \rightarrow \text{GAS}$	3134	Haynes (2015)
	3134.0	This work

Table 2 Comparison between calculated and experimental phase transition enthalpies for iron

	$\Delta H_{\text{BCC} \rightarrow \text{FCC}}$ in J/mol $T = 1183.95\text{K}$	$\Delta H_{\text{FCC} \rightarrow \text{BCC}}$ in J/mol $T = 1664.95\text{K}$	$\Delta H_{\text{BCC} \rightarrow \text{LIQ}}$ in J/mol $T = 1808.99\text{K}$
Smith (1946)	900	–	–
Wallace et al. (1960)	911(±80)	–	–
Anderson and Hultgren (1962)	941(±80)	1088(±210)	–
Dench and Kubaschewski (1963)	900	837	–
Braun and Kohlhaas (1965)	910(±20)	850(±30)	14,400(±300)
Morris et al. (1966)	–	–	13,799(±418)
Vollmer et al. (1966)	–	–	14,400(±200)
Treverton and Margrave (1971)	–	–	13,836(±293)
Normanton et al. (2013)	923(±10)	765(±22)	–
Cezairliyan and McClure (1974)	–	890	–
Rogez and Le Coze (1980)	900(±20)	–	–
Schürmann and Kaiser (1981)	954	770	14,390
Bendick and Pepperhoff (1982)	820	850	–
Johansson (1937)	1339	444	–
Darken and Smith (1951)	900	690	15,355
Kaufman et al. (1963)	879	937	–
Braun and Kohlhaas (1965)	900	950	–
Guillermé et al. (1985)	1013	826	13,807
Desai (1986)	900	850	13,810
This work	928	871	14,145

jump does not occur in the neural network approximation of this work.

Figure 5a shows the enthalpy of Fe_{BCC} and Fe_{FCC} near the transition $\text{Fe}_{\text{BCC}} \rightarrow \text{Fe}_{\text{FCC}}$ at 1184.08 K together with the underlying training data and with the SGTE approximation as a comparison. The relative differences of the training data and H_{N} are shown in Fig. 5b. The neural network approximations obtained in this work lie in good agreement with the measurement data. For the Fe_{BCC} phase the neural network approximation and the SGTE approximation are almost identical. For the Fe_{FCC} phase, the results from this work predict lower values in the metastable regime under 1184.08 K than the SGTE approximation from Bale et al. (2016) and the difference increases with decreasing temperature.

3.2.3 Liquid

The neural network approximation for liquid iron Fe_{LIQ} is based on the work of Desai (1986), who suggests a value for the isobaric heat capacity of liquid iron of 46.632 J/(mol K), and additionally on measurement values for the enthalpy. The optimization procedure for liquid iron is slightly different from the optimization of the solid phases. The isobaric heat capacity of liquids is constant. For the learning algorithm the stable temperature regime of iron between 1809 K

and 3134 K is represented by 100 points in the interval [1809, 3134]. The target value for each of the 100 points is the 46.632 J/(mol K) from Desai (1986). Calculations have shown that this condition alone is not sufficient for a slope of 0 J/(mol K²) of C in of the liquid phase. The error expression for liquid iron is therefore extended by the expression as given in Eq. (3.1) by

$$\frac{dC}{dT} = 0. \quad (3.1)$$

Figure 7a shows the results for the enthalpy of Fe_{BCC} and Fe_{LIQ} near the transition $\text{Fe}_{\text{BCC}} \rightarrow \text{Fe}_{\text{LIQ}}$ at 1808.9 K together with the results from the *FactSage* 7.0 FactPS database (Bale et al. 2016). The relative differences between H_{N} and the measurement data for Fe_{BCC} and Fe are shown in Fig. 7b. The approximation for Fe_{LIQ} above the melting temperature lies in good agreement with the available measurement values and is almost identical to the approximation from the FactPS database (Bale et al. 2016). In the metastable temperature regimes, for Fe_{LIQ} above and for Fe_{BCC} below the melting temperature, the neural network approximation predicts slightly lower values for C as the approximation from Bale et al. (2016).

Fig. 9 Overview over the thermodynamic functions from this work of Fe_{BCC} , Fe_{FCC} , Fe_{LIQ} and Fe_{GAS} between $0 \text{ K} < T < 6000 \text{ K}$. **a** $G(T)$, **b** $H(T)$, **c** $S(T)$, **d** $C(T)$. The solid lines stand for the stable, the dashed lines for the metastable regime (color figure online)

3.2.4 Gaseous phase

The approximation of the gaseous phase is based on the NIST JANAF thermochemical tables (Chase 1998). This means the basis for the approximation of the thermodynamic functions does not consist of measurement values as for the other phases of iron but on already optimized values. Nevertheless, and for the sake of completeness, the gaseous phase is still incorporated in the proposed results. A important aspect of the approximation of the gaseous phase is that the NIST JANAF thermochemical tables provide values for G , H , S and C . Figure 8a–d shows the approximated thermodynamic functions of gaseous iron. The results for the gaseous phase

show that every available thermodynamic quantities can be used for the approximation of the thermodynamic functions

