

Machine Learning Accelerated, High Throughput, Multi-Objective Optimization of Multiprincipal Element Alloys

Tian Guo, Lianping Wu, and Teng Li*

Multiprincipal element alloys (MPEAs) have gained surging interest due to their exceptional properties unprecedented in traditional alloys. However, identifying an MPEA with desired properties from a huge compositional space via a cost-effective design remains a grand challenge. To address this challenge, the authors present a highly efficient design strategy of MPEAs through a coherent integration of molecular dynamics (MD) simulation, machine learning (ML) algorithms, and genetic algorithm (GA). The ML model can be effectively trained from 54 MD simulations to predict the stiffness and critical resolved shear stress (CRSS) of CoNiCrFeMn alloys with a relative error of 2.77% and 2.17%, respectively, with a 12 600-fold reduction of computation time. Furthermore, by combining the highly efficient ML model and a multi-objective GA, one can predict 100 optimal compositions of CoNiCrFeMn alloys with simultaneous high stiffness and CRSS, as verified by 100 000 ML-accelerated predictions. The highly efficient and precise design strategy can be readily adapted to identify MPEAs of other principal elements and thus substantially accelerate the discovery of other high-performance MPEA materials.

1. Introduction

Multiprincipal element alloys (MPEAs) are an emerging class of materials with an array of highly desirable properties that are otherwise hard to achieve in conventional alloys,^[1–3] including wear resistance,^[4] corrosion resistance,^[5,6] a desirable balance between strength and ductility,^[7–10] and high-temperature stability.^[11–13] These remarkable properties have been attributed to high entropy, sluggish diffusion, and severe lattice distortion of MPEAs,^[14–17] which are in turn tied strongly to the compositions of MPEA. For example, it has been shown that the properties of MPEAs with non-equiatomic compositions are often superior to those with equiatomic or near equiatomic compositions.^[18] There exist abundant yet largely unexplored opportunities to achieve optimal material properties of MPEAs by fine-tuning their compositions.

However, designing MPEAs with desirable properties via tailoring compositions remains challenging, largely due to the vast multicompositional parameter space. Experimental

explorations have been mainly empirical and often use a trial-and-error approach far from comprehensive and efficient.^[19,20] The rational design of MPEAs through composition optimization is the key to achieving desirable material properties. In recent years, computational simulation approaches, such as thermodynamic modeling, density functional theory (DFT), and molecular dynamics (MD) simulations, have been used to investigate the phase stability, solidification behavior, and crystallization kinetics of MPEAs to guide the material design.^[21–23] For example, a calculated phase diagram method (e.g., CALPHAD) can accelerate the prediction of stable phases of MPEAs with solid solution phases.^[24] Fundamental insights on composition optimization of MPEAs aside, computational simulation approaches such as DFT and MD are often

time-consuming and cost-prohibitive. Thermodynamic modeling (e.g., CALPHAD) can only predict the phase diagrams of material.^[25,26] It remains a grand challenge to achieve a rational design of MPEAs in their huge multicompositional parameter space with high efficiency and accuracy.

Machine learning (ML) algorithms with multiple processing layers to enable data learning via multiple levels of abstraction have radically improved the state-of-the-art in image/speech recognition, autonomous driving, and drug discovery.^[27,28] Recently, ML has begun to be utilized in materials science research, for example, identifying structural flow defects in disordered solids,^[29] modeling and designing composite materials,^[30–32] discovering inorganic–organic hybrid materials,^[33] and predicting the new stable structure of quaternary Heusler compounds.^[34] For MPEA design, ML has been used to predict the formation of a single-phase solid solution^[35] and investigate the phase selection of MPEAs.^[36] Recently, designing a MPEA with a larger hardness by combining an ML surrogate model with experimental measurement of hardness is also proposed.^[37] However, the supervised ML models used in these studies require a large amount of training data (e.g., at least 155 experimental data^[38]) for model training, severely limiting their scalability to new material compositions due to annotation cost. Moreover, existing ML models can only perform single performance optimization (e.g., hardness). At the same time, engineering material design often strives for a balance of multiple material properties (e.g., being both strong and stiff).

Aiming to address the above challenges of the rational design of MPEAs, we develop a property-orientated MPEA

T. Guo, L. Wu, T. Li
Department of Mechanical Engineering
University of Maryland
College Park, MD 20742, USA
E-mail: lit@umd.edu

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/smll.202102972>.

DOI: 10.1002/smll.202102972

design strategy that coherently integrates MD simulation, few-shot ML, and genetic algorithm (GA) to enable simultaneous optimization of multiple performance indices at high efficiency. As a demonstration of such a strategy, we optimize the stiffness and critical resolved shear stress (CRSS) of CoNiCrFeMn alloys, a representative MPEA.^[39,40] We show that, with the results from MD simulations of CoNiCrFeMn alloys with only 54 sampling compositions and 17 material and structure descriptors of these alloys, our few-shot ML model can be trained to predict the stiffness and CRSS of CoNiCrFeMn alloys with exceptional precision (with a relative error of 2.77% and 2.17%, respectively). Our design strategy is also highly efficient, leading to a 12 600-fold reduction of computation time in comparison with pure MD simulations. Finally, based on the trained ML model and a nondominated sorting GA II (NSGA II),^[41] we obtain the optimal compositions of 100 CoNiCrFeMn alloys with simultaneous high stiffness and high CRSS, which are then validated by ML-accelerated computations of 100 000 randomly selected CoNiCrFeMn compositions. The rational design strategy integrating MD, ML, and GA reported here is facile, ultrafast, low-cost, and universal, which can be readily applicable to other MPEA material systems, and thus holds promise to accelerate the discovery of new MPEAs with high performance in multiple properties at a low cost.

2. Results and Discussion

2.1. Deformation Mechanism of Multiprincipal Element Alloys

Here, we choose CoNiCrFeMn in a face-centered cubic crystalline structure as a model material of MPEA and create 54 different composition profiles of CoNiCrFeMn alloys for the predictions of Young's modulus and CRSS, respectively (Tables S1 and S2, Supporting Information). Next, we carry out grand canonical Monte Carlo (CMC) simulation (see Experimental Section for CMC details) to build the atomic configuration of the CoNiCrFeMn alloy of a given composition profile, as illustrated in **Figure 1A**. Then, we perform MD simulations of the CoNiCrFeMn alloys under uniaxial compression to determine its Young's modulus and under shear loading to determine its CRSS (see Experimental Section for MD details). Results from the MD simulations of the CoNiCrFeMn alloys serve as the training data set for the ML model to be described in detail in a later section.

