


RESEARCH ARTICLE

Open Access



# Application of artificial neural networks for the prediction of interface mechanics: a study on grain boundary constitutive behavior

Mauricio Fernández<sup>1,2\*</sup> , Shahed Rezaei<sup>3</sup>, Jaber Rezaei Mianroodi<sup>4,5</sup>, Felix Fritzen<sup>1</sup> and Stefanie Reese<sup>3</sup>

\*Correspondence:  
mauricio.fernandez@mechbau.uni-  
stuttgart.de;  
fernandez@cps.tu-darmstadt.de  
<sup>1</sup>Institute of Applied Mechanics,  
University of Stuttgart,  
Pfaffenwaldring 7, 70569  
Stuttgart, Germany  
Full list of author information is  
available at the end of the article

## Abstract

The present work aims at the identification of the effective constitutive behavior of  $\Sigma 5$  aluminum grain boundaries (GB) for proportional loading by using machine learning (ML) techniques. The input for the ML approach is high accuracy data gathered in challenging molecular dynamics (MD) simulations at the atomic scale for varying temperatures and loading conditions. The effective traction-separation relation is recorded during the MD simulations. The raw MD data then serves for the training of an artificial neural network (ANN) as a surrogate model of the constitutive behavior at the grain boundary. Despite the extremely fluctuating nature of the MD data and its inhomogeneous distribution in the traction-separation space, the ANN surrogate trained on the raw MD data shows a very good agreement in the average behavior without any data-smoothing or pre-processing. Further, it is shown that the trained traction-separation ANN captures important physical properties and is able to predict traction values for given separations not contained in the training data. For example, MD simulations show a transition in traction-separation behaviour from pure sliding mode under shear load to combined GB sliding and decohesion with intermediate hardening regime at mixed load directions. These changes in GB behaviour are fully captured in the ANN predictions. Furthermore, by construction, the ANN surrogate is differentiable for arbitrary separation and also temperature, such that a thermo-mechanical tangent stiffness operator can always be evaluated. The trained ANN can then serve for large-scale FE simulation as an alternative to direct MD-FE coupling which is often infeasible in practical applications.

**Keywords:** Interfaces, Traction-separation relation, Grain boundary, Molecular dynamics, Machine learning (ML), Artificial neural networks (ANN), Data-driven surrogate

## Introduction

Understanding and predicting the interface behavior plays a major role in the optimum design of many heterogeneous and defectious engineering materials. The range of applications varies from composite structures at the macro-scale, see, e.g., [1, 2], to grain boundaries at the micro-scale, as presented in, e.g., [3]. For example, in laminate composite mate-

rials the modeling of the mechanical behavior at the interface is of great importance, see [4–6]. A cohesive zone formulation is frequently utilized to describe the interface constitutive behavior in terms of traction (projection of the stress tensor on the crack plane) versus separation (displacement jump). For an arbitrary interface, one has to take into account interface plasticity as well as surface elasticity effects, which calls for more advanced and consistent interface models, see, e.g., [7]. For instance, a framework for a cohesive model by considering additional tractions related to membrane-like forces is discussed in [8]. Despite the recent developments, the coupling of different active mechanisms such as plasticity and damage at the interface requires more attention and investigations [9]. An interface model to accurately incorporate grain boundary sliding as well as intergranular fracture was proposed in [10] which is motivated by means of atomistic simulations.

A grain boundary (GB) is an interesting example of an interface in which many complex mechanical phenomena—such as GB sliding—may contribute to the interface behavior, see [11,12]. The accurate and efficient modeling of GB constitutive behavior and capturing all relevant phenomena remains a challenging field in materials science, see, e.g., [13–16]. Many theoretical models are able to capture selected material phenomena, but require substantial numerical treatment in macroscopic simulations and are, therefore, not suitable for practical application. This opens the search for computationally more efficient alternatives. One alternative is the calibration of a data-driven surrogate model for the GB constitutive behavior based on experimental and/or synthetic data. Hereby, the calibration data should be generated as closely as possible to the underlying physics in order to incorporate all relevant phenomena. This can be achieved by atomistic simulations of the GB [10,17,18]. The generated data can then be used for the training of a suitable machine learning approach in order to generate an efficient surrogate model for the GB. This approach is explored in the following.

### **Small scale matters: importance of atomistic simulations**

The macroscopic response of materials is rooted in its discrete nature at the atomic scale [19–21]. In metals, for example, the evolution of an ensemble of point defects (such as vacancies), line defects (e.g. dislocations), surface defects (e.g. GBs) and volumetric defects (such as precipitates) leads to the macroscopic behavior. Evolution of each of these is controlled by motion and interaction of the atoms. Thus, atomistic simulations are proven to be an effective approach to obtain deeper insight into the behavior of a range of materials [22,23]. Capturing relevant atomistic scale details of material properties and transferring them to macroscopic scale is an active and ongoing field of research [24,25]. These include, for example, modeling configurational [26] and compositional [27] aspects of the evolution of defects (dislocations) using atomistically informed phase field model. Underlying atomistic details are especially important when it comes to studies on fracture and damage [28]. A new theory is formulated in [29] for dislocation emission under mode I loading using molecular statics simulations. A multiscale method is developed in [30] for examining fracture of polycrystals by coupling molecular dynamics and mesoscale peridynamics.

Concerning investigations on GBs, molecular dynamics (MD) simulations offer one approach for in depth examinations of the GB phenomena. Due to the computation cost of MD, these simulations are usually limited to two neighboring grains, see, e.g., [31–34].

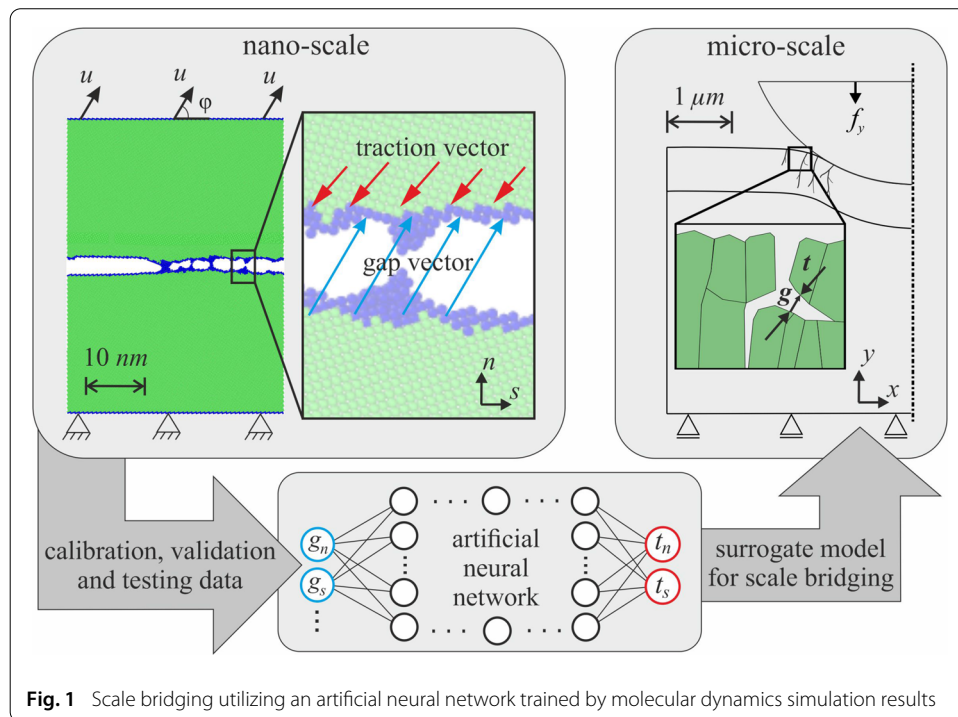
Nevertheless, MD offers the possibility to investigate many GB phenomena. For example, in [35,36] the intergranular fracture process in aluminum is examined and the constitutive relation for the grain boundary debonding is determined. The GB behavior under various mixed-mode loadings is investigated in [18]. In [37], MD is used for the exploration of metal/ceramic interface shear failure. In [10] it is shown that MD reveals several intergranular features under fracture modes I and II while loading as well as unloading of the GB. Further, the same study shows that the MD simulations can very well improve continuum models of GB cohesive zones. In summary, MD simulation data can reflect, in principle, the relevant intrinsic GB phenomena such that data-driven surrogates built upon such information may offer attractive scalable alternatives to MD simulation. These could represent scalability in space and time through coarse-grain continuum models and diffusive time scale models, respectively, thus, reducing the computational cost of a fully resolved atomistic model without losing the important details of small scales.

### **Machine learning methods as an alternative to material models**

The complexities of new materials and cumbersome numerical expenses for the evaluation of the corresponding material models pose several difficulties for the exploration and simulation of novel material combinations at engineering scale. Current computational power and novel machine learning (ML) approaches offer scalable methods which can combine insights of material modeling with the flexibility and efficiency of data-driven surrogates [38]. New ML approaches show promising performance in many engineering fields, as, e.g., computational mechanics [39–42] and structural engineering [43]. At this point, one should also mention data-driven computing [44,45]. In these works, the material behavior is only described by data. The idea is to find the point in the data set which is closest to the constraint set given by kinematical relations, balance laws and boundary conditions (see also [46,47]).

A promising ansatz in multi-scale problems is to incorporate ML models as surrogates for small scale material behavior in macroscopic simulations [38,48–50]. Multiscale analysis of reinforced concrete is conducted in [51], where ANNs are used to approximate the stress vs. crack opening material response based on mesoscale simulations (see also [52]). Radial numerically explicit potentials have been developed in [53] for hyperelastic materials for the acceleration of two-scale problems, which can be improved by recent strategies for the generation of points on hyperspheres, see, e.g., [54]. In [55], ANNs are used in combination with FE simulations in order to capture the hydro-mechanical coupling of porous media. Based on FE simulations, trained ANNs can show excellent results in predicting the effective electrical response of graphene/polymer nanocomposites and in macroscopic computations [56]. An on-the-fly adaptive scheme with error estimators is developed in [57], allowing the flexible switching between highly efficient microscale-trained ANNs and the physics-driven reduced-order model in macroscopic mechanical FE simulations. The wide spectrum of ANN models and the application of them show that ANNs are not only capable of bridging constitutive behavior across scales in materials science but they also offer the opportunity to efficiently utilize trained surrogates in large scale simulations.

Also in the more specific context of interface modelling and multiscale problems, ML has shown promising results. In the recent work [58], a framework is proposed in order to



**Fig. 1** Scale bridging utilizing an artificial neural network trained by molecular dynamics simulation results

generate models for interfaces based on game theory, deep-learning and discrete element modeling. While the approach shows excellent results and applicability for large scale FE simulations, it leaves open the question, if ANNs can be calibrated by noisy/fluctuating data.

Obviously, one may smooth the data for a subsequent calibration of an ANN surrogate. But any data-smoothing or data-processing step can potentially falsify or improve the calibrated constitutive behavior to an unclear degree. Therefore, the central issue whether an ANN surrogate can be calibrated directly based on raw oscillatory data needs to be investigated. Further, from the perspective of macroscopic FE process simulations, it is of interest whether temperature-dependent interface behavior is also traceable. Finally, as remarked in [57], if an ANN is calibrated for the approximation of a material law within a training region, it is necessary to answer the question, whether an evaluation of the trained ANN far outside of the training region yields physical results. This point is important for multiscale problems, since a macroscopic FE computation may call in its integration points for a material behavior which is far outside of the ANN training region.

**Combination of MD simulations and ANNs**

The present work aims at bridging the atomistic scale at the GB level and the continuum scale at the polycrystalline level by calibrating a computationally efficient traction-separation surrogate model for the GB, which can serve for multiscale problems, as illustrated in Fig. 1.

In the present study, proportional loading of the GB is investigated. MD simulation data for varying loading and temperature is taken into account for the calibration of the ANN surrogate. The latter is formulated with the objective of serving macroscopic FE simulations, i.e., the ANN surrogate architecture is chosen such that the surrogate is

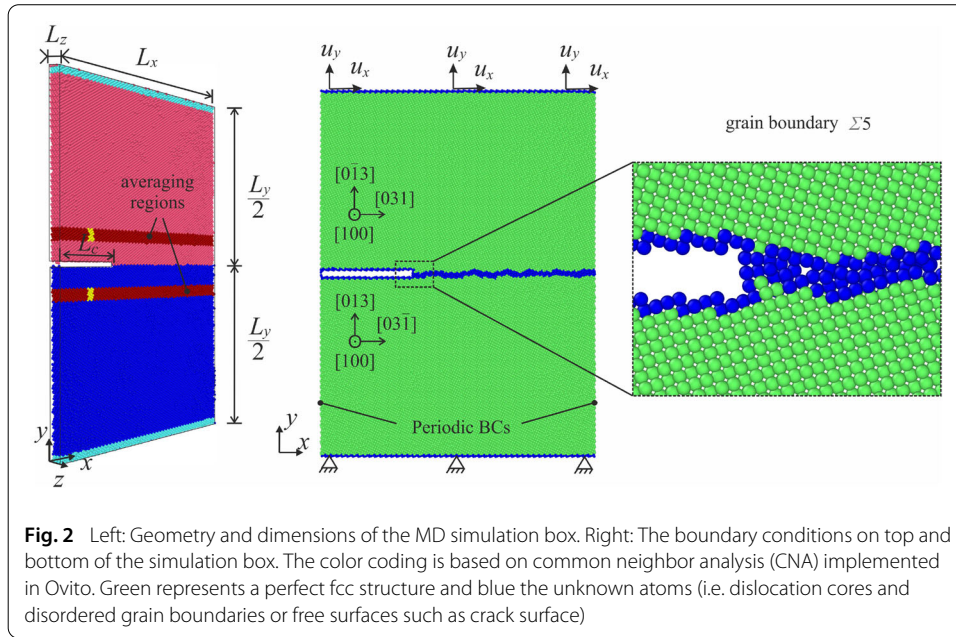
differentiable for all separation and temperature values in order to be able to obtain a tangent operator through automatic differentiation. Further, the ANN surrogate is calibrated directly on the raw MD data which shows—due to the nature of MD—not only strong fluctuating behavior but also a strong inhomogeneity in the separation and in the traction values. The present work shows that the trained ANN can in fact be calibrated from the raw MD data and it then follows very well the average behavior of the separation-traction relation. The best performing ANN is tested against MD simulations not contained in the training data set and it is shown that the ANN exhibits important physical properties after training. A physics-guided model completion is sketched for the surrogate model, such that the completed model stays differentiable and can securely be evaluated in future FE computation at larger scales.

