



A learning-based multiscale method and its application to inelastic impact problems

Burigede Liu^a, Nikola Kovachki^a, Zongyi Li^a, Kamyar Azizzadenesheli^b,
Anima Anandkumar^a, Andrew M. Stuart^a, Kaushik Bhattacharya^{a,*}

^a Division of Engineering and Applied Science, California Institute of Technology, Pasadena CA 91125, United States of America

^b Department of Computer Science, Purdue University, West Lafayette IN 47906, United States of America

ARTICLE INFO

Dataset link: https://github.com/Burigede/Learning_based_multiscale.git

Keywords:

Multiscale modeling
Machine learning
Crystal plasticity

ABSTRACT

The macroscopic properties of materials that we observe and exploit in engineering application result from complex interactions between physics at multiple length and time scales: electronic, atomistic, defects, domains etc. Multiscale modeling seeks to understand these interactions by exploiting the inherent hierarchy where the behavior at a coarser scale regulates and averages the behavior at a finer scale. This requires the repeated solution of computationally expensive finer-scale models, and often *a priori* knowledge of those aspects of the finer-scale behavior that affect the coarser scale (order parameters, state variables, descriptors, etc.). We address this challenge in a two-scale setting where we learn the fine-scale behavior from off-line calculations and then use the learnt behavior directly in coarse scale calculations. The approach builds on the recent success of deep neural networks by combining their approximation power in high dimensions with ideas from model reduction. It results in a neural network approximation that has high fidelity, is computationally inexpensive, is independent of the need for *a priori* knowledge, and can be used directly in the coarse scale calculations. We demonstrate the approach on problems involving the impact of magnesium, a promising light-weight structural and protective material.

1. Introduction

The macroscopic behavior of materials is the end result of mechanisms operating over a wide range of length and time scales, where mechanisms at the larger scales both filter (average) and modulate (set the boundary condition) those at the lower scales (Phillips, 2001). The development and optimization of new material/structural systems therefore require an understanding of the various mechanisms and their interactions across the scales. While each mechanism has been studied by developing models at an individual scale: density functional theory at the electronic scale (Giustino, 2014), molecular dynamics at the atomistic scale (Finnis, 2010), defect models at the nanoscale (Bulatov and Cai, 2013), crystal plasticity at the sub-grain scale (Asaro, 1983), empirical inelastic theories at the engineering scale (Gurtin et al., 2013) etc., recent work has focused on multiscale modeling that seeks to understand the behavior across multiple scales (Fish, 2009; de Borst and Ramm, 2011). The entire range of material behavior is first divided into a hierarchy of scales (Van Der Giessen et al., 2020), the relevant mechanisms at each scale are identified and analyzed using theories/tools based on an individual scale, and the hierarchy is stitched together by passing information between scales. While the mathematical theory of homogenization (Bensoussan et al., 2011; Pavliotis and Stuart, 2008) provides a concrete

* Corresponding author.

E-mail address: bhatta@caltech.edu (K. Bhattacharya).

<https://doi.org/10.1016/j.jmps.2021.104668>

Received 11 June 2021; Received in revised form 30 September 2021; Accepted 6 October 2021

Available online 22 October 2021

0022-5096/© 2021 Elsevier Ltd. All rights reserved.

basis in specialized situations, the underlying conceptual framework has been adopted broadly. Importantly, multiscale modeling has explained experimental observations where empirical models have failed (e.g. strength of solids in extreme conditions Barton et al., 2011).

One widely used approach is the sequential multiscale or parameter passing method which extends the empirical approach by evaluating parameters in the coarse model using information from lower scale models (E, 2011). Examples include training atomistic models from first principles (Cheng et al., 2019), inferring defect kinetics from atomistic simulations (Fu et al., 2005), and fitting macroscopic plasticity models from crystal plasticity calculations (Balasubramanian and Anand, 2002). While the coarse model can be derived in some situations (e.g. linear elasticity Bensoussan et al., 2011), it has to be postulated *a priori* in most situations. Another approach, with greater fidelity, is the concurrent multiscale method that evaluates mechanisms operating at different scales in parallel so that the small and large scale models are computed concurrently. Examples include the Car-Parrinello molecular dynamics (Car and Parrinello, 1985), the quasicontinuum method (Tadmor et al., 1996), and the FE² approach (Feyel and Chaboche, 2000). However, concurrently evaluating the mechanisms across scales is expensive and can exceed the present computational power for analyzing practical engineering problems. Further, it is often necessary to postulate *a priori* the descriptors (state or internal variables, order parameters) by which the coarser and finer scale models communicate. The existence and identification of such descriptors are far from clear, especially in time-dependent phenomena (Bhattacharya, 1999).

In short, the practical implementation of the multiscale modeling of materials suffers from two challenges. The first is that one often needs *a priori* empirical knowledge about the interaction between models at various scales. The second, especially in the concurrent multiscale approach, is the need to repeatedly solve the expensive finer scale model only to use a very small portion of the information. This naturally raises the question: *how can the data generated by repeatedly solving the finer scale model be utilized to create a computationally efficient surrogate of its solution operator, that can directly be used at the coarser scale with no further modeling?*

Machine-learning and especially deep neural networks have been extremely successful in image recognition (LeCun et al., 1995; He et al., 2016) and natural language processing tasks (Goldberg, 2017; Collobert and Weston, 2008). There is also a growing literature on the use of machine-learning methods in multiscale modeling of materials. It has been used in developing atomistic models from quantum mechanical calculations significantly expanding the scope of parameter passing approaches (Marchand et al., 2020; Cole et al., 2020; Wen and Tadmor, 2019). Within continuum models, machine learning has been used to identify a reduced basis (using K-means clustering for example) which is then used to efficiently solve partial differential equations including unit cell problems (Liu et al., 2016; E and Yu, 2018; Raissi et al., 2019). Machine-learning methods have also been used to learn the homogenized behavior of unit cell problems in linear elasticity, and the homogenized properties have been used in larger scale problems (Liu et al., 2019; Liu and Wu, 2019; Saha et al., 2021; Bessa et al., 2017). Similarly, unit cell properties have been learned from atomistic (Xiao et al., 2020) and quantum mechanical calculations (Teh et al., 2021) and used in continuum calculations (Xiao et al., 2020). Turning now to time-dependent phenomena, Mozaffar et al. (2019) have used a recurrent neural network to learn the homogenized stress-strain relation of a representative volume element governed by standard isotropic plasticity for moderate strains (up to 8%).

In this work, we develop a framework to answer the question raised above in problems involving the impact of a polycrystalline inelastic solid in a two-scale setting. We introduce the two-scale problem, and describe our overall approach in Section 2. A critical challenge is that material models are described as partial differential equations that map inputs from one function space (e.g. average strain history) to outputs on another function space (correspondingly, the resulting stress history). While typical approaches use a finite-dimensional subspace obtained by discretization to solve these problems, it is desirable for the learnt map to be independent of the particular discretization or resolution.

We apply the framework to crystal plasticity adapted to include twinning motivated by the light-weight structural material magnesium. Magnesium is the lightest of all structural materials, and has amongst the highest strength-to-weight ratio. Therefore, it is of interest in bio-medical, automotive, and protective applications (Joost and Krajewski, 2017; Kulekci, 2008; Chen et al., 2016). It is hexagonal-closed-packed with a variety of deformation modes including soft basal slip, a relative soft tensile twin, hard pyramidal and prismatic slip systems. The diversity and complexity of the deformation modes make the overall response of polycrystalline medium quite complex and of considerable interest to multiscale modeling (e.g., Chang and Kochmann, 2015; Ravaji and Joshi, 2021; Yaghoobi et al., 2021).

We show the ability of our machine-learned approximation to learn the overall behavior of a representative volume over a wide range of trajectories leading up to strains as large as 50%. We also show how the machine-learned approximation can learn physics like material symmetry and causality from the underlying data. We use the machine-learned approximation to study macroscopic impact problems in Section 4. We show that it is possible to solve macroscopic problems with the fidelity of concurrent multiscale modeling and beyond (because we do not need *a priori* identification of state variables) at a computational cost that is a only few times greater than that of solving the problem with an empirical model and orders of magnitude smaller than solving the problem with concurrent multiscale modeling. We conclude with a brief discussion in Section 5.

2. Broad overview of our approach

Consider a heterogeneous body occupying the region $\Omega \subset \mathbb{R}^d, d = 2, 3$ in the reference configuration. We are interested in situations where the ratio ε of the scale of the heterogeneity to that of the body is small. Let $u : \Omega \rightarrow \mathbb{R}^d$ denote the deformation and $F = \nabla u$ the deformation gradient. The state of the body is described by a set of internal variables $\xi : \Omega \rightarrow \mathbb{R}^m$ and the deformation gradient. The constitutive relation is described by the (Piola–Kirchhoff) stress function $S^e : \mathbb{R}^{d \times d} \times \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}^{d \times d}$ and

Strain + gradient + internal variables \rightarrow stress

a kinetic relation $K^\varepsilon : \mathbb{R}^{d \times d} \times \mathbb{R}^m \times \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}^m$ that describes the evolution of the internal variables. Let $\rho^\varepsilon : \Omega \rightarrow \mathbb{R}$ denote the (referential) density.

