J. Inst. Eng. India Ser. D
https://doi.org/10.1007/s40033-022-00424-z

ARTICLE OF PROFESSIONAL INTERESTS

Exploiting Machine Learning in Multiscale Modelling of Materials

G. Anand[1] · Swarnava Ghosh[2] · Liwei Zhang[3] · Angesh Anupam[4] · Colin L. Freeman[5] · Christoph Ortner[3] · Markus Eisenbach[2] · James R. Kermode[6]

Received: 10 September 2022 / Accepted: 28 October 2022

Abstract Recent developments in efficient machine learning algorithms have spurred significant interest in the materials community. The inherently complex and multiscale problems in Materials Science and Engineering pose a formidable challenge. The present scenario of machine learning research in Materials Science has a clear lacunae, where efficient algorithms are being developed as a separate endeavour, while such methods are being applied as 'black-box' models by others. The present article aims to discuss pertinent issues related to the development and application of machine learning algorithms for various aspects of multiscale materials modelling. The authors present an overview of machine learning of equivariant properties, machine learning-aided statistical mechanics, the incorporation of *ab initio* approaches in multiscale models of materials processing and application of machine learning in uncertainty quantification. In addition to the above, the applicability of Bayesian approach for multiscale modelling will be discussed. Critical issues related to the multiscale materials modelling are also discussed.

Introduction

There has been extensive recent interest in the application of machine learning in diverse fields within materials science and engineering. Numerous review papers [1–6], viewpoints [7, 8], focused articles on material informatics [9, 10] and road-map documents [11, 12] have been published. With the increasing complexity of emergent materials, predictive structure–property correlation models require mechanistic understanding across a wide range of space and timescales. Experimental insight is often limited by the concurrent occurrence of multiscale phenomena. Conventional methods of multiscale simulation rely on passing information from simulations at *ab initio* up to continuum scale through atomistic scale with a sequential flow of information. In this approach, there is inherent extrapolation away from reference data with a consequent introduction of error. The ultimate challenge of multiscale simulation involves the bridging of atomistic observables to engineering-scale design variables, while retaining adequate precision and accuracy. To address these issues, there have been recent advances in terms of application of a wide range of machine learning tools. The direction of multiscale modelling in materials has primarily two main branches. Firstly, there has been significant interest in employing machine learning to predict the electronic

This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05- 00OR22725 with the US Department of Energy (DOE). The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan (http://energy.gov/downloads/doe- public-access-plan).

✉ G. Anand
ganand@metal.iiests.ac.in

[1] Department of Metallurgy and Materials Engineering, Indian Institute of Engineering Science and Technology, Shibpur, Howrah 711103, India

[2] National Centre for Computational Sciences, Oak Ridge National Laboratory, 1 Bethel Valley Rd, Oak Ridge, TN 37831, USA

[3] Department of Mathematics, University of British Columbia, 1984 Mathematics Rd, Vancouver, BC V6T1Z2, Canada

[4] Department of Computer Science, Cardiff Metropolitan University, Llandaff Campus, Cardiff, Wales CF5 2YB, UK

[5] Department of Materials Science and Engineering, University of Sheffield, Mappin St, Sheffield S1 3JD, UK

[6] Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Warwick CV4 7AL, UK

Published online: 28 November 2022

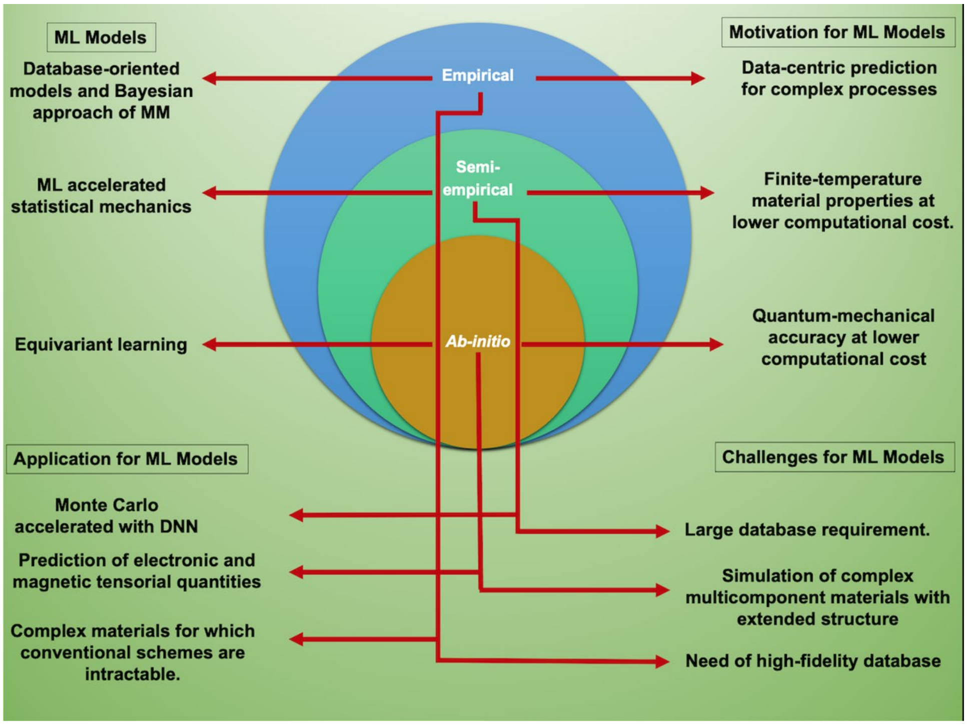
ML Models
Database-oriented models and Bayesian approach of MM
ML accelerated statistical mechanics
Equivariant learning
Empirical
Semi-empirical
Ab-initio
Motivation for ML Models
Data-centric prediction for complex processes
Finite-temperature material properties at lower computational cost.
Quantum-mechanical accuracy at lower computational cost
Application for ML Models
Monte Carlo accelerated with DNN
Prediction of electronic and magnetic tensorial quantities
Complex materials for which conventional schemes are intractable.
Challenges for ML Models
Large database requirement.
Simulation of complex multicomponent materials with extended structure
Need of high-fidelity database

Fig. 1 Schematic showing the various machine learning (ML) models, motivation, application and challenges associated with such ML models for *ab initio*, semi-empirical atomistic approaches and empirical multiscale materials models

structure of materials, whose accuracy is close to that of *ab initio* theory, while providing significant computational efficiency. The second approach involves data-mining-based schemes, where prediction of structure–property correlation is based on development of models based on data. This article summarizes discussions between researchers from both streams to start a discussion on the development of newer methods for ever challenging problems in materials science and engineering. The article follows on from an online conference held in December 2021, entitled ‘Exploiting Machine Learning in Multiscale Modelling of Materials’ (EMLM^3), co-organized by the first and last author of the present article and involving all other authors as attendees [13]. The article is structured into five sections corresponding to distinct themes, followed by a discussion of key challenges remaining in the field. The aim of this article is to document the discussions during EMLM^3 (Fig. 1); readers are referred to the published literature for basics [3, 14–20] and recent advances in machine learning and its application for predicting the structure–property correlations in materials engineering [21–23].