Our MD simulations capture the deformation process of CoNiCrFeMn alloys under uniaxial compression and shear loading and reveal that dislocation slip and twinning are the main deformation mechanisms in both loading conditions. **Figure 1D,E** plots the stress–strain curve and the corresponding deformation snapshots of the CoNiCrFeMn alloy obtained from uniaxial compression and shear loading, respectively. **Figure 1D** shows that the stress in the alloy first increases monotonically as the applied compressive strain increases. Meanwhile, severe lattice distortions or transient atomic “shuffling” lead to more and more atoms departing from equilibrium positions. Then, the stress drops abruptly at a critical strain, which indicates the onset of yielding, featuring the formation of a partial dislocation after a sufficient number of shuffling atoms form in the initially defect-free single crystalline alloy. The emission

of a partial dislocation is accompanied by the formation of a stacking fault. Then, the resultant slip of the partial dislocation on the stacking fault leads to forming a twin boundary (**Figure 1D**). The Young's modulus is defined as the slope of the linear portion of the corresponding compressive stress–strain curve between strains of 0% and 1.4%. The shear loading deformation of the alloy also experiences the emission of dislocation upon yielding, followed by twinning formation upon further straining. The CRSS of the alloy is measured by the peak shear stress upon yielding.

Figure 1F,G plots Young's modulus and CRSS of the 54 CoNiCrFeMn alloys. The alloy with equiatomic composition (i.e., Co₂₀Ni₂₀Cr₂₀Fe₂₀Mn₂₀) with the highest configuration entropy does not have the highest Young's modulus and the CRSS, in comparison with other alloys with non-equiatomic compositions. Even more evident is the strong dependence of the mechanical properties of the CoNiCrFeMn alloys on their compositions. For example, Young's modulus of the alloy decreases dramatically with decreasing concentration of Mn, Cr, and Fe. By contrast, the decreasing concentration of Co and Ni corresponds to increasing Young's modulus. As for CRSS, increasing the concentration of Ni leads to a higher CRSS, while increasing the concentration of Cr, Fe, and Mn causes the decrease of CRSS. The influence of Co on CRSS depends on its concentration: the decrease of Co concentration reduces the CRSS when the Co mole ratio is greater than 70% but increases the CRSS when its mole ratio is less than 70%. The CRSS decrease when the Co mole ratio is greater than 70% is because the unstable stacking energy decrease when the Co mole ratio is greater than 70% as shown in **Figures S1 and S2**, Supporting Information. The first peak along the GFE curves is the unstable stacking energy (USFE), which is the energy barrier for the nucleation of a leading dislocation partial. The lower unstable stacking energy, the lower energy barrier for the nucleation of a leading dislocation, which leads to the lower CRSS.^[42] Such a dependence of mechanical properties of the CoNiCrFeMn alloys on their compositions illustrated in **Figure 1F,G** are by no means comprehensive, given the vast compositional parameter space of the CoNiCrFeMn alloys, yet epitomizes the enormous intricacy of identifying the optimal composition profile of the alloy with desirable mechanical properties. However, as to be shown in the next section, these MD simulation results are sufficient to train the ML model, which in turn can predict the composition–property relation of the CoNiCrFeMn alloys with significant accuracy.

The above MD simulations suggest a strong dependence of mechanical properties of CoNiCrFeMn alloys on their elemental composition and phase structure. To facilitate the prediction of the composition–property relation of the CoNiCrFeMn alloys, we identify 17 feature descriptors as defined in detail in **Table S3**, Supporting Information, and Supporting Information. These descriptors can be categorized into three groups, including six descriptors related to the intrinsic properties that influence the formation of a solid solution, amorphous phase, and intermetallic compound in alloys, three descriptors related to the thermodynamic stability of alloys, and nine descriptors related to the element composition with a focus on lattice distortion, elastic misfit, and dislocation motion based on strengthening theory. Note that all 17 descriptors can be readily