Given u_0, v_0, u^*, s^* , the displacement u^ε and internal variables ξ^ε are given by the solution of the system

$$\nabla \cdot S^\varepsilon = \rho^\varepsilon u_{tt}^\varepsilon \quad \text{on } \Omega \quad (1)$$

$$K^\varepsilon = 0 \quad \text{on } \Omega \quad (2)$$

$$u^\varepsilon(x, 0) = u_0(x), \quad u_t^\varepsilon(x, 0) = v_0(x), \quad \xi^\varepsilon(x, 0) = \xi_0 \quad \text{on } \Omega \quad (3)$$

$$u^\varepsilon(x, t) = u^*(x, t) \quad \text{on } \partial_1 \Omega \quad (4)$$

$$S^\varepsilon(\nabla u^\varepsilon, \xi^\varepsilon, x)n(x) = s^*(x, t) \quad \text{on } \partial_2 \Omega \quad (5)$$

where $\partial_1 \Omega \cup \partial_2 \Omega = \partial \Omega$. (1) is the equation of motion, (2) the kinetic relation that describes the evolution of the internal variables, (3) the initial condition and (4), (5) the boundary conditions. Note that the displacement and internal variables oscillate on a scale smaller than ε and we emphasize this with the superscript. In this work, we consider an almost periodic medium where $S^\varepsilon(F, \xi, x) = S(F, \xi, x, x/\varepsilon)$, $K^\varepsilon(F, \xi, \xi_t, x) = K(F, \xi, \xi_t, x, x/\varepsilon)$, $\rho^\varepsilon(x) = \rho(x, x/\varepsilon)$ where $S : \mathbb{R}^{d \times d} \times \mathbb{R}^m \times \mathbb{R}^m \times Y \rightarrow \mathbb{R}^{d \times d}$, $K : \mathbb{R}^{d \times d} \times \mathbb{R}^m \times \mathbb{R}^m \times \Omega \times Y \rightarrow \mathbb{R}^m$, $\rho : \Omega \times Y \rightarrow \mathbb{R}$ are periodic with period Y ($|Y| = 1$) in their last variable.

We now show that the solution to this problem which has to be resolved on the fine scale ε may be approximated with that of the macroscopic problem

$$\nabla \cdot \bar{S} = \bar{\rho} u_{tt} \quad \text{on } \Omega \quad (6)$$

$$u(x, 0) = u_0(x), \quad u_t(x, 0) = v_0(x) \quad \text{on } \Omega \quad (7)$$

$$u(x, t) = u^*(x, t) \quad \text{on } \Gamma_1 \quad (8)$$

$$\bar{S}n(x) = s^*(x, t) \quad \text{on } \Gamma_2 \quad (9)$$

where the stress \bar{S} and displacement u are smooth on the scale of ε , for an appropriate macroscopic constitutive behavior or closure relation

$$F : \{F(\tau) : \tau \in (0, t)\} \mapsto \bar{S}(t) \quad t \in (0, T) \quad (10)$$

that describes how the macroscopic stress \bar{S} depends on the history of the macroscopic deformation gradient consistent with the fine scale problem.

To do so, we rewrite (1), (4), (5) as

$$\int_{\Omega} S^\varepsilon \cdot \nabla w dx - \int_{\partial_2 \Omega} s^* \cdot w = \int_{\Omega} \rho^\varepsilon u_{tt}^\varepsilon \cdot w dx \quad \forall w \quad (11)$$

and make the two-scale ansatz¹

$$u^\varepsilon(x, t) = u^0(x, t) + \varepsilon u^1(x, x/\varepsilon, t) + \varepsilon^2 u^2(x, x/\varepsilon, t) + \dots \quad (12)$$

$$\xi^\varepsilon(x, t) = \xi^0(x, x/\varepsilon, t) + \varepsilon \xi^1(x, x/\varepsilon, t) + \dots \quad (13)$$

where $u^j(x, y, t)$ is periodic in y with period Y for any $j = 1, 2, \dots$. Note that

$$S^\varepsilon = S(\nabla_x u^0 + \nabla_y u^1, \xi^0, x, y) + \dots, \quad K^\varepsilon = K(\nabla_x u^0 + \nabla_y u^1, \xi^0, \xi_t^0, x, y) + \dots \quad (14)$$

Taking a test function of the form

$$w(x, t) = w^0(x, t) + \varepsilon w^1(x, x/\varepsilon, t) + \varepsilon^2 w^2(x, x/\varepsilon, t) + \dots \quad (15)$$

where $w^j(x, y, t)$ is periodic in y with period Y for any $j = 1, 2, \dots$, we obtain

$$\int_{\Omega} S^\varepsilon \cdot (\nabla_x w^0 + \nabla_y w^1) dx - \int_{\partial_2 \Omega} s^* \cdot w^0 = \int_{\Omega} \rho^\varepsilon u_{tt}^0 \cdot w^0 dx \quad \forall w^0, w^1. \quad (16)$$

Integrating this over Y , we obtain

$$\int_{\Omega} \langle S^\varepsilon \rangle \cdot \nabla_x w^0 dx - \int_{\partial_2 \Omega} s^* \cdot w^0 = \int_{\Omega} \langle \rho^\varepsilon \rangle w_{tt}^0 \cdot w^0 dx \quad \forall w^0, \quad (17)$$

$$\int_{\Omega} \int_Y S^\varepsilon \cdot \nabla_y w^1 dy dx = 0 \quad \forall w^1 \quad (18)$$

where $\langle \cdot \rangle$ denotes the average over the unit cell Y . The first Eq. (17), initial conditions (3)_{2,3}, and boundary conditions (4), (5) define the macroscopic problem (6)–(9). We treat x , $F = \nabla_x u^0$, and t as parameters in the second Eq. (18) and obtain the unit cell problem:

$$\nabla \cdot S(F + \nabla v, \xi, x, y) = 0 \quad \text{on } Y \quad (19)$$

¹ We ignore the displacement boundary condition (4) to simplify the treatment, but can incorporate it using a boundary layer.

$$K(F + \nabla v, \xi, \xi_t, x, y) = 0 \quad \text{on } Y \quad (20)$$

$$\xi(y, 0) = \xi_0(y) \quad \text{on } Y \quad (21)$$

$$v \text{ periodic.} \quad (22)$$

Now, given ξ_0 , the solution to this unit cell problem defines the requisite macroscopic closure relation (10) with $\bar{S} = \langle S^\varepsilon \rangle$ for the macroscopic problem.

It is common to specify the stress function in terms of the Cauchy stress $\sigma = (\det F)^{-1} S F^T$ which is symmetric. Further, the underlying physical models S, K are invariant under a change of frame and so is the map (10). It follows that $F : \{R(\tau)F(\tau), \tau \in (0, t)\} \mapsto R(t)\langle S \rangle(t)$ for any time-dependent rotation $R(t)$. According to the polar decomposition theorem, the deformation gradient $F = RU$ where R is a rotation and U is positive definite-symmetric. Therefore, it suffices to specify the equivalent constitutive behavior or closure relation

$$\bar{F} : \{U(\tau) : \tau \in (0, t)\} \mapsto \langle \sigma \rangle(t) \quad t \in (0, T). \quad (23)$$

Now, the implementation of the multiscale problem above requires the calculation of the map \bar{F} , and therefore the unit cell problem at each macroscopic point x and at each instant t ; this is extremely expensive. Our idea is to learn the macroscopic constitutive behavior using model reduction and deep neural networks following the approach of Bhattacharya et al. (2020) by utilizing data generated by solutions of the unit cell problem over various strain histories obtained from an appropriate probability distribution in the space of strain histories. To do so, we observe that the unit cell problem in fact specifies the map

$$\Psi : \{U(t) : t \in (0, T)\} \mapsto \{\langle \sigma \rangle(t) : t \in (0, T)\}. \quad (24)$$

for some $T > 0$. Further, the underlying equations and therefore this map is causal, i.e., $\langle S \rangle(t)$ depends only on $\{F(\tau) : \tau \in (0, t)\}$. Finally, any map of the form (24) that is causal uniquely defines a map of the form (23). Therefore, we learn Ψ which maps one function to another.

From a function of time to a function of time

The map $\Psi : L_2((0, T); \mathbb{R}^{d(d+1)/2}) \rightarrow L_2((0, T); \mathbb{R}^{d(d+1)/2})$ is one between infinite dimensional Hilbert spaces. However, the data is discretized and standard neural networks are defined as maps between finite-dimensional spaces. Thus we seek a finite dimensional approximations of these infinite-dimensional spaces. Further, we want our approximation and our architecture to be independent of any specific discretization. To that end, we seek maps $p_i : L_2 \rightarrow \mathbb{R}^{d_i}$ and $p_o : L_2 \rightarrow \mathbb{R}^{d_o}$ that reduces (project) the input and output spaces and maps $\ell_i : \mathbb{R}^{d_i} \rightarrow L_2$ and $\ell_o : \mathbb{R}^{d_o} \rightarrow L_2$ that lift them back up such that $p_i \circ \ell_i \approx id$, $p_o \circ \ell_o \approx id$. We then find an approximate map $\psi : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d_o}$ such that

$$\Psi \approx \ell_o \circ \psi \circ p_i. \quad \text{Map in low dim space} \quad (25)$$

In this work, we use principal component analysis (PCA) (Wold et al., 1987) to specify the maps p_i, p_o and a fully-connected deep neural network with M layers to approximate ψ :

$$\psi(s) = W_M \dots \omega(W_2 \omega(W_1(s) + b_1) + b_2) \dots + b_M, s \in \mathbb{R}^{d_i}, \quad (26)$$

where W_1, \dots, W_M are the weight matrices, b_1, \dots, b_M are the bias vectors, and ω is the scaled exponential linear unit (SELU) (Klambauer et al., 2017) with scaling constants $\chi = 1.67$ and $\beta = 1.05$. We approximate the PCA specified maps by a standard SVD algorithm and the weight and bias parameters of the neural network with standard stochastic gradient based minimization techniques (Bhattacharya et al., 2020).

This learnt approximate map replaces the constitutive relation in a macroscopic integrator. The approach is summarized in Algorithm 1.

Remarks

1. The proposed approach does not require any explicit macroscopic constitutive relation. Instead, the constitutive behavior is implicitly defined by the unit cell problem (19)–(22). We learn the solution of this problem using a neural network, and this neural network acts as the constitutive relation in macroscopic problems. Similarly the proposed approach does not require any macroscopic internal variable (descriptor of the macroscopic state of the material) (Rice, 1971). These aspects are distinct from almost all other approaches, traditional and multiscale.

In the traditional approach, we do not consider the full problem but postulate the existence of an empirical macroscopic internal variable $\zeta : \Omega \rightarrow \mathbb{R}^{m'}$ and macroscopic constitutive relations $\bar{S} : \mathbb{R}^{d \times d} \times \mathbb{R}^{m'} \times \Omega \rightarrow \mathbb{R}^{d \times d}$, $\bar{K} : \mathbb{R}^{d \times d} \times \mathbb{R}^{m'} \times \mathbb{R}^{m'} \times \Omega \rightarrow \mathbb{R}^{m'}$, and to solve the macroscopic problem (6)–(9) supplemented with the macroscopic kinetic relations $\dot{\bar{K}} = 0$, $\zeta(x, 0) = 0$ on Ω . It is computationally inexpensive, but relies on limited empirical knowledge and makes no explicit use of the micro-scale physics.