Multiscale Materials Modelling

Machine Learning and Uncertainty Quantification in Multiscale Simulations

The behaviour of materials is a collective outcome of the numerous mechanisms that play out across a hierarchy of length and timescales. In a *concurrent* multiscale analysis scheme [24], the behaviour at each material point at each time of the higher coarse scale model is informed by the solution of the lower fine scale model. Though this idea has been incorporated in FE^2 concurrent two-scale approach [25], it is prohibitively expensive to extend it to multiple scales. A more widely used technique is the *sequential* or parameter passing multiscale approach, where parameters from lower fine scale simulations are passed as inputs of the higher coarse-scale simulations. The intricacies of material response at each scale lead to uncertainty in the final engineering application. Therefore, it is important that the integral uncertainties are accurately quantified. However, direct computation of these integral uncertainties is prohibitive [26]. Traditionally, there has been limited data bridging multiple scales. However, owing to recent developments in experimental techniques and computational methods, the present researchers are transforming from a data deficient to a data-rich field. Additionally, there has been landmark advances in machine learning and uncertainty quantification which facilitate learning from lower scale material behaviour with quantified uncertainty. These developments open up several avenues of investigation. To elucidate, by using recent advances in new materials modelling methods, experimental data and supervised learning methods for multiscale modelling of materials? How does material uncertainties propagate across multiple scales and how efficiently quantify these? Figure 2 shows a schematic outline for answering these two questions. The data from simulations of a fine scale problem can be used to build a machine learning/data-driven models. This model can be validated against experimental data, and the model and uncertainty will then be passed as inputs to the coarse scale problem. See [27] for developing a machine learning approach for mapping the electronic fields to the local deformation fields using a neural network.

Incorporating *ab initio* Approaches in Multiscale Modelling of Materials Processing

Properties of materials are linked to their processing, which in turn is linked to their nano-/micro-structure. Though these links are important for the synthesis of new materials with desired functionalities, it is not understood the physical phenomena that link the processing route to the synthesized microstructure and to the target property. However, recent advances in efficient computational methods allow us to establish these links and guide synthesis and processing of new materials. These can also provide a deeper understanding of the complex mechanisms that play out during the process. A sequential multiscale modelling approach is to utilize first-principles calculations to evaluate the free energy surface of each phase over different range of processing parameters [28]. These free-energy surfaces can be used as inputs to meso-scale phase field simulations for simulating microstructure evolution during processing. The resulting microstructures can be homogenized and passed as inputs to continuum finite element simulations. See Fig. 3 for a schematic describing this idea. Recent advances in computational methods for large-scale first-principles simulations of defects in materials [29] can be used to accurately calculate defect generation and evolution during processing.

Bayesian Approaches to Multiscale Modelling of Materials

Surrogate Models: Simplifying Multiscale Modelling of Materials

The three main steps of a typical Bayesian data analysis are (i) composing the model, (ii) fitting the composed model to the available data, (iii) refining the model through continuously monitoring its fit and making a comparison with other models [30]. The first step, composing a model, usually involves the usage of prior knowledge about the system. The second step fitting the model is the core of Bayesian inference and is often performed using Monte Carlo sampling algorithms. The main aim of this step is to evaluate the posterior probability distribution of the model parameters [31].

Multiscale modelling of the composite material system can be achieved using an approach like mean field homogenization [32]. Due to the increased number of parameters in the nonlinear range, it is unlikely that a unique parameter set will be able to reproduce all the experimental conditions due to the intrinsic limitations of the model. In such scenarios, Bayesian inference is a plausible option for estimating the parameters of the proposed multiscale model. In this approach, the posterior probability distribution function takes into account all the aforementioned uncertainties reasonably. As per the Bayesian philosophy, the posterior distribution function of the parameters represents the correction of the prior distribution through a likelihood function calculated using the available experimental data. The Bayesian inference method has been used to estimate the parameters of nonlinear, homogenous materials in single-scale settings [33–35]. Nonetheless, the inference in a multiscale paradigm

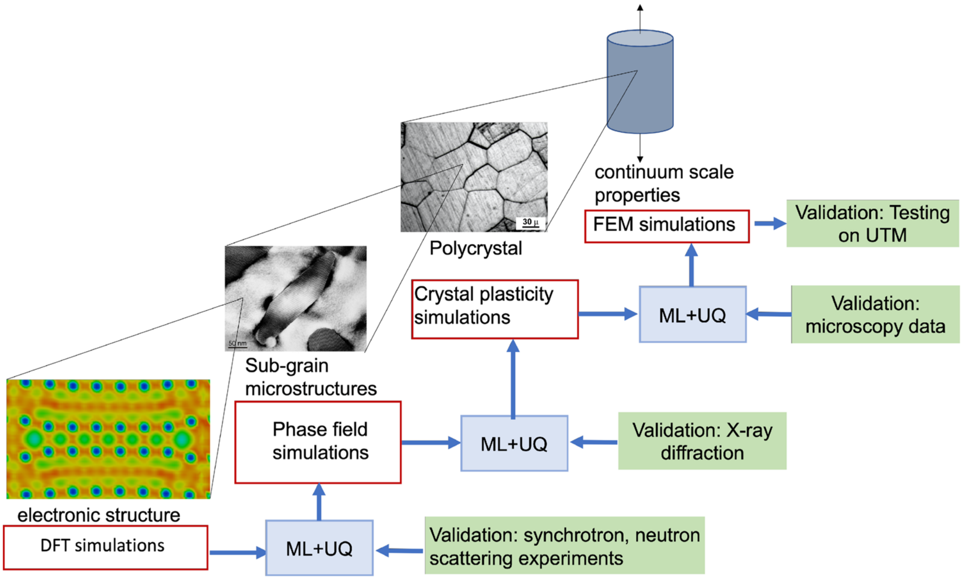
continuum scale
properties
FEM simulations
Validation: Testing on UTM
Polycrystal
30 μ
Crystal plasticity simulations
ML+UQ
Validation: microscopy data
Sub-grain microstructures
50 nm
Phase field simulations
ML+UQ
Validation: X-ray diffraction
electronic structure
DFT simulations
ML+UQ
Validation: synchrotron, neutron scattering experiments

Fig. 2 Schematic overview of machine learning and uncertainty quantification for materials modelling across the scales

