Semi-supervised generative approach to point-defect formation in chemically disordered compounds: application to uranium-plutonium mixed oxides

Maciej J. Karcz,†1,2,* Luca Messina,†1,‡ Eiji Kawasaki,‡ Serenah Bathellier,† Didier Ballerier,† and Emeric Bourasseau†

1CEA, DES, IRESNE, DEC, Cadarache, F-13108 Saint-Paul-Lez-Durance, France
2Université Paris-Saclay, CEA, List, F-91120, Palaiseau France

(Dated: November 23, 2022)

Machine-learning methods are nowadays of common use in the field of material science. For example, they can aid in optimizing the physicochemical properties of new materials, or help in the characterization of highly complex chemical compounds. An especially challenging problem arises in the modeling of chemically disordered solid solutions, for which some properties depend on the distribution of chemical species in the crystal lattice. This is the case of defect properties of uranium-plutonium mixed oxides nuclear fuels. The number of possible configurations is so high that the problem becomes intractable if treated with direct sampling. We thus propose a machine learning approach, based on generative modeling, to optimize the exploration of this large configuration space. A probabilistic, semi-supervised approach using Mixture Density Network is applied to estimate the concentration of thermal defects in (U,Pu)O₂. We show that this method, based on the prediction of the density of states of formation energy of a defect, is computationally much more cost-efficient compared to other approaches available in the literature.

I. INTRODUCTION

Multi-component solid solutions have raised much interest due to the great versatility of their applications and the possibilities of optimizing their parameters. By altering the atomic composition it is possible to tune the desired properties and increase, e.g., their irradiation resistance [1, 2], or to modify their elastic or electronic [3–5] properties. However, the atomic-scale study of such compounds can become especially challenging for those characterized by chemical disorder [6]. Since different types of atoms occupy the same crystal lattice, they can be distributed in many different ways. This is particularly noticeable in the field of high-entropy alloys (HEA) [7–9]. For the properties that depend on the distribution of chemical species in the crystal lattice, such as the properties of point defects [10], the number of possible configurations is so high, that the exhaustive exploration of this configuration space becomes extremely computationally expensive. The same challenge arises in the modeling of chemically disordered mixed oxides (MOX) nuclear fuels. For instance, in the case of (U,Pu)O₂ mixed oxides, since U and Pu are occupying the same crystal lattice, there is a large number of possible configurations which makes the study of this compound difficult with current atomic-scale simulation methods [11, 12].

A very common approach to the calculation of properties of chemically disordered solid solutions is the use of special quasirandom structures (SQS) [13]. Using SQS entails that the most disordered structures are expected to be the most probable ones, which is true for ideally disordered solutions, i.e., solutions with negligible mixing enthalpy. However, if the solution is not perfectly ideal, the most disordered structures are not necessarily the most probable ones. Even in the case of perfectly disordered solutions, SQS provides only a few structures, not allowing for the exploration of the whole configuration space. The insufficient sampling of the partition function prevents one from obtaining correct ensemble averages, which makes this approach not very reliable. Chemical disorder can also be approached using Markov Chain Monte Carlo (MCMC) techniques. For instance, in the work of Takoukam-Takoundjou et al. [13], MCMC was applied to compute the average thermodynamic properties of (U,Pu)O₂ over a large number of sampled configurations. However, due to the large size of the configuration space, ensuring the ergodicity of the Markov process of the MCMC calculations requires a very high amount of computational resources [14]. Additionally, while it was possible to perform such a study with an interatomic potential [11], more reliable \textit{ab initio} methods such as density functional theory (DFT) [12] are too computationally expensive to perform such extensive sampling. Another approach is to attempt a systematic exploration of the configuration space starting from a small system and progressively increasing its size, as shown in the study of Bathellier et al. [15] on the defect formation energy in (U,Pu)O₂ using CRG potential [11]. Nevertheless, such a method quickly becomes impractical, due to the exponentially increasing computational cost. In Bathellier’s work, systematic exploration was done up to the second cationic nearest neighbors around the defect: a system consisting of 18 cations. If this approach were to be extended to a system of 42 cations instead of 18, it would take approximately 21 trillion hours of CPU time. Because of that, the systematic exploration was limited only to the first and second cationic nearest neighbors around the defect.

Previous works in the literature on the properties of

* maciej.karcz@cea.fr
1 luca.messina@cea.fr
chemically disordered solid solutions were restricted to
the study of either very few configurations or relatively
small systems. However, an alternative approach that
could go beyond those limitations can be proposed with
machine learning (ML). ML has spread out as an inter-
disciplinary field in various domains. It has been proven
as a useful tool in constructing ML potentials [16–18], ac-
celerating the modeling of battery materials [19, 20], aug-
menting the exploration of material properties [21, 22],
among many other applications. ML has also been ap-
plied to the study of stability of chemically disordered
compounds. In the work of Yaghoobi and Alaei [23], en-
ergy computations of bulk crystalline structures of five
HEAs were performed using kernel ridge regression [24]
and neural network, trained on DFT data. However, such
an approach, as with more conventional methods, is lim-
ited by the size of the configuration space and may not
be suitable for systems for which access to data is more
computationally expensive. This is the case of (U,Pu)O2
where DFT calculations must use the Hubbard correction
(DFT+U) to take into account the strong correlations
among the U$^{4+}$ and Pu$^{4+}$ 5f electrons [25]. Moreover,
this cost would have to be paid again, whenever such a
study was to be extended to other defects and properties,
or for every change of interatomic potential or chosen ab
initio method.