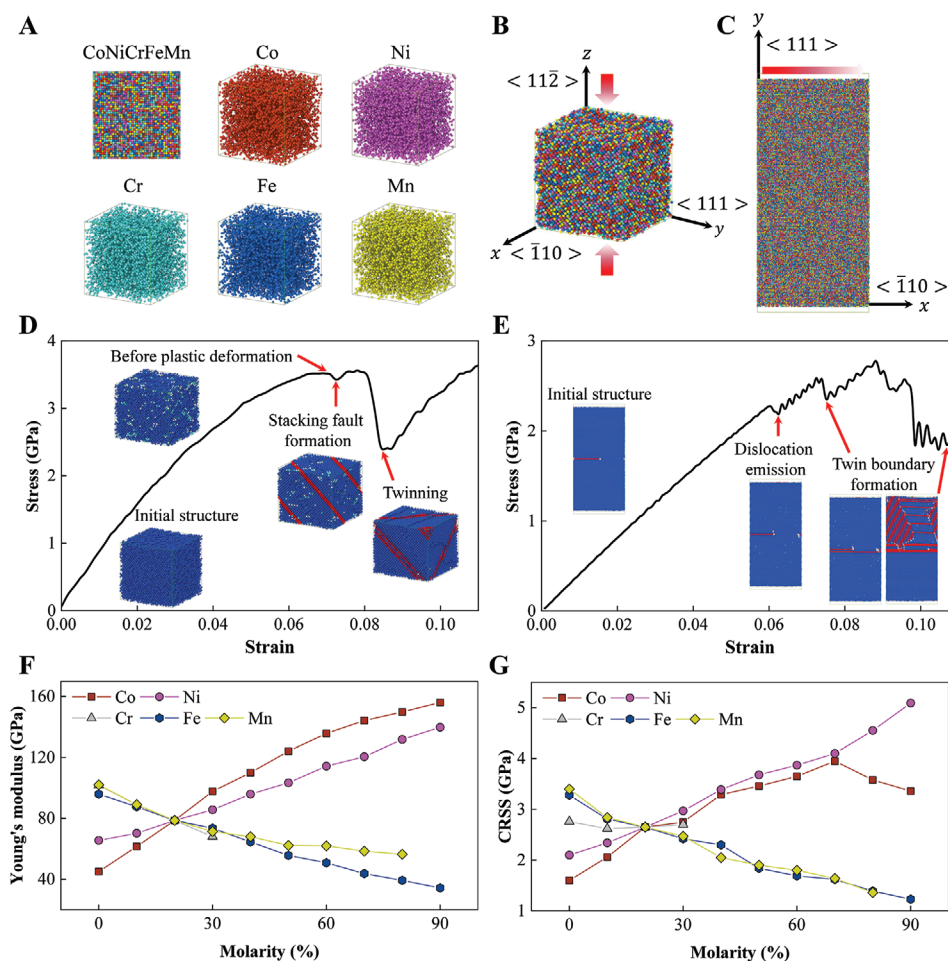


Figure 1. Molecular dynamics simulations of the deformation process of CoNiCrFeMn alloys. A) Atomic configuration and spatial distribution of Co, Ni, Cr, Fe, and Mn atoms in a $\text{Co}_{20}\text{Ni}_{20}\text{Cr}_{20}\text{Fe}_{20}\text{Mn}_{20}$ alloy, obtained by a Monte Carlo simulation. B,C) Schematic of the uniaxial compressive simulation model (B) and the shear loading simulation model (C) of $\text{Co}_{20}\text{Ni}_{20}\text{Cr}_{20}\text{Fe}_{20}\text{Mn}_{20}$ alloy. D,E) Simulated compressive (D) and shear (E) stress–strain curves of $\text{Co}_{20}\text{Ni}_{20}\text{Cr}_{20}\text{Fe}_{20}\text{Mn}_{20}$ alloy and its corresponding deformation snapshots. The atoms colored in blue, green, red, and white correspond to face-centered cubic structure, body-centered cubic structure, hexagonal closest packed structure, and other crystalline structures, respectively. F) Young's modulus of 54 CoNiCrFeMn alloys defined in Table S1, Supporting Information, and G) CRSS of the 54 CoNiCrFeMn alloys defined in Table S2, Supporting Information.

computed from available data.^[43] The values of the 17 descriptors of the 54 CoNiCrFeMn alloys are listed in Tables S4 and S5, Supporting Information.

2.2. Few-Shot Learning Algorithm

We establish the ML model based on a few-shot learning (FSL) algorithm^[44] to constitute the mapping relationship between the 17 feature descriptors and target mechanical properties (i.e., Young's modulus [E] and CRSS). Figure S3, Supporting Information, describes the computational flow and mechanism of the FSL algorithm using image classification as an example, with the aim to attribute the query data image to one of three categories: butterfly, horse, and turtle. A support set consists of one representative image for each of the three categories.

The network consists of three modules (i.e., a data pairing module, a mapping module, and a comparison module) and a classifier. The function of the data pairing module is to use one

query data and the data from the support set to produce three sample pairs. Sample pairs will be the input of the mapping module, which extracts abstract features by multiple convolutional layers and produces three sets of image features for the next module. The comparison module evaluates the matching score of abstract features of the image features. It produces a set of feature maps to represent the similarity between the query data and the support data. The classifier at the end of the network uses the feature maps obtained by the comparison module to make a logic regression to categorize the query data. The categorized result is determined by the category with the highest matching score.

2.3. Machine Learning-Based Prediction of Composition–Property Relation of CoNiCrFeMn Alloys

We next apply this FSL algorithm, which learns to compare the molarity change in compositions, to constitute the relationship