The outline of the paper is as follows. In section “[Atomistic simulation and constitutive behavior of grain boundaries](#)” all details of the MD simulations are described. Then, in section “[Artificial neural networks for surrogate modeling of the constitutive behavior of grain boundaries](#)”, the architecture and training approach of ANNs for the problem at hand are illustrated and the best performing ANN is compared to the MD results. The paper ends with section “[Summary and outlook](#)”.

### **Atomistic simulation and constitutive behavior of grain boundaries**

In the current work, atomistic simulations of intergranular crack growth are considered as the physical model of damage evolution. Traction-separation values extracted from MD simulations in this section are used to train the ML surrogate model discussed in the next section. For this purpose, face-centered cubic (fcc) Al, modeled using the embedded atom method (EAM) of [59], is chosen as the metal of interest. The embedded atom potential of [59] is shown to reproduce accurate defect energy values close to density functional theory (DFT) calculations.

A simulation box, see Fig. 2, with dimensions  $(L_x, L_y, L_z) = (32.0, 42.3, 1.2)$  nm is defined. The box is periodic along  $x$  and  $z$  directions while it has free surfaces along  $y$  direction, resulting in an infinitely large GB, thus, eliminating any surface effect on the interface. The upper half of the box, from  $y = L_y/2$  to  $y = L_y$ , is filled with fcc Al atoms with lattice orientation of  $x||[031]$ ,  $y||[0\bar{1}3]$  and  $z||[100]$ . The lower half is filled with fcc Al atoms with lattice orientation of  $x||[03\bar{1}]$ ,  $y||[013]$  and  $z||[100]$ . The system’s internal energy is then minimized using overdamped dynamics [60] for 5000 time steps of size 2fs. To ensure fully relaxed grain boundary structure, the simulation box temperature is kept at 300 K for 10,000 steps under isothermal-isobaric (NP $\theta$ ) ensemble, where N, P and  $\theta$  represent number of atoms, pressure and temperature, respectively, with zero  $T_{xx}$ ,  $T_{zz}$  and  $T_{xz}$  stress components. All other stress components are also zero due to free surface. During another 10,000 NP $\theta$  steps, the system temperature is reduced from 300 K to the defined load temperature. Relaxing the GB structure at elevated temperature is necessary for accelerating the energy minimization as well as pushing the system out of the local minima. Note that a simple static energy minimization is usually not enough to reach a fully relaxed atomic structure at the GB. Once the grain boundary structure is in a relaxed state, a small crack is created at the boundary by removing atoms inside the region of  $0 < x < L_c$  and  $L_y/2 - 0.4 \text{ nm} < y < L_y/2 + 0.4 \text{ nm}$ . Note that the initial crack is embedded on the GB to act as a defect on the boundary. As it is shown in [10], the initial crack length will



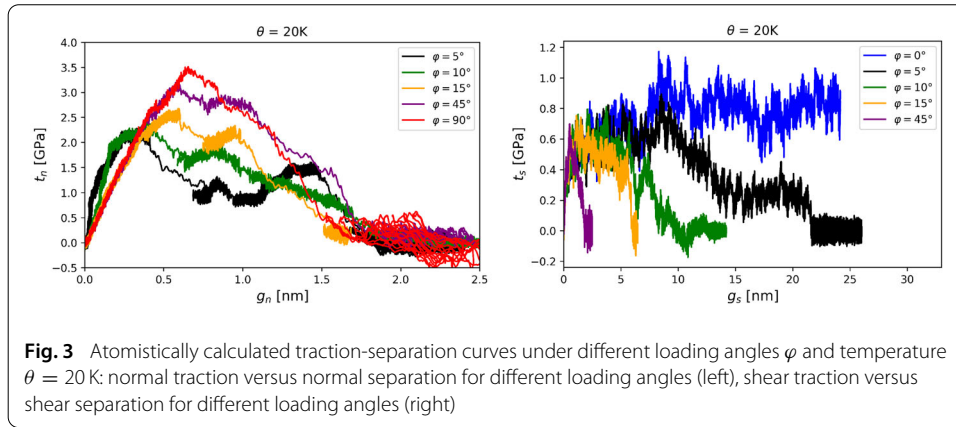
**Fig. 2** Left: Geometry and dimensions of the MD simulation box. Right: The boundary conditions on top and bottom of the simulation box. The color coding is based on common neighbor analysis (CNA) implemented in Ovito. Green represents a perfect fcc structure and blue the unknown atoms (i.e. dislocation cores and disordered grain boundaries or free surfaces such as crack surface)

affect the results, as it should. This corresponds to a defected GB which is usually the case in real applications. Depending on the material of the interest and the production procedure, GBs will often include defects such as nanoscopic crack, precipitates, voids, etc. More accurate characterization of the GB quality requires experimental observations and depend on the engineering material of interest which is beyond the scope of the current work. Once the initial crack is created, the system is relaxed during 10,000 NPθ steps under zero stress. Five atomic layers at lower section of the box (shown as cyan in Fig. 2) are fixed, while the top five layers are moved in *x* and *y* directions depending on the load angle. The load is applied with the velocity of  $5 \times 10^{-3}$  nm/ps and the simulation roughly lasts 100,000 steps under the canonical (NVθ) ensemble, where N, V and θ represent number of atoms, volume and temperature, respectively. The system temperature is kept constant at the desired load temperature using the velocity rescale method [61]. This is necessary to dissipate energy release during defect (crack) growth.

Following [10], traction values are calculated close to the interface to eliminate the influence of atomic structure irregularity at the crack surface in calculating the stresses (see brown regions in Fig. 2). These regions are chosen close to the GB in order to reduce the effect of elastic deformation between the brown regions. It shall be noted that the accurate definition of the stress tensor at atomic scale is subject to debate [62]. In this work, we adopt the virial definition [62,63], due to its simplicity and low computation cost. The kinetic term of the virial stress is ignored since the simulations are limited to low temperatures, thus, the structural changes are the main contributors to the stress. Including the kinetic term will change the stress results by 1.8% at the highest temperature case considered in this study. Thus, the virial stress

$$T = -\frac{1}{V} \sum_{\substack{ij \\ i \neq j}} f_{ij} \otimes r_i \tag{1}$$

is calculated, where  $r_i$  and  $f_{ij}$  are position of atom *i* and force vector acting on atom *i* from atom *j*, respectively. *V* is the volume of the considered region and the summation



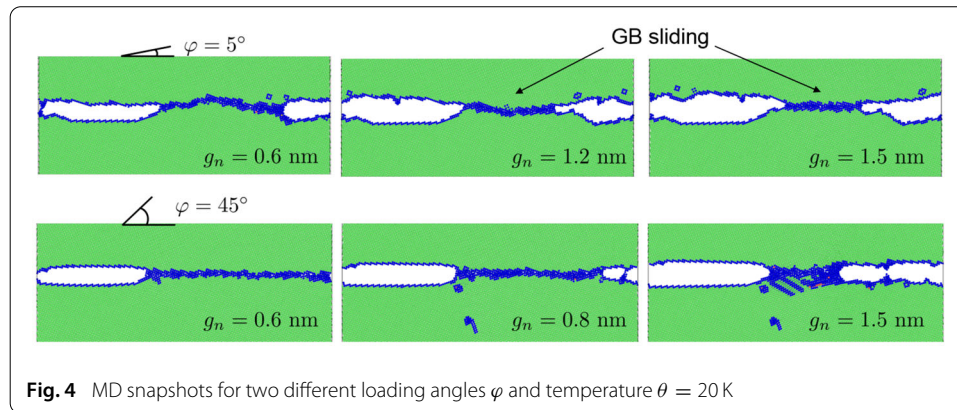
is over all atoms in the region. Normal ( $T_{yy}$ ) and shear ( $T_{xy}$ ) components of the stress tensor in (1) are considered as the corresponding traction values in mode I ( $t_n$ ) and II ( $t_s$ ), respectively. The effective gap vector contains the nominal normal  $g_n = \delta\bar{y} - \delta\bar{y}_0$  and transversal (shear)  $g_s = \delta\bar{x} - \delta\bar{x}_0$  separations, defined through the average atomic displacements

$$\delta\bar{y} = \frac{\sum_i^{N_{bt}} y_i}{N_{bt}} - \frac{\sum_i^{N_{bb}} y_i}{N_{bb}}, \quad \delta\bar{x} = \frac{\sum_i^{N_{yt}} x_i}{N_{yt}} - \frac{\sum_i^{N_{yb}} x_i}{N_{yb}} \quad (2)$$

in the measurement regions, where  $N_{bt}$  and  $N_{bb}$  are the numbers of atoms in top and bottom brown regions in Fig. 2, respectively. Analogously,  $N_{yt}$  and  $N_{yb}$  are the numbers of atoms in top and bottom yellow regions in Fig. 2, respectively. Since the system is periodic in  $x$  direction, simply averaging  $x$  position of atoms in the whole brown region will result in constant value of  $L_x/2$ . Thus, the yellow region inside the brown one is defined to track the shear separation. Note that subscript 0 in the normal and shear gap definition denotes the initial distance between two regions under zero load. Load angles ( $\varphi$ ) in the range of 0 to 180° and load temperatures ( $\theta$ ) of 20, 40, 60 and 80 K are considered in this study (see (14) and (15)). The atomistic simulations are performed using LAMMPS [64]. Ovito [65] is employed for visualization and post-processing. A few examples of traction-separation curves from MD simulations are shown in Fig. 3 for different load angles  $\varphi$  (see Fig. 4) and temperature  $\theta = 20$  K.

As illustrated in the left plot in Fig. 3, pure mode I ( $\varphi = 90^\circ$ ) shows the highest normal traction (of about 3.5 GPa) before crack growth is started. The highest stress is slightly reduced when the load is tilted towards 45°. However, reducing the load angle further to 15° dramatically reduces the peak traction to about 2.5 GPa. Considering shear traction-separation curves (Fig. 3 right), the case off mode II ( $\varphi = 0^\circ$ ) shows almost ideal plastic deformation without any crack growth (confirmed by observing the atomic structure during the loading). This is the case of grain boundary sliding. Tilting the load to 15°, the final separation, at which virtually no shear traction is observed, drops to 6nm and for 45° to 2.4 nm.

Furthermore, results of 5, 10 and 15° load angles (Fig. 3, left) show an increase of the normal traction (similar to hardening) after the initial drop of the traction (i.e. starting around  $g_n \approx 0.7$  nm until the second traction drop at  $g_n \approx 1.5$  nm for the case of 5°). As it is seen in Fig. 3, the length of the hardening-like region is reduced as the load angle increases from 5 to 10 and 15°. In the limit of normal load ( $\varphi = 90^\circ$ ), there is no observable



**Fig. 4** MD snapshots for two different loading angles  $\varphi$  and temperature  $\theta = 20$  K

hardening region and secondary traction drop, thus the traction-separation follows the classical model of interface decohesion. Inspecting the corresponding atomic structures (see Fig. 4), it is revealed that the hardening behavior in the mixed mode loading is due to GB sliding and partial healing of the initial crack. In other words, it seems that GB sliding, in particular at low angle loads, makes the interface stronger, due to atoms sliding on top of each other and filling the initial gaps at the GB and crack, compared to the same GB if the sliding was prevented (or not correctly included in the larger scale models).

Snapshots of the atomic structure for loading cases of  $5^\circ$  and  $45^\circ$  are shown in Fig. 4. As it is seen, in case of  $45^\circ$ , after reaching maximum normal traction at  $g_n \approx 0.6$  nm, the crack grows relatively rapidly due to the normal component of the load. Note also that the defect density in the remaining connected domain raises significantly which hints at a high local defect density. In Fig. 4, bottom right, structures resembling dislocation cell structures can be identified. In addition, the GB plane, in this case, remains straight. However, in the case of  $5^\circ$  load, the GB plane, due to the excessive amount of sliding, does not remain straight. This, as it is seen in the top row of Fig. 4, leads to the hardening region in the traction behavior observed in Fig. 3.

As it is seen in Fig. 3, the traction curves of the MD simulations show strongly oscillatory behavior. These fluctuations (although partly due to the small system size) are natural and typical at atomic scales, when damage and plastic behaviour is observed locally. Tracking these fluctuations is necessary for physical modeling of plasticity and damage. Unfortunately, the huge computational costs of full atomistic simulations prevent the application of these methods to larger systems, hence the necessity of transferring appropriately up-scaled data to the coarse-grained models as explained in the following section.

## Artificial neural networks for surrogate modeling of the constitutive behavior of grain boundaries

### Choice of ANNs and formulation of the surrogate model

The calibration of general material behavior is a challenging task in materials science. In general, nonlinear, path-dependent material behavior with characteristic material symmetries must be considered. For the present work, artificial neural networks [66] are taken into consideration for the calibration of a surrogate model for the effective material law of the grain boundary. For history-dependent functions, naturally, recurrent neural networks (RNNs) offer attractive alternatives, but require enormous amounts of training paths of



standardized lengths, which is a highly non-trivial and, furthermore, overly costly task if MD simulations serve as data source for the training of the networks. Furthermore, the path-dependency poses major challenges, such that the investigation of path-dependent functions from highly oscillating data, as provided by MD simulations, is out of the scope of the present work. Nevertheless, if at least proportional loading is considered, then a corresponding surrogate model for the solely state-dependent (i.e., path-independent) traction vector  $\mathbf{t}(\mathbf{g}, \theta)$  can drastically reduce the computational costs of multiscale simulations, which is the objective in the remainder of the present work. Based on these arguments, in the following we choose feedforward neural networks (FFNNs) for the approximation of the homogenized/up-scaled material law  $\mathbf{t}(\mathbf{g}, \theta)$  based on MD simulation data.