Exploration of the configuration space requires a
method that is sufficiently efficient so it can be easily
applied to changing environments and different prop-
ties. In the search for an approach that could allow for
such exploration, we will focus in this work on the bound
Schottky defects (BSD) properties in (U,Pu)O$_2$; namely
the calculation of their formation energy and equilibrium
concentration. A schematic representation of the three
types of BSD defects in (U,Pu)O$_2$ fluorite structure is
shown in Fig. 1. In UO$_2$, positron annihilation spec-
troscopy measurements coupled with first-principles cal-
culations have shown that BSD is a predominant radia-
tion damage feature [26, 27] and also a favorable site for
fission gas trapping [28, 29]. This indicates that bound
Schottky defects are among the most stable defects in
uranium oxides. There is, however, very scarce data
about the properties of BSD or other point defects in
(U,Pu)O$_2$ in the literature. In recent years, there was
a systematic progression in the sampling methods of the
configuration space in MOX nuclear fuels [13, 15, 30].
However, the limited sampling of the configuration space,
that was possible to achieve so far, provided only a par-
tial insight into the defect formation energy in (U,Pu)O$_2$
and was insufficient to allow for the calculation of equi-
librium concentrations.

In this work we will face these challenges using ML
methods of generative type [31] - aiming to approxi-
mate the true, hidden probability distribution function
(PDF) of the desired properties. The advantage of gen-
erative methods over more common applications of ML
is the possible gain in computational time. In fact, by
choosing a family of surrogate distributions, or a network
architecture that resembles the characteristic of original
data, it is possible to obtain an accurate approximation
of the desired properties with a relatively small amount
of samples. There are several different types of generative
methods. For example, generative adversarial networks
(GAN) [32] allow for the construction of probability dis-
butions by framing an objective as a supervised learn-
ing problem with two competing sub-models. Other ex-
amples are variational autoencoders (VAE) [33], or deep
belief networks (DBN) [34], both allowing for the recon-
struction of complex distributions. Still, giving the model
too many degrees of freedom for such a task very often
results in the need for large training databases. This,
however, is not always necessary. In our study of BSD,
we consider the formation energy to be mainly influen-
ced by the atoms closest to the defect, as shown in Bathel-
lier’s work [15], and we treat the impact of the further
atoms on the energy calculations as a random noise that
we assume is normally distributed. Therefore, we model
the formation energy distribution of the BSD defect as
a Gaussian random variable [35], using Mixture Density
Network (MDN) [36] in a semi-supervised fashion, as we
explain in more detail in Sec. II D. We show that this
approach is very robust and cost-efficient, allowing for
a significant extension of the range of influence of local
atomic configurations, compared to the previous study
[15].

This paper is organized as follows. Mathematical for-
mulas for the concentration of thermal defects, calcula-
tion of formation energy, and used descriptors are given
in Sec. II. Description of used databases, the architecture
of the used MDN model, and the results of the calcula-
tions are in Sec. III.

II. METHODS

A. Concentration of thermal defects

The concentration of vacancy type of defects can be
expressed using Gibbs formation free energy $G_d(T)$:

![FIG. 1. Representation of three types of bound Schottky
defects in the fluorite structure of (U,Pu)O$_2$, with the oxygen
cation vacancies marked as V$_a$ or V$_c$ accordingly]
\[ C_d(T) = \left\langle \exp \left( -\frac{G_d(T)}{k_B T} \right) \right\rangle = \left\langle \exp \left( -\frac{(H_d - T S_d)}{k_B T} \right) \right\rangle, \tag{1} \]

so it becomes:

\[ C_d(T) = \left\langle \exp \left( \frac{S_d}{k_B} \exp \left( -\frac{H_d}{k_B T} \right) \right) \right\rangle. \tag{2} \]

Here \( H_d \) is the formation enthalpy, \( S_d \) the formation entropy, \( T \) the temperature, \( d \) a type of defect, and \( k_B \) the Boltzmann constant. The left part of the Eq. (2) \( \exp \left( \frac{S_d}{k_B} \right) \) is independent of temperature (in the harmonic approximation), where \( S_d \) can be computed from the phonon spectrum of the system. In this work, we will neglect \( S_d \), due to its second-order contribution to the calculations of \( C_d(T) \), and focus on the approximation of enthalpy \( H_d \). Let us define an atomic configuration \( x_c \in X \) as a given atomic supercell, where \( X \in \mathbb{R}^{3N} \) is the space of atomic configurations and \( c \) denotes the type of cation that was removed from the supercell to form the Schottky defect. From now on, \( H_d = E_d^J(x_c) \) and we will be referring to formation enthalpy as formation energy. The average defect concentration can be written as:

\[ C_d(T) = \left\langle \exp \left( -\frac{E_d^J(x_c)}{k_B T} \right) \right\rangle = \sum_{i=1}^{N} w(x_c^i) \exp \left( -\frac{E_d^J(x_c^i)}{k_B T} \right). \tag{3} \]

\( E_d^J(x_c) \) is the formation energy of a defect of type \( d \) in the configuration \( x_c \) and \( N \) denotes the number of configurations in \( X \). Here \( w(x_c) \) is the normalized weight that corresponds to the Gibbs-Boltzmann distribution and depends on the configuration energy \( E(x_c) \) as follows:

\[ w(x_c) = \frac{\exp \left( -\frac{E(x_c)}{k_B T} \right)}{Z}, \tag{4} \]

where \( Z \) is the partition function, calculated over all of the configurations \( N \) in the configuration space \( X \):

\[ Z = \sum_{j=1}^{N} \exp \left( -\frac{E(x_c^j)}{k_B T} \right). \tag{5} \]

As this work focuses on the study of bound Schottky defects, from this point on, in the following formulas, \( d \) will be replaced with BSD.