Such an approach can be justified using homogenization theory in linear elasticity and with limitations (away from long wavelength instabilities) in finite elasticity, i.e. in theories where we do not have any internal variables. However, the rigorous justification remains an open problem in models with internal variables including plasticity. Indeed, the strict statement, in particular, the existence of a properly defined macroscopic internal variable, is likely false with pinning, memory effects, mixing of energetic and kinetic terms etc., but may hold in some approximate sense. Further, even if the original full problem depends on the rate of change of the internal variable, the macroscopic constitutive functions may depend on the history of the deformation.

In the parameter passing multiscale approach, we solve the unit cell problem under various conditions and use the solutions to complement empirical data in the development of \bar{S}, \bar{K} . This is an exercise in regression, and well-suited for machine learning (Liu

Algorithm 1: Overview of the learning based multiscale modeling.**Off-line (once for any material)**

Solve the unit cell problem for various trajectories $U(t), t \in (0, T)$

Use this data to train the approximation (25)

Online (for each simulation)*Initialize*

Discretize the macroscopic problem in space using finite elements.

Discretize the macroscopic problem in time and use an explicit macroscopic integrator.

while $t_n \leq T$ **do**

At each quadrature point:

The integrator provides the deformation history $\{F_m\}_{m \leq n}$.

Use the polar decomposition theorem $F_m = Q_m U_m, m = 1, \dots, n$.

Generate a trajectory $U(t)$ consistent with the history $\{u_m\}_{m \leq n}$.

Use the learnt approximation (25) to obtain the Cauchy stress history $\langle \sigma \rangle(t)$.

Return the appropriately oriented stress $Q_n \langle \sigma \rangle(t_n) Q_n^T$ to the integrator.

Update the deformation using the macroscopic integrator.

end

et al., 2019). This approach can be relatively computationally inexpensive and it incorporates some microscale physics. However, it still relies on limited empirical knowledge and postulates the existence and identification of the macroscopic internal variable (descriptors) ζ and constitutive functions \tilde{S}, \tilde{K} .

In the *concurrent multiscale approach*, we integrate the macroscopic problem as in our proposed method, but solve the unit problem at each quadrature point for every point in the time interval of interest with initial data ξ_0 based on the macroscopic internal variable ζ . Thus, it does not postulate the existence of \tilde{S}, \tilde{K} , but requires the *a priori* identification of a macroscopic internal variable ζ and the relation between ξ and ζ . This can be challenging because the approach entails the ‘inverse homogenization’ problem of generating a representative state of the microscopic internal variables given a macroscopic internal variable. Although the method makes extensive use of the microscopic physics with little need for empirical input, it is extremely computationally expensive.

2. We can, in principle, learn Ψ directly from experimental data (e.g. Jordan et al., 2020 for the viscoelastic response of polypropylene) rather than the solution of a lower scale model on a unit cell if sufficient experimental data were available. Similarly, we do not need to solve the unit cell problem, but use an approximate method at that scale. We demonstrate this later by replacing the unit cell problem with a Taylor model i.e. assuming that $v = 0$ in the unit cell problem.

3. We use PCA for our model reduction p_i, p_o and a deep neural network for the approximate map Ψ . We can replace these with other model reduction (e.g. auto-encoders, k-means clustering) and machine learning architectures (e.g. convolutional networks, random features).

4. The model-reduction step, implemented in this work with PCA, is an essential component of the proposed framework for a number of reasons. The first is conceptual. We seek to approximate Ψ in (24) that is a mapping from one infinite dimensional function space to another. Since neural nets is a map from one finite dimensional space to another, there is an essential need for a model reduction step. Indeed, any discretization would be a model reduction. The question then is to find a good finite dimensional space (discretization) for given data, and PCA is a natural choice.

The second reason is that we want the approximation to be independent of the discretization that is used to generate the data to train the approximation. We may want to accumulate training data from multiple sources and they may not all be generated using the same discretization. More importantly in the current work, we want to use the approximation in macroscale simulations, and we want to choose the discretization (time-step) of the macroscale simulation to be relevant to that calculation and not be limited to the discretization (time-step) used to generate the data. It is shown in Bhattacharya et al. (2020) that when a neural network is trained directly from the discretized data without using the PCA, the quality of the approximation is discretization dependent and there can be large errors when a neural network trained at one discretization is used at another (Figures 4.3(a) and 4.6(a) in Bhattacharya et al., 2020). In contrast, the approach presented above with the PCA leads to a discretization-independent approximation as we show in Fig. 1(d).

Third, the size of training data required for accurately training a neural network depends on the input and output dimensions. Conversely, for a given size of the data set, the error of the neural network approximation increases with the dimensions. So model reduction through PCA actually helps control the overall error of the approximation when the size of the training data is fixed. It is shown in Figures 4.3(a) and 4.6(a) in Bhattacharya et al. (2020) that while the error initially decreases with increasing reduced dimension, it eventually increases beyond a certain size when the size of the training data is held fixed.

The final reason is practical. We are interested in highly nonlinear phenomena like crystal plasticity where one needs fine time-steps for accuracy. This means that the data have very high dimensions. In the example we study subsequently, we use a temporal resolution of 10^5 to generate the data, and it is simply not practically feasible to obtain the data necessary to train a neural network between \mathbb{R}^{10^5} and \mathbb{R}^{10^5} . More importantly, it is not necessary, as we shall show in our example, since the PCA error is small compared to the overall error.

We refer the reader to Theorems 3.1, 3.4 and 3.5 of [Bhattacharya et al. \(2020\)](#) for a characterization of the relative errors.

5. The proposed approach requires that we train a neural network over the entire duration T of any simulation, and this may be unknown *a priori*. There are two reasons for this requirement. First, we do not carry any information about the state of the material, but only the average deformation gradient versus the average stress. Second, we do not make any assumption on fading memory. Learning these aspects from the data gathered by the simulations of the unit cell is interesting, but beyond the scope of the current work.

6. The cost of the approach has two components. The first is the one-time off-line cost of generating the data and training the neural network. This typically scales linearly with the period over which the neural network needs to be trained, the number of simulations in the training data set, and the number of epochs of training required. The latter two may also depend on the first, and this remains to be understood. In the examples we study, the number of simulations required to train a neural network is similar to the number of quadrature points in a typical sample. For this reason, our off-line cost is similar to that of a single simulation of the concurrent multiscale approach. The second is the online cost of evaluating the neural network during the simulation. This is typically small compared to the cost of the macroscopic time integration, but can scale quadratically with the period of the simulation since the evaluation of the stress at any instant requires the evaluation of the entire trajectory. Our examples show that the cost of the evaluation is a few times the cost of evaluating an empirical constitutive relation; however, we are able to take large time steps with the macroscopic integrator since we know the entire trajectory in our approach.

7. Our approach requires us to store the history of deformation. However, since we use model reduction, we do not need to store the trajectory for every time step but only require sampling it sufficiently to reconstruct the history. This enables us to manage memory.

3 Learning crystal plasticity

We demonstrate our approach by studying the inelastic deformation of polycrystalline solids.

3.1 The unit cell problem: crystal plasticity with twinning

A polycrystal is a medium made of a collection of disjoint subdomains or grains. Each grain is made of the same material, but the orientation of the grain may differ with respect to a reference frame. We specify the texture (number, size and orientations) of the grains in the unit cell by a (piece-wise) rotation-valued orientation function $Q : Y \rightarrow SO(d)$. The behavior of the material at a point y is given by that of a reference material rotated by $Q(y)$: $S(F, \xi, x, y) = S_r(FQ(y), \xi_{Q(y)})$, $K(F, \xi, \dot{\xi}, x, y) = K_r(FQ(y), \xi_{Q(y)}, \dot{\xi}_{Q(y)})$ where ξ_Q is the action of the rotation Q on the internal variable ξ , and S_r, K_r describe the behavior in the reference material.

Crystalline solids can undergo plastic or inelastic deformation governed by slip on one of n_s slip systems and twinning on one of n_t twin systems. The kinetic relation K_r describes how the internal variables — the crystallographic orientation Q , the total inelastic deformation gradient $F_{in} \in \mathbb{R}^{d \times d}$, the slip activity $\gamma = \{\gamma_\alpha\}_{\alpha=1}^{n_s}$ in the n_s slip systems, and the twin volume fractions $\lambda = \{\lambda_\beta\}_{\beta=1}^{n_t}$ in the n_t twin systems that satisfy $\lambda_\beta \in [0, 1]$, $\sum_\beta \lambda_\beta = 1$ — evolve. The details of the model following [Chang and Kochmann \(2015\)](#) are provided in [Appendix](#). We study two versions of this model.

2DFFT The first is in two dimensions, has two slip systems and no twinning i.e., $n_s = 2, n_t = 0$ (details in [Appendix](#)). The initial texture consists of 16 grains generated using periodic Voronoi tessellation ([Fritzen et al., 2009](#)). The corresponding full-field unit cell problem is solved using a fast Fourier transformation scheme following [Vidyasagar et al. \(2017\)](#).

3DTaylor The second is in three dimensions and motivated by magnesium which is of current interest as a lightweight structural material. The detailed slip and twin systems, and the associated parameters are given in [Appendix](#). We do not solve the unit cell problem, but use the Taylor averaging assumption that the deformation gradient is uniform in the unit cell ($v = 0$) ([Kocks et al., 2000](#)). We use an initial texture of 128 randomly oriented grains.

3.2 Learning crystal plasticity

The first task is to generate the data, and this requires sampling $L_2((0, T); \mathbb{R}^{d(d+1)/2})$. We seek a strain path that is smooth but changes direction arbitrarily. To that end, we divide $(0, T)$ into N intervals $\Delta t_n = t_n - t_{n-1}, n = 1, \dots, N$ where $0 = t_0 < t_1 < \dots < t_N = T$ and set $(U_{ij})_n = (U_{ij})_{n-1} + v_n U_{max} \sqrt{\Delta t_n}$, $i, j = 1, \dots, d, i \leq j$ where $v_n \in \{-1, 1\}$ follow a Rademacher distribution. We take $U_{ij}(t)$ to be the cubic Hermite interpolation of $\{(t_n, (U_{ij})_n)\}$. We take $N = 10$ and study both fixed and random time intervals.