The notation of this section follows standard vector-matrix-notation from a data perspective. Physical tensors, e.g.,  $\mathbf{t}$ , are addressed by the corresponding data (vector components), e.g.,  $\underline{t}$ . With respect to the GB, normal and shear components of a physical tensor are marked by subscripts  $n$  and  $s$ , respectively. Hereby, the normal component of a related physical quantity is the first vector component, i.e.,  $n \leftrightarrow 1$ , and the shear components corresponds to the second vector components, i.e.,  $s \leftrightarrow 2$ . We then address the normal and shear components of the gap  $\underline{g} = (g_n, g_s)$  and traction vector  $\underline{t} = (t_n, t_s)$  as  $g_n \leftrightarrow g_1$ ,  $g_s \leftrightarrow g_2$ ,  $t_n \leftrightarrow t_1$  and  $t_s \leftrightarrow t_2$ . The load state  $\underline{s} = (\underline{g}, \theta) = (g_1, g_2, \theta) = (s_1, s_2, s_3) \in \mathbb{R}^3$ , i.e.  $s_1 = g_1$ ,  $s_2 = g_2$  and  $s_3 = \theta$ , represents the arguments of the unknown traction vector  $\mathbf{t}(\mathbf{g}, \theta) \leftrightarrow \underline{t}(\underline{s}) \in \mathbb{R}^2$ . The identification of a surrogate model  $\hat{\underline{t}}(\underline{s}) : \mathbb{R}^3 \mapsto \mathbb{R}^2$ ,  $\hat{\underline{t}}(\underline{s}) \approx \underline{t}(\underline{s})$ , is addressed by FFNNs. Based on an initial suitable rescaling  $\underline{r}(\underline{s})$  of the given load state  $\underline{s}$ , the FFNN for the present work is described as

$$\underline{z}^{[0]} = \underline{r}(\underline{s}) \in \mathbb{R}^3, \quad \underline{z}^{[i]} = a^{[i]}(\underline{W}^{[i]} \underline{z}^{[i-1]} + \underline{b}^{[i]}) \in \mathbb{R}^{n^{[i]}}, \quad i = 1, \dots, N \quad (3)$$

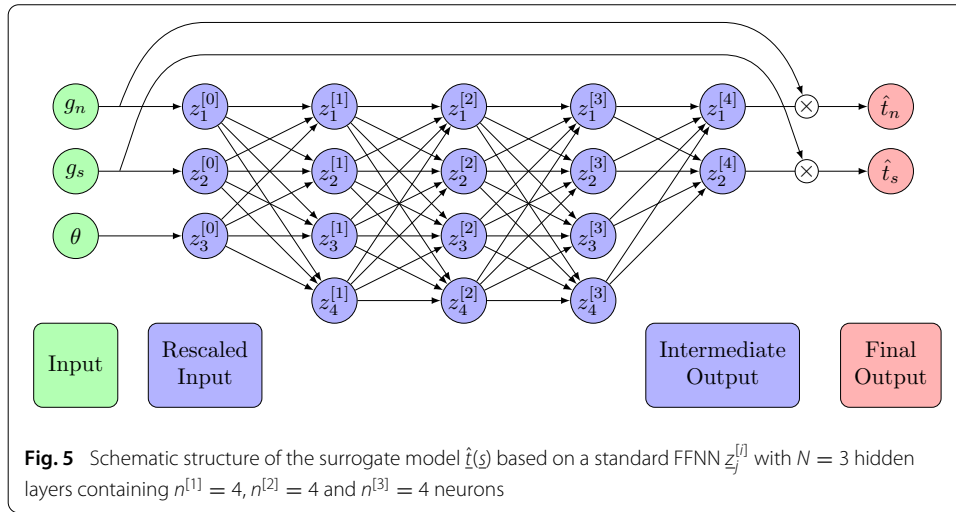
and the surrogate model  $\hat{\underline{t}}(\underline{s})$  for the traction law is then defined as

$$\hat{\underline{t}}(\underline{s}) = \begin{bmatrix} \hat{t}_n(\underline{s}) \\ \hat{t}_s(\underline{s}) \end{bmatrix} = \begin{bmatrix} \hat{t}_1(\underline{s}) \\ \hat{t}_2(\underline{s}) \end{bmatrix} = \begin{bmatrix} s_1 z_1^{[N+1]} \\ s_2 z_2^{[N+1]} \end{bmatrix} \in \mathbb{R}^2, \quad \underline{z}^{[N+1]} = \underline{W}^{[N+1]} \underline{z}^{[N]} \in \mathbb{R}^2, \quad (4)$$

such that by construction  $\hat{\underline{t}}(\underline{s})$  vanishes for  $\underline{g} = \underline{0}$  for all temperatures. Further, the ansatz (4) ensures  $\hat{t}_s = 0$  for  $g_s = 0$  independent of  $g_n$  and  $\hat{t}_n = 0$  for  $g_n = 0$  independent of  $g_s$ , which are properties relevant for crack opening modes I and II. The FFNN output  $\underline{z}^{[N+1]}$  can be interpreted as the secant stiffness/slope of the traction with respect to the gap vector  $\underline{g}$ . While the secant stiffness  $\underline{z}^{[N+1]}$  is the output of a standard FFNN, the surrogate  $\hat{\underline{t}}(\underline{s})$  can be considered as ResNet-like structure, which builds upon feature re-use and has shown very satisfactory results in the ANN literature, see, e.g., [67] or [68]. The layers of the standard FFNN  $\underline{z}^{[i]}$  for  $i = 1, \dots, N$  are referred to as hidden layers, while the intermediate output  $\underline{z}^{[N+1]}$  is referred to as the output layer of the standard FFNN. The FFNN is characterized by the parameters:

1. number of hidden layers  $N$ ,
2. number of neurons  $n^{[i]}$  of each hidden layer  $i = 1, \dots, N$ ,
3. element-wise activation function  $a^{[i]}(x)$  of each hidden layer  $i = 1, \dots, N$ ,
4. weights  $\underline{W}^{[i]} \in \mathbb{R}^{n^{[i]} \times n^{[i-1]}}$  of each layer  $i = 1, \dots, N + 1$ , and
5. biases  $\underline{b}^{[i]} \in \mathbb{R}^{n^{[i]}}$  of each hidden layer  $i = 1, \dots, N$ .

As an illustrative example, the surrogate model  $\hat{\underline{t}}(\underline{s})$  with a FFNN with  $N = 3$ ,  $n^{[1]} = 4$ ,  $n^{[2]} = 4$ ,  $n^{[3]} = 4$  is schematically depicted in Fig. 5.



The determination of the parameters of a network is a highly non-trivial problem common to all ANNs. First, the number of hidden layers, number of neurons and activation functions are fixed and then the weights and biases of the FFNN are determined through the minimization of a suitable error measure with respect to provided data. The identification of the meta-parameters, i.e., the optimal number of hidden layers and number of neurons per layer, requires extensive testing, often referred to as architecture sweeping in Machine Learning. The choice of activation functions is guided by the problem at hand and by the requirements on the final surrogate model. For the present problem setting, since the final surrogate model  $\hat{t}(s)$  is needed to be differentiable for all states which is due to the necessity of a tangent operator in future multiscale simulations, the following activation functions are taken into consideration:

1. Softplus  $a(x) = \text{sp}(x) = \log(1 + \exp(x))$ ,
2. Hyperbolic tangent  $a(x) = \tanh(x) = (\exp(2x) - 1)/(\exp(2x) + 1)$ .

The softplus function is positive, unbounded and offers a smooth approximation of the widely used rectified linear unit  $a(x) = \max(x, 0)$ , which mimics the behavior of a neuron/unit only reacting if the input signal  $x$  is greater than the threshold, i.e.  $x \geq 0$ . The hyperbolic tangent offers a smooth and saturating transition for values in the interval  $(-1, 1)$  mimicking a neuron/unit yielding positive or negative output depending on the input signal  $x$ . The chosen activation functions have their most important transition zones around  $x = 0$ . Therefore, the rescaling  $r(s)$  in (3) should bring a given state to this transition zone. The present work simply considers a rescaling with component-wise mean and standard deviation with respect to a given dataset of states.

The activation functions can, in general, be either chosen for each layer individually or the same one is used for all layers. In the present work, we use a single activation function for all hidden layers and we keep the number of neurons per layer constant, i.e.,  $a^{[i]}(x) = a(x)$  and  $n^{[i]} = n$  for all  $i = 1, \dots, N$ . However, the weights and biases are allowed to differ from layer to layer. The architecture of the FFNN explored in the sequel is, therefore, parameterized by

$$A = \{N, n, a(x)\}. \tag{5}$$

For fixed architecture, i.e., given  $N$ ,  $n$  and  $a(x)$ , the weights and biases of all layers, all together simply addressed from now on as  $\underline{W}$  and  $\underline{b}$  for a compact notation, respectively, need calibration. In order to do so, we consider three disjoint groups of datasets:

- Calibration dataset

$$D^C = \{\underline{s}^{C(1)}, \dots\} \subset \mathbb{R}^3, \quad D_{\underline{t}}^C = \{\underline{t} \in \mathbb{R}^2 : \underline{t} = \underline{t}(\underline{s}), \underline{s} \in D^C\}, \quad (6)$$

- Validation dataset

$$D^V = \{\underline{s}^{V(1)}, \dots\} \subset \mathbb{R}^3, \quad D_{\underline{t}}^V = \{\underline{t} \in \mathbb{R}^2 : \underline{t} = \underline{t}(\underline{s}), \underline{s} \in D^V\}, \quad (7)$$

- Testing dataset

$$D^T = \{\underline{s}^{T(1)}, \dots\} \subset \mathbb{R}^3, \quad D_{\underline{t}}^T = \{\underline{t} \in \mathbb{R}^2 : \underline{t} = \underline{t}(\underline{s}), \underline{s} \in D^T\}. \quad (8)$$

The number of samples of a dataset  $D$  will be denoted as  $\#(D)$ . We consider the mean squared error (MSE) of the surrogate  $\hat{\underline{t}}(\underline{s})$  with respect to a dataset  $D$

$$MSE(D) = \frac{1}{2} \sum_{i=1}^2 \frac{1}{\#(D)} \sum_{\underline{s} \in D} [t_i(\underline{s}) - \hat{t}_i(\underline{s})]^2 \quad (9)$$

and the coefficient of determination  $R^2$  (pronounced: R squared) with respect to a dataset  $D$

$$R^2(D) = \frac{1}{2} \sum_{i=1}^2 \left[ 1 - \frac{\sum_{\underline{s} \in D} [t_i(\underline{s}) - \hat{t}_i(\underline{s})]^2}{\sum_{\underline{s} \in D} [t_i(\underline{s}) - \bar{t}_i(D)]^2} \right], \quad \bar{\underline{t}}(D) = \frac{1}{\#(D)} \sum_{\underline{s} \in D} \underline{t}(\underline{s}) \quad (10)$$

The MSE offers the possibility to minimize the average square error. Thereby, the model is calibrated aiming at an accurate average behavior. This is particularly useful in the presence of noisy data, such as the MD simulation data of this work. Note that errors in the tractions in tangential and normal direction are treated equally which can be considered as an isotropic error measure. The  $R^2$  score measures the component-wise average quality of the prediction of the surrogate  $\hat{\underline{t}}(\underline{s})$  compared to the simple average over the data set  $\bar{\underline{t}}(D)$ . Hereby, one should note that the term  $\sum_{\underline{s} \in D} [t_i(\underline{s}) - \bar{t}_i(D)]^2$  for  $i \in \{1, 2\}$  corresponds to the (scaled) component-wise sample variance for the dataset  $D$ . Of course, the MSE is bounded from below by 0, while  $R^2$  is bounded from above by 1, such that these values indicate a “good quality” of a surrogate model. It is shortly remarked that  $R^2$  is not the square of any quantity, but a simple measure commonly used in ML for the quality assessment of a model. More explicitly,  $R^2$  can be negative, e.g., consider the data  $D_{\underline{t}} = \{(1, 1), (2, 2), (3, 3)\}$  and some model predictions  $\hat{D}_{\underline{t}} = \{(3, 3), (2, 2), (1, 1)\}$ , which yield  $R^2(D) = -3$ , i.e., the trained model is worse than simply using the average  $\bar{\underline{t}}(D)$  over the data for new predictions.

The training of the FFNN with fixed architecture is performed as follows: First, the calibration dataset  $D^C$  is used to build the rescaling function  $r(\underline{s})$  in (3) as follows ( $i \in \{1, 2, 3\}$ )

$$r_i(\underline{s}) = \frac{s_i - \mu_i}{\sigma_i}, \quad \mu_i = \frac{1}{\#(D^C)} \sum_{\underline{s} \in D^C} s_i, \quad \sigma_i = \frac{1}{\#(D^C)} \sum_{\underline{s} \in D^C} (s_i - \mu_i)^2, \quad (11)$$

i.e.,  $r(\underline{s})$  has then zero mean and component-wise a standard deviation of 1 with respect to  $D^C$ . Next, we employ the MSE with respect to  $D^C$  as the objective function  $\psi = MSE(D^C)$ , referred to as loss function in the ANN literature. Then, the weights  $\underline{W}$  and biases  $\underline{b}$  of

the FFNN are determined through the minimization of the loss function  $\psi$ :

$$\min_{\underline{w}, \underline{b}} \psi. \quad (12)$$

It is important to note that the weights and biases are updated based on the minimization of  $\psi$ , i.e., exclusively based on the calibration dataset  $D^C$ . In particular,  $D^V$  and  $D^T$  do not influence the weights and biases until now. The dataset  $D^V$  is used during the minimization of  $\psi$  not to update the weight and biases, but (i) to keep track of the validation measures  $MSE(D^V)$  and  $R^2(D^V)$  and (ii) to terminate the minimization of the loss function  $\psi$  if these validation measures do not improve over a determined number of iterations over the complete calibration data (referred to as epochs in the ANN literature). The use of validation data ( $D^V$ ) alongside the actual calibration data ( $D^C$ ) is well-established today. Besides sensing stagnation of the loss function, corresponding to local or global minima of  $\psi$ , a key responsibility of the validation set is the detection and avoidance of overfitting, which corresponds to a rise of the loss function  $\psi$  on the control group  $D^V$  despite decaying  $\psi$  on the calibration set  $D^C$ . The identification of the weights and biases based on calibration and validation data is referred to as training of the network. After training, the network should undergo final testing. This can not be carried out with  $D^C$  or  $D^V$ , since the network is optimized for  $D^C$  and is indirectly influenced by  $D^V$  due to the training termination. Therefore, the final testing of the trained network is performed based on the supplementary dataset  $D^T$ , which the network has never “seen”. In this work, we classify trained networks based on their respective scores for  $R^2(D^T)$  after training. One should note that  $D^T$  should be of substantial size in order to allow for thorough testing. On the other hand, as the input data is scarce due to the expensive MD simulations, the overall amount of available data is limited. A compromise must, thus, be made in order to still have sufficient calibration data while retaining representative validation and testing data.

Naturally, different architectures should be tested. An architecture sweep is referred to as the optimization (12) for every architecture of the architecture set

$$\mathcal{A} = \{A_1, \dots\}, \quad A_i = \{N_i, n_i, a_i(x)\}, \quad (13)$$

at what the minimization (12) is performed for each architecture several times with random initialization of the weights and biases. The best performing architectures in terms of  $R^2(D^T)$  are then considered as candidates for the surrogate  $\hat{t}(\underline{s})$ .