**B. Formation energy of bound Schottky defect**

Vacancy-type (here BSD) defect formation energy in concentrated disordered alloys [37, 38] can be calculated as:

\[ E_{\text{BSD}}^J(x_c) = E_{\text{BSD}}(x_c) - E(x_c) + \mu_c + 2\mu_O - k_B T \log(y_c), \tag{6} \]

where \( c \) can be either U or Pu, \( \mu \) is the chemical potential of U, Pu, or O, and \( y_c \) is the cation \( c \) concentration in the supercell, so \( y_U = 1 - y_{Pu} \). Here \( E(x_c) \) is the energy of a reference, bulk supercell, with the cation of type \( c \) in the place of the cation vacancy. \( E_{\text{BSD}}(x_c) \) is the energy of a supercell with the cation \( c \) and two oxygen atoms removed.

If we consider the case of an ideal solution, where local order can be neglected, the chemical potentials are given by [15, 37]:

\[ \begin{align*}
\mu_U^0 + 2\mu_O^0 &= \frac{E(\text{UO}_2)}{N_C} + k_B T \log(y_U), \\
\mu_{Pu}^0 + 2\mu_O^0 &= \frac{E(\text{PuO}_2)}{N_C} + k_B T \log(y_{Pu}).
\end{align*} \tag{7} \]

\( E(\text{UO}_2) \) and \( E(\text{PuO}_2) \) are the energies of the pure \( \text{UO}_2 \) and \( \text{PuO}_2 \) supercells respectively. \( \mu_U^0, \mu_{Pu}^0 \) and \( \mu_O^0 \) are the chemical potentials of U, Pu, and O in the perfectly disordered solution. By \( N_C \) is denoted the number of cations in the supercell. This formula is valid under the assumption that the effect of defects or short-range order on the chemical potentials is negligible. In other words, the system with or without defect is treated as an ideal solid solution [13, 39]. The formation energy can now be expressed in a simplified form [15]:

\[ \begin{align*}
E_{\text{BSD}}^J(x_U) &= E_{\text{BSD}}(x_U) - E(x_U) + \frac{E(\text{UO}_2)}{N_C}, \\
E_{\text{BSD}}^J(x_{Pu}) &= E_{\text{BSD}}(x_{Pu}) - E(x_{Pu}) + \frac{E(\text{PuO}_2)}{N_C}.
\end{align*} \tag{8, 9} \]

In principle, the difference between Eq. (8) and Eq. (9) comes from a choice of the energy reference, which, in the case of ideal solutions, should not affect the equilibrium defect concentration, given by Eq. (3). Therefore, choosing either one of the above formulas for \( E_{\text{BSD}}^J(x_U) \) or \( E_{\text{BSD}}^J(x_{Pu}) \) is equally valid and should give the same results from the perspective of \( C_{\text{BSD}}(T) \) calculation.

Since the chemical potentials are written under the assumption that local order can be neglected, an indication of its effect might be the difference between the calculations of Eq. (3) with \( E_{\text{BSD}}(x_U) \) or \( E_{\text{BSD}}(x_{Pu}) \) reference. In the Appendix A we show that calculations performed using both formulas yield similar results, indicating that local order has a limited effect in (U, Pu)O\(_2\) according to the used CRG potential. Therefore, all of the calculations in this work were done for the BSD vacancy type defect.
of defect under the assumption of an ideal solution, and using Eq. (8) to compute the BSD formation energy.

C. Calculation of concentration of thermal defects

Equation (3) can be rewritten as:

\[
C_{BSD}(T) = \frac{1}{N} \sum_{i=1}^{N} w(x^i_x)f(x^i_x),
\]

where \( f(x^i_x) = \exp \left( -\frac{E_{BSD}(x^i_c)}{k_B T} \right) \). By neglecting local order, we can make an assumption that \( w(x^i_x) \) is constant and equal to \( \frac{1}{N} \), similarly to what was done in Bathellier’s work [15]. In other words, we assume that every configuration has the same formation energy \( E(x^i_x) \), so we can write:

\[
C_{BSD}(T) = \frac{1}{N} \sum_{i=1}^{N} f(x^i_x).
\]

If we denote the probability distribution of configuration by \( p(x^i_x) \), we could use the Monte Carlo approximation to estimate the concentration of defects:

\[
C_{BSD}(T) = \mathbb{E}_{x^i_x \sim p(x^i_x)}[f(x^i_x)] \approx \frac{1}{M} \sum_{i=1}^{M} f(x^{(i)}_x),
\]

where every \( x^i_x \) has the same \( p(x^i_x) \) and \( M \) is the number of samples used for the approximation so \( M \neq N \) and in particular we can choose \( M \ll N \). Accurate calculation of \( C_{BSD}(T) \) requires many samples \( M \) of configurations \( x^i_x \), and for each \( x^i_x \), \( E_{BSD}(x^i_c) \) must be computed (e.g., by DFT or interatomic potentials), which makes this task very computationally demanding.

\( E_{BSD}(x^i_x) \) can be calculated for every \( x^i_x \) and for each configuration, there is exactly one formation energy. Since for the calculation of \( C_{BSD}(T) \) only resulting energies of \( E_{BSD}(x^i_c) \) are needed, we can rewrite Eq. (11) as an ensemble average directly of the formation energies:

\[
C_{BSD}(T) = \frac{1}{N} \sum_{i=1}^{N} f(x^{(i)}_x) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N'} f_E(E^{(i,j)}_x) \delta(E^{(i,j)}_x - E_{BSD}(x^i_x))
= \frac{1}{N} \sum_{i=1}^{N'} \left( \frac{1}{N} \sum_{i=1}^{N} \delta(E^{(i,j)}_x - E_{BSD}(x^i_x)) \right) f_E(E^{(i,j)}_x)
= \sum_{j=1}^{N'} g(E^{(i,j)}_x) f_E(E^{(i,j)}_x),
\]