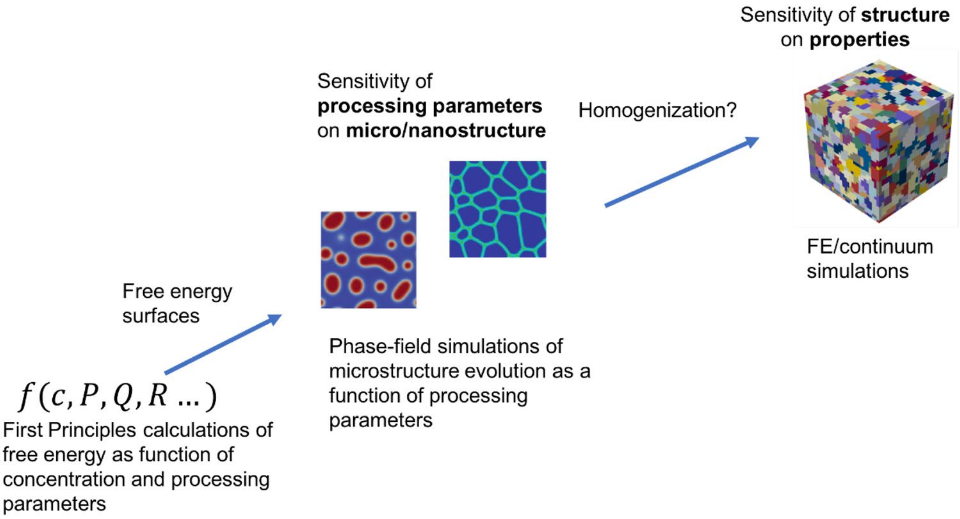
Sensitivity of **structure** on **properties**
Sensitivity of **processing parameters** on **micro/nanostructure**
Homogenization?
FE/continuum simulations
Free energy surfaces
Phase-field simulations of microstructure evolution as a function of processing parameters
$f(c, P, Q, R \ldots)$
First Principles calculations of free energy as function of concentration and processing parameters

Fig. 3 Schematic overview of the link between materials processing and the resulting mechanical properties

is comparatively more difficult due to its intrinsic computational requirements.

The Bayesian inference can however be made computationally more efficient in a typical nonlinear scenario by incorporating the concept of surrogate modelling [32]. Surrogate modelling framework is employed in situations wherein the model is too complex mainly due to a large number of parameters or a very higher order of differential equations. Although a model is deemed as an abstraction of the underlying system, it can also introduce an abstraction of abstraction through a suitable surrogate modelling mechanism. Formulation of a suitable surrogate model capable of generating satisfactory simulation in the operating region can significantly improve the computational efficiency of a model. All the mentioned traits of surrogate modelling make it a useful tool in the multiscale modelling of materials. A more detailed overview of the surrogate modelling approach in the context of material science can be found here [36]. The chosen surrogate model should ideally be feasible for high-dimensional nonlinear mapping. The feature set corresponding to the input–output data is pivotal in deciding the nonlinear base function, and thereafter, a linear combination of the chosen base model can be used to formulate the surrogate model. An artificial neural network (ANN) is instrumental in developing a surrogate model because, in theory, an ANN can carry out high-dimensional nonlinear mapping provided an adequate number of hidden layers and neurons are considered. For example, an ANN-based surrogate model is trained with various micro-structure geometrical and material parameters and Bayesian inference is used for parameter estimation [32]. In another such example, the same sort of surrogate modelling, employing ANN is applied for multiscale modelling of carbon-based nanocomposites [37]. The selection of a surrogate model type can be difficult as there are numerous options, and therefore, a detailed review [38] can be referred by the practitioner to make an informed decision.

In addition to a popular choice, ANN, for surrogate modelling in multi-scale modelling of materials, a more transparent but highly efficient nonlinear dynamical systems model class, such as Nonlinear Auto-regressive Moving Average with Exogenous Inputs (NARMAX) [39]. Despite possessing a good nonlinear mapping capabilities, an ANN lacks transparency and can be termed as a ‘black box’ model, whereas a NARMAX model can be used to tease apart the characteristics of the various dynamical elements present within the original multi-scale model representation of the system. The parameter estimation of a NARMAX model can be done using either a deterministic or probabilistic approach, but the later one would be a preferred option because the posterior probability distribution of the parameters can accommodate the systems’ uncertainties to a fair extent.

ABC in Multi-scale Modelling

Approximate Bayesian Computation (ABC) is a data-driven computation method originating from the classical theory of Bayesian statistical inference. The likelihood function under the Bayesian theory represents the probability of the observed data under a model. The model in this context is statistical in nature. Therefore, the likelihood function is the quantification of the joint probability of the observed data as a function of the parameters corresponding to the selected model. The derivation of likelihood function in analytical sense is mostly straightforward in simple models. Nonetheless, the same does not apply to more complex models. In such cases, deriving an analytical formula corresponding to likelihood function can become computationally expensive or sometimes even near impossible. The ABC harnesses the modern days high-performance computing infrastructure to bypass the necessity of evaluating likelihood function. Essentially, an ABC algorithm employs a simulation approach combined with Monte Carlo sampling for returning the posterior distribution of model parameters [31, 40].

The ABC methodology is inspired from simple rejection algorithm, but practitioners often employ an efficient version of this algorithm such as, ABC-Markov Chain Monte Carlo (ABC-MCMC) or ABC-Sequential Monte Carlo (ABC-SMC). In an ABC algorithm, at every iteration, the parameter set is sampled from a probability distribution, which simulates the model. The data generated through simulation are simplified using a chosen summary statistics which is compared to the observed summary statistics to decide whether to accept or reject the sampled parameters [41].

The ABC method has a great potential in multi-scale modelling of materials in almost all the scenarios wherever it intend to perform Bayesian inference. For example, the parameter estimation of models described in the literature [32, 37] can also be performed by applying ABC instead of following the suggested surrogate modelling route. A multi-scale modelling of permeability and the estimation of parameters based on Bayesian inference is discussed in the literature [42]. The permeability field is an important factor pertaining to a reservoir performance. However, instead of using a Multilevel Markov Chain Monte Carlo (MLMCMC), ABC is used, which has a superior performance compared to the MCMC approach. This study bolsters the application of ABC in multi-scale modelling. Some studies on the usage ABC in multi-scale modelling of materials and their comparison with the more established framework, such as surrogate modelling, would be an interesting area of investigation for the community. The further details related to ABC techniques including some practical challenges are succinctly described here [43].

Machine Learning of Equivariant Atomic Properties

Equivariant Atomic Properties

In this section the authors review the recent adoption of machine-learning methodologies for the representation and approximation of equivariant atomic properties. Generally speaking, an atomic property is a mapping from the local atomic configuration $\boldsymbol{\sigma}_i = \{\sigma_j\}_{r_{ij}<r_{\text{cut}}}$ within a structure $\{\sigma_i\}_i$ (molecule or material) to a scalar or tensor that may satisfy certain symmetries, where r_{cut} is a characteristic lengthscale beyond which interactions can be truncated. From these "local" atomic properties one can then build models for the entire structure. For example, if $\sigma_i = (\boldsymbol{r}_i, z_i)$ describes an atom with position $\boldsymbol{r}_i$ and atomic number z_i, then the potential energy for a material or molecule can be written as

$$\mathcal{V}(\{\sigma_i\}_i) = \sum_i V_i(\boldsymbol{\sigma}_i)$$

where the site energy V_i is an atomic property that is *invariant* under translations, rotations, reflections and permutations of the atomic environment (freezing the centre atom).