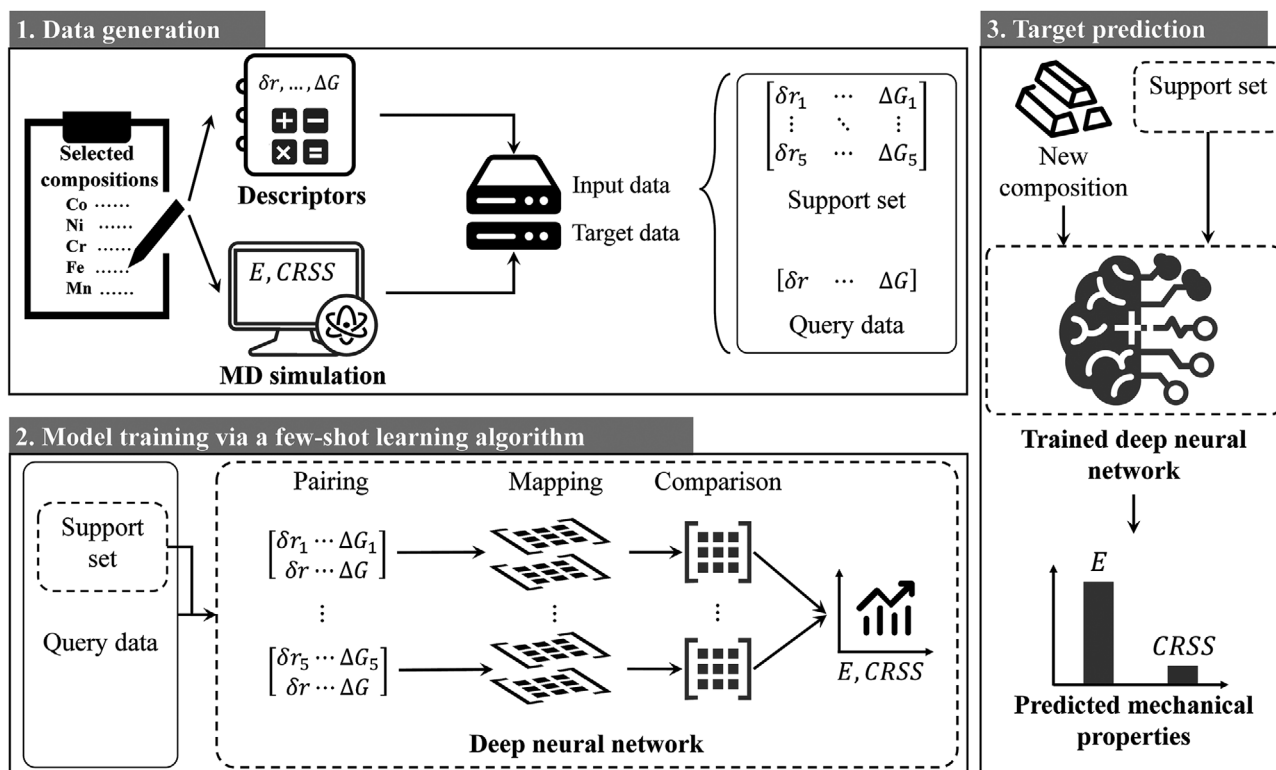


Figure 2. ML model accelerates the mechanical prediction of MPEA material properties. The diagram describing the training strategy for ML models involving data generation and deep neural network framework and prediction workflow.

between compositions and target properties. Two ML models are established to predict E and CRSS, respectively. **Figure 2** illustrates the implementation of the ML-based prediction, which contains three steps:

Step 1. Data generation: The training data is from the selected compositions of CoNiCrFeMn alloys and their MD results. The input data of the ML model for Young's modulus and CRSS are screened by a descriptor screening process.^[45] We first select the descriptors with an absolute value of Pearson correlation coefficient (see Supporting Information) greater than 0.2. Then, recursive elimination and exhaustive screening (see Figure S4, Supporting Information) are employed to identify key descriptors that most significantly affect the Young's modulus and CRSS of MPEA, respectively (Tables S8 and S9, Supporting Information). Finally, we build two training datasets (including inputs and targets data) for ML model training. In each training dataset, the maximum molarities of each metallic element in a composition are selected to build the support set. In addition, we randomly generate ten compositions of CoNiCrFeMn alloys (Table S6 and S7, Supporting Information) and calculate their descriptors and mechanical properties using the same approach as the compositions of CoNiCrFeMn alloys in the training data set. Results from these ten CoNiCrFeMn alloys with random compositions serve as the testing data for ML model evaluation.

Step 2. Models training via a FSL algorithm: The proposed ML model in this study is composed of multiple processing layers to extract the features of input data with multiple levels of abstraction in each processing layer (Figure 2). The model architecture is similar to that described in Figure S3, Supporting Information,

including a pairing module, a mapping module, a comparison module, and a classifier. The detailed network architecture and parameters of the ML model are further illustrated in Figure S5, Supporting Information. ML model training is carried out for two mechanical properties (i.e., E and CRSS) independently. For each model training, we set the learning rate to 0.001 and stop the training process in the 300th epoch, as the fluctuation of loss function value in the testing dataset tends to be stable (Figure S6, Supporting Information). For the two trained ML models, the relative prediction error in the testing dataset for E and CRSS is 2.77% and 2.17%, respectively (further detailed in Figure 3A,B). The exceptional precision of ML prediction for E and CRSS using a limited number of training data verifies the high effectiveness of the FSL algorithm.

Step 3. Target prediction: The ML models trained in Step 2 are used to predict the mechanical properties of CoNiCrFeMn alloys with new compositions. When a new composition is queried, its corresponding descriptors and the support set built in Step 1 are fed into the trained ML model. Then, the corresponding mechanical property can be obtained with a high prediction precision and at an extremely low computational cost.

2.4. High Accuracy and Efficiency of the Trained Machine Learning Models in Property Prediction

Figure 3A,B compares the MD simulation results and the predictions of the best performing ML model on E and CRSS of the ten CoNiCrFeMn alloys in the testing data set (Tables S6

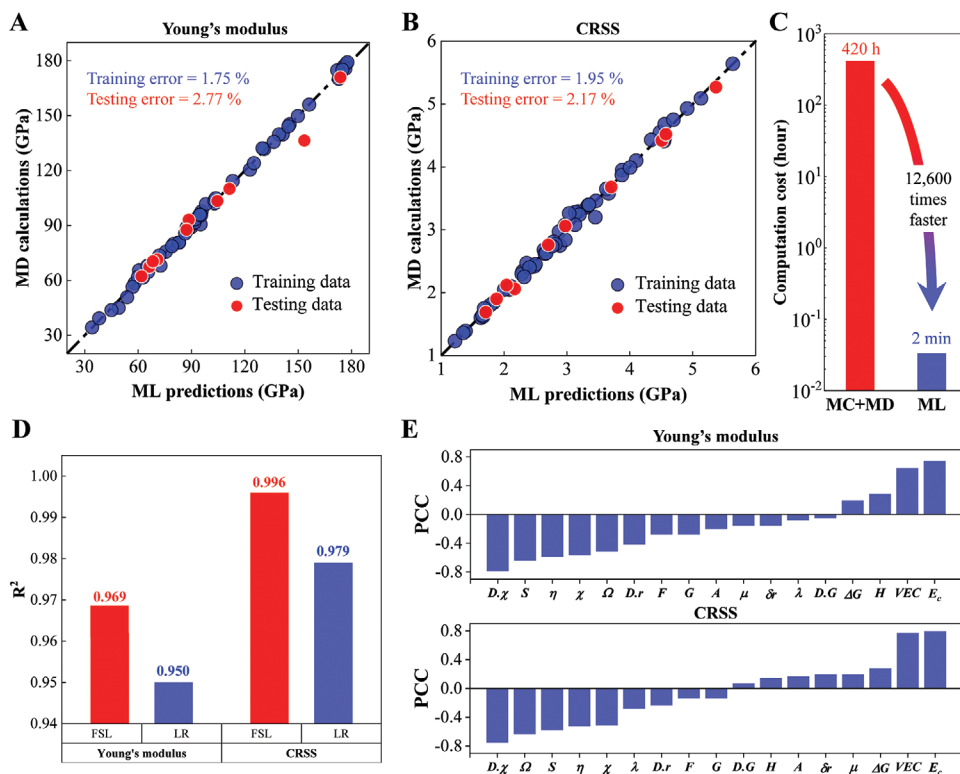


Figure 3. Accuracy, efficiency, and goodness of fit of ML-based prediction of mechanical properties of CoNiCrFeMn alloys. A,B) Comparison between the MD simulation results and the predictions of the best performing trained ML model on E and CRSS of the ten CoNiCrFeMn alloys in the testing dataset, showing a high prediction accuracy of the trained ML models. C) High efficiency of ML model prediction of mechanical properties of CoNiCrFeMn alloys, with a 12 600-fold acceleration from the calculation using MD simulations. D) Comparison of the goodness of fit between the proposed FSL model and LR model, with the same training/testing data. E) Pearson correlation coefficients between each descriptor (17 in total) and mechanical properties of CoNiCrFeMn alloys (E and CRSS), with the value of 1 meaning perfect positive correlation, -1 meaning perfect negative correlation, and 0 meaning no correlation between a given descriptor and a given mechanical property.