where \( E^i_x \) is a value of the function \( E^i_{BSD}(x_c) \) for a given \( x_c \) and \( f_E(E^i_x) = \exp \left( -\frac{E^i_x}{k_B T} \right) \). While every configuration has just one formation energy, one energy can correspond to multiple configurations. Therefore \( g(E^{(i,j)}_x) \) expresses the density of states that describes the number of occurrences of a given energy \( E^i \) and \( N' \) is the number of energies \( E^i \), so \( N \geq N' \). Since \( g(E^{(i,j)}_x) \geq 0 \) \( \forall j \) and \( \sum_j g(E^{(i,j)}_x) = 1 \) we can write:

\[
C_{BSD}(T) = \mathbb{E}_{E^i \sim p(E^i)}[f_E(E^i_x)] \approx \frac{1}{M} \sum_{i=1}^{M} f_E(E^{(i)}_x),
\]

where \( p(E^i) = g(E^i) \), so:

\[
p(E^i) = \frac{1}{N} \sum_{i=1}^{N} \delta(E^i - E_{BSD}(x^i_x)).
\]

With the Monte Carlo approximation, instead of sampling \( p(x^i_x) \), \( C_{BSD}(T) \) could be obtained by sampling \( p(E^i) \). Since the real \( p(E^i) \) is not known, we try to estimate it with an ML model, with the goal of using a method that provides a good estimation of \( p(E^i) \) with much less data than what we would need, e.g., for a direct Monte Carlo simulation as in Eq. (12). In other words, this is a way to shift the cost of Monte Carlo approximation of \( C_{BSD}(T) \) with \( p(x^i_x) \), to the cost of \( p(E^i) \) estimation.

D. Semi-supervised estimation of probability distribution of formation energy

Probability distribution \( p(E^i) \) expresses what is the distribution of formation energy over all possible configurations. In this work, the latter will be expressed with descriptors as proposed in the previous work [15]. Each configuration is described by its descriptor \( k \), later also called “mode”, representing the atomic environment around the BSD type of defect. It consists of \( N^a_{Pu} \in \mathbb{N} \) (\( c \rightarrow \text{cation} \)) which is the varying Pu content in the first sphere of influence near the cation vacancy (0-12 Pu atoms) and \( N^a_{Pu} \in \mathbb{N} \) (\( a \rightarrow \text{anion} \)) - varying number of Pu atoms near oxygen vacancies (0-6 Pu atoms for BSD2 and BSD3, 0-5 for BSD1). Since \( N^a_{Pu} \) considers all atoms in the first sphere of influence (see Sec. III A), Pu atoms that are close to oxygen vacancies \( N^a_{Pu} \) are also counted in \( N^c_{Pu} \).

We can rewrite \( p(E^i) \) as:

\[
p(E^i) \approx \sum_k p(E^i|k')p(k').
\]

Here \( p(E^i|k) \) is the conditional probability of energy \( E^i \) given descriptor \( k \) and \( p(k) \) is its probability. Given the previously undertaken assumption that every configuration has the same probability, as in Eq. (11), we can
compute \( p(k) \) analytically by counting the total number of configurations included in each \( k \):

\[
p(k) = \frac{K}{P},
\]

where \( K \) is the number of configurations included in a descriptor \( k \), and \( P \) is the total number of configurations corresponding to all possible permutations of U and Pu in the first sphere of influence (1nn).

Estimation of probability distribution \( p(E^f) \), defined in Eq. (16), can be approached in a semi-supervised way. Thanks to the undertaken assumption that local order can be neglected, the unsupervised part of Eq. (16) is calculated analytically. However, the likelihood of supervised part \( p(E^f|k) \) is not known and has to be addressed differently. As it is not possible to calculate \( p(E^f) \) for the whole spectrum of configurations due to the high computational cost, we thus aim at computing the approximation \( p_\theta(E^f) \) of \( p(E^f) \):

\[
p_\theta(E^f) = \sum_i p_\theta(E^f|i) p(k^i).
\]

Therefore, the task of estimating \( p(E^f) \) comes down to the approximation \( p_\theta(E^f|k) \) of \( p(E^f|k) \) for each descriptor \( k \). As we are expecting the impact of atoms on the formation energy calculations to fade with distance and that the closest atoms have the biggest influence, we can decompose \( E^f \) into two main parts:

\[
E^f = E^f(N_{Pu}^c, N_{Pu}^a) + \epsilon = E^f(k) + \epsilon.
\]

By \( E^f(k) \) is denoted the impact on the formation energy calculations of \( E^f \) that comes from the closest 12 neighboring atoms to the defect. Here we assume \( \epsilon \) to be a random normally distributed noise coming from the more distant neighbors. In this example, one can consider \( E^f(k) \) as a factor that mostly influences the mean, and \( \epsilon \) the variance of calculated energies. We can partially observe this characteristic of the data using the databases provided by the previous work [15]. In Fig. (2) we can see the projection of the descriptor space, given the configurations of the 12 closest atoms to the BSD defect, where energies for each descriptor tend to cluster together creating visible “modes”.

Based on the assumption described in Eq. (19) we approximate \( p(E^f|k) \) using a Mixture Density Network. As the MDN model predicts a distribution that is a Gaussian mixture, then the approximation \( p_\theta(E^f|k) \) of \( p(E^f|k) \), can be expressed as:

\[
p_\theta(E^f|k) = \sum_j \pi_j \mathcal{N}(E^f|\mu_j, \sigma_j^2).
\]

Here \( j \) is the number of Gaussian distributions used in a mixture, \( \pi_j \) is the mixing coefficient of a mixture, \( \mu_j \) and \( \sigma_j^2 \) corresponds to the mean and variance accordingly of each component of the mixture. The component density can be described as follows:

\[
\mathcal{N}(E^f|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(E^f-\mu)^2}{2\sigma^2}}.
\]

To capture the similarity between \( p_\theta(E^f) \) and \( p(E^f) \), a standard approach using Kullback-Leibler divergence (KL) was applied [22, 40, 41]. A more detailed explanation is included in Appendix B.