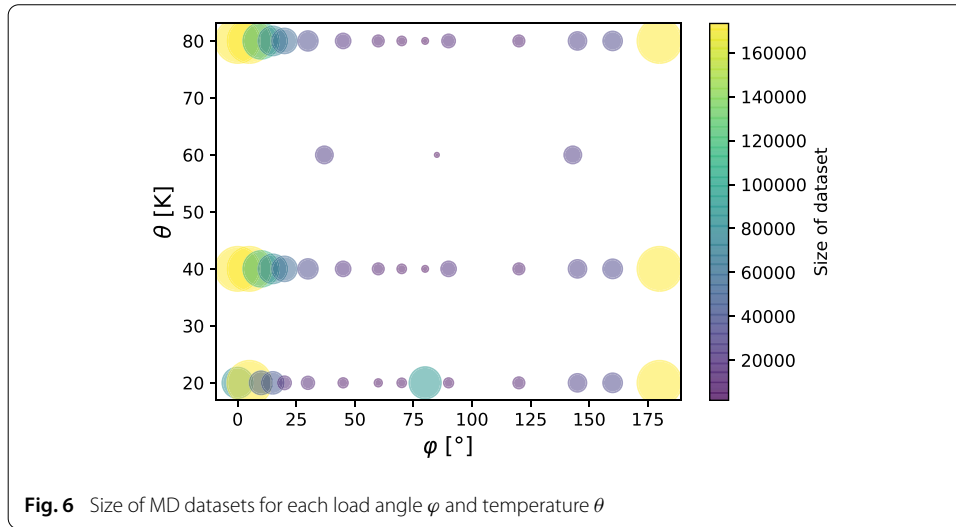
### Explicit datasets and training of surrogate model

The MD simulation data is organized in the following two datasets  $D_1^{MD}$  and  $D_2^{MD}$

$$\begin{aligned} D_1^{MD} &= \{\underline{s} \in \mathbb{R}^3 : \underline{s} = (g_1, g_2, \theta) \text{ for load angle } \varphi \in D_1^\varphi \text{ and temperature } \theta \in D_1^\theta\}, \\ D_1^\varphi &= \{0^\circ, 5^\circ, 10^\circ, 15^\circ, 20^\circ, 30^\circ, 45^\circ, 60^\circ, 70^\circ, 80^\circ, 90^\circ, 120^\circ, 145^\circ, 160^\circ, 180^\circ\}, \\ D_1^\theta &= \{20 \text{ K}, 40 \text{ K}, 80 \text{ K}\}. \end{aligned} \quad (14)$$

$$\begin{aligned} D_2^{MD} &= \{\underline{s} \in \mathbb{R}^3 : \underline{s} = (g_1, g_2, \theta) \text{ for load angle } \varphi \in D_2^\varphi \text{ and } \theta \in D_2^\theta\}, \\ D_2^\varphi &= \{37^\circ, 85^\circ, 143^\circ\}, \quad D_2^\theta = \{60 \text{ K}\}. \end{aligned} \quad (15)$$

The reason for the distinction between  $D_1^{MD}$  and  $D_2^{MD}$  will be motivated in this section. The dataset  $D_1^{MD}$  is comprised by a total of 2,600,341 data points, while  $D_2^{MD}$  contains 54,421 data points. Due to the nature of MD simulations, the corresponding  $\underline{g}$  and  $\underline{t}$  data are fluctuating. Furthermore, the number of recorded equilibrium states in each MD



simulation varies depending on the load angle  $\varphi$  and temperature  $\theta$ . This is illustrated in Fig. 6. These properties of the MD datasets pose two challenges for the training of a surrogate model: (i) noisy input and output quantities ( $\underline{g}$  and  $\underline{t}$ , respectively) and (ii) oversampling in certain regions of the input- and output-space.

While noisy data is critical in direct interpolation approaches, the FFNNs used in our study are calibrated against the MSE which represents an average error measure. The MD training data can, therefore, be left untouched, i.e., no data post-processing is required. This is an important point, since one may also apply smoothing procedures on the MD data in order to try to improve the calibration of a surrogate model. But, for the MD data of this investigation, not only the output  $\underline{t}$ , but also the input  $\underline{g}$  would have to be smoothed, each of them independently. Every smoothing algorithm would then falsify the connection between input and output to some arbitrarily unclear degree, depending on arbitrary smoothing parameters (choice of ansatz functions, size of smoothing window,...) and assumptions on the relationship between input and output. The present work, therefore, regards the direct usage of the raw MD data as an unbiased feature of the calibration of FFNNs as a surrogate model.

For fixed architecture of a FFNN, its training is performed with the objective of minimizing the loss function, i.e. the MSE over the calibration data, according to (12). Here, the second challenge—the oversampling for some  $\varphi$  and  $\theta$  in the MD datasets—needs to be considered. If, e.g., for  $\theta = 20$  K the number of recorded gap vectors  $\underline{g}$  and corresponding traction vectors  $\underline{t}$  for  $\varphi = 5^\circ$  is one order of magnitude larger than for  $\varphi = 60^\circ$ , cf. Figure 6, then the optimization of the loss function will be biased towards the bigger data set leaving certain temperatures and load angles at a lower coverage. This perspective makes clear that some sort of homogeneous data reduction is to be considered in the definition of  $D^{C,V,T}$ .

Aiming for (i) a sufficient coverage of the material behavior during the training and (ii) a challenging final testing of the surrogate, the dataset  $D_1^{\text{MD}}$  is considered for the construction of the calibration and validation datasets  $D^C$  and  $D^V$ , while  $D_2^{\text{MD}}$  has been deliberately designed for the test dataset  $D^T$ . The test dataset  $D_2^{\text{MD}} = D^T$  for final evaluation is not a subset of  $D_1^{\text{MD}}$  and not even defined on the load directions implicitly defining  $D_1^{\text{MD}}$

but at intermediate load angles and temperature. This is in contrast to standard training procedures of ANNs, where, usually, the split  $D^{C,V,T}$  is not even mentioned or  $D^{C,V,T}$  are extracted from one common dataset, which has been shuffled randomly. For such approaches, due to the extraction of points from a common randomly shuffled dataset, the corresponding  $D^{C,V,T}$  datasets do not differ substantially such that the prediction quality of a trained network shows no significant deviation comparing  $D^C$  and  $D^T$ . This could easily be achieved by the union of  $D_1^{MD}$  and  $D_2^{MD}$ , random shuffling of the data pairs and corresponding percentual definition of  $D^{C,V,T}$ . But, in contrast to such standard approaches, the present work explicitly aims at a more challenging testing of the trained networks and, therefore, chooses to consider the two distinct datasets  $D_1^{MD}$  and  $D_2^{MD}$  and to define  $D^T$  through  $D_2^{MD}$ . Further, the MD simulations allow only for few discrete load directions. Hence, the generalization capabilities of trained networks should be tested more thoroughly to prevent unintended overfitting along the training directions.

Based on the previously discussed arguments, the FFNNs are trained in a two-phase procedure as follows:

- **Phase 1**—optimizing using homogeneous number of samples

P1.1 For every load angle  $\varphi$  and temperature  $\theta$ , extract 1500 random data pairs  $(g, t)$  per load path from the available MD data, which then constitute the corresponding reduced datasets  $D_{1,2}^{MD,r}$ . The reduced data sets contain then  $\#(D_i^{MD,r}) = \#(D_i^\varphi) \times \#(D_i^\theta) \times 1500$  data points, for  $i = 1, 2$ , i.e.,  $\#(D_1^{MD,r}) = 67,500$  and  $\#(D_2^{MD,r}) = 4,500$  with  $\#(D_1^{MD,r})/\#(D_1^{MD}) \approx 0.0260$  and  $\#(D_2^{MD,r})/\#(D_2^{MD}) \approx 0.0827$ .

P1.2 Define 70% of  $D_1^{MD,r}$  as  $D^C$ , 30% of  $D_1^{MD,r}$  as  $D^V$  and  $D^T = D_2^{MD,r}$ .

P1.3 Run the architecture sweep for the architecture set  $\mathcal{A}$  with

$$\begin{aligned} N &\in \{2, 3, 4, 5, 6\}, \\ n &\in \{4, 5, 6, 7, 8, 16, 32\}, \\ a(x) &\in \{\text{sp}(x), \tanh(x)\}, \\ \mathcal{A} &= \{2, 3, 4, 5, 6\} \times \{4, 5, 6, 7, 8, 16, 32\} \times \{\text{sp}(x), \tanh(x)\}. \end{aligned} \quad (16)$$

Each architecture is initialized four times, each time with new random initial values for the weights and biases of the network.

P1.4 Save trained FFNNs with the five best  $R^2(D^T)$ .

- **Phase 2**—refinement using variable number of samples per load path

P2.1 Use the complete MD data, i.e., define 70% of  $D_1^{MD}$  as  $D^C$ , 30% of  $D_1^{MD}$  as  $D^V$  and  $D^T = D_2^{MD}$ .

P2.2 Retrain the trained FFNNs of Phase 1 with the new  $D^{C,V}$ .

P2.3 Save trained FFNNs with the three best  $R^2(D^T)$ .

It should be pointed out, that the approach taken in Phase 1 aims at an initial homogeneous data reduction with equal number of points for each  $\varphi$  and  $\theta$ . In Phase 2, the complete datasets are taken into account for final tuning of the best performing networks.

The corresponding FFNNs were implemented and trained with Google's TensorFlow (v1.12.0) with Python3 (v3.4.4). For the training routines, whole batch training was carried out and TensorFlow's ADAM optimizer was used, see [69] and Tensorflow's documen-

**Table 1 Best performing FFNN sorted by  $R^2(D^T)$  after Phase 2**

ID	Architecture $\{N, n, a(x)\}$	Phase 2		
		$R^2(D^T)$	$R^2(D^C)$	$R^2(D^V)$
FFNN1	{4, 8, tanh(x)}	0.9475	0.9756	0.9756
FFNN2	{4, 7, tanh(x)}	0.9442	0.9736	0.9736
FFNN3	{3, 8, tanh(x)}	0.9401	0.9727	0.9727

tation of ADAM for details. The learning rate  $\eta$  of the ADAM optimizer was decreased every 10% of the maximum number of training epochs by a factor  $\rho$ . Note that each reduction of the learning rate leads to a reinitialization of the ADAM solver. This can help to overcome local minima and stagnation. Training was terminated if the validation  $MSE(D^V)$  did not improve over 30% of the maximum number of epochs. The state of the network with the lowest  $MSE(D^V)$  during training was returned for each architecture. Then the trained FFNN was tested with  $R^2(D^T)$ . During Phase 1, 10,000 epochs, a starting learning rate of  $\eta = 0.1$  and  $\rho = 0.7$  were considered, yielding a final learning rate of  $\eta = 0.1 \times (0.7^9) \approx 0.004$ . Phase 2 was performed with 100 epochs, an initial  $\eta = 0.002$  and  $\rho = 0.95$ . The best performing FFNNs obtained are tabulated in Table 1.

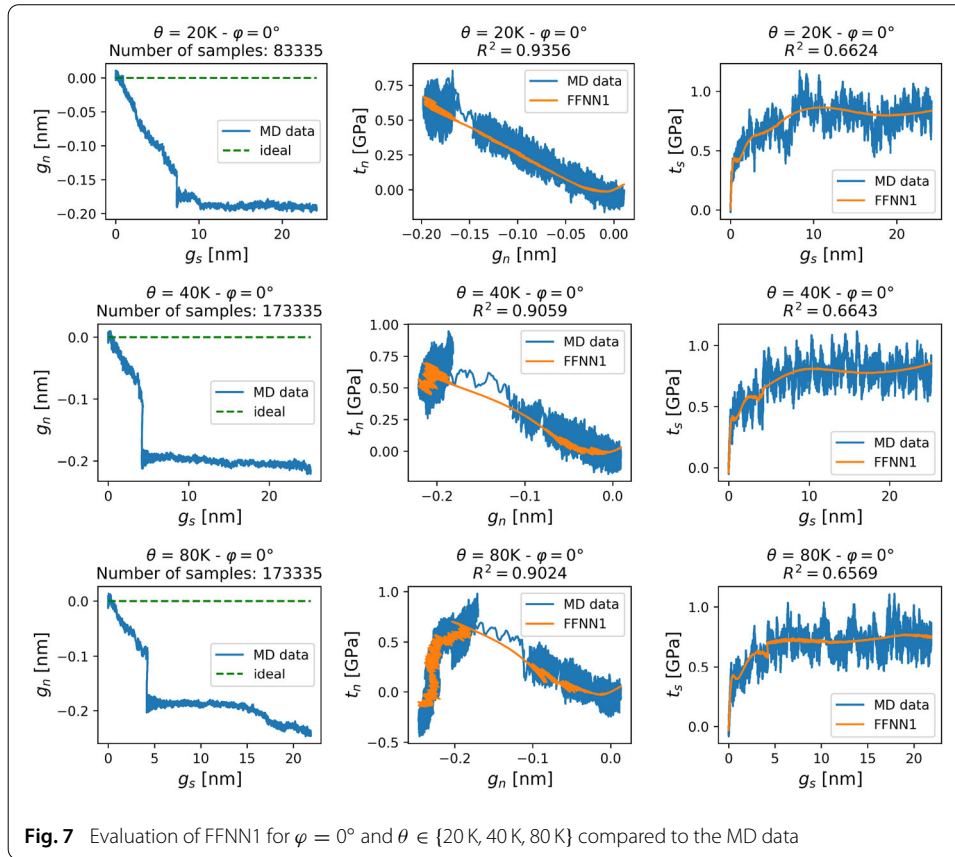
In Table 1, the best performing network of the present work reaches a maximum  $R^2(D^T)$  of 0.9475, which is assumed to be still acceptable due to the highly oscillatory MD data. The network FFNN1 of Table 1 is comprised of  $N = 4$  hidden layers with  $n = 8$  neurons each and it is built upon the  $\tanh(x)$  activation function. FFNN1 is chosen as the surrogate model for further inspection. The corresponding  $\hat{t}^{\text{FFNN1}}(s)$  is simply addressed as  $\hat{t}(s)$ .

#### Selected predictions of the surrogate model

For a more intuitive mechanical interpretation of the following evaluation of the surrogate model, in this section we switch back the subscript notation  $n$  and  $s$  for normal and shear components, , i.e.  $1 \leftrightarrow n$  and  $2 \leftrightarrow s$ , respectively, are used.