3.2.5 Further remarks

For the optimization of the neural network model values for H , C and S were used. Values for G were used indirectly to approximate transition temperatures. This demonstrates the models ability to use any of these quantities for the approximation of thermodynamic functions, which is clearly an advantage over the polynomial-based models. Figure 8a–d shows the curves of G , H , S and C for the whole considered

temperature range and for all of the different phases of iron. The dashed lines indicate the metastable regimes of respective phase. Table 1 lists the calculated transition temperatures of iron for the different phase transitions together with available literature data. The calculated values from this work lie in good agreement with the literature data. Additionally Table 2 lists the enthalpies of transformation for the different phase transformations of iron. The calculated values from this work are in good agreement with the data from the literature. It is worth to mention that these values are also part of the training data and are learned by the network during the optimization phase. Nevertheless, these values show clearly that the proposed method can be used for the approximation of the thermodynamic functions of unary systems and the obtained results are self-consistent.

4 Summary and outlook

This work presents a new model for the approximation of thermodynamic functions of unary systems based on artificial neural networks. For iron as a complex example the different thermodynamic functions were successfully approximated. The comparison with literature data and already existing approximations of the thermodynamic functions of pure iron shows the suitability of the proposed method. The proposed method solves the underlying optimization problem with a minimum user input and almost automatically. One major disadvantage of the proposed method is the black-box character of ANNs. As a consequence the ability of the network to deliver correct values can only be verified through trial and error and not by investigating the optimized network itself. The extension of the proposed method on material systems with more than one constituents and the investigation in how far the beneficial properties of the proposed method can be extended are the main questions for further investigations (Fig. 9).

Acknowledgements Open Access funding provided by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The author Maximilian Länge declares that he has no conflict of interest. This article does not contain any studies with human participants performed by any of the authors.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson PD, Hultgren R (1962) The thermodynamics of solid iron at elevated temperatures. *Trans Metall Soc AIME* 224:842–845
- Avrutskiy VI (2017) Enhancing approximation abilities of neural networks by training derivatives. [arXiv:1712.04473v2](https://arxiv.org/abs/1712.04473v2)
- Awbery JH, Griffiths E (1940) The thermal capacity of pure iron. *Proc R Soc Lond Ser A* 174:1–15
- Bale CW, Bélisle E, Chartrand P et al (2016) FactSage thermochemical software and databases, 2010–2016. *Calphad* 54:35–53
- Bendick W, Pepperhoff W (1982) On the α/γ phase stability of iron. *Acta Metall* 30:679–684
- Braun M, Kohlhaas R (1965) Die spezifische Wärme von Eisen, Kobalt und Nickel im Bereich hoher Temperaturen. *Phys Status Solidi B* 12:429–444
- Cezairliyan A, McClure JL (1974) Thermophysical measurements on iron above 1500 K using a transient (subsecond) technique. *J Res Natl Stand Sec A* 78A:1
- Chase MW Jr (1998) NIST-JANAF thermochemical tables, 4th edn. American Chemical Society, New York, American Institute of Physics for the National Institute of Standards and Technology, Washington
- Chase MW, Ansara I, Dinsdale AT et al (1995) Thermodynamic models and data for pure elements and other end members of solutions. *Calphad* 19:437–447
- Chen Q, Sundman B (2001) Modeling of thermodynamic properties for Bcc, Fcc, liquid, and amorphous iron. *J Phase Equilib* 22:631–644
- Darken LS, Smith RP (1951) Thermodynamic functions of iron. *Ind Eng Chem* 43:1815–1820
- Dench WA, Kubaschewski O (1963) Heat capacity of iron at 800°C to 1420°C. *J Iron Steel Inst* 201:140–143
- Desai PD (1986) Thermodynamic properties of iron and silicon. *J Phys Chem Ref Data* 15:967–983
- Dinsdale AT (1991) SGTE data for pure elements. *Calphad* 15:317–425
- Esser H, Baerlecken E-F (1941) Die wahre spezifische Wärme von reinem Eisen und Eisen-Kohlenstoff-Legierungen von 20 bis 1100°. *Arch Eisenhüttenw* 14:617–624
- Eucken A, Werth H (1930) Die spezifische Wärme einiger Metalle und Metallegierungen bei tiefen Temperaturen. *Z Anorg Allg Chem* 188:152–172
- Griffiths EH, Griffiths E (1914) The capacity for heat of metals at low temperatures. *Proc R Soc Lond* 90:557–560
- Guillermet AF, Gustafson P, Hillert M (1985) The representation of thermodynamic properties at high pressures. *J Phys Chem Solids* 46:1427–1429
- Günther P (1917) Untersuchungen über die spezifische Wärme bei tiefen Temperaturen. *Ann Phys (Leipzig)* 356:828–846
- Haynes WM (2015) CRC handbook of chemistry and physics, 95th edn. CRC Press, Hoboken
- Hemmat Esfe M (2017) Designing an artificial neural network using radial basis function (RBF-ANN) to model thermal conductivity of ethylene glycol–water-based TiO₂ nanofluids. *J Therm Anal Calorim* 127:2125–2131
- Hemmat Esfe M, Yan W-M, Afrand M, Sarraf M, Toghraie D, Dahari M (2016) Estimation of thermal conductivity of Al₂O₃/water (40%)–ethylene glycol (60%) by artificial neural network and correlation using experimental data. *Int Commun Heat Mass* 74:125–128
- Hornik K (1991) Approximation capabilities of multilayer feedforward networks. *Neural Netw* 4:251–257