III. RESULTS

A. Description of databases

In this work, atomic configurations are built as 2592 atoms supercells (864 cations and 1728 anions) of (U, Pu)O\(_2\) with 50\% Pu concentration. This corresponds to a 6 × 6 × 6 time replication of a primitive cell of a fluorite structure. The central fluorite cell with different types of BSD defect is shown in Fig. 1. All atomic configurations are built with the same reference supercell and they only differ from each other in their atomic compositions in the closest proximity of the defect. In this way, we are limiting the size of the configurations space only to the possible configurations of atoms closest to the defect. First, the reference supercell, without a defect, was built by a random distribution of Pu and U in the crystal lattice. Then, small ensembles of atoms centered around the defect were generated separately, called spheres of influence, and replaced with the atoms in the center of the reference supercell. Spheres differ from each other in their cationic composition and were generated with progressively increased size with the radius changing from the one including only the closest first nearest neighbor (1nn) atoms, to the one including up to the fourth nearest...
TABLE I. Summary of the number of configurations in the 1-4nn formation energy databases of different types of BSD defects. The different amount of produced configurations for BSD1 (1440) and BSD2 or BSD3 (1470) comes from the fact that the number of modes is different for BSD1 (48) than for BSD2 or BSD3 (49). This is the consequence of different placement of oxygen atoms in BSD1 which can have a maximum of 5 neighboring Pu atoms, contrary to 6 in BSD2 or BSD3 (see Fig. 1).

| Database | 1nn  | 2nn  | 3nn  | 4nn  |
|----------|------|------|------|------|
| BSD1     | 4096 | 1440 | 1440 | 1440 |
| BSD2     | 4096 | 1470 | 1470 | 1470 |
| BSD3     | 4096 | 22668| 1470 | 1470 |

neighbor (4nn) atoms. For example, the 1nn sphere of influence consists of 12 atoms, that correspond only to the first nearest neighbors, while 4nn sphere of influence consists of 54 atoms: 12 atoms from 1nn, 6 from 2nn, 24 from 3nn, and 12 from 4nn. In each sphere, 12 atoms from 1nn were selected in a way to cover all possible Pu permutations near cationic and oxygen vacancies, according to \( N_{Pu}^\alpha \) and \( N_{Pu}^\beta \) of a selected type of BSD defect. Atoms of 2-4nn were distributed randomly. In this way, the calculation of the formation energy is based on the energy predictions from the 1nn atoms, where 2-4nn atoms are treated as an additional noise, as explained previously in Eq. (19). In such a manner, 1-4nn databases of configurations were generated accordingly to the size of used spheres. The formation energies of BSD were obtained via classical molecular statics (or 0-K energy minimization) implemented in the LAMMPS code [42], with the use of the Cooper-Rushton-Grimes (CRG) potential [11].

The procedure for preparing the databases is the same as in Bathellier et al. [15]. With the exception of 1nn BSD1-3 and 2nn BSD3, they consist of 30 configurations for each permutation of 1nn Pu atoms. The summary is shown in Table I.

**B. MDN architecture and training**

A general structure of the used ML model is shown in Fig. 3. Due to the description of data given as in Eq. (19) it was possible to achieve a trade-off between training cost and model accuracy with a single-Gaussian architecture of the MDN model. It consists of 2 inputs in an input layer, one hidden layer consisting of 10 neurons, and 3 outputs in the output layer - mean \( \mu \), variance \( \sigma^2 \), and mixing coefficient \( \pi \): 63 parameters in total. Models with more hidden layers and predicting parameters for more Gaussian distributions were tested as well, but we did not observe a performance improvement. Also, smaller models require less data for the optimization of the free parameters. For the hidden layer, a hyperbolic tangent activation function is used.

According to Eq. (18), descriptor space should be sampled from \( p(k) \). Instead, we build training datasets from an equal amount of examples of randomly chosen configurations corresponding to each mode \( k \) (according to each unique pair of \( N_{Pu}^\alpha \) and \( N_{Pu}^\beta \)). Then the resulting predictions for each \( k \) are normalized with analytically calculated \( p(k) \), as explained in Procedure 1. We have found that the minimum number of configurations needed for MDN training is \( \approx 200 \) as shown in Appendix C. In addition, there was no significant improvement in the accuracy of MDN models trained on datasets containing more than \( \approx 1000 \) configurations. Even though the predictions of the MDN models are more accurate when trained on larger datasets, it does not necessarily entail much improvement in the \( C_{BSD}(T) \) computations. As we show in the next section, roughly 200 configurations can be sufficient for calculations of ensemble averages, as is the case of the \( C_{BSD}(T) \). However, a higher precision will be necessary for the examination of the range of influence of the local environment on the BSD formation energy. Therefore, for the next experiments, training datasets were built by drawing 4 or 20 samples at random, from each mode, which corresponds to 192 and 960 samples for BSD1 and 196 and 980 samples for BSD2 and BSD3.

**C. Prediction of formation energy and defect concentration calculations**

Models were trained with previously defined 1-4nn databases for three types of BSD defects. The MDN architecture used in this work produces a single Gaussian distribution for every entry point. As shown in Eq. (16), the density of states \( p(E) \) can be then constructed as a
sum of all Gaussian distributions predicted for a given input set. The difference between the weights of normal distributions in the Gaussian mixture comes from the analytically calculated \(p(k)\) part of Eq. (16). The whole procedure, together with the later described \(C_{BSD}(T)\) estimation, is summarized in Procedure 1.