*Predictions for  $D_1^{\text{MD}}$*  We first examine the predictions with respect to the dataset  $D_1^{\text{MD}}$ , cf. (14), from which the calibration and validation datasets  $D^C$  and  $D^V$  are both extracted. Evaluation of the FFNN1  $\hat{t}(s)$  for the load angles  $\varphi \in \{0^\circ, 30^\circ, 90^\circ, 145^\circ\} \subset D_1^\varphi$  for all training temperatures  $\theta \in \{20 \text{ K}, 40 \text{ K}, 80 \text{ K}\}$  is depicted in Figs. 7, 9, 11 and 12. Hereby, the gap vectors  $g$  of the MD simulations (which do not always strictly follow  $\varphi$ ) have been used in order to evaluate the FFNN1 predictions.

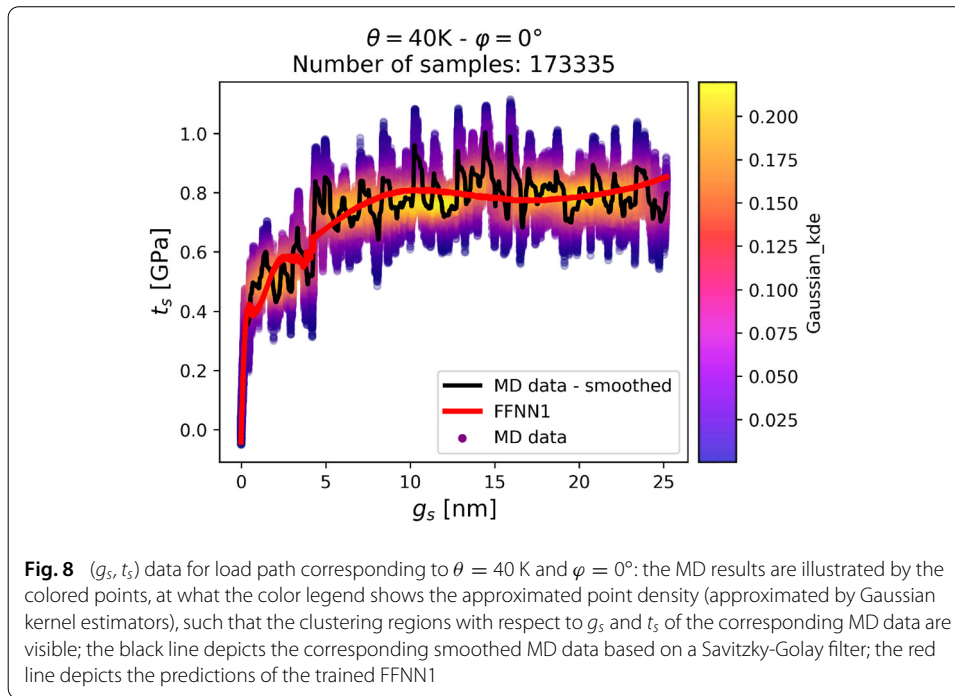
In Fig. 7 the evaluation of FFNN1 for  $\varphi = 0^\circ$  shows the performance of FFNN1 close to mode II behavior. It should be remarked that the MD data shows oscillatory behavior for  $\underline{t}$  and  $\underline{g}$ . Further, due to measuring approach of the gap vector close to the vicinity of the GB, cf. Sect. , the measured  $\underline{g}$  (depicted in blue in the left plots in Fig. 7) does not exactly follow the load angle  $\varphi = 0^\circ$  (depicted by the green dashed line in the left plots in Fig. 7). Due to the non-vanishing  $g_n$ , the material law is evaluated, at what the material response possesses steep gradients in the vicinity of the origin. The calibrated FFNN1 is evaluated along the recorded MD path, which jumps back and forward w.r.t.  $g_n$ , yielding the oscillatory curves in the middle plots of Fig. 7. The oscillatory displayed behavior is not a problem of FFNN1 (which is single-valued), but a consequence of the evaluation of the oscillatory non-ideal gap vector path of the MD simulations. Despite the non-ideal



path of the MD gap vector, the prediction of FFNN1 (depicted in orange in Fig. 7), seems to capture the average behavior of the grain boundary even for large shear offsets of the opposing crack surfaces ( $g_s$ ).

First, it should be stressed again, that the FFNNs of the present work have been trained directly on the oscillating raw MD data based on the MSE. Consider the load path for  $\theta = 40\text{ K}$  for the load angle  $\varphi = 0^\circ$  (see, also, Fig. 7, right column, middle plot) contained in  $D_1^{\text{MD}}$ , cf. (14). For this case, the  $(g_s, t_s)$  data is illustrated in more detail in Fig. 8. Here, a Gaussian kernel density estimator has been applied to the MD data in order to visualize the sample clustering along the load path. Further, a Savitzky-Golay (SG) filter for a smoothing window of 1/40-th of the number of samples and a polynomial order of 3 have been used. It should be stressed that the SG filter smooths the  $g_s$  and the  $t_s$  separately. The SG filter yields the black curve displayed in Fig. 8. The SG filtered data follows the effective behavior, at what due to the small smoothing window length (1/40-th of the number of samples) an oscillatory behavior is still visible. This can be suppressed by, e.g. choosing an even larger smoothing window size, which is one of many arbitrary smoothing parameters. In Fig. 8 it is visible that the trained network FFNN1 is able to approximate the average behavior of the effective traction law in a smooth fashion, without any biased data processing. The visualization shown in Fig. 8 clearly states that, assuming a normal distribution for the oscillations, the FFNN finds almost the maximum likelihood approximation of the input data. In terms of the prediction quality based on the  $R^2$  values based solely on the  $(g_s, t_s)$  data displayed in Fig. 8, the SG filtered data yields  $R^2 \approx 0.8942$ , while FFNN1 reaches a

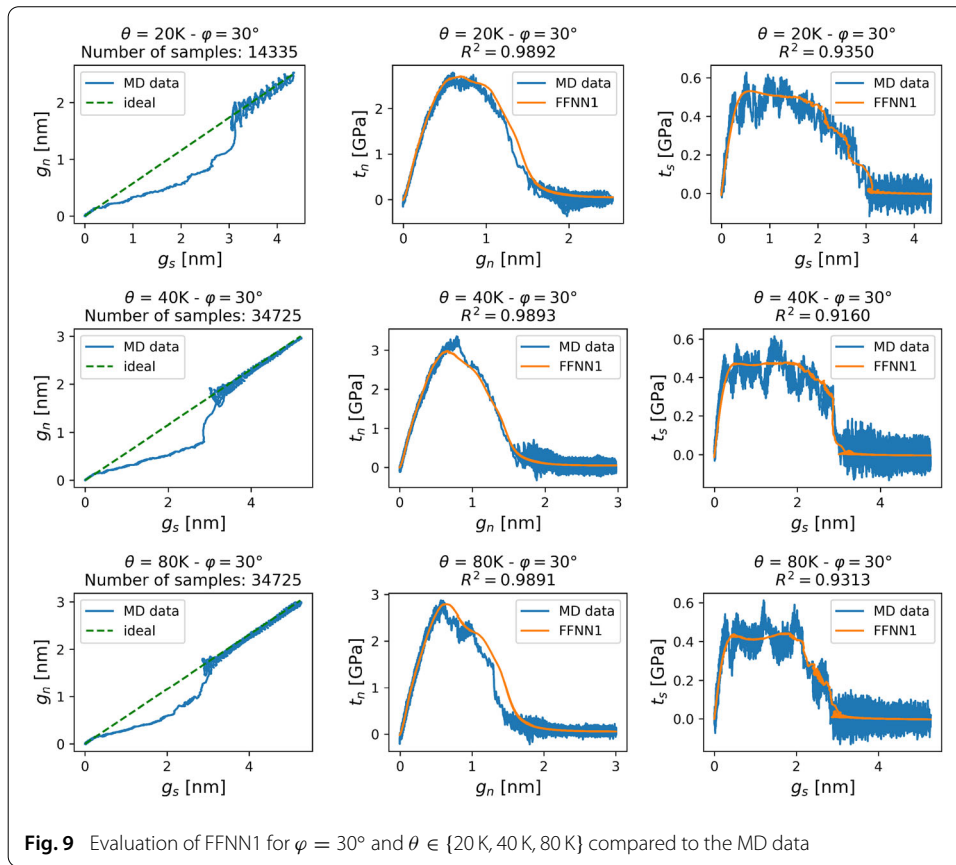




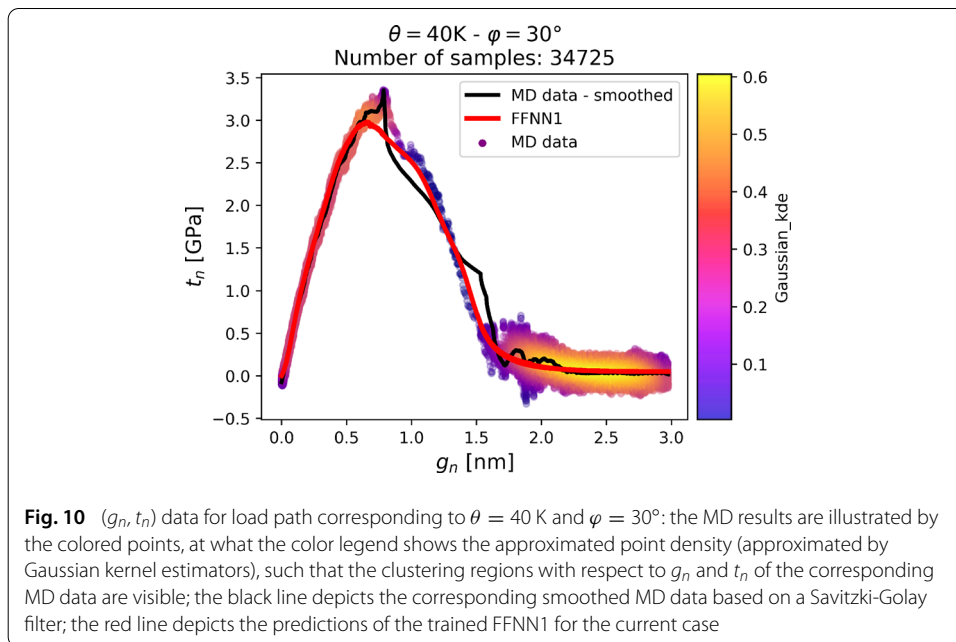
value of  $R^2 \approx 0.6643$ . This is due to rather homogeneous distribution of  $g_s$  and the highly noisy  $t_s$  values for the current case. This shows that high  $R^2$  values in the prediction of the MD data for  $\varphi = 0^\circ$  is not only challenging for neural networks, but even for flexible smoothing filters.

For increasing load angle  $\varphi$ , a mixture between mode I and mode II is to be accounted for. Such a scenario is depicted in Fig. 9 for  $\varphi = 30^\circ$ . Again, as remarked in the discussion of Fig. 7, the measured gap vector of the MD simulations does not follow the ideal direction at all simulation steps, see left plots in Fig. 9. The FFNN1 is able to extract the corresponding average behavior of the normal and shear component of the traction vector up to failure (represented by virtually zero tractions  $\hat{t}(s)$ ). Furthermore, the surrogate model  $\hat{t}(s)$  remains close to zero for post-critical gap openings, which highlights the consistency of the surrogate model. It should be noted, that the surrogate model  $\hat{t}(s)$ , cf. (4), has been constructed only to vanish identically for  $g_n = 0\text{ nm}$  or  $g_s = 0\text{ nm}$ . The behavior of the trained FFNN1 almost vanishing after material failure was not specifically accounted for. This behavior is captured through the trained network solely by the provided MD data, despite of its highly noisy nature.

We shortly consider the case for  $\theta = 40\text{ K}$  and  $\varphi = 30^\circ$  and the  $(g_n, t_n)$  data displayed in Fig. 10 in more detail. The same parameters as for  $\theta = 40\text{ K}$  and  $\varphi = 0^\circ$  (yielding Fig. 8) have been applied for the Gaussian kernel density estimator and the SG filter. In Fig. 10 it immediately becomes visible that up to the peak value of  $t_n$  the sample density is relatively homogeneous with respect to  $g_s$  and only small oscillations of  $t_n$  are observed. Then, debonding events occur very quickly, such that only a few  $g_n$  points with decreasing  $t_n$  are available. After material failure, the normal tractions  $t_n$  remain in the vicinity of zero with a concentrated, but highly noisy behavior. The corresponding SG filtered  $(g_n, t_n)$  data is depicted through the black line in Fig. 10. The SG filter yields data points close to the effective behavior but shows a weak performance in the material failure phase. This

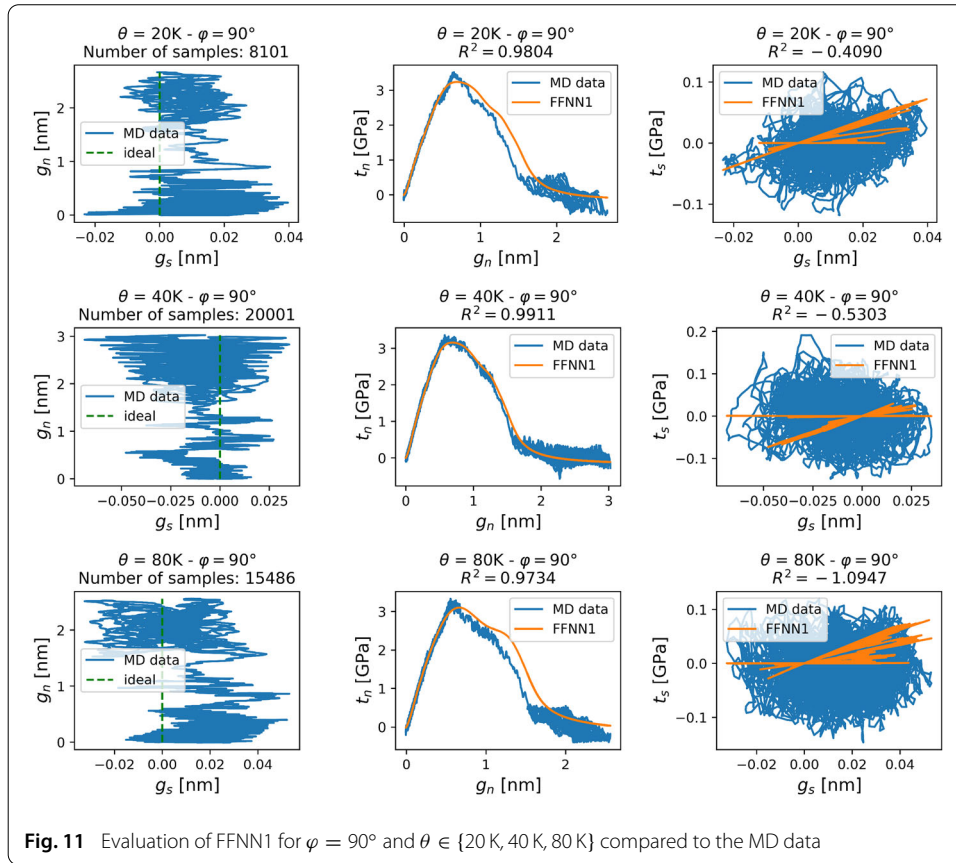


**Fig. 9** Evaluation of FFNN1 for  $\varphi = 30^\circ$  and  $\theta \in \{20\text{ K}, 40\text{ K}, 80\text{ K}\}$  compared to the MD data