We begin with 2DFFT. We generate 6000 strain paths $U(t)$ using random time steps and solve the unit cell problem for the average stress $\langle \sigma \rangle(t)$ for each of these paths using a spatial resolution of 64×64 and 1000 time steps. We then down sample the data to 200 time steps. A single sample of data consists of the pair $\{U(t), \langle \sigma \rangle(t)\}$. We reserve 2000 samples for testing and use various parts of the remaining 4000 for training. We use a PCA dimension of 32×3 for both the input and output spaces (here $d(d+1)/2 = 3$). We define the test error as the L_2 norm of the error in predicted stress history normalized by the L_2 norm of the stress history.

The results are shown in [Fig. 1](#). [Fig. 1\(a\)](#) shows that the test error (averaged over all test specimens) decreases with increasing amounts of training data and training epochs reaching an average of 5% for a training size of 3200 in 400 epochs. [Figs. 1\(b,c\)](#) shows the input and output (both truth and approximation) for typical test and training samples with a neural network trained over 3200

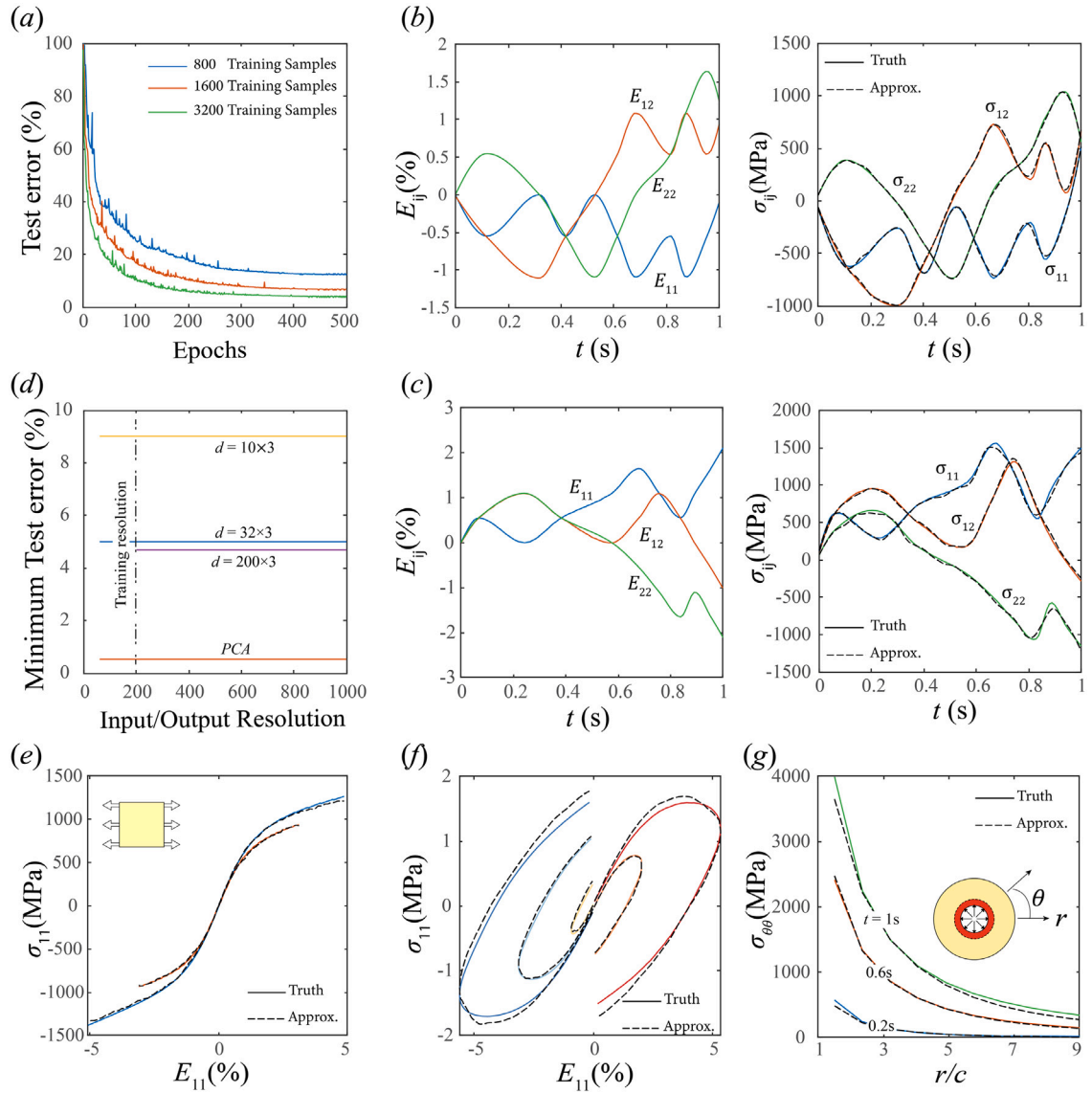


Fig. 1. Deep learning approximation of 2D crystal plasticity. (a) Average test error for various training sample sizes and training epochs. (b) Typical sample from training set. (c) Typical sample from test set, (d) Test error vs. resolution of the input and output. (e) Test for uniaxial strain loading. (f) Test for uniaxial strain load-unload. (g) Test for shear (cavity expansion).

samples and 500 epochs. We conclude that our model reduction approach is able to learn a very accurate approximation of the map Ψ .

Fig. 1(d) shows that we learn the behavior of the continuum model and is not tuned to any specific discretization used at the micro level. For a given continuous input $U(t)$, we generate additional test samples (both input and output) with various time discretization from 50 to 1000 steps, and compare the error of the approximation trained with the original training set. The figure shows that the error is independent of this resolution. The figure also shows that the error depends on the dimension of the PCA reduction, decreasing with increasing dimension till it saturates at the dimension of the training data. Finally, the figure shows the error due to PCA alone, and shows that the error of the learnt model is a few times that of the error of PCA.

Figs. 1(e-g) show that the network trained using our protocol with random time steps provides a very accurate approximation of the map Ψ in strain paths commonly encountered in practice. These include uniaxial strain $U_{11} = f(t)$, $U_{22} = 1/f(t)$ and $U_{23} = 0$ in Fig. 1(e) for loading only when $f(t) = ct$, $c \in (0, 1)$ and Fig. 1(f) for loading unloading where $f(t) = ct$, $t \in (0, 0.5)$; $f(t) = c(t - 0.5)$, $t \in (0.5, 1)$ and for Fig. 1(g) shear $U_{11} = (U_{22})^{-1} = 1/\sqrt{1 + ct^2}$ (that we encounter in cavity expansion).

We use the same sampling of strain paths to obtain data in 3D, but using Taylor averaging.

Table 1

Table of all training we have conducted.

Case number	Maximum strain	Loading protocol		Size of data set		PCA dimension	PCA error (%)		Total error (%)	
		Training	Test	Training	Test		Input	Output	Training	Test
2DFFT: Two dimensions, Full solution of unit cell problem										
1	10%	R	R	3200	1000	32x3	0.08	0.13	5	5
2	10%	R	F	3200	1000	32x3	0.08	0.13	5	5
3	10%	R	U	3200	100	32x3	0.08	0.13	5	9
4	10%	R	L-U	3200	100	32x3	0.08	0.13	5	10
5	10%	R	C	3200	100	32x3	0.08	0.13	5	8
6	10%	F	F	3200	1000	32x3	1.30E-03	0.02	5	5
7	10%	F	R	3200	1000	32x3	1.30E-03	0.02	5	33
8	10%	F	U	3200	1000	32x3	1.30E-03	0.02	5	30
9	10%	L-U	R	900	100	32x3	1.30E-03	0.07	1.2	90
10	10%	R	R	3200	1000	32x3	1.30E-03	0.02	5	5

3DTaylor: Three dimensions, Taylor approximation of unit cell problem

11	10%	R	R	3200	1000	100x6	0.18	0.46	5	5
12	10%	R	F	3200	1000	100x6	0.18	0.46	5	4
13	10%	R	U	3200	100	100x6	0.18	0.46	5	8
14	10%	F	F	3200	1000	100x6	0.011	0.05	4	4
15	10%	F	R	3200	1000	100x6	0.011	0.05	4	30
16	50%	R	R	4000	2000	100x6	1.2	1.1	3.7	17
17	50%	R	R	8000	2000	100x6	1.2	1.1	3.1	15
18	50%	R	R	16000	2000	100x6	1.2	1.1	2.8	13
19	50%	R	R	30000	6000	100x6	1.2	1.1	2.5	10
20	50%	R	R	30000	6000	60x6	1.2	1.1	2.5	11
21	50%	R	R	30000	6000	30x6	1.2	1.1	2.5	15
22	50%	R	U	30000	100	100x6	1.2	1.1	2.5	12
23	50%	R	L-U	30000	100	100x6	1.2	1.1	2.5	13
24	50%	R	C	30000	100	100x6	1.2	1.1	2.5	12
25	10%	R	R	3200	1000	100x6	0.18	0.45	4	4

F: Fixed step size, R: Random step size, U: Uniaxial, L-U: Loading-unloading, C: Cavity expansion, m = 2 except case 10 and 25 where m = 5

Table 2

Comparison of computation cost (wall-clock time in seconds)

Method	Problem	Taylor anvil (3167 elements)	Plate impact (11552 elements)
Traditional constitutive law (Johnson-Cook)		50	262
	Online Computation	435	2362
Input-output map (this work)	Off-line data (one time)	5.9×10^6	
	Off-line training (one time)	6.0×10^4 (GPU)	
Concurrent multiscale, Taylor averaging (est.)		1.2×10^8	3.9×10^8
Concurrent multiscale, full solution (est.)		3.2×10^{12}	7.6×10^{12}

We have conducted various tests of both the 2DFFT and 3DTaylor and the results are gathered in Table 1. Cases 1–5, 11–12, 22–24 show that training with random time-steps is effective even when tested against other strain paths in both 2DFFT and 3DTaylor. Fixed time step does well against data with fixed time step but poorly against other data (case 5–7, 13–14). Comparison between case 11 and 19 shows that we need more training data as the maximum strain increases. Cases 19–21 shows that increasing the PCA dimension reduces the error. Putting these together, we conclude that using random time steps is effective even when tested against other strain paths, while using fixed time steps only performs well on data generated with the same fixed time step but poorly on different data.