and S7, Supporting Information). We define the mean relative error (e) and the coefficient of determination (R^2) to evaluate the accuracies and the goodness of fit of the trained ML models, as follow:

$$e = \frac{1}{10} \sum_{i=1}^{10} \left(\frac{y_i^{\text{MD}} - y_i^{\text{ML}}}{y_i^{\text{MD}}} \right) \times 100\% \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{10} (y_i^{\text{ML}} - \bar{y}^{\text{ML}})(y_i^{\text{MD}} - \bar{y}^{\text{MD}})}{\sqrt{\sum_{i=1}^{10} (y_i^{\text{ML}} - \bar{y}^{\text{ML}})^2} \sqrt{\sum_{i=1}^{10} (y_i^{\text{MD}} - \bar{y}^{\text{MD}})^2}} \quad (2)$$

where y_i^{MD} and y_i^{ML} represent the mechanical properties of the i th composition obtained by MD simulation and ML prediction, respectively, with the summation over the ten compositions in the testing dataset (see Tables S4 and S5, Supporting Information, for details). The best performing ML models can predict mechanical properties of CoNiCrFeMn alloys with high accuracy (goodness of fit), with a mean relative error and a coefficient of determination of 2.77% and 0.969 for E and 2.17% and 0.996 for CRSS, respectively. We further compare the performances of the proposed FSL with a linear regression (LR) model, which is trained with the same training/testing data as the FSL model (see Experimental Section for details) as shown

in Figure 3D. Our proposed FSL model has higher goodness of fit than that of the LR model for both E and CRSS. Furthermore, we compare our ML prediction results with experimental results. According to our ML prediction, the values of CRSS of the following MPEAs show a decreasing order of CoNiCr > CoNiCrMn > CoNiMn > CoNiFeMn > CoNi > Ni, which agrees with the experimental measurements of the yield stress of these MPEAs.^[46] Most alloys measured in experiments are polycrystalline, while the material modeled in our MD simulation is single crystalline. The Young's modulus computed in our simulations is in the [100] directions, which is smaller than that of the single crystal in other crystalline directions.^[47] As a result, the calculated Young's modulus in our study is shown to be overall smaller than the experimental measurements. Nevertheless, we find that the order of Young's modulus of a set of alloys in [100] direction predicted by our MD simulations (MnFeNi < CrFeNi < MnCoNi < CrCoNi) agrees well with that of polycrystalline alloys of the same compositions measured in experiments,^[48] which suggests the validity of our MD simulation results. Such an agreement also suggests that the ML-predicted CRSS is a good indicator of the yield stress of the corresponding MPEA.

Figure 3E plots the Pearson correlation coefficients (with the value of 1 meaning perfect positive correlation, -1 meaning perfect negative correlation, and 0 meaning no correlation)

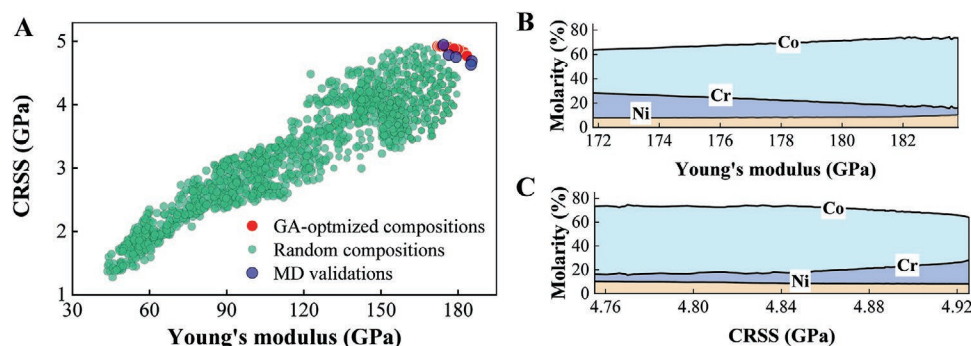


Figure 4. ML-based multi-objective optimization of CoNiCrFeMn alloys. A) Distribution of the resulting Young's modulus and CRSS of the Pareto optimal compositions (red dots), which are validated by MD simulations (blue dots). Light green dots plot the ML predicted mechanical properties of CoNiCrFeMn alloys with random composition, clearly demonstrating the high effectiveness and efficiency of the GA-optimized compositions (For visual clarity, only 2000 out of the 100 000 randomly selected cases are plotted in this figure). B,C) The variation of the molarity of each element along the high Young's modulus line (B) and the high CRSS line (C).

between the mechanical properties of CoNiCrFeMn alloys (i.e., E and CRSS) and the 17 descriptors of the alloys, respectively. These coefficients are based on the MD simulation results from the training dataset (as listed in Tables S8 and S9, Supporting Information). For the prediction of E and CRSS using the best performing trained ML models, the descriptors with an absolute value of Pearson correlation coefficient greater than 0.2 and screened by recursive elimination and exhaustive screening, are selected as the inputs of the ML models.