**Fig. 10** ( $g_n, t_n$ ) data for load path corresponding to  $\theta = 40\text{ K}$  and  $\varphi = 30^\circ$ : the MD results are illustrated by the colored points, at what the color legend shows the approximated point density (approximated by Gaussian kernel estimators), such that the clustering regions with respect to  $g_n$  and  $t_n$  of the corresponding MD data are visible; the black line depicts the corresponding smoothed MD data based on a Savitzki-Golay filter; the red line depicts the predictions of the trained FFNN1 for the current case

occurs due to the separate smoothing of  $g_n$  and  $t_n$ , which then falsifies the relation to an unclear/arbitrary degree. This could be enhanced, based on possibly suitable assumptions for the current process, with an even smaller smoothing window size (i.e, smaller than



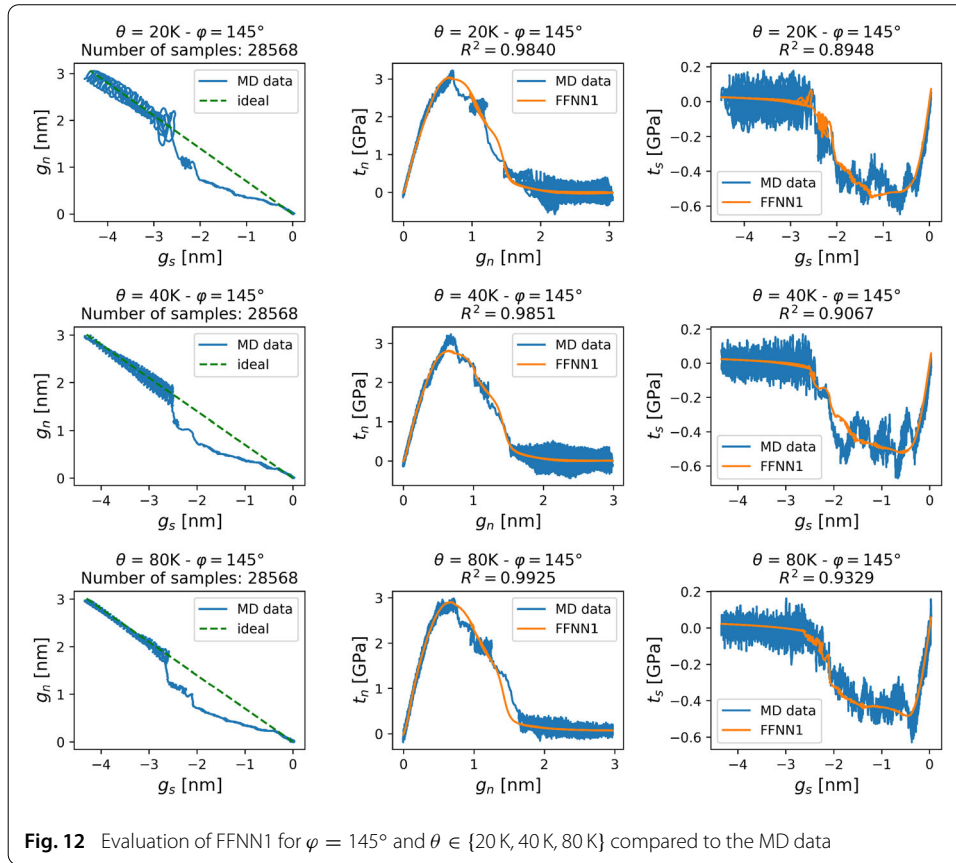
**Fig. 11** Evaluation of FFNN1 for  $\varphi = 90^\circ$  and  $\theta \in \{20\text{ K}, 40\text{ K}, 80\text{ K}\}$  compared to the MD data

1/40th of the number of samples) and, possibly, with alternative polynomial order (i.e., other than 3). This shows that for the problem at hand, a smoothing algorithm needs to be tuned for each load path individually due to inhomogeneous point distribution and region dependent noise of the MD results. The calibrated FFNN1 shows a satisfactory approximation of the current load path, even in the material failure phase. As for the discussion corresponding to Fig. 8, the visualization Fig. 10 also strongly indicates, that FFNN1 offers an excellent alternative close to the maximum likelihood approximation of the data, assuming a normal distribution for the oscillations.

At  $\varphi = 90^\circ$  pure mode I is considered. In Fig. 11 it can be seen that FFNN1 is able to follow the qualitative behavior up to failure (see middle plots in Fig. 11), while the shear component of the traction vector stays around zero due to the construction of  $\hat{t}(s)$ , cf. (4).

As depicted in Fig. 12, it can be seen that the network is also able to capture the effective behavior for  $\varphi = 145^\circ$ . Load angles beyond  $\varphi = 90^\circ$  are also of interest since—based on the topology of the grain boundary and the dependency on the load direction—a non-symmetric behavior could be observed.

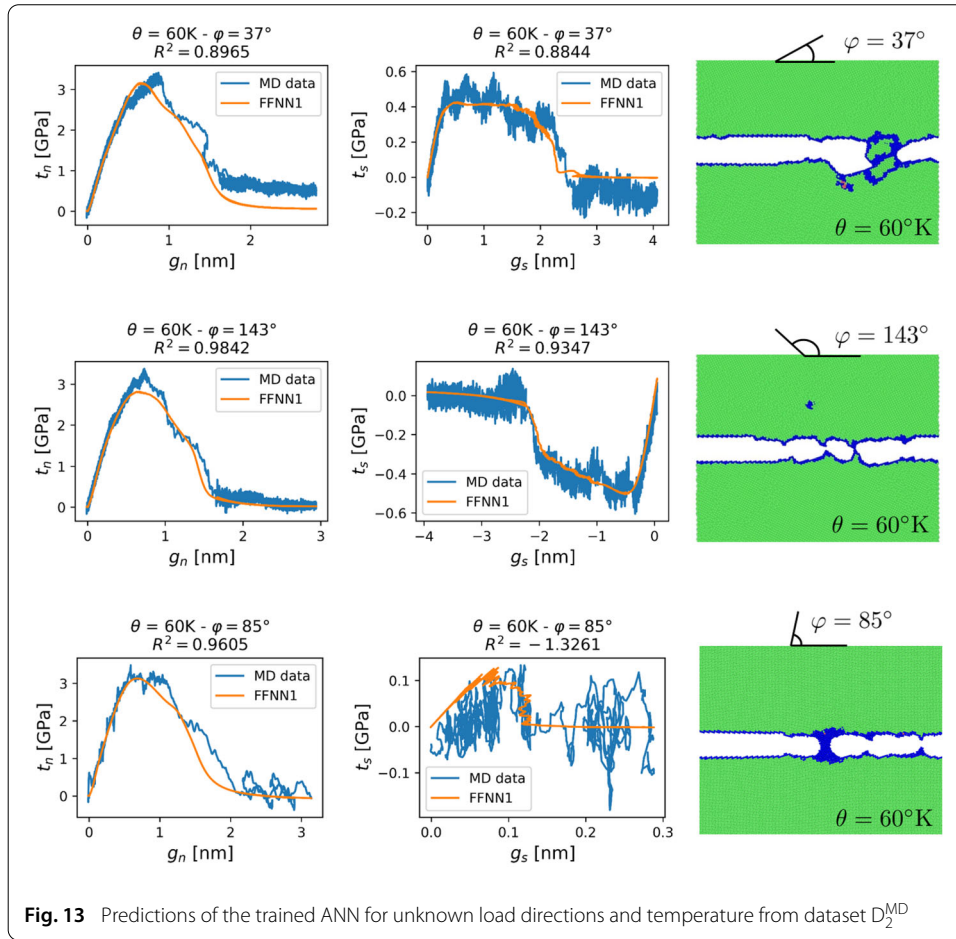
*Prediction for  $D_2^{\text{MD}}$*  Next, we look at the prediction quality of FFNN1 for the cases in the test dataset  $D_2^{\text{MD}}$ , cf. (15). Note that these are the load and temperature conditions that were not used during the training of the ANN. Fig. 13 shows the traction-separation prediction of FFNN1 on top of the MD results. As it is seen in Fig. 13, ANN predicts relatively good traction-separation curves for the cases not included in the training. In particular, the peak traction values, as well as final separation values are captured quite



well. In most of the curves, ANN predictions follow the shape of the curves in details. Note that the details of the traction-separation curves are due to atomic scale events during crack opening, as shown in right column of the figure.

*Evaluation of the surrogate model for ideal load angles* While the previous comparisons somewhat confirm that the MD data is well reproduced, the intrinsic appeal of the surrogate is that it accepts virtually arbitrary inputs  $\underline{g}$ . In Fig. 14 the evaluation for ideal gap vectors, i.e.,  $\underline{g}$  follows the ideal load angle  $\varphi^{\text{ideal}}$  exactly, is illustrated for different temperatures.

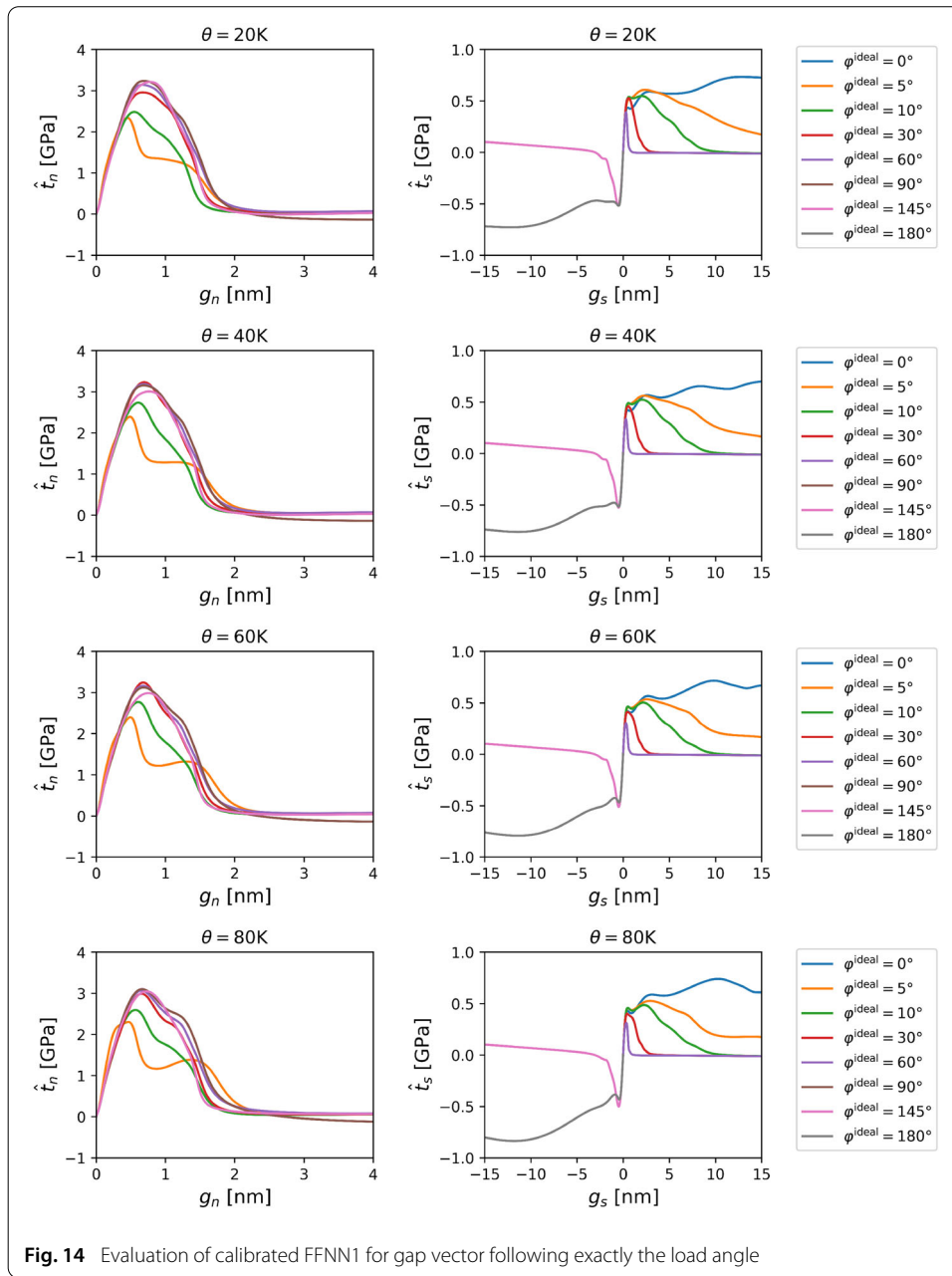
It should be noted that the evaluations presented in Fig. 14 are closely related to the predictions with respect to the MD data, but are not equal, since the gap vectors of the MD data do not follow the load angle at all times (see, e.g., left plots of Figs. 7 and 9). For example, evaluation of the trained surrogate  $\hat{t}(\underline{g})$  for the gap vectors corresponding to the load angle  $\varphi = 0^\circ$ , as shown in Fig. 7, is comparable to the corresponding evaluation of  $\hat{t}(\underline{g})$  for the ideal load angle  $\varphi^{\text{ideal}} = 0^\circ$ , depicted by the blue lines in Fig. 14. Hereby, it should be noted that  $\hat{t}_n$  vanishes identically for  $\varphi^{\text{ideal}} = 0^\circ$ . The evaluation of  $\hat{t}(\underline{g})$  for  $\varphi = 30^\circ$ , see Fig. 9, shows a good agreement with the evaluation for  $\varphi^{\text{ideal}} = 30^\circ$ , see red curves in Fig. 14. For ideal gap vector the trained  $\hat{t}$  seems to retain the effective behavior provided by the MD data. The extracted GB behavior is highly complex, e.g., the load curve for  $\varphi^{\text{ideal}} = 5^\circ$  in Fig. 14 seems counterintuitive, but due to the complexities of atomic bond formations at the GB, such behavior is possible, cf. Fig. 3 and discussion leading to Fig. 4. Analytical approaches capturing this behavior would require extensive investigation



**Fig. 13** Predictions of the trained ANN for unknown load directions and temperature from dataset  $D_2^{MD}$

and appropriate models. The current data-driven approach inherits this behavior based exclusively on the available MD data. Further, the surrogate model captures the transition behavior between pure mode II and pure mode I loading quite well. For pure mode II, i.e.,  $\varphi^{ideal} = 0^\circ$ ,  $\hat{t}_n = 0$  holds and  $\hat{t}_s$  stays in the vicinity of a constant value even for large  $g_s$ . This behavior resembles the ideally plastic shearing of a metal. For increasing ideal angle  $\varphi^{ideal}$ , material failure manifests itself at smaller magnitudes of  $g_s$ , while  $\hat{t}_n$  stabilizes fairly rapidly for  $\varphi^{ideal} \in [30^\circ, 145^\circ]$ . After material failure, tractions predicted by the surrogate drop to the vicinity of zero and show a stable behavior.