Finally cases 9 and 23 show that the approximation error is independent of the rate exponent.

Causality Recall that causality is essential to specify the necessary constitutive information (23) from mapping (24) that we approximate. While our model and the resulting data is causal, the architecture we use is not restricted to be causal. Fig. 2

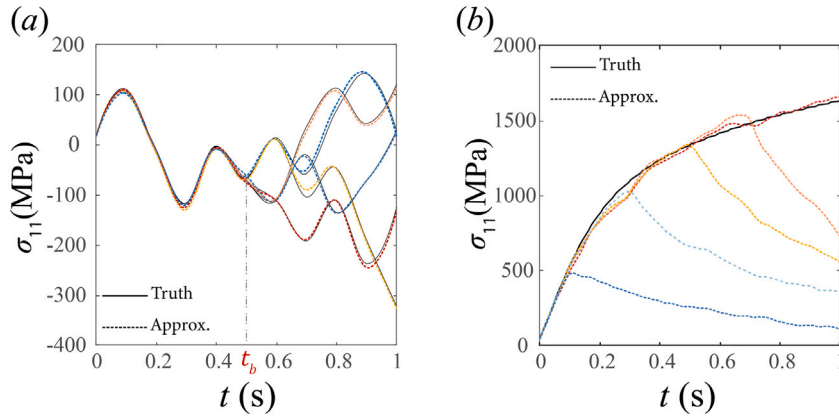


Fig. 2. Causality. (a) Test against five strain paths that diverge at $t = 0.5$. (b) Test against five strain paths that diverge at different instances of time.

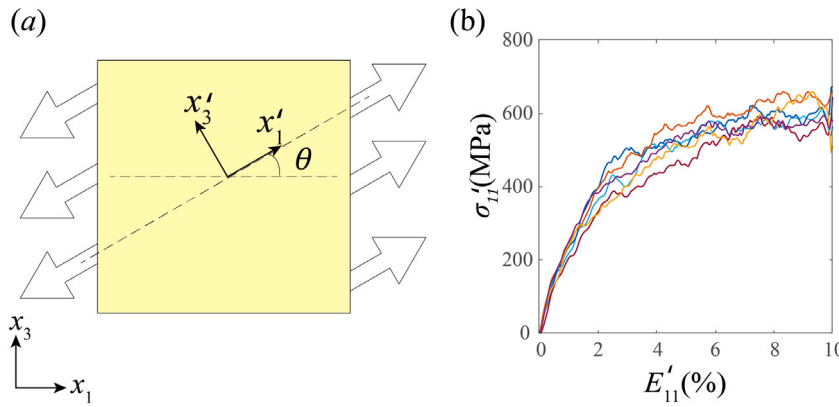


Fig. 3. Isotropy. (a) Schematic of the boundary condition. (b) Tensile stress at local coordinate with θ in the range of $[0, 30^\circ, 60^\circ, \dots, 360^\circ]$.

demonstrates that the approximation trained from data that is causal automatically learns causality in the stress–strain relationship. We consider two sets of test samples. In the first set, we consider five strain paths that are identical for $t \in (0, 0.5)$ and distinct for $t \in (0.5, 1)$. We observe in Fig. 2(a) that the approximation returns an identical stress response for $t \in (0, 0.5)$ and distinct stress responses for $t \in (0.5, 1)$. In the second set, we consider five (uniaxial) strain paths where the strain increases linearly for $t \in (0, t_i)$ for varying $t_i, i = 1, \dots, 5$ and is then held constant. Fig. 2(b) shows that in all cases, the stress in the i th path follows identical paths to those seeing the same data until time t_i , and then diverges.

Isotropy We have chosen 128 randomly oriented grains to generate our 3DTaylor data, and therefore we expect the overall behavior to be almost isotropic. However, our architecture does not impose this. We now show that the approximation automatically learns the isotropy from the data. We test the neural network trained with the 3D Taylor model (case 19 in Table 1) by applying uniaxial tensile loading in the local (material coordinate) $x'_1 - x'_3$ formulated by rotating the global coordinate system $x_1 - x_3$ with an angle θ perpendicular to the x_2 direction. Fig. 3 shows the stress–strain relation predicted by the trained approximation for various rotation angles, and these are almost identical. Thus, the trained approximation learns isotropy from the data.

4 Application to impact problems

We now use the trained neural network for macroscopic calculations of two classical impact problems, both in three dimensions. We implement the network trained using the 3DTaylor unit cell calculations (case 19 of Table 1) as a material model (“VUMAT”) in the commercial finite element package (ABAQUS, 2014). We emphasize that the neural network is only trained once for all the calculations presented below.

4.1 Taylor anvil test

A magnesium cylindrical impactor (of height $H_T = 5$ mm and diameter $D_T = 1$ mm) traveling with an initial velocity $V = 200$ m/s impacts a rigid friction-less wall at time $t = 0$ as shown schematically in Fig. 4(a). Fig. 4(c) shows the von Mises stress $\sigma_M =$

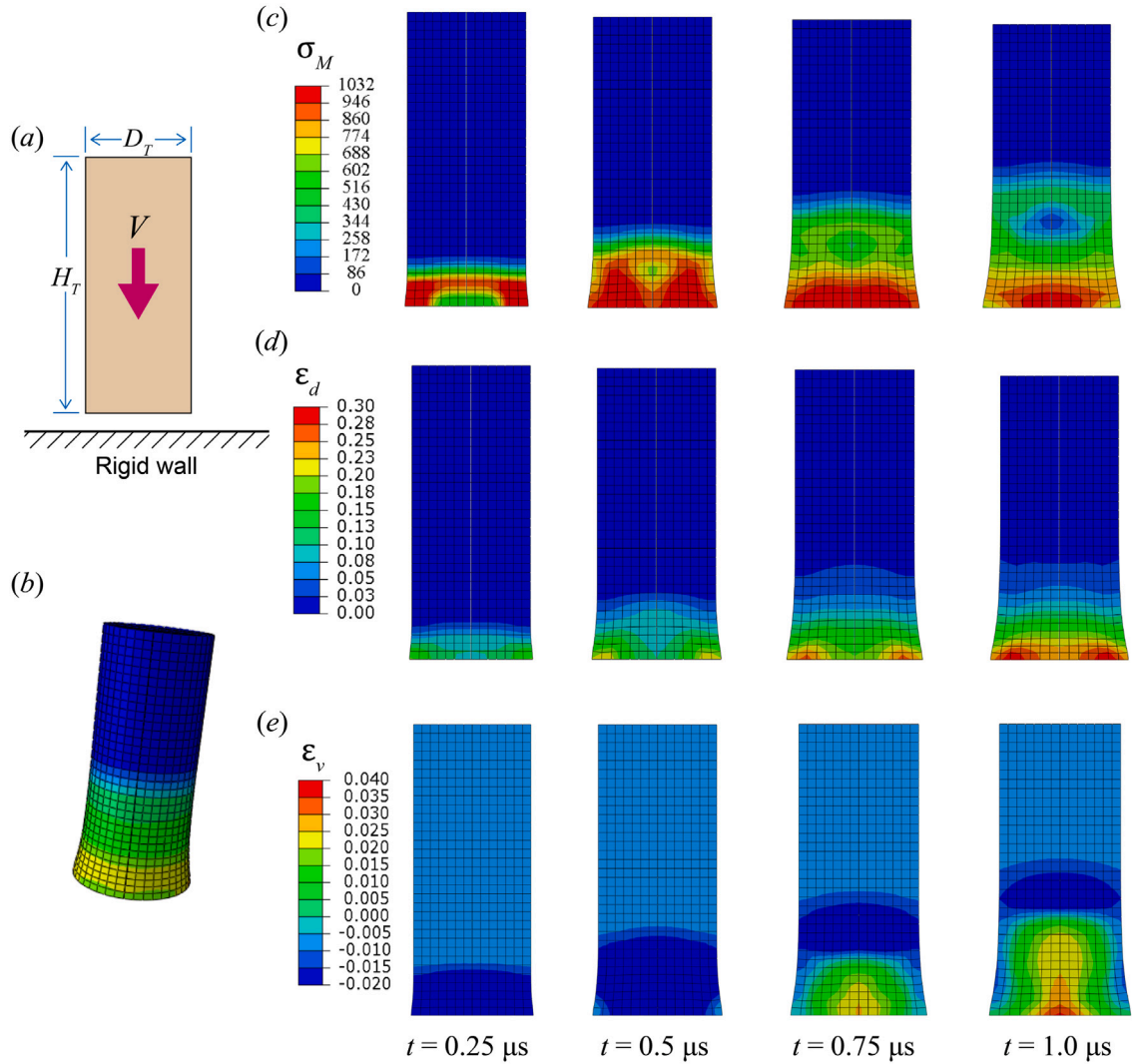


Fig. 4. Typical results of a Taylor anvil test (a) Schematic. (b) Deformation of the impactor at $t = 1 \mu s$. (c-e) Snapshots an axial cross-section with the von Mises stress measure (c), deviatoric strain measure (d) and volumetric strain measure (e).

$\sqrt{3/2}|\sigma - (\text{tr}\sigma)/3I|$ (where $|\cdot|$ denotes the Fröbenius matrix norm), Fig. 4(d) the deviatoric strain measure $e_d = |F^T F / (\det F)^2 - I|/2$ that indicates the evolution of the plastic deformation, and Fig. 4(e) the volumetric strain measure $e_v = \det(F) - 1$ that indicates the longitudinal elastic wave. Since the material is isotropic and geometry axisymmetric, the results are shown for an axial section. Upon impact, an elastic wave first propagates into the impactor followed by a region of plastic deformation. We also have a release wave moving in radially from the sides, leading to a complex radial distribution of the plastic deformation.