More generally, most atomic properties one encounters have equivariant symmetries under those operations. The charge and magnetic dipoles, forces, or a complete Hamiltonian or friction operator. For example, $\mathcal{V} \in \mathbb{R}^3$ could represent the charge dipole on a molecule with $V(\boldsymbol{\sigma}_i)$ the atomic contribution from atom i.

Such general equivariant symmetries can be expressed as

$$V(\boldsymbol{\sigma}_i) = \text{SYM}_Q\big[V(Q\boldsymbol{\sigma}_i)\big], \qquad \forall Q \in O(3), \tag{1}$$

where $Q\boldsymbol{\sigma}_i$ denotes the action of the generalized rotation matrix Q on the atomic environment $\boldsymbol{\sigma}_i$ and SYM_Q denotes the action of Q on the property V. (Here, the orthogonal group $O(3)$ was focused, but in principle the ideas generalize very broadly.) For example, if V denotes a site energy potential, then $\text{SYM}_Q V = V$. If $V \in \mathbb{R}^3$ denotes a dipole or a force, then $\text{SYM}_Q V = Q^T V$.

Equivariant Features and Linear Models

The simplest way to approximate V is to represent it by a linear model, i.e. it may expand

$$V(\boldsymbol{\theta};\boldsymbol{\sigma}_i) = \boldsymbol{\theta} \cdot \boldsymbol{\mathcal{B}}(\boldsymbol{\sigma}_i) = \sum_k \theta_k \boldsymbol{\mathcal{B}}_k(\boldsymbol{\sigma}_i),$$

where $\{\boldsymbol{\mathcal{B}}_k\}_k$ are tensorial-valued features (or, a basis) exactly satisfying the desired equivariance,

$$\boldsymbol{\mathcal{B}}_k(\boldsymbol{\sigma}_i) = \text{SYM}_Q(\boldsymbol{\mathcal{B}}_k(Q[\boldsymbol{\sigma}_i)]), \forall Q \in O(3).$$

Numerous constructions of such features have been proposed, including, e.g. the invariant Behler, Parinello symmetry functions [44], the SOAP descriptor [45–47], Spectral Neighbor Analysis Potential (SNAP) [48], Moment Tensor Potentials (MTPs) [49, 50], or most recently the Atomic Cluster Expansion (ACE) [51–55]. Moreover, there are natural and intimate connections between all of these approaches that have been explored in depth in [56], where also references to further constructions of equivariant features can be found. It refers to the references above for further details and only point out that constructing equivariant features of atomic environments is now a fairly well-understood task, which can be performed efficiently and reliably, and said features can be evaluated in a numerically robust and efficient manner.

Nonlinear Models

A feature vector $\boldsymbol{\mathcal{B}} = (\boldsymbol{\mathcal{B}}_k)_k$ may also be used to construct nonlinear models of the form

$$V(\boldsymbol{\theta};\boldsymbol{\sigma}_i) = \mathcal{F}(\boldsymbol{\theta};\boldsymbol{\mathcal{B}}(\boldsymbol{\sigma}_i)),$$

where $\mathcal{F}$ would typically be represented in terms of an artificial neural network [44] or a Gaussian process [45]. Another class of nonlinear models of increasing importance are equivariant neural networks (ENNs), which are easiest to think about as message passing networks [57–59] in which each layer t is represented as an atomic structure $\{\sigma_i^{(t)}\}_i$ but now with "hidden" features attached to a state, $\sigma_i^{(t)} = (\boldsymbol{r}_i, z_i, \boldsymbol{\xi}_i^{(t)})$. The "message" is an equivariant mapping

$$\boldsymbol{\xi}_i^{(t+1)} = V^{(t+1)}(\boldsymbol{\sigma}_i^{(t)}).$$

In the output layer, the features may then be pooled to extract for example a global property. Thus, in this setting, each node in the network is an equivariant atomic property.

In general, these nonlinear models give rise to such an immense freedom in their design that f will not give any further details, but direct the reader to the cited references for further information.

Parameter Estimation

Ignoring the more complex case of ENNs for the moment, both linear and nonlinear models in the end give rise to parameterized atomic properties $V(\boldsymbol{\theta};\sigma_i)$ that satisfy certain equivariance conditions. The authors estimate the parameters $\boldsymbol{\theta} = (\theta_k)_k$ from first-principles datasets by fitting the parameters, usually through a least squares approach.

A critical challenge one often encounters is that the atomic property V cannot be directly observed, but only indirect observations can be made. For example, a site energy is not an observable property, but total potential energy and forces can be obtained from first-principles calculations. In general, the present researchers therefore seek to estimate parameters by minimizing a least squares loss function

$$L(\boldsymbol{\theta}) = \sum_j w_j \|O_j(V) - y_j\|^2 + \Phi(\boldsymbol{\theta}),$$

where O_j is an observation, y_j the value to which the observation is fitted, w_j a weight, and Φ a general regularization term. If the parameterization V and the observations O_i are linear and the regularizer Φ is quadratic, then minimizing L gives rise to a *linear least squares* problem for which efficient and robust direct solution techniques exist. Otherwise the minimization must be performed iteratively using, for example, (stochastic) gradient descent methods, or quasi-Newton methods.

By means of example, suppose that V is a linear parameterization of a site energy or general atomic property and that an observation O_i represents the evaluation of the corresponding total property on a structure $\boldsymbol{\sigma}^{(j)} = \{\boldsymbol{\sigma}_i^{(j)}\}$, e.g. the total energy or the net dipole. In that case,

$$\begin{aligned} O_j(V) &= \sum_i V(\boldsymbol{\theta};\boldsymbol{\sigma}_i^{(j)}) = \sum_i \sum_k \theta_k \boldsymbol{\mathcal{B}}_k(\boldsymbol{\sigma}_i^{(j)}) \\ &= \sum_k \theta_k \sum_i \boldsymbol{\mathcal{B}}_k(\boldsymbol{\sigma}_i^{(j)}). \end{aligned}$$

This shows how general observations can be composed of local features and also highlights the linear structure in the case of linear models and linear observations.

Applications

The machinery outlined above, but with numerous modifications to fine-tune and specialize, has been used with increasing success in the atomistic modelling community, for a broad range of applications. Early on the main focus was on the construction of highly transferrable interatomic potentials with (near-) ab initio accuracy [48, 50, 51, 55, 57, 60–62]. More recently, there have been studies on adapting it to learning equivariant Hamiltonians [63, 64], wave functions [65], or dielectric and magnetic tensorial properties [66].