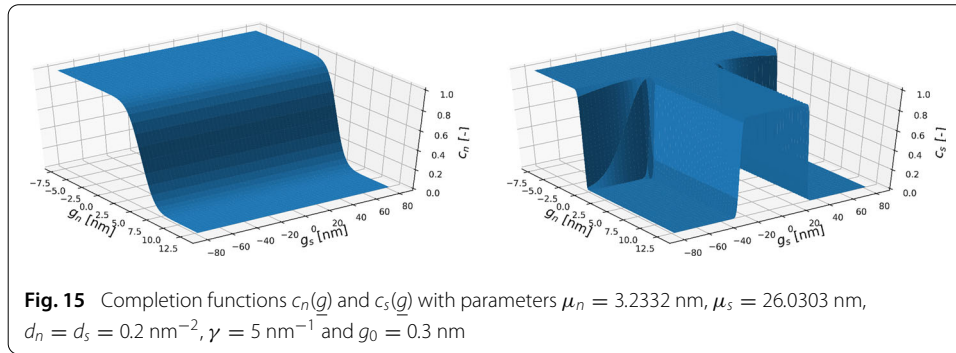
*Physics-guided surrogate completion* Finally, in view of future multiscale simulation, the behavior of the calibrated surrogate for input quantities outside of the training region needs to be addressed. In [57], calibrated FFNNs were used in two-scale problems as surrogates for the microscopic non-linear hyperelastic material law. There, during the macroscopic simulation, the surrogate was called in several Gauss points of the FE computation for input quantities outside of the training region. Due to the fact that in a general macroscopic or multi-scale simulation, the loading of a material point in terms of strain or displacement input cannot be foreseen, it is not possible to identify a general training region and train surrogates for all possible scenarios. As remarked by [57], the usage of a surrogate outside of the training region can be unreliable. In such cases, one may call for alternative physics-guided but computationally more expensive models, e.g.,



**Fig. 14** Evaluation of calibrated FFNN1 for gap vector following exactly the load angle

reduced-order models. Concerning the present work, it is clear that in future grain scale simulations, the surrogate  $\hat{\underline{t}}(s)$  most probably will be called for larger gap vector values than provided by the available MD simulation data. Hereby, based on the obvious behavior of material failure,  $\hat{\underline{t}}$  should drop asymptotically to zero for large  $\|g\|$ , except for pure mode II since friction remains active. Due to the clear limit behavior of the function to be approximated, the present work simply proposes to consider a physics-guided completion of the trained surrogate  $\hat{\underline{t}}(s)$  with respect to the gap vector as follows

$$\hat{\underline{t}}^c(s) = \begin{bmatrix} \hat{t}_n(s)c_n(\underline{g}) \\ \hat{t}_s(s)c_s(\underline{g}) \end{bmatrix} \tag{17}$$



with the completion functions

$$\begin{bmatrix} c_n(\underline{g}) \\ c_s(\underline{g}) \end{bmatrix} = \begin{bmatrix} \exp\left(-d_n \text{sp}(g_n - \mu_n)^2\right) \\ \exp\left(-d_s \chi_s(g_n) [\text{sp}(g_s - \mu_s) + \text{sp}(-g_s - \mu_s)]^2\right) \end{bmatrix}, \tag{18}$$

$$\chi_s(g_n) = \frac{1}{2}(\tanh(\gamma[g_n - g_0]) + 1), \quad \begin{bmatrix} \mu_n \\ \mu_s \end{bmatrix} = \begin{bmatrix} \max_{D_1^{\text{MD}} \cup D_2^{\text{MD}}} g_n \\ \max_{D_1^{\text{MD}} \cup D_2^{\text{MD}}; \varphi \geq 5^\circ} |g_s| \end{bmatrix}.$$

The parameter  $\mu_n$  is defined in (18) as the maximum observed  $g_n$  over all MD simulations. The parameter  $\mu_s$  defined in (18) corresponds to the maximum observed absolute value of  $g_s$  for  $\varphi \geq 5^\circ$ . The completion functions  $c_n$  and  $c_s$  ensure that for values beyond  $\mu_{n/s}$  an exponential decay of the tractions is activated through the softplus function  $\text{sp}(x)$ , at what the decay parameters  $d_{n/s} > 0$  can be chosen as required. The completion function  $c_s$  further aims at the deactivation of the exponential decay for  $g_n = 0$  (pure mode II) through the transition function  $\chi_s(g_n)$  with parameters  $\gamma, g_0 > 0$ . The proposed completion functions for the traction-separation problem at hand are differentiable, such that the completed surrogate  $\hat{t}^c(\underline{s})$  remains differentiable at all load states. Of course, sharp, non-smooth completion functions based on  $\max(x, 0)$  (instead of the softplus function in (18)) and the unit step function (instead of the transition  $\chi_s$  in (18)) could be considered as well, but differentiability would then be lost. This would be a clear disadvantage from the perspective of multiscale simulations and needed tangent operators. In Fig. 15, the completion functions are visualized for the parameters  $\mu_n = 3.2332$  nm,  $\mu_s = 26.0303$  nm,  $d_n = d_s = 0.2$  nm<sup>-2</sup>,  $\gamma = 5$  nm<sup>-1</sup> and  $g_0 = 0.3$  nm. For the chosen  $\mu_n$ , in the left plot of Fig. 15 one can see that the completion function  $c_n(\underline{g})$  for  $\hat{t}_n(\underline{s})$  changes from values close to 1 smoothly to values close to 0 for  $g_n$  passing  $\mu_n$ . This property then ensure a definite asymptotic behavior of  $\hat{t}_n^c = \hat{t}_n(\underline{s})c_n(\underline{g})$  towards 0 for large normal separation  $g_n$ . The completion function  $c_s(\underline{g})$  for  $\hat{t}_s(\underline{s})$  allows for  $g_n > g_0$  and  $|g_s| < \mu_s$  a normal evaluation of the model, while for  $|g_s| > \mu_s$ , due to  $c_s(\underline{g}) \rightarrow 0$ , an asymptotic behavior of  $\hat{t}_s^c(\underline{s}) = \hat{t}_s(\underline{s})c_s(\underline{g})$  towards 0 is obtained. For  $g_n = 0 < g_0$ , the asymptotic decay is turned off for all  $g_s$  and mode II behavior trained in  $\hat{t}_s(\underline{s})$  is allowed for.

### Summary and outlook

The present study investigates the identification of a surrogate ANN model for the effective traction-separation at the GB interface, based on MD simulation data. Intensive MD simulations at the atomic scale have been performed for varying loading conditions and for various temperatures. Despite the highly fluctuating and inhomogeneous nature of the traction-separation values of the MD simulations, ANNs have been shown to be able to

extract the effective material behavior. This has been achieved without any smoothing of the MD data. The ResNet like architecture of the trained ANNs with a standard FFNN as secant has shown satisfactory quality, as well as a stable physical behavior, even for large separation values. In view of a usage of the calibrated model well outside of the training range, a physics-guided model completion has been proposed. The surrogate completion extends the surrogate to arbitrarily large separation values and offers a secure evaluation in macroscopic FE simulations. The completed surrogate is differentiable by design, such that a tangent operator is always computable for arbitrary separation and temperature.

The present work focused on proportional loading and varying temperature. For future work it is, therefore, important to further investigate the abilities of ML approaches in interface problems with more dependencies (temperature, history and more) and to explicitly address the question if the approach works on the inhomogeneous, raw/unbiased data or necessitates a pre-processing of it before training—therefore, assuming the corresponding error and falsified material response. The consideration of more dependencies in the effective material law naturally offers a more accurate prediction with better physical insights, see, e.g., [58,70]. Concerning GBs specifically, not only temperature, but also loading rates (see, e.g., [71,72]) and GB misorientation with corresponding dislocation pile ups (see, e.g. [34,73]) could be considered in advanced ML approaches. Modeling general GBs under different load/temperature and rate conditions is a high dimensional problem and well beyond the current computational power. However, employing the material symmetry together with known special GB types and mechanisms involved will reduce the number of atomistic simulations required to train the ANN. If the special cases are selected wisely, the ANN will be able to fill the gaps, as presented in the current work. Note that besides crack growth or GB sliding, other mechanisms, such as GB migration, could be activated depending on the GB type. For example, as shown in [10], a  $\Sigma 3$  twin boundary will migrate under mode II loading. Thus,  $\Sigma 3$  is one of the important boundaries to be included in future work for ANN training on full range of GBs. Hereby, based on either homogenization relations or upscaled ad-hoc models, new homogenization-theory inspired ANN approaches as the one of [50] may offer a promising start. Additionally, future developments in macroscopic FE simulations of polycrystals addressing the comparison between classical interface models and ML approaches are of high interest not only from the physical prediction perspective but also from the computationally efficiency point of view, specially for non-proportional loading.

#### Acknowledgements

Vivid discussions within the scope of Cluster of Excellence SimTech (DFG EXC310 and EXC2075) regarding machine learning and data-driven model surrogation are highly appreciated by M. Fernández and F.Fritzen. S. Rezaei and J. Rezaei Mianroodi acknowledge the support of Jianan Gu in conducting the molecular dynamics simulations.

#### Authors' contributions

MF and FF designed, implemented and trained the artificial neural networks of the present work. SR, JR M and SR initiated the methodology of the present work and designed the molecular dynamics simulations. All authors contributed to the manuscript in writing and proof-reading. All authors read and approved the final manuscript.

#### Funding

The contributions of M. Fernández and F. Fritzen are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – FR2702/6 and FR2702/8 – in the scope of the Emmy-Noether and Heisenberg funding lines. The contributions of S. Rezaei and S. Reese are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through Subproject A6 of the Transregional Collaborative Research Center SFB/TRR 87. The contribution of J. Rezaei Mianroodi is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through



Subproject M5 in the Priority Programme 1713.

#### Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Institute of Applied Mechanics, University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany, <sup>2</sup>Cyber-Physical Simulation Group, Department of Mechanical Engineering, Technische Universität Darmstadt, Dolivostraße 15, 64293 Darmstadt, Germany, <sup>3</sup>Institute of Applied Mechanics, RWTH Aachen University, Mies-van-der-Rohe-Str. 1, 52074 Aachen, Germany, <sup>4</sup>Material Mechanics, RWTH Aachen University, Schinkelstrasse 2, 52062 Aachen, Germany, <sup>5</sup>Microstructure Physics and Alloy Design, Max-Planck-Institut für Eisenforschung, Max-Planck-Straße 1, 40237 Düsseldorf, Germany.

Received: 4 November 2019 Accepted: 19 December 2019

Published online: 28 January 2020

#### References

- Sørensen BF, Gamstedt EK, Østergaard RC, Goutianos S. Micromechanical model of cross-over fibre bridging—prediction of mixed mode bridging laws. *Mech Mater.* 2008;40(4):220–34. <https://doi.org/10.1016/j.mechmat.2007.07.007>.
- Vossen BG, Schreurs PJG, van der Sluis O, Geers MGD. Multi-scale modeling of delamination through fibrillation. *J Mech Phys Solids.* 2014;66:117–32. <https://doi.org/10.1016/j.jmps.2014.01.009>.
- Benedetti I, Aliabadi MH. A three-dimensional cohesive-frictional grain-boundary micromechanical model for intergranular degradation and failure in polycrystalline materials. *Comput Methods Appl Mech Eng.* 2013;265:36–62. <https://doi.org/10.1016/j.cma.2013.05.023>.
- Leo CVD, Luk-Cyr J, Liu H, Loeffel K, Al-Athel K, Anand L. A new methodology for characterizing traction-separation relations for interfacial delamination of thermal barrier coatings. *Acta Mater.* 2014;71:306–18. <https://doi.org/10.1016/j.actamat.2014.02.034>.
- van der Sluis O, Iwamoto N, Qu J, Yang S, Yuan C, van Driel WD, Zhang GQ. Advances in delamination modeling of metal/polymer systems: atomistic aspects. In: Morris JE, editor. *Nanopackaging: nanotechnologies and electronics packaging*. Cham: Springer; 2018. p. 129–83.
- Rezaei S, Arghavani M, Wulfinghoff S, Kruppe NC, Brögelmann T, Reese S, Bobzin K. A novel approach for the prediction of deformation and fracture in hard coatings: comparison of numerical modeling and nanoindentation tests. *Mech Mater.* 2018;117:192–201. <https://doi.org/10.1016/j.mechmat.2017.11.006>.
- Javili A, Steinmann P, Mosler J. Micro-to-macro transition accounting for general imperfect interfaces. *Comput Methods Appl Mech Eng.* 2017;317:274–317. <https://doi.org/10.1016/j.cma.2016.12.025>.
- Ottosen NS, Ristinmaa M, Mosler J. Framework for non-coherent interface models at finite displacement jumps and finite strains. *J Mech Phys Solids.* 2016;90:124–41. <https://doi.org/10.1016/j.jmps.2016.02.034>.
- Rezaei S, Mianroodi J R, Khaledi K, Reese S. A nonlocal method for modeling interfaces: numerical simulation of decohesion and sliding at grain boundaries. *Comput Methods Appl Mech Eng.* 2019.
- Rezaei S, Jaworek D, Mianroodi JR, Wulfinghoff S, Reese S. Atomistically motivated interface model to account for coupled plasticity and damage at grain boundaries. *J Mech Phys Solids.* 2019;124:325–49. <https://doi.org/10.1016/j.jmps.2018.10.015>.
- Meyers MA, Mishra A, Benson DJ. Mechanical properties of nanocrystalline materials. *Progr Mater Sci.* 2006;51(4):427–556. <https://doi.org/10.1016/j.pmatsci.2005.08.003>.
- Wei Y, Su C, Anand L. A computational study of the mechanical behavior of nanocrystalline fcc metals. *Acta Mater.* 2006;54(12):3177–90. <https://doi.org/10.1016/j.actamat.2006.03.007>.
- Möller JJ, Bitzek E, Janisch R, ul Hassan H, Hartmaier A. Fracture ab initio: a force-based scaling law for atomistically informed continuum models. *J Mater Res.* 2018;33:3750–61. <https://doi.org/10.1557/jmr.2018.384>.
- Ma A, Roters F, Raabe D. On the consideration of interactions between dislocations and grain boundaries in crystal plasticity finite element modeling—theory, experiments, and simulations. *Acta Mater.* 2006;54(8):2181–94. <https://doi.org/10.1016/j.actamat.2006.01.004>.
- van Beers PRM, McShane GJ, Kouznetsova VG, Geers MGD. Grain boundary interface mechanics in strain gradient crystal plasticity. *J Mech Phys Solids.* 2013;61(12):2659–79. <https://doi.org/10.1016/j.jmps.2013.08.011>.
- Xu T, Stewart R, Fan J, Zeng X, Yao A. Bridging crack propagation at the atomistic and mesoscopic scale for bcc-fe with hybrid multiscale methods. *Eng Fract Mech.* 2016;155:166–82. <https://doi.org/10.1016/j.engfracmech.2015.12.015>.
- Warner DH, Sansoz F, Molinari JF. Atomistic based continuum investigation of plastic deformation in nanocrystalline copper. *Int J Plast.* 2006;22(4):754–74. <https://doi.org/10.1016/j.jplas.2005.04.014>.
- Elzas A, Thijssen B. Cohesive laws describing the interface behaviour of iron/precipitate interfaces under mixed loading conditions. *Mech Mater.* 2019;129:265–78. <https://doi.org/10.1016/j.mechmat.2018.10.010>.
- Farkas D, Petegem SV, Derlet PM, Swygenhoven HV. Dislocation activity and nano-void formation near crack tips in nanocrystalline Ni. *Acta Mater.* 2005;53(11):3115–23. <https://doi.org/10.1016/j.actamat.2005.02.012>.
- Qiu R-Z, Li C-C, Fang T-H. Mechanical properties and crack growth behavior of polycrystalline copper using molecular dynamics simulation. *Phys Script.* 2017;92(8):085702. <https://doi.org/10.1088/1402-4896/aa7c2c>.
- Molinari J-F, Aghababaei R, Brink T, Frérot L, Milanese E. Adhesive wear mechanisms uncovered by atomistic simulations. *Friction.* 2018;6:245–59. <https://doi.org/10.1007/s40544-018-0234-6>.