We repeat the calculation with various impact velocities and impactor geometries. Fig. 5(a) shows the effect of the initial velocity V on the deformation of the impactor. The elastic wave propagates with similar velocity but increasing intensity, while the plastic deformation increases with increasing velocity. The effect of changing the radius while keeping the height the same is shown in Fig. 5(b) for an impact velocity of $V = 200$ m/s. We see that the difference in the release wave from the sides changes the stress and plastic deformation distributions.

4.2 Projectile impact on a plate

A cylindrical projectile of radius $D_p = 2$ mm traveling at a velocity $V = 200$ m/s impacts a large magnesium plate of thickness $H_p = 1$ mm which is simply supported far away from the point of impact — see Fig. 6(a). The impacting cylinder is assumed to be rigid and infinitely dense compared to the plate and thus unaffected by the impact. The deviatoric strain measure e_d and the von Mises strain σ_M are shown in Fig. 6(c) and (d) respectively. An elastic wave followed by a plastic wave propagates into the

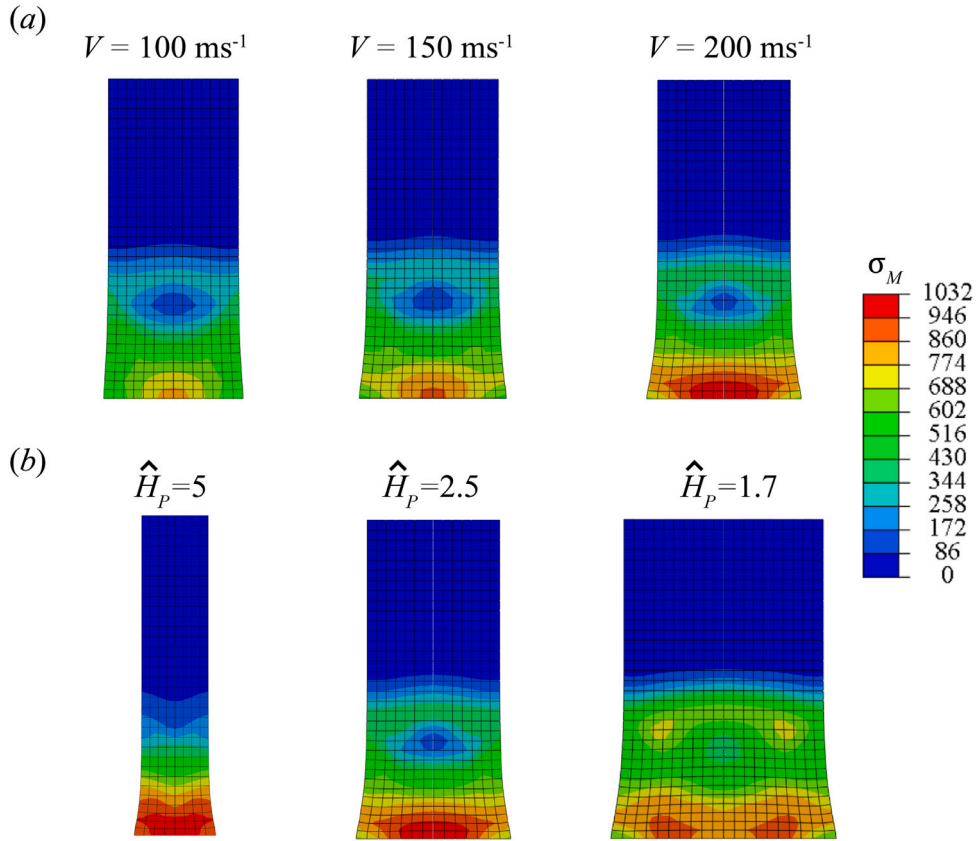


Fig. 5. Parametric study of the Taylor anvil test. Axial cross-section with the von Mises stress measure at $t_s = 1$ for (a) different impact velocities V and (b) various aspect ratios $\hat{H}_p = H_T/D_T$.

plate and is reflected from the free-face. Subsequently, the plate becomes a wave guide with a radially expanding elastic and plastic waves as the impactor penetrates into the plate.

Fig. 7 shows the results when we repeat the simulation with various impact velocities (Fig. 7(a)) and plate thicknesses (Fig. 7(b)). The radially propagating elastic wave has the same velocity but increases in intensity with increasing impact velocity and decreasing plate thickness. As the plate becomes thinner, we observe a change in the deformation mode with thin plates deforming in bending. In particular, the plate begins to separate from the impactor at the center and the amount of plastic deformation is reduced.

4.3 Computational effort

The computational cost of the proposed method is compared with other approaches for the examples studied above in Table 2. All calculations were performed on a single core of Intel Skylake CPU (2.1 GHz) except the neural network training which was done on a NVIDIA P100 GPU with 3584 CUDA cores with a 1.3 MHz clock (roughly, the Skylake CPU cost is bounded from above by 10^3 times the GPU cost). We do not conduct concurrent multiscale simulations, but we may estimate the cost as a product of the cost of a unit cell problem, the number of elements and the number of time steps of the macroscopic problem.

We find that the online cost of our method is only about ten times that of the computational cost of using an empirical constitutive law but orders of magnitude smaller than that of a concurrent multiscale method (with Taylor averaging). Further, the one-time off-line cost of generating and training our approximation is comparable to that of a single calculation using a concurrent multiscale method.

5 Discussion

We have presented a framework that enables the specification of the macroscopic constitutive behavior (or closure relation) using microscopic computations combined with machine learning. In this framework, data is generated by the solution of a fine scale model and is then used to train a surrogate which emulates the model's solution operator. The surrogate is then directly incorporated into coarse scale calculations. We have demonstrated in problems involving the impact of polycrystalline inelastic solids that this approach can solve macroscopic problems with the fidelity of concurrent multiscale modeling and beyond (because

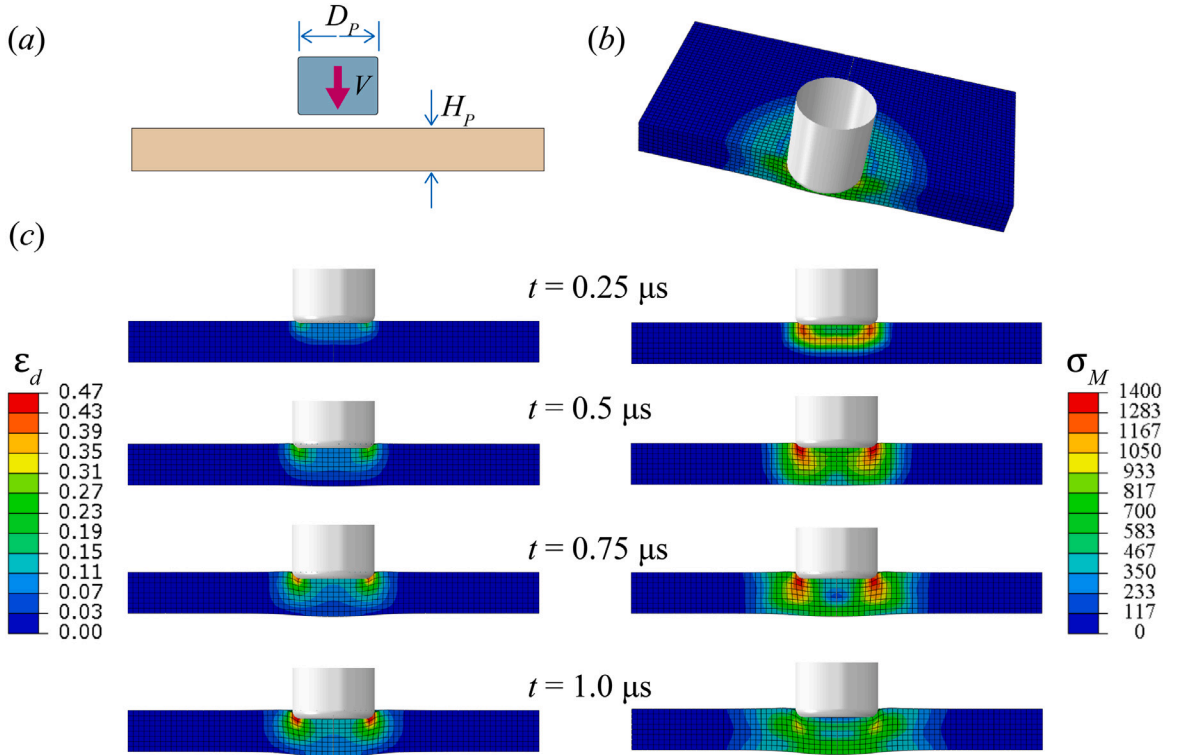


Fig. 6. Typical results of a projectile impact on a plate (a) Schematic. (b) Deformation in 3D. (c-e) Snapshots an axial cross-section with the von Mises stress measure (c), deviatoric strain measure (d) and volumetric strain measure (e).

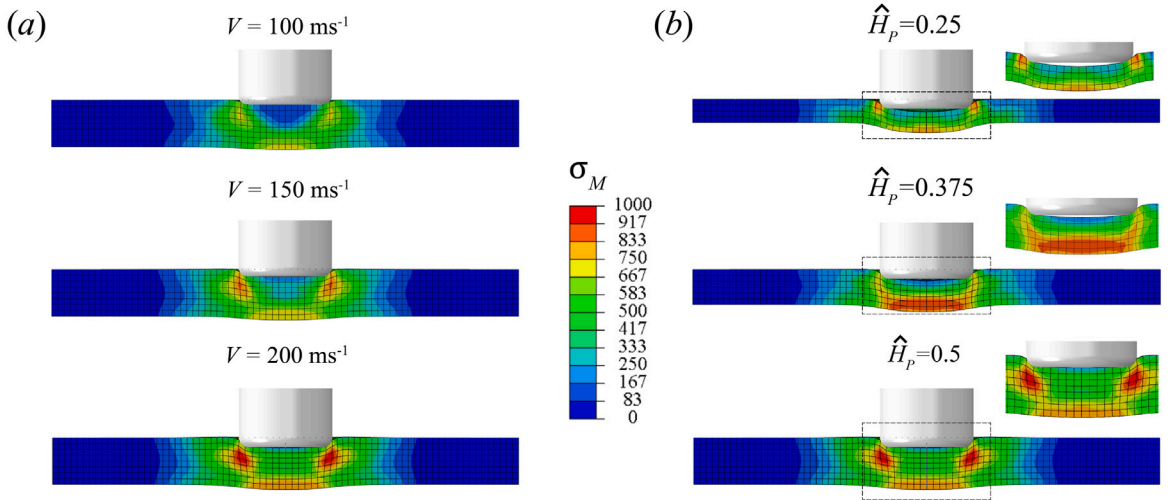


Fig. 7. Parametric study of the plate impact test. Axial cross-section with von Mises stress measure at $t = 1$ s for various (a) impact velocities V and (b) plate thicknesses $\hat{H}_p = H_p/D_p$.

we do not need a priori identification of state variables) at a few times the computational cost of solving the problem with an empirical model.