- Johansson CH (1937) Thermodynamisch begründete Deutung der Vorgänge bei der Austenit-Martensit-Umwandlung. *Arch Eisenhuettenwes* 11:241–251
- Kaufman L, Clougherty E, Weiss R (1963) The lattice stability of metals—III. *Iron Acta Metall* 11:323–335
- Keesom WH, Kurelmeyer B (1939) The atomic heat of iron from 1.1 to 20.4° K. *Physica* 6:633–647
- Kelley KK (1943) The specific heat of pure iron at low temperatures. *J Chem Phys* 11:16–18
- Kollie TG, McElroy DL, Barisoni M, Brooks CR (1969) Pulse calorimetry using a digital voltmeter for transient data acquisition. ORNL Report No. ORNL-4380
- Kraftmakher YA, Romashina TY (1965) Specific heat of iron near the curie point. *Fiz Tverd Tela* 7:2532
- Lederman FL, Salamon MB, Shacklette LW (1974) Experimental verification of scaling and test of the universality hypothesis from specific-heat data. *Phys Rev B* 9:2981–2988
- Lukas HL, Sundman B, Fries SG (2007) Computational thermodynamics: the CALPHAD method. Cambridge University Press, Cambridge
- Morris JP, Foerster EF, Schulz CW, Zellar GR (1966) Use of a diphenyl ether calorimeter in determining the heat of fusion of iron. U.S. Bureau of Mines Rep. Invest, 6723
- Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: 27th international conference on machine learning
- Normanton AS, Bloomfield PE, Sale FR, Argent BB (2013) A calorimetric study of iron–cobalt alloys. *Met Sci J* 9:510–517
- Orehotsky JL, Schröder K (1970) A high temperature heat exchange calorimeter. *J Phys E* 3:889–891
- Pallister PR (1949) The specific heat and resistivity of high-purity iron up to 1250 C. *J Iron Steel Inst* 161:87–90
- Reddy BPN, Reddy PJ (1974) Low temperature heat capacities of some ferrites. *Phys Status Solidi A* 22:219–223
- Riedmiller M, Braun H (1993) A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: IEEE international conference on neural networks, pp 586–591
- Rodebush WH, Michalek JC (1925) The atomic heat capacities of iron and nickel at low temperatures. *J Am Chem Soc* 47:2117–2121
- Rogez J, Le Coze J (1980) Description et étalonnage d'un calorimètre adiabatique à balayage (800 K – 1800 K). *Rev Phys Appl* 15:341–351
- Roslyakova I, Sundman B, Dette H, Zhang L, Steinbach I (2016) Modeling of Gibbs energies of pure elements down to 0 K using segmented regression. *Calphad* 55:165–180
- Saunders N, Miodownik AP (1998) CALPHAD (calculation of phase diagrams): a comprehensive guide, vol 1. Pergamon, Oxford
- Schröder K, MacInnes WM (1969) Simple pulse heating method for specific heat measurements. *J Phys E* 2:959–962
- Schürmann E, Kaiser H-P (1981) Entwicklung und Überprüfung eines Hochtemperaturkalorimeters. *Arch Eisenhuettenwes* 52:99–101
- Shanks HR, Klein AH, Danielson GC (1967) Thermal properties of Armco iron. *J Appl Phys* 38:2885–2892
- Simon F, Swain RC (1922) Untersuchungen über die spezifische Wärme bei tiefen Temperaturen. *Ann Phys Leipz* 373:241–280
- Smith RP (1946) Equilibrium of iron–carbon alloys with mixtures of CO–CO₂ and CH₄–H₂. *J Am Chem Soc* 68:1163–1175
- Stepakoff GL, Kaufman L (1968) Thermodynamic properties of h.c.p. iron and iron–ruthenium alloys. *Acta Metall* 16:13–22
- Theano Development Team (2016) Theano: a Python framework for fast computation of mathematical expressions. CoRR
- Treverton JA, Margrave JL (1971) Thermodynamic properties by levitation calorimetry III. The enthalpies of fusion and heat capacities for the liquid phases of iron, titanium, and vanadium. *J Chem Thermodyn* 3:473–481
- Tsuchiya M, Izumiyama M, Imai Y (1971) Free energy of the solid-phase of pure iron. *J Jpn Inst Met* 35:839–845
- Vollmer O, Kohlhaas R, Braun M (1966) Notizen: Die Schmelzwärme und die Atomwärme im schmelzflüssigen Bereich von Eisen, Kobalt und Nickel. *Z. Naturforsch, Teil A* 21
- Wallace DC, Sidles PH, Danielson GC (1960) Specific heat of high purity iron by a pulse heating method. *J Appl Phys* 31:168–176

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