22. Van Swygenhoven H, Derlet PM. Grain-boundary sliding in nanocrystalline fcc metals. *Phys Rev B*. 2001;64:224105. <https://doi.org/10.1103/PhysRevB.64.224105>.
23. Schiøtz J, Jacobsen KW. A maximum in the strength of nanocrystalline copper. *Science*. 2003;301(5638):1357–9. <https://doi.org/10.1126/science.1086636>.
24. Bitzek E, Kermode JR, Gumbsch P. Atomistic aspects of fracture. *Int J Fract*. 2015;191(1):13–30. <https://doi.org/10.1007/s10704-015-9988-2>.
25. Beyerlein IJ, Xu S, Llorca J, El-Awady JA, Mianroodi JR, Svendsen B. Alloy design for mechanical properties: conquering the length scales. *MRS Bull*. 2019;44(04):257–65. <https://doi.org/10.1557/mrs.2019.67>.
26. Mianroodi JR, Svendsen B. Atomistically determined phase-field modeling of dislocation dissociation, stacking fault formation, dislocation slip, and reactions in fcc systems. *J Mech Phys Solids*. 2015;77:109–22. <https://doi.org/10.1016/j.jmps.2015.01.007>.
27. Mianroodi JR, Shanthraj P, Kontis P, Gault B, Svendsen B, Raabe D. Atomistic phase field chemomechanical modeling of solute segregation and dislocation-precipitate interaction in Ni–Al–Co. *Acta Mater*. 2019;175(2018):1–30. <https://doi.org/10.1016/j.actamat.2019.06.008>.
28. Giesa T, Pugno NM, Wong JY, Kaplan DL, Buehler MJ. What's inside the box?—length-scales that govern fracture processes of polymer fibers. *Adv Mater*. 2014;26(3):412–7. <https://doi.org/10.1002/adma.201303323>.
29. Andric P, Curtin WA. New theory for mode I crack-tip dislocation emission. *J Mech Phys Solids*. 2017;106:315–37. <https://doi.org/10.1016/j.jmps.2017.06.006>.
30. Gur S, Sadat MR, Frantzikonis GN, Bringuier S, Zhang L, Muralidharan K. The effect of grain-size on fracture of polycrystalline silicon carbide: a multiscale analysis using a molecular dynamics-peridynamics framework. *Comput Mater Sci*. 2019;159:341–8. <https://doi.org/10.1016/j.commatsci.2018.12.038>.
31. Spearot DE, Jacob KI, McDowell DL. Non-local separation constitutive laws for interfaces and their relation to nanoscale simulations. *Mech Mater*. 2004;36(9):825–47. <https://doi.org/10.1016/j.mechmat.2003.08.002>.
32. Zhou XW, Zimmerman JA, Reedy ED, Moody NR. Molecular dynamics simulation based cohesive surface representation of mixed mode fracture. *Mech Mater*. 2008;40(10):832–45. <https://doi.org/10.1016/j.mechmat.2008.05.001>.
33. Paliwal B, Cherkaoui M. An improved atomistic simulation based mixed-mode cohesive zone law considering non-planar crack growth. *Int J Solids Struct*. 2013;50(20):3346–60. <https://doi.org/10.1016/j.ijsolstr.2013.06.002>.
34. Elzas A, Thijssse B. Dislocation impacts on iron/precipitate interfaces under shear loading. *Modell Simul Mater Sci Eng*. 2016;24(8):085006. <https://doi.org/10.1088/0965-0393/24/8/085006>.
35. Yamakov V, Saether E, Phillips DR, Glaessgen EH. Molecular-dynamics simulation-based cohesive zone representation of intergranular fracture processes in aluminum. *J Mech Phys Solids*. 2006;54(9):1899–928. <https://doi.org/10.1016/j.jmps.2006.03.004>.
36. Yamakov V, Saether E, Glaessgen EH. Multiscale modeling of intergranular fracture in aluminum: constitutive relation for interface debonding. *J Mater Sci*. 2008;43(23):7488–94. <https://doi.org/10.1007/s10853-008-2823-7>.
37. Fu XQ, Liang LH, Wei YG. Modeling of atomistic scale shear failure of Ag/MgO interface with misfit dislocation network. *Comput Mater Sci*. 2019;170:109151. <https://doi.org/10.1016/j.commatsci.2019.109151>.
38. Mudunuru MK, Panda N, Karra S, Srinivasan G, Chau VT, Rougier E, Hunter A, Viswanathan HS. Surrogate models for estimating failure in brittle and quasi-brittle materials. *Appl Sci*. 2019;9:2706. <https://doi.org/10.3390/app9132706>.
39. Versino D, Tonda A, Bronkhorst CA. Data driven modeling of plastic deformation. *Comput Methods Appl Mech Eng*. 2017;318:981–1004. <https://doi.org/10.1016/j.cma.2017.02.016>.
40. Oishi A, Yagawa G. Computational mechanics enhanced by deep learning. *Comput Methods Appl Mech Eng*. 2017;327:327–51. <https://doi.org/10.1016/j.cma.2017.08.040>.
41. Capuano G, Rimoli JJ. Smart finite elements: a novel machine learning application. *Comput Methods Appl Mech Eng*. 2019;345:363–81. <https://doi.org/10.1016/j.cma.2018.10.046>.
42. Bock FE, Aydin RC, Cyron CJ, Huber N, Kalidindi SR, Klusemann B. A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Front Mater*. 2019;6:110. <https://doi.org/10.3389/fmats.2019.00110>.
43. Salehi H, Burgueño R. Emerging artificial intelligence methods in structural engineering. *Eng Struct*. 2018;171:170–89. <https://doi.org/10.1016/j.engstruct.2018.05.084>.
44. Kirchdoerfer T, Ortiz M. Data-driven computational mechanics. *Comput Methods Appl Mech Eng*. 2016;304:81–101. <https://doi.org/10.1016/j.cma.2016.02.001>.
45. Kirchdoerfer T, Ortiz M. Data driven computing with noisy material data sets. *Comput Methods Appl Mech Eng*. 2017;326:622–41. <https://doi.org/10.1016/j.cma.2017.07.039>.
46. Nguyen LTK, Keip M-A. A data-driven approach to nonlinear elasticity. *Comput Struct*. 2018;194:97–115. <https://doi.org/10.1016/j.compstruc.2017.07.031>.
47. Eggersmann R, Kirchdoerfer T, Reese S, Stainier L, Ortiz M. Model-free data-driven inelasticity. *Comput Methods Appl Mech Eng*. 2019;350:81–99. <https://doi.org/10.1016/j.cma.2019.02.016>.
48. Haj-Ali R, Pecknold DA, Ghaboussi J, Voyiadjis GZ. Simulated micromechanical models using artificial neural networks. *J Eng Mech*. 2001;127(7):730–8. [https://doi.org/10.1061/\(ASCE\)0733-9399\(2001\)127:7\(730\)](https://doi.org/10.1061/(ASCE)0733-9399(2001)127:7(730)).
49. Reimann D, Nidadavolu K, ul Hassan H, Vajragupta N, Glasmachers T, Junker P, Hartmaier A. Modeling macroscopic material behavior with machine learning algorithms trained by micromechanical simulations. *Front Mater*. 2019;. <https://doi.org/10.3389/fmats.2019.00181>.
50. Liu Z, Wu CT, Koishi M. A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Comput Methods Appl Mech Eng*. 2019;345:1138–68. <https://doi.org/10.1016/j.cma.2018.09.020>.
51. Unger JF, Könke C. Neural networks as material models within a multiscale approach. *Comput Struct*. 2009;87(19):1177–86. <https://doi.org/10.1016/j.compstruc.2008.12.003>.
52. Hashash YMA, Jung S, Ghaboussi J. Numerical implementation of a neural network based material model in finite element analysis. *Int J Numer Methods Eng*. 2004;59(7):989–1005. <https://doi.org/10.1002/nme.905>.
53. Fritzen F, Kunc O. Two-stage data-driven homogenization for nonlinear solids using a reduced order model. *Eur J Mech A Solids*. 2018;69:201–20. <https://doi.org/10.1016/j.euromechsol.2017.11.007>.

54. Kunc O, Fritzen F. Generation of energy-minimizing point sets on spheres and their application in mesh-free interpolation and differentiation. *Adv Comput Math* (accepted for publication, Sep. 8, 2019),2019;1–27.
55. Wang K, Sun W. A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning. *Comput Methods Appl Mech Eng.* 2018;334:337–80. <https://doi.org/10.1016/j.cma.2018.01.036>.
56. Lu X, Giovanis DG, Yvonnet J, Papadopoulos V, Detrez F, Bai J. A data-driven computational homogenization method based on neural networks for the nonlinear anisotropic electrical response of graphene/polymer nanocomposites. *Comput Mech.* 2018;64:307–21. <https://doi.org/10.1007/s00466-018-1643-0>.
57. Fritzen F, Fernández M, Larsson F. On-the-fly adaptivity for nonlinear twoscale simulations using artificial neural networks and reduced order modeling. *Front Mater.* 2019;6:75. <https://doi.org/10.3389/fmats.2019.00075>.
58. Wang K, Sun W. Meta-modeling game for deriving theory-consistent, microstructure-based traction-separation laws via deep reinforcement learning. *Comput Methods Appl Mech Eng.* 2019;346:216–41. <https://doi.org/10.1016/j.cma.2018.11.026>.
59. Mishin Y, Farkas D, Mehl M, Papaconstantopoulos D. Interatomic potentials for monoatomic metals from experimental data and ab initio calculations. *Phys Rev B.* 1999;59(5):3393–407. <https://doi.org/10.1103/PhysRevB.59.3393>.
60. Sheppard D, Terrell R, Henkelman G. Optimization methods for finding minimum energy paths. *J Chem Phys.* 2008;128(13):134106. <https://doi.org/10.1063/1.2841941>.
61. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys.* 2007;126(1):014101. <https://doi.org/10.1063/1.2408420>.
62. Admal NC, Tadmor EB. A unified interpretation of stress in molecular systems. *J Elast.* 2010;100(1–2):63–143. <https://doi.org/10.1007/s10659-010-9249-6>.
63. Thompson AP, Plimpton SJ, Mattson W. General formulation of pressure and stress tensor for arbitrary many-body interaction potentials under periodic boundary conditions. *J Chem Phys.* 2009;131(15):1–6. <https://doi.org/10.1063/1.3245303>.
64. Plimpton S. Fast parallel algorithms for short-range molecular. *Dynamics.* 1995;117:1–42. <https://doi.org/10.1006/jcph.1995.1039>.
65. Stukowski A. Visualization and analysis of atomistic simulation data with OVITO—the Open Visualization Tool. *Modell Simul Mater Sci Eng.* 2010;18(1):015012. <https://doi.org/10.1088/0965-0393/18/1/015012>.
66. Géron A. Neural networks and deep learning, 2019.
67. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).* 2016; p. 770–8. <https://doi.org/10.1109/CVPR.2016.90>.
68. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).* 2017; p. 2261–9. <https://doi.org/10.1109/cvpr.2017.243>.
69. Kingma D P, Ba J L. Adam: A method for stochastic optimization. In: *3rd international conference on learning representations, ICLR 2015—conference track proceedings*, p. 1–15. 2015. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
70. Jung S, Ghaboussi J. Neural network constitutive model for rate-dependent materials. *Comput Struct.* 2006;84(15):955–63. <https://doi.org/10.1016/j.compstruc.2006.02.015>.
71. Ahmed N, Hartmaier A. Mechanisms of grain boundary softening and strain-rate sensitivity in deformation of ultrafine-grained metals at high temperatures. *Acta Mater.* 2011;59(11):4323–34. <https://doi.org/10.1016/j.actamat.2011.03.056>.
72. Li X, Roth CC, Mohr D. Machine-learning based temperature- and rate-dependent plasticity model: application to analysis of fracture experiments on dp steel. *Int J Plast.* 2019;118:320–44. <https://doi.org/10.1016/j.ijplas.2019.02.012>.
73. Broedling NC, Hartmaier A, Gao H. A combined dislocation–cohesive zone model for fracture in a confined ductile layer. *Int J Fract.* 2006;140(1):169–81. <https://doi.org/10.1007/s10704-005-6025-x>.

## Publisher's Note

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

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---