Besides the high accuracy of the prediction of mechanical properties of CoNiCrFeMn alloys, it is worth noting that our ML model can drastically accelerate the property prediction with high efficiency. For comparison, we time the central processing unit hours (CPU-h) to evaluate the computation costs of MD-based simulations and ML-based predictions of the mechanical properties. For MD-based simulations, the calculation of one stress-strain curve for a given CoNiCrFeMn alloy composition is performed using large-scale atomic molecular massively parallel simulator (LAMMPS) on a 40-core supercomputer (Intel E5-2680, 2.80 GHz), which on average takes ≈ 10.5 CPU-h, equivalent to ≈ 420 CPU-h on a single-core CPU. By contrast, ML-based predictions take only ≈ 2 min, starting from ML model training to model testing and targeting property prediction on a single-core CPU. In other words, the ML-based prediction of the mechanical properties is more than 12 600 times faster than the MD-based simulation. Such a remarkable reduction of computational cost by ML-accelerated material discovery approach effectively overcomes the challenge of low efficiency of MD simulations in identifying high-performance CoNiCrFeMn alloys.

2.5. Machine Learning-Based Multi-Objective Optimization of CoNiCrFeMn Alloys

To search for CoNiCrFeMn alloys with high stiffness and CRSS, we use the NSGA II^[41] to identify optimal compositions. The workflow is illustrated in Figure S7, Supporting Information. We randomly generate 6000 CoNiCrFeMn alloy compositions as the initial data and feed the initial data into the trained ML model to predict their E and CRSS. We

identify all nondominated data (defined as that no other data exceeds its E and CRSS simultaneously) in the initial data and divide them into different levels using the fast nondominated sorting approach. From the highest to lowest levels, we select the data in turn until the size of the population reaches 100, as illustrated in Figure S7, Supporting Information. In these 100 compositions, we randomly swap their molarities of elements (i.e., crossover processing) to create a new set of compositions (6000 in total), named offspring compositions. Then, the molarities of elements in each offspring composition are altered randomly (i.e., mutation processing) to obtain the mutation compositions, which inherit the superior mechanical properties of their previous generation compositions. The obtained 6000 mutation compositions serve as the initial data for the next loop of property optimization. The results of the optimization in each generation are recorded. The above optimization loop is repeated until the variations in both the E and the CRSS between two adjacent generations are less than 0.01 GPa. The GA-optimization algorithm is implemented by a python package Pymoo.^[49]

The 100 GA-optimized compositions, defined as the Pareto front (as listed in Table S10, Supporting Information), can guide the experimental design of CoNiCrFeMn alloys to achieve both high strength and high stiffness. All 100 optimal alloy compositions are considered equally good; that is, no composition has both a higher E and a higher CRSS than those of all 100 optimal alloy compositions. Figure 4A plots the predicted E and CRSS of the 100 optimal compositions, with the red dots defining the high stiffness and the high CRSS compositions. Figure 4B,C plots the variation of the element molarity of the 100 optimal compositions along the high stiffness line and the high CRSS line, respectively. It is shown that both high stiffness and high CRSS of CoNiCrFeMn alloys are essentially dominated by the Co, Cr, and Ni concentration, which agrees with recent experimental results.^[50] According to the experimental measurement, mechanical properties (e.g., the yield strength, ultimate tensile, and elongation) of the CoNiCr MPEA are superior to those of the CoNiCrFeMn MPEA. For the same grain size, the true tensile stress and resolved shear stress for the onset of twinning in the CoNiCr MPEA are determined to be higher than those of CoNiCrFeMn MPEA.^[51]

To validate the effectiveness of the GA-optimized compositions,^[52] we generate 100 000 compositions of CoNiCrFeMn alloys randomly and predict their E and CRSS via the trained ML models. As shown in Figure 4A, it is evident that the mechanical properties of all 100 000 compositions appear within the Pareto front, and only very few compositions have mechanical properties that approach the Pareto front. These high throughput ML predictions sufficiently validate the high effectiveness and demonstrate the high efficiency of the GA-optimized compositions. As another evidence of the effectiveness of the GA-optimized composition prediction, we perform MD simulations of CoNiCrFeMn alloys under uniaxial compression and shear loading, with 5 out of the 100 GA-optimized compositions (see Table S11, Supporting Information). The MD-predicted E and CRSS are plotted in Figure 4A, showing an exceptional agreement with ML-predicted values (with a mean relative error of 2.73% and 1.52% for E and CRSS, respectively) further demonstrating the effectiveness of GA optimization and ML prediction.

The quantitative relationships between the molarity of the element and key mechanical properties of CoNiCrFeMn MPEA, as shown in Figure 4B,C, can be used as design guidelines for pursuing MPEA with desirable mechanical properties. The research approach of ML accelerated, high throughput, multi-objective optimization of CoNiCrFeMn MPEA demonstrated in the present study can be readily applicable to enable the discovery of high performance MPEA of other compositions. We therefore call for further explorations of these fertile opportunities.

3. Experimental Section

Monte Carlo and Molecular Dynamics Simulation: The Monte Carlo (MC) and MD simulations described in the main text were performed using modified embedded atom method potential,^[53] as implemented in the LAMMPS.^[54] The potential was adequate for simulations of the mechanical behavior of a solid solution alloy. All microstructures and deformation mechanisms were examined with the adaptive common neighbor analysis^[55,56] and visualized using OVITO.^[57] Before starting the computation of any specific materials properties, the authors performed a MC simulation for 500 000 MC steps at 1200 K to obtain the equilibrium atomic configuration of alloy. Periodic boundary conditions were applied to all three coordinate directions, and atomic vibration and change in sample dimensions were allowed during the simulation.

Uniaxial Compression Simulations: A schematic representation of an FCC Co₂₀Ni₂₀Cr₂₀Fe₂₀Mn₂₀ MPEA for the uniaxial compression model is shown in Figure 1B. All simulated MPEA were in the shape of a cube with edge lengths $l_x = l_y = l_z = 7.2$ nm, in the three orthogonal directions $x = [100]$, $y = [010]$, and $z = [001]$, respectively. The total number of atoms in each simulation cell was 32 000. Periodic boundary conditions were employed in the three coordinate directions. After atoms in each simulation cell were initialized, the simulation cell was energetically minimized by the conjugate gradient algorithm to obtain a stable single-phase structure. A constant NPT ensemble^[58] was used to relax the samples under zero pressure at 300 K for 10 ps. Finally, the alloys were subjected to uniaxial compression in the z -direction with the system temperature maintained at 300 K and constant pressure kept in other directions. An acceptable constant strain rate $1 \times 10^8 \text{ s}^{-1}$ in MD modeling and a constant integration time step of 0.001 ps were applied during compression. Statistically volume-averaged stress obtained from the Virial theorem was calculated at every 0.001 strain increment to track the strain–stress responses of alloys.