Accelerating Statistical Mechanics Using Machine Learning

In addition to understanding the ground state of materials, it is important to be able to describe the finite temperature behaviour of materials. To obtain these properties from a microscopic atomistic model of a material, statistical mechanics need to be employed. The connection between these two scales is provided, in the case of a canonical ensemble, by a temperature-dependent probability distribution of the occupation of the individual micro-states, i.e. the points in phase space, that is the Boltzmann distribution

$$p(\xi;\beta) = \frac{e^{-\beta H(\xi)}}{\int_\Omega e^{-\beta H(\xi')} \mathrm{d}\xi'}$$

where $\xi \in \Omega$ represents a microstate in the high-dimensional phase space Ω of the system, $\beta = 1/k_B T$ is the inverse temperature and $H(\xi)$ is the Hamiltonian. Thus, the expected values of observables A are temperature-dependent averages weighted with this distribution.

$$\langle A \rangle_\beta = \int_\Omega A(\xi) p(\xi;\beta) \mathrm{d}\xi$$

The methods of choice for sampling this huge phase space of materials at finite temperature are Monte Carlo algorithms, starting with the original Metropolis algorithm [67] and including further improvements in sampling that have achieved significant gains in the scaling and sampling properties for ranges of temperatures [68–70].

Traditionally Monte Carlo sampling has been applied to model systems where the energy is fast and easy to evaluate. To capture the subtleties of the interactions in real materials, it is desirable to be able to combine the statistical mechanics with first-principles density functional theory calculations that capture the quantum mechanical origins of the energy of the materials due to electron interactions. Yet these calculations are usually prohibitively expense due to the cost of the single energy evaluations. Thus this direct approach of combining classical Monte Carlo simulations with DFT calculations, has been applied only in few cases of carefully chosen system, thus as the calculations of the Curie temperature in iron-based materials [71, 72] or in binary solid solution alloys [73]. To be able to increase the number and complexity of materials that can be investigated, it is necessary to reduce the number of first-principles energy evaluations. This can be achieved by constructing surrogate models from first-principles data that can be evaluated with a significant reduction of computational resources. A traditional approach to constructing such a model for solid solution alloys has been the cluster expansion method [74, 75].

With the form of the cluster expansion ansatz, there are many choices to be made about which clusters to include in the model. This includes both the number of sites included in these clusters as well as the spatial extend of these interactions. These problems share a noticeable similarity to the hyperparameter optimization in many machine learning approaches. Thus the relation between the model that is supported by a given dataset size and the desired error in finite temperature Monte Carlo calculations can be controlled by combining these to methods as demonstrated in an application of the Wang–Landau Monte Carlo algorithm to FeCo [76]. When moving from binary alloys as in the previous example to more complex materials, such as concentrated multi-component solid solution alloys or high entropy alloys, the problems of identifying the hyper-parameters of a model that maximizes the information that can be extracted from a training dataset while avoiding over-fitting becomes even more acute. To avoid overfitting it is common to introduce a regularization parameter to penalize large regression coefficients. For a simple pair interaction model subset of the cluster expansion approach the use of a Bayesian information criterion allows the identification of the ideal model size for a given training set that maximizes the model likelihood while minimizing the risk of overfitting. This robust data-driven approach for finding models for the energy of high entropy alloys will also indicate the need for larger datasets if the required model accuracy cannot be supported by an existing dataset size [77].

A different approach to building surrogate models for the energy of alloy configurations is provided by the construction of deep artificial neural networks (DNNs) that takes the site occupation by the atomic species (for multi-component alloys as a one-hot encoding) as an input vector. As DNNs usually have a significantly larger parameter space than the physically inspired cluster expansion models, the model training would require prohibitively large training sets to avoid the possibility of the model being stuck in unphysical models. An approach that can at the same time reduce that dataset sizes and stabilizes the deep-learning predictions of the energies in solid solution alloys utilizes the additional information that is available from the underlying first-principles calculations that generated the training data in the first place. These calculations provide in addition to the energy of the system also the distribution of charge and, if the material is magnetic, the magnetic moments in the material. This allows the application of multitask learning [78]. By constructing a network architecture that in addition to the energy also predicts the these additional physical properties, the DNN model can be guided to a more robust prediction of the physical properties as the additional knowledge acts as regularizer driven by the physics of the system [79].

Finally, the model construction approaches described above can be combined with Monte Carlo simulations. Yet the original training dataset might not cover the phase space sufficiently to ensure the accurate description of the statistical mechanics of complex multi-component alloys where both transitions between ordered and disordered phases as well as phase separation between different compositional phases might occur. To capture this behaviour the model can be actively refined by recording the phase space points that are visited during a simulation run and identifying new phase space regions that were not sufficiently covered by the initial training data. By enriching the training set with new first-principles calculations for these points and retraining the model can be refined. This enables a self-consistent workflow that alternates the first-principles data generation, model retraining and Monte Carlo simulations until a the process has converged to a model that provides a stable prediction for the finite temperature statistical mechanics of the material [80].

Discussion

The present article provides a pedagogical overview of the basics of the multiscale materials modelling. The challenges of sequential multiscale materials modelling have been highlighted along with some possible solutions involving uncertainty quantification along with experimental validation at each stage of the multiscale model. Modelling materials processing presents an additional challenge, as such a scenario is inherently dynamic in nature. Such problem requires the incorporation of the influence of the material defects, as well. It has been proposed that free energy surfaces derived from the *ab initio* approaches can be input to the mesoscale microstructure development model, such as phase field model (PFM). The PFM can provide the input to the property evaluation model such as finite element model. However, development of accurate free energy surface with defects can be a challenging problem. Such a challenge can be addressed through machine learned interatomic potentials, which can generate the potential energy landscape with the quantum-mechanical accuracy with limited computational cost [81].

Highly complex problems involving numerous variables pertaining to structure–property correlations have been earlier attempted through database-oriented approaches [82, 83]. However, recent interest in the Bayesian approach to develop the surrogate models for structure–property correlations can be alternative way forward. The development of surrogate models requires the coarse graining of parameter space. The conventional approach of the mean-field homogenization is being replaced with the Bayesian inference through application of ANN as well as NARMAX approach for the determination of feature set of the input–output database. The Bayesian approach has already being increasingly applied to single-scale problems such as alloy design [84], grain growth [85], as well as property predictions [86–88]. However, the application of such a approach for multiscale materials modelling problems has been limited [32].

The authors have also briefly reviewed recent developments using equivariant features derived from atomic environments. This is now a well-understood approach, which can be performed efficiently and reliably, in a numerically robust and efficient manner to provide input representations for linear and nonlinear machine learning models suitable for a range of applications from interatomic potentials through to direct surrogate models of electronic structure calculations.

The determination of finite temperature properties through statistical mechanics-based approach conventionally relies on the sampling of the phase space through Metropolis Monte Carlo method. In such as method, energy calculation is a bottleneck step. The surrogate model for the energy calculation step can be either developed by employing physically inspired cluster expansion approach or by DNN approach. The DNN approach is a promising approach for the development of energy calculation surrogate. However, it required the large dataset covering ideally the full phase space. As mentioned in Sect. 5, such a challenge can be solved by the development of the self-consistent workflow involving iterative *ab initio* database generation, model refitting and Monte Carlo simulations. This approach of dynamic database generation is closely related to the MODNet approach, where low data problem has been tackled with feature selection based on normalized mutual information [89]. The issue of low data becomes particularly challenging for multicomponent materials, where again efficient schemes are being developed [90–92].