We demonstrated the framework using crystal plasticity, but it can easily be extended to other continuum multiscale phenomena including composite materials, phase transitions, stress-assisted diffusion, and discrete dislocation dynamics. We can use our framework in an iterated manner to study hierarchical phenomena involving multiple scales (Liu et al., 2021).

We conclude with a discussion of ideas to build on this framework. Our approach requires us to generate data over the entire interval of time to capture the memory of multiscale systems, and this may prove limiting in very long-time computations.

However, in typical phenomena, this memory fades and therefore one may only need a limited history. Such fading memory may be incorporated using a recurrent neural network (RNN) structure (Medsker and Jain, 2001). This eliminates the need to track the entire history, but potentially adds to the cost of training. This is similar to the approach of Mozaffar et al. (2019).

A closely related question is whether RNN and related methods can use the data to discover an underlying Markovian macroscopic model by identifying appropriate macroscopic internal variables, delay kernels etc., which accurately describe the history-dependence. This would be exciting because it not only reduces computational cost but provides new insights into the physics.

We have studied examples where the microscopic model can be studied in the long-wavelength limit so that the unit cell problem is over-damped. The extension to situations like granular materials and molecular dynamics where inertia is critical at the microscale remains open.

Our proposed approach requires training a new approximation for every starting material. In our example of crystal plasticity, we need to generate data and perform training for every initial texture, every set of parameters in the single crystal model. It will be useful if we can learn this as a part of the training, i.e., extend the map $\tilde{F} : \{U(\tau) : \tau \in (0, t), \xi_0\} \mapsto \langle \sigma \rangle(t) \quad t \in (0, T)$. This has been successfully demonstrated in simple problems like Darcy flow (Bhattacharya et al., 2020), and remains a work in progress.

An important question is the inverse problem of inferring single crystal properties based on polycrystalline response. This has recently been addressed in Liu et al. (2021) using the framework of uncertainty quantification. A key step in that approach is modeling the polycrystalline response, and that work used parameter passing. We can replace the parameter passing with the neural network approximation proposed here for improved fidelity. A deeper approach is to include the single crystal parameters in the list of input variables, approximate the map from the single crystal parameters and deformation history to the stress history and use this as the basis for the inverse problem. This remains a work in progress.

Finally, in this work we have generated our training data on strain paths that are smoothed random walks in strain space. This was motivated by the fact that we expect strain paths to be fluctuating but relatively smooth in time in the problems we study. We have also shown that this sampling is in fact effective in the problems we study. However, the optimal distribution from which one should sample training data is an open question and an active subject of research. An alternate approach is to use active learning where new training data is introduced on-line as the macroscopic calculation proceeds. It is unclear how effective the approach we propose in this work would be with active learning since the architecture depends on the data through the model reduction step. So extending the approach to active learning is an important issue for the future.

CRediT authorship contribution statement

Burigede Liu: Conceived the work, Developed the framework, Lead in implementing the framework, Obtaining the numerical results, Discussions during the course of this work and in interpreting the results, Lead in drafting the manuscript, Finalizing. **Nikola Kovachki:** Conceived the work, Developed the framework, Discussions during the course of this work and in interpreting the results, Finalizing. **Zongyi Li:** Discussions during the course of this work and in interpreting the results, Finalizing. **Kamyar Azizzadenehsheli:** Discussions during the course of this work and in interpreting the results, Finalizing. **Anima Anandkumar:** Discussions during the course of this work and in interpreting the results, Finalizing. **Andrew M. Stuart:** Conceived the work, Developed the framework, Discussions during the course of this work and in interpreting the results, Finalizing. **Kaushik Bhattacharya:** Conceived the work, Developed the framework, Discussions during the course of this work and in interpreting the results, Lead in drafting the manuscript, Finalizing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and scripts needed to evaluate the conclusions of this paper are available in the GitHub repository “Learning based multiscale” (https://github.com/Burigede/Learning_based_multiscale.git).

Acknowledgments

We are grateful to Dennis Kochmann for discussion and for providing us with the 2DFFT and the 3D Taylor code to generate the data. This research was sponsored by the Army Research Laboratory, United States and was accomplished under Cooperative Agreement Number W911NF-12-2-0022. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. ZL is supported by the Kortschak Scholars Program. AA is supported in part by Bren endowed chair and De Logi grant. AMS is also partially supported by NSF, United States grant DMS-1818977.

Appendix. Crystal plasticity with twinning

We consider the constitutive framework developed by [Chang and Kochmann \(2015\)](#). The internal variables are the crystallographic orientation $Q \in SO(d)$, total inelastic deformation gradient $F_{\text{in}} \in \mathbb{R}^{d \times d}$, the slip activity $\gamma = \{\gamma_\alpha\}_{\alpha=1}^{n_s}$ in the n_s slip systems and twin volume fractions $\lambda = \{\lambda_\beta\}_{\beta=1}^{n_t}$ in the n_t twin systems that satisfy $\lambda_\beta \in [0, 1]$, $\sum_\beta \lambda_\beta = 1$. We introduce secondary internal variables (accumulated plastic activity) $\{e_\alpha\}_{\alpha=1}^{n_s}$ by integrating

$$\dot{e}_\alpha = |\dot{\gamma}_\alpha|. \quad (\text{A.1})$$

To specify the constitutive functions S and K , we specify a stored energy density

$$W(F, F_{\text{in}}, e, \lambda, y) = W_e(F F_{\text{in}}^{-1}) + W_p(e) + W_t(\lambda) \quad \text{where} \quad (\text{A.2})$$

$$W_e(A) = \frac{G}{2} \left(\frac{\text{tr} A^T A}{(\det A)^{2/3}} - 3 \right) + \lambda_e (\det A - 1)^2, \quad (\text{A.3})$$

$$W_p(e) = \frac{1}{2} e \cdot H e + \sum_{\alpha=1}^{n_s} \sigma_\alpha^\infty (e_\alpha + \frac{\sigma_\alpha^\infty}{h_\alpha} \exp(-\frac{h_\alpha e_\alpha}{\sigma_\alpha^\infty})), \quad (\text{A.4})$$

$$W_t(\lambda) = \lambda \cdot \mathcal{K} \lambda + \sum_{\beta=1}^{n_t} \frac{1}{2} h_\beta \lambda_\beta^2 \quad (\text{A.5})$$

and dissipation functions

$$D_p(\dot{\gamma}) = \sum_{\alpha=1}^{n_s} \frac{\tau_{0,\alpha} \dot{\gamma}_{0,\alpha}}{m_\alpha + 1} \left(\frac{\dot{\gamma}_\alpha}{\dot{\gamma}_{0,\alpha}} \right)^{m_\alpha + 1}, \quad (\text{A.6})$$

$$D_t(\dot{\lambda}) = \sum_{\beta=1}^{n_t} \frac{\tau_{0,\beta} \dot{\lambda}_{0,\beta}}{m_\beta + 1} \left(\frac{\dot{\lambda}_\beta}{\dot{\lambda}_{0,\beta}} \right)^{m_\beta + 1}. \quad (\text{A.7})$$

The stress function is specified as

$$S(F, F_{\text{in}}, e, \lambda) = \frac{\partial W}{\partial F}(F, F_{\text{in}}, e, \lambda) = \frac{\partial W_e}{\partial A} \bigg|_{F, F_{\text{in}}, e, \lambda, y} F_{\text{in}}^{-1}. \quad (\text{A.8})$$

Note that we have chosen an isotropic elastic law for convenience and it does not explicitly depend on position y .

The kinetic relations are specified as

$$\dot{F}_{\text{in}} F_{\text{in}} = L_p + L_t \quad \text{where} \quad (\text{A.9})$$

Table A.1
Parameters used for magnesium in the crystal-plasticity model.

	Parameter	Value	Unit
Basal $\langle a \rangle$	h_a	7.1	GPa
	σ_a^∞	0.7	MPa
	h_{ij}	0	MPa
	m_a	0.5	–
	$\dot{\gamma}_{0,a}$	$1.0 \cdot 10^5$	s^{-1}
Prismatic $\langle a \rangle$	h_a	40	GPa
	σ_a^∞	170	MPa
	h_{ij}	20	MPa
	m_a	0.5	–
	$\dot{\gamma}_{0,a}$	$1.0 \cdot 10^5$	s^{-1}
Pyramidal $\langle c + a \rangle$	h_a	30	GPa
	σ_a^∞	200	MPa
	h_{ij}	25	MPa
	$\tau_{0,a}$	50.5	MPa
	m_a	0.5	–
	$\dot{\gamma}_{0,a}$	$1.0 \cdot 10^5$	s^{-1}
Tensile twin	h_β	7	MPa
	k_{ij}	40	GPa
	m_β	0.5	–
	$\dot{\lambda}_{0,\beta}$	$1.0 \cdot 10^5$	s^{-1}
	γ_t	0.129	–
Elastic Lamé Moduli	λ_e	24	GPa
	G	25	GPa
Density	ρ	$1.0 \cdot 10^4$	kg m^{-3}

$$L_p = \sum_{\alpha=1}^{n_s} \dot{\gamma}_{\alpha} \left[(1 - \sum_{\beta=1}^{n_t} \lambda_{\beta}) s_{\alpha} \otimes m_{\alpha} + \sum_{\beta=1}^{n_t} \lambda_{\beta} s_{\alpha\beta} \otimes m_{\alpha\beta} \right], \quad (\text{A.10})$$

$$L_t = \gamma_t \sum_{\beta=1}^{n_t} \dot{\lambda}_{\beta} s_{\beta} \otimes m_{\beta}, \quad (\text{A.11})$$

$$0 \in \frac{\partial}{\partial \gamma_{\alpha}} (W + D_p), \quad (\text{A.12})$$

$$0 \in \frac{\partial}{\partial \lambda_{\beta}} (W + D_t). \quad (\text{A.13})$$

Crucially s_{α} , m_{α} , b_{α} , m_{α} depend on position. Note that the final two kinetic relations are written as differential inclusions because the derivative is not smooth (W_p is specified in terms of e).