Shear Loading Simulations: To calculate the CRSS of the alloy, the simulation cell shown in Figure 1C was used, with edge lengths of 18.5, 37.8, and 1.8 nm in the three orthogonal crystallographic directions, $x = [\bar{1}10]$, $y = [111]$, and $z = [1\bar{1}2]$, respectively. The total number of atoms in each simulation cell was 103 380. The authors used the method proposed by Osetsky and Bacon,^[59] inserting a single straight dislocation in a perfect fcc crystal. Simple shear loading was achieved by displacing the top 10 Å thick slab in the x -direction while the bottom slab was fixed. The shear strain step and strain rate were 3×10^{-5} and $\approx 10^8 \text{ s}^{-1}$, respectively. The MD simulations were performed using an NVT ensemble with a Nose–Hoover thermostat at 300 K. Periodic boundary conditions were employed in all three coordinate directions.

Machine Learning Models Training: The FSL model in this study had a layer-by-layer construction, which contained multiple convolutional, max pooling, and full connection layers. The detailed construction and parameters are plotted in Figure S2, Supporting Information. The weights and bias in each hidden layer (all layers but input/output layers) were initialized randomly with a normal distribution. The authors implemented their FSL models in PyTorch.^[60] The training process was performed using the Adam algorithm^[61] with a fixed learning rate of 0.001. Mean squared error was used in each training process as the loss function and relative deviation as the evaluation function. The training was stopped in the 200th epoch to avoid model overfitting because the value of relative deviation in the testing dataset did not decline anymore. The LR model and the support vector machine model were implemented by Scikit-learn. The same training/testing data were used for both FSL and LR models. All source code, trained models, and MD simulation results are available in https://github.com/TienGwo/HEA_prediction.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

T.G. and L.W. contributed equally to this work. The authors acknowledge the University of Maryland supercomputing resources (<http://hpcc.umd.edu>) and Maryland Advanced Research Computing Center (MARCC) made available for conducting the research reported in this paper.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that supports the findings of this study are available in the Supporting Information of this article.

Keywords

machine learning, molecular dynamic simulations, multi-objective optimization, multiprincipal element alloys