In addition to the above, during EMLM^3 focused discussions on various issues related to the application of machine learning (ML) to various problems such as reproducibility of calculations, database-oriented materials discovery and challenges in dealing with long-ranged electrostatic interactions through ML approaches were carried out.

The reproducibility studies on ML-based methodologies are scarce, in comparison with the concerted efforts to report the reproducibility of plane-wave *ab initio* calculations [93]. The attempt to ensure the reproducibility in the atomistic calculations has been through the interatomic potential repository of NIST, which stores the numerous classical potential published in the literature. Another effort to document the classical interatomic interaction is the OpenKIM framework [94], which not only stores the interatomic potentials, but also verifies the interatomic potential for code integrity and predictive capability. Recent interest in applying machine learning in materials science has led to a significant volume of publications. However, it is important to note that it is becoming increasingly difficult to reproduce calculations. In principle, the results of the calculation may be reproducible, if the code, training datasets and environment of the calculator are properly shared. But, the repetition of the calculations is generally not pursued and present scenario of peer-reviewed journal publications does not emphasize the data and code sharing with the scientific results.

The curation of datasets and allied information of ML models would be critical for not only their justifiable application, but also helpful in the identification of the inherent limitation of these models. Such an approach is in line with the traditional approaches [95, 96]. It is envisaged that the significant interest in applying ML to materials problem would gradually lead to the fifth paradigm of scientific discovery.

Funding The authors are thankful to UKIERI and DST for funding the Partnership Development Workshop. The authors are also thankful to US-DOE-ORNL for funding. This research used resources of the Oak Ridge Leadership Computing Facility, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The authors acknowledge the Network Builder Grant from Cardiff Met University.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C.W. Park, A. Choudhary, A. Agrawal, S.J. Billinge et al., Recent advances and applications of deep learning methods in materials science. npj Comput. Mater. **8**(1), 1–26 (2022)
2. E.-W. Huang, W.-J. Lee, S.S. Singh, P. Kumar, C.-Y. Lee, T.-N. Lam, H.-H. Chin, B.-H. Lin, P.K. Liaw, Machine-learning and high-throughput studies for high-entropy materials. Mater. Sci. Eng. R. Rep. **147**, 100645 (2022)
3. G.L. Hart, T. Mueller, C. Toher, S. Curtarolo, Machine learning for alloys. Nat. Rev. Mater. **6**(8), 730–755 (2021)
4. D. Morgan, R. Jacobs, Opportunities and challenges for machine learning in materials science. Annu. Rev. Mater. Res. **50**, 71–103 (2020)
5. J. Gubernatis, T. Lookman, Machine learning in materials design and discovery: examples from the present and suggestions for the future. Phys. Rev. Mater. **2**(12), 120301 (2018)
6. Y. Liu, T. Zhao, W. Ju, S. Shi, Materials discovery and design using machine learning. J. Mater. **3**(3), 159–177 (2017)
7. J. Westermayr, M. Gastegger, K.T. Schütt, R.J. Maurer, Perspective on integrating machine learning into computational chemistry and materials science. J. Chem. Phys. **154**(23), 230903 (2021)
8. B. Meredig, Five high-impact research areas in machine learning for materials science. Chem. Mater. **31**(23), 9579–9581 (2019)
9. D. Jha, V. Gupta, L. Ward, Z. Yang, C. Wolverton, I. Foster, W.-K. Liao, A. Choudhary, A. Agrawal, Enabling deeper learning on big data for materials informatics applications. Sci. Rep. **11**(1), 1–12 (2021)
10. R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects. npj Comput. Mater. **3**(1), 1–13 (2017)
11. H. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M.A. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, et al., Roadmap on machine learning in electronic structure. Electron. Struct. **4**(2), 023004 (2022)
12. J.F. Rodrigues, L. Florea, M.C. de Oliveira, D. Diamond, O.N. Oliveira, Big data and machine learning for materials science. Discov. Mater. **1**(1), 1–27 (2021)
13. G. Anand, J.R. Kermode, Exploiting Machine Learning in Multiscale Modelling of Materials. https://warwick.ac.uk/fac/sci/wcpm/emlm2021
14. S. Alexander, S. Bawabe, B. Friedman-Shaw, M.W. Toomey, The physics of machine learning: an intuitive introduction for the physical scientist. arXiv:2112.00851 (2021)
15. J. Wei, X. Chu, X.-Y. Sun, K. Xu, H.-X. Deng, J. Chen, Z. Wei, M. Lei, Machine learning in materials science. InfoMat **1**(3), 338–358 (2019)
16. D. Morgan, R. Jacobs, Opportunities and challenges for machine learning in materials science. arXiv:2006.14604 (2020)
17. C. Gao, X. Min, M. Fang, T. Tao, X. Zheng, Y. Liu, X. Wu, Z. Huang, Innovative materials science via machine learning. Adv. Func. Mater. **32**(1), 2108044 (2022)
18. P. Huembeli, J.M. Arrazola, N. Killoran, M. Mohseni, P. Wittek, The physics of energy-based models. Quantum Mach. Intell. **4**(1), 1–13 (2022)
19. J.R. Cendagorta, J. Tolpin, E. Schneider, R.Q. Topper, M.E. Tuckerman, Comparison of the performance of machine learning models in representing high-dimensional free energy surfaces and generating observables. J. Phys. Chem. B **124**(18), 3647–3660 (2020)
20. K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science. Nature **559**(7715), 547–555 (2018)
21. R.E. Goodall, A.A. Lee, Predicting materials properties without crystal structure: Deep representation learning from stoichiometry. Nat. Commun. **11**(1), 1–9 (2020)
22. R.E. Goodall, A.S. Parackal, F.A. Faber, R. Armiento, A.A. Lee, Rapid discovery of stable materials by coordinate-free coarse graining. Sci. Adv. **8**(30), 4117 (2022)
23. S.I.P. Tian, A. Walsh, Z. Ren, Q. Li, T. Buonassisi, What information is necessary and sufficient to predict materials properties using machine learning? arXiv:2206.04968 (2022)
24. J.D. Lee, J. Li, Z. Zhang, L. Wang, In: S.A. Meguid, G.J. Weng, (eds.) Sequential and Concurrent Multiscale Modeling of Multiphysics: From Atoms to Continuum, pp. 1–38. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-52794-9_1
25. F. Feyel, A multilevel finite element method (fe2) to describe the response of highly non-linear structures using generalized continua. Comput. Methods Appl. Mech. Eng. **192**(28), 3233–3244 (2003). https://doi.org/10.1016/S0045-7825(03)00348-7. Multiscale Computational Mechanics for Materials and Structures
26. L.J. Lucas, H. Owhadi, M. Ortiz, Rigorous verification, validation, uncertainty quantification and certification through concentration-of-measure inequalities. Comput. Methods Appl. Mech. Eng. **197**(51–52), 4591–4609 (2008)
27. Y.S. Teh, S. Ghosh, K. Bhattacharya, Machine-learned prediction of the electronic fields in a crystal. Mech. Mater. **163**, 104070 (2021). https://doi.org/10.1016/j.mechmat.2021.104070
28. S. Ghosh, K. Bhattacharya, Influence of thermomechanical loads on the energetics of precipitation in magnesium aluminum alloys. Acta Mater. **193**, 28–39 (2020). https://doi.org/10.1016/j.actamat.2020.03.007
29. S. Ghosh, K. Bhattacharya, Spectral quadrature for the first principles study of crystal defects: application to magnesium. J. Comput. Phys. **456**, 111035 (2022). https://doi.org/10.1016/j.jcp.2022.111035
30. A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian data analysis, 2nd edn. Chapman & Hall. CRC Texts in Statistical Science (2004)
31. M.A. Beaumont, Approximate Bayesian computation. Annu. Rev. Stat. Appl. **6**, 379–403 (2019)
32. L. Wu, K. Zulueta, Z. Major, A. Arriaga, L. Noels, Bayesian inference of non-linear multiscale model parameters accelerated by a deep neural network. Comput. Methods Appl. Mech. Eng. **360**, 112693 (2020)
33. T. Most, in *Reliability and Optimization of Structural Systems*, ed. by D. Straub (CRC Press, London, 2010)
34. S. Madireddy, B. Sista, K. Vemaganti, A Bayesian approach to selecting hyperelastic constitutive models of soft tissue. Comput. Methods Appl. Mech. Eng. **291**, 102–122 (2015)
35. H. Rappel, L.A. Beex, J.S. Hale, L. Noels, S. Bordas, A tutorial on Bayesian inference to identify material parameters in solid mechanics. Arch. Comput. Methods Eng. **27**(2), 361–385 (2020)