**Procedure 1: \(C_{BSD}(T)\) estimation with MDN**

**Data:** \(D = \{(k, E^f_0)\}_{k=1}^{L}\)

1. Find the parameters \(\theta^*\) by training \(p_0(E^f)\) on database \(D\), e.g., as in Appendix B;
2. For each unique \(k\) in \(D\) do
   3. Evaluate \([\mu, \sigma^2, \pi]\) from trained MDN model;
   4. Calculate \(p(k)\) as in Eq. (17);
   5. \(V_{mix} \leftarrow [\mu, \sigma^2, \pi \ast p(k)]\);
6. end
7. Sample \(M\) energies \(E^f\) from Gaussian mixture \(V_{mix}\);
8. Monte Carlo approximate \(C_{BSD}(T)\) as in Eq. (14);

\(V_{mix}\) is a vector of Gaussian parameters \([\mu, \sigma^2, \pi \ast p(k)]\) corresponding to each of the descriptors \(k\). Note, that while \(\pi\) changes the amplitude between distinct Gaussian distributions in each vector \([\mu, \sigma^2, \pi \ast p(k)]\), \(p(k)\) changes the amplitudes between the different vectors. For the case when only 1 Gaussian distribution is generated per descriptor \(k\), \(\pi = 1\) and the amplitudes of each element of the mixture \(V_{mix}\) are equal to \(p(k)\).

To visualize MDN predictions, a comparison between distributions produced by different MDN models for configurations with \(N_{Pu} = 6\) is shown in Fig. 4. The decreasing height of visible “peaks” for each distribution comes from the increasing noise added by the 2-4nn spheres. As expected, further spheres of influence have a decreasingly lower impact on the formation energy calculation. Here, the 4nn atoms do not appear to have a great impact on the density of states in comparison to the 1-3 nn atoms. Hence the boundary, beyond which further atoms start to have negligible influence on \(E_{BSD}^f\), is 3nn sphere of influence, corresponding to a radius of 7 Å from the defect in the \((U, Pu)O_2\) fluorite structure, consistently with what was found in Bathellier’s work [15]. The same conclusions apply to the other types of BSD defects.

Next, the obtained probability distributions were used to compute the BSD concentration by Monte Carlo sampling, as in Eq. (14), for a range of temperatures between 500K and 2000K. The comparison is shown in Fig. 5. For the computations we have used the 1nn database of Bathellier et al. [15] of BSD3 formation energy consisting of 4096 configurations. Because the available 1nn database is complete (consisting of all possible configurations of U and Pu atoms in the first sphere of influence), it was possible to compare defect concentrations calculated from the MDN predictions with the direct calculation (DC) of \(C_{BSD}(T)\) with Eq. (11). Using only 1nn atoms is, however, a weak approximation of \(C_{BSD}(T)\) and was used only for the evaluation of the 1nn MDN model’s performance. Orange and green curves in Fig. 5 represent the Monte Carlo approximation of \(C_{BSD}(T)\), done with the probability distribution predicted by the MDN model trained on 196 and 980 configurations from the 1nn database accordingly. Based on the \(C_{BSD}(T)\) calculations, an effective formation energy \(E_{eff}^f\) can be calculated as the negative slope of each curve:

\[
C_{BSD}(T) = \exp \left( -\frac{E_{eff}^f}{k_BT} \right),
\]

\[
E_{eff}^f = -k_BT \log (C_{BSD}(T)).
\]
Table II. Concentration of defects per atomic site $C_{BSD}(T)$ for selected temperatures in $(U, Pu)O_2$ 50% Pu, and effective formation energy $E_{eff}$ calculated as a slope of $C_{BSD}(T)$ data in logarithmic scale (Eq. (22)). The name of the database indicates: the considered sphere of influence, source of energy samples, type of defect, and the number of configurations used for the computation of $C_{BSD}(T)$ using direct calculation (DC) or MDN approach. Values of the MDN predictions are given as a mean ± 2 times standard deviation, calculated from the predictions of 100 MDN models.

| Database  | 500K     | 1000K    | 1500K    | $E_{eff}$[eV] |
|----------|----------|----------|----------|---------------|
| 1nn, DC, BSD3, 4096 | $2.51 \times 10^{-48}$ | $6.84 \times 10^{-25}$ | $6.25 \times 10^{-17}$ | 4.70          |
| 1nn, MDN, BSD3, 980 | $(2.50 \pm 0.47) \times 10^{-48}$ | $(6.82 \pm 0.25) \times 10^{-25}$ | $(6.24 \pm 0.12) \times 10^{-17}$ | $7.0 \pm 0.01$ |
| 1nn, MDN, BSD3, 196 | $(2.52 \pm 0.44) \times 10^{-48}$ | $(6.83 \pm 0.32) \times 10^{-25}$ | $(6.25 \pm 0.16) \times 10^{-17}$ | $7.0 \pm 0.01$ |
| 3nn, MDN, BSD3, 196 | $(4.28 \pm 1.58) \times 10^{-48}$ | $(8.31 \pm 0.87) \times 10^{-25}$ | $(7.04 \pm 0.45) \times 10^{-17}$ | $6.7 \pm 0.02$ |
| 3nn, MDN, BSD2, 196 | $(2.88 \pm 0.88) \times 10^{-51}$ | $(2.64 \pm 0.28) \times 10^{-26}$ | $(7.43 \pm 0.50) \times 10^{-18}$ | $5.0 \pm 0.01$ |
| 3nn, MDN, BSD1, 192 | $(2.06 \pm 0.75) \times 10^{-59}$ | $(1.74 \pm 0.20) \times 10^{-30}$ | $(1.12 \pm 0.08) \times 10^{-20}$ | $5.8 \pm 0.02$ |

To estimate the variance of the MDN model parameters and its effect on the predicted defect concentration, 100 models were trained on different training datasets. The resulting $E_{eff}$ and $C_{BSD}(T)$ calculations for selected temperatures are shown in Table II.