Received: May 24, 2021

Revised: August 7, 2021

Published online: September 15, 2021

- [1] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, S.-Y. Chang, *Adv. Eng. Mater.* **2004**, 6, 299.
- [2] X. Zhang, J. Yao, B. Liu, J. Yan, L. Lu, Y. Li, H. Gao, X. Li, *Nano Lett.* **2018**, 18, 4247.
- [3] E. P. George, D. Raabe, R. O. Ritchie, *Nat. Rev. Mater.* **2019**, 4, 515.
- [4] M. G. Poletti, G. Fiore, F. Gili, D. Mangherini, L. Battezzati, *Mater. Des.* **2017**, 115, 247.
- [5] J. Ding, A. Inoue, Y. Han, F. L. Kong, S. L. Zhu, Z. Wang, E. Shalaan, F. Al-Marzouki, *J. Alloys Compd.* **2017**, 696, 345.
- [6] S. S. Nene, M. Frank, K. Liu, S. Sinha, R. S. Mishra, B. A. McWilliams, K. C. Cho, *Scr. Mater.* **2019**, 166, 168.
- [7] B. Schuh, F. Mendez-Martin, B. Volker, E. P. George, H. Clemens, R. Pippan, A. Hohenwarter, *Acta Mater.* **2015**, 96, 258.
- [8] Y. P. Lu, X. Z. Gao, L. Jiang, Z. N. Chen, T. M. Wang, J. C. Jie, H. J. Kang, Y. B. Zhang, S. Guo, H. H. Ruan, Y. H. Zhao, Z. Q. Cao, T. J. Li, *Acta Mater.* **2017**, 124, 143.
- [9] X. Y. Li, L. Lu, J. G. Li, X. Zhang, H. J. Gao, *Nat. Rev. Mater.* **2020**, 5, 706.
- [10] S. Y. Peng, Y. J. Wei, H. J. Gao, *Proc. Natl. Acad. Sci. USA* **2020**, 117, 5204.
- [11] A. Karati, K. Guruvidyathri, V. S. Hariharan, B. S. Murty, *Scr. Mater.* **2019**, 162, 465.
- [12] Y. J. Wei, Y. Q. Li, L. C. Zhu, Y. Liu, X. Q. Lei, G. Wang, Y. X. Wu, Z. L. Mi, J. B. Liu, H. T. Wang, H. J. Gao, *Nat. Commun.* **2014**, 5, 8.
- [13] Y. Zou, J. M. Wheeler, H. Ma, P. Okle, R. Spolenak, *Nano Lett.* **2017**, 17, 1569.
- [14] K. Y. Tsai, M. H. Tsai, J. W. Yeh, *Acta Mater.* **2013**, 61, 4887.
- [15] Y. J. Liang, L. J. Wang, Y. R. Wen, B. Y. Cheng, Q. L. Wu, T. Q. Cao, Q. Xiao, Y. F. Xue, G. Sha, Y. D. Wang, Y. Ren, X. Y. Li, L. Wang, F. C. Wang, H. N. A. Cai, *Nat. Commun.* **2018**, 9, 8.
- [16] E. Ma, *Scr. Mater.* **2020**, 181, 127.
- [17] S. Y. Liu, Y. J. Wei, *Extreme Mech. Lett.* **2017**, 11, 84.
- [18] M. J. Yao, K. G. Pradeep, C. C. Tasan, D. Raabe, *Scr. Mater.* **2014**, 72–73, 5.
- [19] Q. Ding, Y. Zhang, X. Chen, X. Fu, D. Chen, S. Chen, L. Gu, F. Wei, H. Bei, Y. Gao, M. Wen, J. Li, Z. Zhang, T. Zhu, R. O. Ritchie, Q. Yu, *Nature* **2019**, 574, 223.
- [20] K. Huang, Z. B. Zhao, H. F. Du, P. Du, H. Wang, R. Y. Wang, S. Lin, H. H. Wei, Y. Z. Long, M. Lei, W. Guo, H. Wu, A. C. S. Sustain, *Chem. Eng.* **2020**, 8, 6905.
- [21] D. C. Ma, B. Grabowski, F. Kormann, J. Neugebauer, D. Raabe, *Acta Mater.* **2015**, 100, 90.
- [22] C. Jiang, B. P. Uberuaga, *Phys. Rev. Lett.* **2016**, 116, 5.
- [23] Y. Lederer, C. Toher, K. S. Vecchio, S. Curtarolo, *Acta Mater.* **2018**, 159, 364.
- [24] O. N. Senkov, J. D. Miller, D. B. Miracle, C. Woodward, *Nat. Commun.* **2015**, 6, 10.
- [25] G. B. Olson, C. J. Kuehmann, *Scr. Mater.* **2014**, 70, 25.
- [26] R. Machaka, *Comp. Mater. Sci.* **2021**, 188, 110244.
- [27] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, 521, 436.
- [28] J. Schmidhuber, *Neural Networks* **2015**, 61, 85.
- [29] E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, A. J. Liu, *Phys. Rev. Lett.* **2015**, 114, 5.
- [30] C. T. Chen, G. X. Gu, *MRS Commun.* **2019**, 9, 556.
- [31] G. X. Gu, C. T. Chen, M. J. Buehler, *Extreme Mech. Lett.* **2018**, 18, 19.
- [32] G. X. Gu, C. T. Chen, D. J. Richmond, M. J. Buehler, *Mater. Horiz.* **2018**, 5, 939.
- [33] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Molloy, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature* **2016**, 533, 73.
- [34] K. Kyoungdoc, L. Ward, H. Jiangang, A. Krishna, A. Agrawal, C. Wolverton, *Phys. Rev. Mater.* **2018**, 2, 123801.
- [35] F. Tancret, I. Toda-Caraballo, E. Menou, P. Diaz-Del-Castillo, *Mater. Des.* **2017**, 115, 486.
- [36] N. Islam, W. J. Huang, H. L. L. Zhuang, *Comp. Mater. Sci.* **2018**, 150, 230.
- [37] J. M. Rickman, H. M. Chan, M. P. Harmer, J. A. Smeltzer, C. J. Marvel, A. Roy, G. Balasubramanian, *Nat. Commun.* **2019**, 10, 10.
- [38] C. Wen, Y. Zhang, C. X. Wang, D. Z. Xue, Y. Bai, S. Antonov, L. H. Dai, T. Lookman, Y. J. Su, *Acta Mater.* **2019**, 170, 109.
- [39] M. H. Tsai, J. W. Yeh, *Mater. Res. Lett.* **2014**, 2, 107.
- [40] H. Wang, D. K. Chen, X. H. An, Y. Zhang, S. J. Sun, Y. Z. Tian, Z. F. Zhang, A. G. Wang, J. Q. Liu, M. Song, S. P. Ringer, T. Zhu, X. Z. Liao, *Sci. Adv.* **2021**, 7, 9.
- [41] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, *IEEE Trans. Evol. Comput.* **2002**, 6, 182.
- [42] W. Li, H. Fan, J. Tang, Q. Wang, X. Zhang, J. A. El-Awady, *Mater. Sci. Eng., A* **2019**, 763, 138143.
- [43] A. M. Halpern, *J. Chem. Educ.* **2012**, 89, 592.
- [44] F. Sung, Y. X. Yang, L. Zhang, T. Xiang, P. H. S. Torr, T. M. Hospedales, arXiv:1711.06025, **2018**.
- [45] H. Zhang, H. Fu, X. He, C. Wang, L. Jiang, L.-Q. Chen, J. Xie, *Acta Mater.* **2020**, 200, 803.
- [46] Z. Wu, H. Bei, G. M. Pharr, E. P. George, *Acta Mater.* **2014**, 81, 428.
- [47] H. Zhang, X. Sun, S. Lu, Z. Dong, X. Ding, Y. Wang, L. Vitos, *Acta Mater.* **2018**, 155, 12.
- [48] G. Laplanche, P. Gadaud, C. Barsch, K. Demtroder, C. Reinhart, J. Schreuer, E. George, *J. Alloy Compd.* **2018**, 746, 244.
- [49] J. Blank, K. Deb, *IEEE Access* **2020**, 8, 89497.
- [50] G. Laplanche, A. Kostka, C. Reinhart, J. Hunfeld, G. Eggeler, E. P. George, *Acta Mater.* **2017**, 128, 292.
- [51] G. Laplanche, A. Kostka, O. M. Horst, G. Eggeler, E. P. George, *Acta Mater.* **2016**, 118, 152.
- [52] G. Liu, G. J. Zhang, F. Jiang, X. D. Ding, Y. J. Sun, J. Sun, E. Ma, *Nat. Mater.* **2013**, 12, 344.
- [53] W.-M. Choi, Y. H. Jo, S. S. Sohn, S. Lee, B.-J. Lee, *npj Comput. Mater.* **2018**, 4, 1.
- [54] S. Plimpton, *J. Comput. Phys.* **1995**, 117, 1.
- [55] D. Faken, H. Jonsson, *Comp. Mater. Sci.* **1994**, 2, 279.
- [56] H. Tsuzuki, P. S. Branicio, J. P. Rino, *Comput. Phys. Commun.* **2007**, 177, 518.
- [57] A. Stukowski, *Modell. Simul. Mater. Sci. Eng.* **2010**, 18, 7.
- [58] W. G. Hoover, *Phys. Rev. A* **1986**, 34, 2499.
- [59] Y. N. Osetsky, D. J. Bacon, *Modell. Simul. Mater. Sci. Eng.* **2003**, 11, 427.
- [60] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. M. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. J. Bai, S. Chintala, in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, (Eds: H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, R. Garnett) Curran Associates, Inc., New York **2019**, p. 8024.
- [61] D. P. Kingma, J. Ba, arXiv:1412.6980, **2014**.