A Appendix / supplemental material

A.1 Model hyperparameters

We use the same model hyperparameters for CDVAE as in Xie et al. [2022]. The encoder is DimeNet++ with 128 hidden channels and 4 interaction blocks. The decoder is GemNet-dT, a variant of the model designed for direct force prediction. Fully connected submodules are composed of 2 hidden layers, 256 channels each. The model has 5.1 million parameters in total.

A.2 Execution time of experiments

MD simulations were conducted using LAMMPS on an Intel i7 11800-H CPU. The creation of data with local search involved optimizing a total of 183 structures, and conducting 400 relaxations and bulk modulus computations. The wall time of the procedure was 16 hours and 3 minutes. The autoencoder optimization was done on an NVIDIA A100 GPU. Sampling involved Langevin dynamics with noise being annealed throughout 50 steps. We increased the number of model inferences on each step to 200 (compared to 100 in Xie et al. [2022]). The initial run on P-CDVAE involved sampling at 7 different breakpoints at varying intervals, 2 of which were disregarded. Sampling of a full batch (128 structures). Structures generated during optimization 7 hours and 22 minutes. Relaxation of those structures with LAMMPS, including those far from the relaxed state and therefore disregarded in the analysis, took 6 hours and 34 minutes.

A.3 Generation task

Fig.1.(a-b) show the CDVAE and P-CDVAE performance in denoising the newly generated materials into the correct crystal phases, evaluated with the Ovito package - black lines indicate the correct boundary between BCC and FCC phases, as described in the literature [Wróbel et al., 2015]. Qualitatively, the P-CDVAE model performs better in terms of correct crystal phase, which is in correspondence with the reconstruction scores in the section 4.1. We show the distribution of the 500 generated configurations against the points in the dataset in Figure.A.3.(c), which shows that the vast part of the ternary phase diagram is covered. Afterwards, we did spin-polarized DFT calculations, and based on the formation energy results, we found great agreement with the cluster expansion method (CE) in terms of the stability of the materials for NiFeCr structures [Wróbel et al., 2015]. This benchmarks the method designed in this work promising for the study of higher than three component alloys.

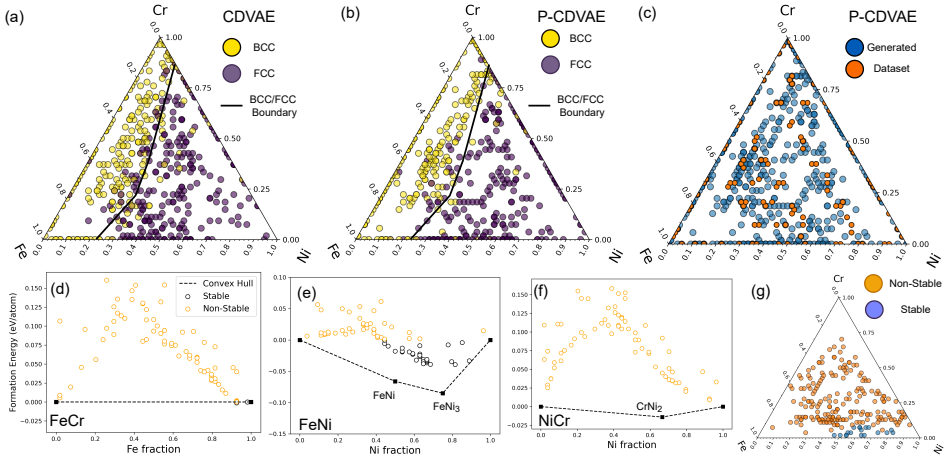

Figure 1: Material generation performance. **(a-b)** summarise both CDVAE and P-CDVAE models' ability to denoise newly generated structures to the *ground truth* CP, determined by CALPHAD calculations shown in Figure 2. **(c)** demonstrates the distribution of the newly generated materials w.r.t the points covered in the initial dataset. In **(d-e)** the stable materials, validated in DFT, are shown for all the binary materials, and the corresponding ternary plot is presented in **(g)**.

A.4 Optimization task

Table 1: The fraction of structures generated in optimization whose bulk modulus falls within the top 10, 20 or 30% of values present in the training data. For NiFeCr-Random, the optimized are compared to the entire train set. For NiFeCr-LS, only the portion of the training data crafted with local search is considered for comparison.

MODEL	TOP 10%	TOP 20%	TOP 30%
NIFECR-RANDOM	22%	32%	50%
NIFECR-20	59%	65%	68%

A.5 Reconstruction scores with Behler-Parrinello vectors

To cover the random ordering of the atoms at a specific point in the composition space, we need to choose relatively large examples that have the freedom to form arbitrary permutations of all the atoms. However, we have observed that perfect reconstruction of the entire structure is unlikely. Therefore, we use Behler-Parinello (BP) [Behler and Parrinello, 2007] vectors, which are embeddings of local geometry for each atom , describing its local environment, to measure the similarity between the ground truth and the reconstructed configurations. In this work, PANNA: Properties from Artificial Neural Network Architectures [Lot et al., 2020], is used as an external package for calculation of the modified version of Behler–Parrinello (mBP) descriptors [Smith et al., 2017]. The mBP representation generates a fixed-size vector, G[s], the "G-vector", for each atom in each configuration of the dataset. Each G-vector describes the environment of the corresponding atom of the configuration to which it belongs, up to a cutoff radius R_c. In terms of the distances R_{ij} and R_{ik} from the atom i to its neighbors j and k and the angles between the neighbors θ_{jik}. The radial and angular G-vectors are given by:

$$G_i^{rad}[s] = \sum_{i \neq j} e^{-\eta(R_{ij}-R_s)^2} f_c(R_{ij}) \tag{1}$$

$$\begin{aligned} G_i^{ang}[s] = 2^{1-\zeta} \sum_{j,k \neq i} &[1 + cos(\theta_{ijk} - \theta_s)]^\zeta \\ &\times e^{-\eta[\frac{1}{2}(R_{ij}+R_{ik})-R_s]^2} f_c(R_{ij}) f_c(R_{ik}) \end{aligned} \tag{2}$$

where the smooth cutoff function (which includes the cutoff radius R_c) is given by:

$$f_c(R_{ij}) = \begin{cases} \frac{1}{2}\left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1\right], & R_{ij} \leq R_c \\ 0, & R_{ij} > R_c \end{cases} \tag{3}$$

and η, ζ, θ_s and R_s are parameters, different for the radial and angular parts. The choice of the cutoff value is made so that it covers up to three nearest neighbours of the center atom in the BCC NiFeCr, which has an average lattice constant of $a = 3$ Å, and thus the third nearest neighbour's distance is $a \times \sqrt{2} = 4.24$ Å. This automatically covers up to the 4th nearest neighbour of the FCC crystal. We use Euclidean distance between G-vectors of two atoms as a measure of their dissimilarity. To assess the overall dissimilarity of two structures, we compute the minimum assignment between the G-vectors of their respective atoms. Then the average of the distances between all atom pairs in such assignment is taken, and we refer to that value as "G-distance" between a pair of structures.

A.6 Dataset creation

MD specification. We create NiFeCr structures in a range of elementary compositions, validated to be stable based on modified embedded-atom Model (MEAM) interatomic potentials developed by Wu et al. [2017] and Lee et al. [2001]. The mentioned MEAM potentials were validated by experimental data and CALPHAD (Calculation of Phase Diagrams) calculations (for detailed information on the potentials see the references). Apart from the specific compositions studied in Wu et al. [2017], we added very small noise to the compositions of stable structures (less than 2 atomic percent), such that the final trained model has seen more data. The dataset creation workflow is presented as a diagram

in Fig. SM.2 and is available as in (`https://github.com/Amirhossein4131/AlloGen.git`). The dataset creation starts by creating a crystal matrix (either BCC or FCC structure), based on the composition. The initial matrix is selected based on the composition of the atoms as shown in Fig. SM.2(b). Afterwards, the target structure with the specific composition is created by random substitution of second and third atom types with the first atom type in the matrix.

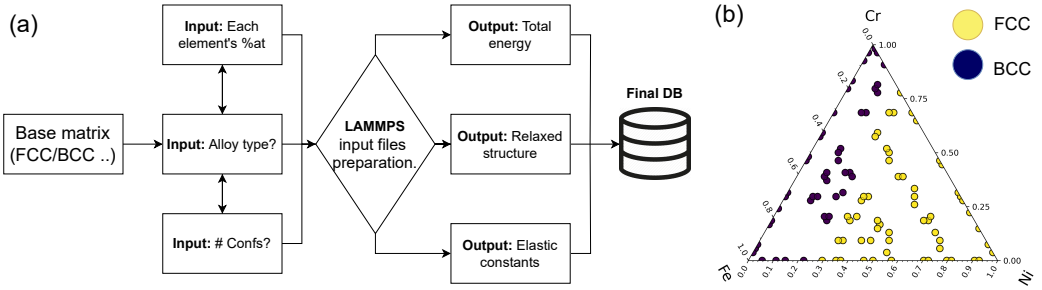

Figure 2: **(a)** Dataset creation workflow. **(b)** Training dataset created based on CALPHAD calculations and experimental data [Wu et al., 2017]

For each elementary composition, 20 random configurations of a given composition are added to the dataset. `LAMMPS` software [Thompson et al., 2022] is used for the MD calculations, as well as the calculation of the formation energies and the bulk modulus. The mentioned properties are calculated after conjugate gradient (CG) relaxation of the system at temperature T=0K.

Dataset augmentation with local search

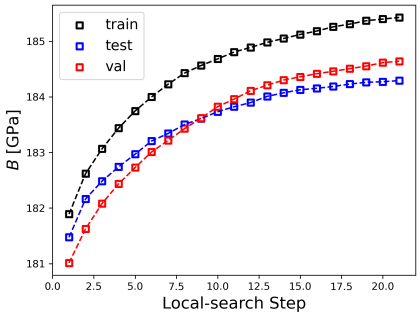

Figure 3: The changes of the mean bulk modulus of the respective portions of train, validation, and test set during local search.

To augment the dataset with optimized property materials, 10% of examples from NiFeCr-Random's training, validation, and test examples were randomly chosen, optimized with the local search method explained in the main text, and added to their respective partitions of the data. Fig. SM.3 shows the property optimizations versus local search iterations.

DFT validation. The DFT spin-polarized calculations were performed with the `Vasp` package [Kresse and Furthmüller, 1996, Kresse and Hafner, 1993, Kresse and Furthmüller, 1996], using PAW PBE exchange-correlation functional [Blöchl, 1994, Kresse and Joubert, 1999] and an intial magnetic moment of NIONS$\times$5.0 (NIONS is the number of atoms in the cell). The Brillouin zone was sampled using a maximum $kpoint$ spacing of $0.5\ ^{-1}$ on Γ-centred Monkhorst-Pack grids [Monkhorst and Pack, 1976] and plane-wave cutoff energy of $520\ eV$. Smearing with the spreading of $0.05\ eV$ was introduced within the Methfessel-Paxton method [Methfessel and Paxton, 1989] to help convergence.