In the 2D calculations, we consider two orthogonal slip systems $s_1 = (1, 0, 0)$, $m_1 = (0, 1, 0)$ and $s_2 = (0, 1, 0)$, $m_2 = (0, -1, 0)$ and no twinning. The shear modulus G and Lamé's constant λ is chosen to be 19 GPa and 24 GPa respectively, while the initial yield strength τ_0 and reference slip rate γ_0 is 100 MPa and 1 s^{-1} . The strain rate sensitivity are chosen to be $m = 0.5$.

We list the parameters that we use for the 3D calculations in Table A.1.

References

2014. ABAQUS user manual: ABAQUS theory guide.
- Asaro, R.J., 1983. Crystal plasticity. *J. Appl. Mech.* 50, 921–934.
- Balasubramanian, S., Anand, L., 2002. Elasto-viscoplastic constitutive equations for polycrystalline fcc materials at low homologous temperatures. *J. Mech. Phys. Solids* 50, 101–126.
- Barton, N.R., Bernier, J.V., Becker, R., Arsenlis, A., Cavallo, R., Marian, J., Rhee, M., Park, H.-S., Remington, B.A., Olson, R.T., 2011. A multiscale strength model for extreme loading conditions. *J. Appl. Phys.* 109, 073501.
- Bensoussan, A., Lions, J.-L., Papanicolaou, G., 2011. Asymptotic Analysis for Periodic Structures. American Mathematical Society.
- Bessa, M.A., Bostanabad, R., Liu, Z., Hu, A., Apley, D.W., Brinson, C., Chen, W., Liu, W.K., 2017. A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality. *Comput. Methods Appl. Mech. Engrg.* 320, 633–667.
- Bhattacharya, K., 1999. Phase boundary propagation in a heterogeneous body. *Proc. R. Soc. Lond. A* 455, 757–766.
- Bhattacharya, K., Hosseini, B., Kovachki, N.B., Stuart, A.M., 2020. Model reduction and neural networks for parametric pdes. *SMAI J. Comput. Math.* 7, 121–157.
- de Borst, R., Ramm, E., 2011. Multiscale Methods in Computational Mechanics. Springer, Heidelberg.
- Bulatov, V., Cai, W., 2013. Computer Simulations of Dislocations. Oxford University Press.
- Car, R., Parrinello, M., 1985. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* 55, 2471–2474.
- Chang, Y., Kochmann, D.M., 2015. A variational constitutive model for slip-twinning interactions in hcp metals: Application to single- and polycrystalline magnesium. *Int. J. Plast.* 73, 39–61.
- Chen, X., Geng, Y., Pan, F., 2016. Research progress in magnesium alloys as functional materials. *Rare Metal Mater. Eng.* 45, 2269–2274.
- Cheng, T., Jaramillo-Botero, A., An, Q., Ilyin, D.V., Naserifar, S., Goddard, W.A., 2019. First principles-based multiscale atomistic methods for input into first principles nonequilibrium transport across interfaces. *Proc. Natl. Acad. Sci.* 116, 18193–18201.
- Cole, D.J., Mones, L., Csányi, G., 2020. A machine learning based intramolecular potential for a flexible organic molecule. *Faraday Discuss.* 224, 247–264.
- Collobert, R., Weston, J., 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th International Conference on Machine Learning, pp. 160–167.
- E, W., 2011. Principles of Multiscale Modeling. Cambridge University Press.
- E, W., Yu, B., 2018. The deep ritz method: A deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Statist.* 6, 1–12.
- Feyel, F., Chaboche, J.-L., 2000. Fe2 multiscale approach for modelling the elastoviscoplastic behaviour of long fibre SiC/ti composite materials. *Comput. Methods Appl. Mech. Engrg.* 183, 309–330.
- Finnis, M., 2010. Crystals, Defects and Microstructures: Modeling Across Scales. Oxford University Press.
- Fish, E., 2009. Multiscale Methods: Bridging the Scales in Science and Engineering. Oxford University Press, Oxford.
- Fritzen, F., Böhlke, T., Schnack, E., 2009. Periodic three-dimensional mesh generation for crystalline aggregates based on voronoi tessellations. *Comput. Mech.* 43, 701–713.
- Fu, C.-C., Dalla Torre, J., Willaime, F., Bocquet, J.-L., Barbu, A., 2005. Multiscale modelling of defect kinetics in irradiated iron. *Nature Mater.* 4, 68–74.
- Giustino, F., 2014. Materials Modelling using Density Functional Theory: Properties and Predictions. Oxford University Press.
- Goldberg, Y., 2017. Neural network methods for natural language processing. *Synth. Lect. Hum. Lang. Technol.* 10, 1–309.
- Gurtin, M., Fried, E., Anand, L., 2013. The Mechanics and Thermodynamics of Continua. Oxford University Press.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Joost, W., Krajewski, P., 2017. Towards magnesium alloys for high-volume automotive applications. *Scr. Mater.* 128, 107–112.
- Jordan, B., Gorji, M.B., Mohr, D., 2020. Neural network model describing the temperature-and rate-dependent stress-strain response of polypropylene. *Int. J. Plast.* 135, 102811.
- Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S., 2017. Self-normalizing neural networks. In: Advances in Neural Information Processing Systems. pp. 971–980.
- Kocks, U., Tome, C., Wenk, H.-R., 2000. Texture and Anisotropy. Cambridge University Press.
- Kulekci, M., 2008. Magnesium and its alloys applications in automotive industry. *Int. J. Adv. Manuf. Technol.* 39, 851–865.
- LeCun, Y., Bengio, Y., et al., 1995. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Netw.* 3361, 1995.
- Liu, Z., Bessa, M., Liu, W.K., 2016. Self-consistent clustering analysis: an efficient multi-scale scheme for inelastic heterogeneous materials. *Comput. Methods Appl. Mech. Engrg.* 306, 319–341.
- Liu, B., Sun, X., Bhattacharya, K., Ortiz, M., 2021. Hierarchical multiscale quantification of material uncertainty. *J. Mech. Phys. Solids* 153, 104492.
- Liu, Z., Wu, C., 2019. Exploring the 3D architectures of deep material network in data-driven multiscale mechanics. *J. Mech. Phys. Solids* 127, 20246.
- Liu, Z., Wu, C., Koishi, M., 2019. A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Comput. Methods Appl. Mech. Engrg.* 345, 1138–1168.
- Marchand, D., Jain, A., Glensk, A., Curtin, W., 2020. Machine learning for metallurgy I. A neural-network potential for al-cu. *Phys. Rev. Mater.* 4, 103601.

- Medsker, L.R., Jain, L.C., 2001. Recurrent Neural Networks. CRC Press.
- Mozaffar, M., Bostanabad, R., Chen, W., Ehmann, K., Cao, J., Bessa, M., 2019. Deep learning predicts path-dependent plasticity. *Proc. Natl. Acad. Sci. USA* 116, 26414–26420.
- Pavliotis, G., Stuart, A., 2008. Multiscale Methods: Averaging and Homogenization. Springer Science & Business Media.
- Phillips, R., 2001. Crystals, Defects and Microstructures: Modeling Across Scales. Cambridge University Press.
- Raissi, M., Perdikaris, P., Karniadakis, G., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707.
- Ravaji, B., Joshi, S.P., 2021. A crystal plasticity investigation of grain size-texture interaction in magnesium alloys. *Acta Mater.* 208, 116743.
- Rice, J.R., 1971. Inelastic constitutive relations for solids: an internal-variable theory and its application to metal plasticity. *J. Mech. Phys. Solids* 19, 433–455.
- Saha, S., Gan, Z., Cheng, L., Gao, J., Kafka, O.L., Xie, X., Li, H., Tajdari, M., Kim, H.A., Liu, W.K., et al., 2021. Hierarchical deep learning neural network (HiDeNN): An artificial intelligence (AI) framework for computational science and engineering. *Comput. Methods Appl. Mech. Engrg.* 373, 113452.
- Tadmor, E.B., Ortiz, M., Phillips, R., 1996. Quasicontinuum analysis of defects in solids. *Phil. Mag. A* 73, 1529–1563.
- Teh, Y.S., Ghosh, S., Bhattacharya, K., 2021. Machine-learned prediction of the electronic fields in a crystal. *Mech. Mater.* 163, 104070.
- Van Der Giessen, E., Schultz, P.A., Bertin, N., Bulatov, V.V., Cai, W., Csányi, G., Foiles, S.M., Geers, M.G., González, C., Hütter, M., et al., 2020. Roadmap on multiscale materials modeling. *Modelling Simulation Mater. Sci. Eng.* 28, 043001.
- Vidyaasagar, A., Tan, W.L., Kochmann, D.M., 2017. Predicting the effective response of bulk polycrystalline ferroelectric ceramics via improved spectral phase field methods. *J. Mech. Phys. Solids* 106, 133–151.
- Wen, M., Tadmor, E.B., 2019. Hybrid neural network potential for multilayer graphene. *Phys. Rev. B* 100, 195419.
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemometr. Intell. Lab. Syst.* 2, 37–52.
- Xiao, S., Hu, R., Li, Z., Attarian, S., Björk, K.-M., Lendasse, A., 2020. A machine-learning-enhanced hierarchical multiscale method for bridging from molecular dynamics to continua. *Extreme Learn. Mach. Deep Learn. Netw.* 32, 14359–14373.
- Yaghoobi, M., Voyiadjis, G., Sundararaghavan, V., 2021. Crystal plasticity simulation of magnesium and its alloys: A review of recent advances. *Crystals* 11, 435.