As it can be seen in Fig. 5, MDN succeeded in calculations of $C_{BSD}(T)$ even with only 196 configurations for training. Larger training databases improve the model accuracy but from the perspective of calculating ensemble averages as $C_{BSD}(T)$ and their effective formation energies $E_{eff}$ this results in a small difference of about $\approx 10$ meV. It is important to note that 10 meV is well above the accuracy that one can expect from empirical interatomic potentials such as CRG.

Finally, $C_{BSD}(T)$ calculations were done for all 3 types of BSD defects, as shown in Fig. 6. Each graph was obtained using the MDN model trained on $\approx 200$ configurations from 3nn databases. Similarly as before, to estimate the variance of the MDN models parameters 100 models were trained. The results for selected temperatures are presented in Table II.

The differences between the relative positions of the curves in Fig. 6 come from the difference in BSD stability predicted by the CRG potential. According to the calculations done in this work and Bathelier et al., [15], BSD3 is the most stable among other BSD defects with the lowest formation energy. This results in the higher concentration of BSD3 defects that can be observed in Fig. 6 (blue curve). The same results about the stability of BSD defects were obtained in Balboa et al., [43], where the BSD formation energies were computed using CRG potential by averaging the energies of seven different configurations. Different results were obtained in Cheik Njifon [44], where the BSD formation energies were calculated using the DFT+U method, suggesting BSD2 as the most stable of BSD defects. However, D. Bathelier showed that this mismatch is due to the small DFT supercell size (96 atoms) used in Cheik Njifon’s work and that a match between DFT and CRG is found when using 324-atom DFT supercells. Additionally, it is important to note that in Cheik Njifon’s work only one SQS supercell was used.

It would be interesting to compare the computations of $C_{BSD}(T)$ in $(U, Pu)O_2$ done with the MDN method in this work with experimental measurements, however, to the authors’ knowledge, there is at present no such study in the literature.

IV. DISCUSSIONS AND CONCLUSIONS

This work shows that the semi-supervised application of the Mixture Density Network (MDN) is a robust approach to evaluate local-atomic dependent properties in chemically disordered $(U, Pu)O_2$ nuclear fuels. In our study, based on CRG potential [11], we have shown that by shifting the cost of sampling atomic configurations to the approximation of the density of states it is possible to access $(U, Pu)O_2$ properties at a much lower cost. That allowed us to provide, for the first time, an approximation of the equilibrium concentration of Bound Schottky Defect (BSD) defects in $(U, Pu)O_2$ nuclear fuels. It was possible to do this computation through the estimation
of the density of states of BSD formation energy by using roughly 200 configurations. As a comparison, to characterize the formation energy of BSD3 in the 2nn sphere of influence in Bathellier et al. [15], 22668 configurations were required. Assuming, as a rough estimate, 5 minutes per CRG energy minimization of 2592 atoms supercell, that translates to 17 hours of CPU time for the MDN approach, instead of 1889 hours as shown in [15]. Now, those two examples concern only one specific type of BSD defect. If we were to inspect, for example, other types of defects, or change the interatomic potential, the cost mentioned above would have to be paid again. Furthermore, the increase in the size of the system, the main limitation present in the previous studies and the dominant factor increasing the computational cost, does not affect the performance of the MDN models. Due to the chosen local description of atomic configurations, the training cost remains the same regardless of the radius of considered spheres of influence. Thanks to that, we were able to determine that the 3nn sphere of influence is a limit beyond which further atoms start to have negligible influence on $E_{BD}^f$ calculations. This conclusion confirms quantitatively what was shown in the previous study [15]. Additionally, the MDN method allows to compute distributions of formation energy for 3nn and 4nn but in general also further spheres of influence, which was not possible in [15].

The presented results show that the approximation of local properties influenced by the chemical disorder can be efficiently approached through the semi-supervised estimation of their probability distribution functions. The reduced amount of calculations needed to obtain satisfactorily accurate estimations with the MDN models makes it possible to envisage the direct use of DFT to produce the training databases. This will be explored in future work.

This approach is not, however, limited to the BSD defect properties in (U, Pu)O$_2$. The same methodology could also be applied to study other defects and atomic-scale properties, such as formation entropies, migration energies, or attempt frequencies [45], but not only. The presented approach will also be useful to address atomic-scale properties of high-entropy alloys (HEA) [8], where the same problem of characterization of the vast configuration space can be often encountered.

V. ACKNOWLEDGMENTS

This research contributes to the CEA SIMU/PICI2 project. It received its funding from the CEA FOCUS/EJN program. The authors wish to express their gratitude to M. Nastar and T. Schuler for their aid in the definition of the mathematical framework for the formation energy calculation and fruitful discussions.

Appendix A: Quantification of impact of local order

The used formula for chemical potentials in Eq. (7) relies on the assumption that local order could be neglected. If it is true, then both formulas would yield the same results for the $C_{BD}(T)$ calculations. The BSD concentration obtained by direct calculation including all data in two datasets of 1nn from [15] were compared: configurations where U or Pu atom was removed to form a BSD3 type vacancy defect. As can be seen in Fig. 7, both formulas give similar results in defect concentration calculations.