36. A. Pandey, R. Pokharel, Machine learning based surrogate modeling approach for mapping crystal deformation in three dimensions. Scr. Mater. **193**, 1–5 (2021)
37. S. Pyrialakos, I. Kalogeris, G. Sotiropoulos, V. Papadopoulos, A neural network-aided Bayesian identification framework for multiscale modeling of nanocomposites. Comput. Methods Appl. Mech. Eng. **384**, 113937 (2021)
38. R. Alizadeh, J.K. Allen, F. Mistree, Managing computational complexity using surrogate models: a critical review. Res. Eng. Des. **31**(3), 275–298 (2020)
39. S.A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-temporal Domains* (Wiley, New York, 2013)
40. M. Sunnåker, A.G. Busetto, E. Numminen, J. Corander, M. Foll, C. Dessimoz, Approximate Bayesian computation. PLoS Comput. Biol. **9**(1), 1002803 (2013)
41. K. Csilléry, M.G. Blum, O.E. Gaggiotti, O. François, Approximate Bayesian computation (abc) in practice. Trends Ecolo. Evol. **25**(7), 410–418 (2010)
42. N. Guha, X. Tan, Multilevel approximate Bayesian approaches for flows in highly heterogeneous porous media and their applications. J. Comput. Appl. Math. **317**, 700–717 (2017)
43. M.A. Beaumont, W. Zhang, D.J. Balding, Approximate Bayesian computation in population genetics. Genetics **162**(4), 2025–2035 (2002)
44. J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys. Rev. Lett. **98**(14), 146401 (2007)
45. A.P. Bartók, M.C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys. Rev. Lett. **104**(13), 136403 (2010)
46. A.P. Bartók, R. Kondor, G. Csányi, On representing chemical environments. Phys. Rev. B **87**(18), 184115 (2013)
47. A. Grisafi, D.M. Wilkins, G. Csányi, M. Ceriotti, Symmetry-adapted machine learning for tensorial properties of atomistic systems. Phys. Rev. Lett. **120**(3), 036002 (2018)
48. A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, G.J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. J. Comput. Phys. **285**, 316–330 (2015)
49. I. Novoselov, A. Yanilkin, A. Shapeev, E. Podryabinkin, Moment tensor potentials as a promising tool to study diffusion processes. Comput. Mater. Sci. **164**, 46–56 (2019)
50. A.V. Shapeev, Moment tensor potentials: a class of systematically improvable interatomic potentials. Multiscale Model. Simul. **14**(3), 1153–1173 (2016)
51. R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials. Phys. Rev. B **99**(1), 014104 (2019)
52. R. Drautz, Atomic cluster expansion of scalar, vectorial, and tensorial properties including magnetism and charge transfer. Phys. Rev. B **102**(2), 024104 (2020)
53. G. Dusson, M. Bachmayr, G. Csanyi, R. Drautz, S. Etter, C. van der Oord, C. Ortner, Atomic cluster expansion: completeness, efficiency and stability. J. Comput. Phys. **454**, 110946 (2022)
54. Y. Lysogorskiy, C.V.D. Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner et al., Performant implementation of the atomic cluster expansion (pace) and application to copper and silicon. npj Comput. Mater. **7**(1), 1–12 (2021)
55. A. Seko, A. Togo, I. Tanaka, Group-theoretical high-order rotational invariants for structural representations: application to linearized machine learning interatomic potential. Phys. Rev. B **99**(21), 214108 (2019)
56. F. Musil, A. Grisafi, A.P. Bartók, C. Ortner, G. Csányi, M. Ceriotti, Physics-inspired structural representations for molecules and materials. Chem. Rev. **121**, 9759–9815 (2021)
57. S. Batzner, A. Musaelian, L. Sun, M. Geiger, J.P. Mailoa, M. Kornbluth, N. Molinari, T.E. Smidt, B. Kozinsky, E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nat. Commun. **13**(1), 2453 (2022). https://doi.org/10.1038/s41467-022-29939-5
58. B. Anderson, T.S. Hy, R. Kondor, Cormorant: covariant molecular neural networks. Adv. Neural Inf. Process. Syst. **32** (2019)
59. M. Haghighatlari, J. Li, X. Guan, O. Zhang, A. Das, C.J. Stein, F. Heidar-Zadeh, M. Liu, M. Head-Gordon, L. Bertels, et al., Newtonnet: a newtonian message passing network for deep learning of interatomic potentials and forces. arXiv:2108.02913 (2021)
60. B. Onat, C. Ortner, J.R. Kermode, Sensitivity and dimensionality of atomic environment representations used for machine learning interatomic potentials. J. Chem. Phys. **153**(14), 144106 (2020)
61. C. van der Oord, G. Dusson, G. Csányi, C. Ortner, Regularised atomic body-ordered permutation-invariant polynomials for the construction of interatomic potentials. Mach. Learn. Sci. Technol. **1**(1), 015004 (2020)
62. I. Batatia, D.P. Kovács, G.N. Simm, C. Ortner, G. Csányi, Mace: higher order equivariant message passing neural networks for fast and accurate force fields. arXiv:2206.07697 (2022)
63. J. Nigam, M.J. Willatt, M. Ceriotti, Equivariant representations for molecular Hamiltonians and n-center atomic-scale properties. J. Chem. Phys. **156**(1), 014115 (2022)
64. L. Zhang, B. Onat, G. Dusson, G. Anand, R.J. Maurer, C. Ortner, J.R. Kermode, Equivariant analytical mapping of first principles hamiltonians to accurate and transferable materials models. arXiv:2111.13736 (2021)
65. O. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. Smidt, K.-R. Müller, SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. Adv. Neural Inf. Process. Syst. **34** (2021)
66. V.H.A. Nguyen, A. Lunghi, Predicting tensorial molecular properties with equivariant machine-learning models. arXiv:2202.01449 (2022)
67. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087 (1953)
68. R. Swendsen, J.-S. Wang, Replica Monte Carlo simulation of spin-glasses. Phys. Rev. Lett. **57**, 2607–2609 (1986)
69. F. Wang, D.P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states. Phys. Rev. Lett. **86**(10), 2050–2053 (2001)
70. A.C.K. Farris, Y.W. Li, M. Eisenbach, Histogram-free multicanonical Monte Carlo sampling to calculate the density of states. Comput. Phys. Commun. **235**, 297–304 (2019)
71. M. Eisenbach, C.-G. Zhou, D.M. Nicholson, G. Brown, J. Larkin, T.C. Schulthess, A scalable method for ab initio computation of free energies in nanoscale systems. In: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. SC '09, pp. 64–1648. ACM, New York, NY, USA (2009)
72. M. Eisenbach, D.M. Nicholson, A. Rusanu, G. Brown, First principles calculation of finite temperature magnetism in Fe and Fe_3C. J. Appl. Phys. **109**(7), 07–138 (2011)
73. S.N. Khan, M. Eisenbach, Density-functional Monte-Carlo simulation of CuZn order-disorder transition. Phys. Rev. B **93**(2), 024203 (2016)
74. J.M. Sanchez, F. Ducastelle, D. Gratias, Generalized cluster description of multicomponent systems. Physica A **128**(1), 334–350 (1984). https://doi.org/10.1016/0378-4371(84)90096-7
75. C. Wolverton, A. Zunger, Ising-like description of structurally relaxed ordered and disordered alloys. Phys. Rev. Lett. **75**, 3162–3165 (1995). https://doi.org/10.1103/PhysRevLett.75.3162