Appendix B: Probability distribution function of formation energy

Let us define a dataset $\mathcal{D}$ as a collection of $L$ energies, $L \leq N$, where $N$ is the total number of configurations in a configuration space:

$$\mathcal{D} = \{E_{i}^f\}_{i=1}^{L}, E_{i}^f \sim p(E_{i}^f).$$

The formation energy of a defect can be described with a density of states probability distribution function $p(E_{i}^f)$ as defined in Eq. (15). Since it is not possible to calculate the $p(E_{i}^f)$ for the whole spectrum of configurations, we therefore aim at computing the approximation $p_{\theta}(E_{i}^f)$, as in Eq. (18), of the true distribution $p(E_{i}^f)$. To capture the similarity between these 2 distributions, the Kullback-Leibler divergence can be used:

![Figure 7](image-url)
\begin{equation}
KL(p||p_\theta) = \sum_j p(E \theta, j) \log \left( \frac{p(E \theta, j)}{p_\theta(E \theta, j)} \right), \quad (\text{B2})
\end{equation}

over all possible formation energies \( E \). We are looking for the set of parameters \( \theta^* \) that corresponds to the minimal value of KL divergence:

\begin{equation}
\theta^* = \arg_{\theta} \min[KL(p||p_\theta)]. \quad (\text{B3})
\end{equation}

We can rewrite KL divergence as an expected value:

\begin{equation}
KL(p||p_\theta) = \sum_i p(E \theta, i) \log \left( \frac{p(E \theta, i)}{p_\theta(E \theta, i)} \right)
= \sum_i \left[ - \log(p_\theta(E \theta, i)) - \log(p(E \theta, i)) \right] + H_p,
\end{equation}

where \( H_p \) is the Shannon entropy. As \( H_p \) is not dependent on \( \theta \) the minimal value of KL divergence can be written as:

\begin{equation}
\theta^* = \arg_{\theta} \min[\mathbb{E}_{E \sim p(E)}[- \log(p_\theta(E^\theta))]]. \quad (\text{B5})
\end{equation}

Computing \( \theta^* \) comes down to computing the minimal negative log-likelihood (using, e.g., the gradient descent techniques):

\begin{equation}
\theta^* = \arg_{\theta} \min \mathcal{L}(\theta, D), \quad \mathcal{L}(\theta, D) = - \frac{1}{M} \sum_{i=1}^M \log p_\theta(E^\theta(i)). \quad (\text{B6})
\end{equation}

We therefore minimize the loss function, Eq. (B6), to find the most optimal set of parameters \( \theta^* \), that allows us to build a probability distribution \( p_\theta(E^\theta) \) that approximates \( p(E^\theta) \). Calculation of \( C_{\text{BSD}}(T) \) can then be done by Monte Carlo approximation similarly as in Eq. (14), but by sampling \( p_\theta(E^\theta) \):

\begin{equation}
C_{\text{BSD}}(T) \approx \mathbb{E}_{E \sim p_\theta(E^\theta)}[f_E(E^\theta)] \approx \frac{1}{M} \sum_{i=1}^M f_E(E^\theta(i)). \quad (\text{B7})
\end{equation}

Appendix C: Verification of MDN architecture

MDN performance was tested in terms of mean negative log-likelihood for different training databases. The results for 3nn and 4nn databases of BSD3 in (U, Pu)O\(_2\) 50% Pu can be seen in Fig. 8. Similar results were obtained for all types of BSD defects. It can be seen that the minimal number of configurations needed for MDN training is \( \approx 200 \). In addition, further increase in the training dataset does not significantly improve the accuracy of the MDN model predictions.

[1] O. El-Atwani, N. Li, M. Li, A. Devaraj, J. Baldwin, M. M. Schneider, D. Sobieraj, J. S. Wröbel, D. Nguyen-Manh, S. A. Maloy, et al., Outstanding radiation resistance of tungsten-based high-entropy alloys, Science Advances 5, eaav2002 (2019).

[2] S. Xia, X. Yang, T. Yang, S. Liu, and Y. Zhang, Irradiation resistance in Al \( x \) CoCrFeNi high entropy alloys, Jom 67, 2340 (2015).

[3] J. Kim, H. Kwon, J.-H. Kim, K.-M. Roh, D. Shin, and H. D. Jang, Elastic and electronic properties of partially ordered and disordered Zr (C1-\( x \)) solid solution compounds: a first principles calculation study, Journal of Alloys and Compounds 619, 788 (2015).

[4] B. Gheboul, M. Gheboul, M. Fatmi, T. Chihi, Z. Heiba, and S. Boucetta, Structural, Elastic, and Electronic Properties of CuCl\(_ x \)Br (1-x) Compounds under Pressure, 

FIG. 8. (Color online) MDN training performance given the varying number of training samples. Figures a) and b) consider models trained on the 3nn and 4nn databases accordingly. The training was done for BSD3, (U, Pu)O\(_2\) 50% Pu.
[38] K. Li, C.-C. Fu, M. Nastar, F. Soisson, and M. Y. Lavrentiev, Magnetochemical effects on phase stability and vacancy formation in fcc Fe-Ni alloys, arXiv preprint arXiv:2203.04688 (2022).

[39] J.-F. Vigier, P. M. Martin, L. Martel, D. Prieur, A. C. Scheinost, and J. Somers, Structural investigation of \((U_{0.7}Pu_{0.3})O_{2-x}\) mixed oxides, Inorganic Chemistry 54, 5358 (2015).

[40] J. M. Joyce, Kullback-Leibler divergence, in International encyclopedia of statistical science (Springer, 2011) pp. 720–722.

[41] S. L. Brunton and J. N. Kutz, Data-driven science and engineering: Machine learning, dynamical systems, and control (Cambridge University Press, 2022).

[42] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in’t Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, \textit{et al.}, Lammps—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, Computer Physics Communications 271, 108171 (2022).

[43] H. Balboa, L. Van Brutzel, A. Chartier, and Y. Le Bouar, Damage characterization of \((U, Pu)O_2\) under irradiation by molecular dynamics simulations, Journal of Nuclear Materials 512, 440 (2018).

[44] I. Cheik Njifon, Modélisation des modifications structurales, électroniques et thermodynamiques induites par les défauts ponctuels dans les oxydes mixtes à base d’actinides \((U, Pu)O_2\), Ph.D. thesis, Aix-Marseille (2018).

[45] H. Mehrer, Diffusion in solids: fundamentals, methods, materials, diffusion-controlled processes (Springer Science & Business Media, 2007) pp. 55–103.