76. Z. Pei, M. Eisenbach, S. Mu, G.M. Stocks, Error controlling of the combined cluster-expansion and Wang–Landau Monte-Carlo method and its application to FeCo. Comput. Phys. Commun. **235**, 95–101 (2019)
77. J. Zhang, X. Liu, S. Bi, J. Yin, G. Zhang, M. Eisenbach, Robust data-driven approach for predicting the configurational energy of high entropy alloys. Mater. Des. **185**, 108247 (2020)
78. R. Caruana, Multitask learning. Mach. Learn. **28**, 41–75 (1997)
79. M. Lupo Pasini, Y.W. Li, J. Yin, J. Zhang, K. Barros, M. Eisenbach, Fast and stable deep-learning predictions of material properties for solid solution alloys. J. Phys. Condens. Matter **33**(8), 084005 (2020)
80. X. Liu, J. Zhang, J. Yin, S. Bi, M. Eisenbach, Y. Wang, Monte Carlo simulation of order-disorder transition in refractory high entropy alloys: a data-driven approach. Comput. Mater. Sci. **187**, 110135 (2021)
81. T. Mueller, A. Hernandez, C. Wang, Machine learning for interatomic potential models. J. Chem. Phys. **152**(5), 050902 (2020)
82. L. Monostori, A. Márkus, H. Van Brussel, E. Westkämpfer, Machine learning approaches to manufacturing. CIRP Ann. **45**(2), 675–712 (1996)
83. H. Bhadeshia, R. Dimitriu, S. Forsik, J. Pak, J. Ryu, Performance of neural networks in materials science. Mater. Sci. Technol. **25**(4), 504–510 (2009)
84. M. Barnett, M. Senadeera, D. Fabijanic, K. Shamlaye, J. Joseph, S. Kada, S. Rana, S. Gupta, S. Venkatesh, A scrap-tolerant alloying concept based on high entropy alloys. Acta Mater. **200**, 735–744 (2020)
85. D. Weisz-Patrault, S. Sakout, A. Ehrlacher, Energetic upscaling strategy for grain growth. ii: probabilistic macroscopic model identified by Bayesian techniques. Acta Mater. **210**, 116805 (2021)
86. S.-G. Kim, S.-H. Shin, B. Hwang, Machine learning approach for prediction of hydrogen environment embrittlement in austenitic steels. J. Mater. Res. Technol. **19**, 2794–2798 (2022)
87. J. Jung, J.I. Yoon, H.K. Park, J.Y. Kim, H.S. Kim, Bayesian approach in predicting mechanical properties of materials: application to dual phase steels. Mater. Sci. Eng. A **743**, 382–390 (2019)
88. Z. Vangelatos, H.M. Sheikh, P.S. Marcus, C.P. Grigoropoulos, V.Z. Lopez, G. Flamourakis, M. Farsari, Strength through defects: a novel Bayesian approach for the optimization of architected materials. Sci. Adv. **7**(41), 2218 (2021)
89. P.-P. De Breuck, G. Hautier, G.-M. Rignanese, Materials property prediction for limited datasets enabled by feature selection and joint learning with MODNet. npj Comput. Mater. **7**(1), 1–8 (2021)
90. J.P. Darby, J.R. Kermode, G. Csányi, Compressing local atomic neighbourhood descriptors. arXiv:2112.13055 (2021)
91. L. Barroso-Luque, J.H. Yang, G. Ceder, Sparse expansions of multicomponent oxide configuration energy using coherency and redundancy. Phys. Rev. B **104**(22), 224203 (2021)
92. I. Kaliuzhnyi, C. Ortner, Optimal evaluation of symmetry-adapted *n*-correlations via recursive contraction of sparse symmetric tensors. arXiv:2202.04140 (2022)
93. K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I.E. Castelli, S.J. Clark, A. Dal Corso et al., Reproducibility in density functional theory calculations of solids. Science **351**(6280), 3000 (2016)
94. E.B. Tadmor, R.S. Elliott, J.P. Sethna, R.E. Miller, C.A. Becker, The potential of atomistic simulations and the knowledgebase of interatomic models. JOM **63**(7), 17 (2011)
95. D.M. Duffy, J.H. Harding, Simulation of organic monolayers as templates for the nucleation of calcite crystals. Langmuir **20**(18), 7630–7636 (2004)
96. C.L. Freeman, J.H. Harding, D.J. Cooke, J.A. Elliott, J.S. Lardge, D.M. Duffy, New forcefields for modeling biomineralization processes. J. Phys. Chem. C **111**(32), 11943–11951 (2007)