



## Coupling mixture designs, high-throughput experiments and machine learning for accelerated exploration of multinary systems



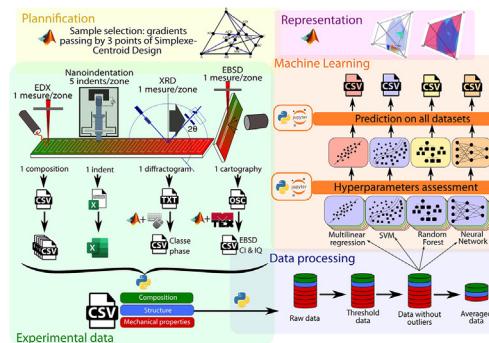
Elise Garel, Jean-Luc Parouty, Hugo Van Landeghem, Marc Verdier, Florence Robaut, Stéphane Coindeau, Raphaël Boichot\*

*Univ. Grenoble Alpes, CNRS, Grenoble INP, SIMAP, 38000 Grenoble, France*

### HIGHLIGHTS

- We propose combining Mixture Design and Machine Learning to efficiently explore the Nb-Ti-Zr-Cr-Mo system in thin films.
- Among all the models tested, Neural Networks are the more accurate to predict phase class, hardness and Young's modulus.
- Results show no clear indication of positive cocktail effect beyond three elements.
- The optimal mechanical properties are systematically found far from the equimolar composition.

### GRAPHICAL ABSTRACT



### ARTICLE INFO

#### Article history:

Received 28 March 2023

Revised 12 May 2023

Accepted 30 May 2023

Available online 02 June 2023

#### Keywords:

Combinatorial design

PVD

Machine learning

High Throughput Experiment

### ABSTRACT

Current societal challenges, such as climate change and resource depletion, highlight an unprecedented need for disruptive innovation in materials science. Significant breakthroughs are expected in multinary materials whose efficient exploration necessitates dedicated strategies. The exploration of a refractory high entropy alloy Nb-Ti-Zr-Cr-Mo is proposed here as test case for a new strategy. Based on the proven methodology of mixture design and on combinatorial thin film metallurgy, the composition space is explored by a limited number of chosen gradients to build an alloy library comprising hardness and ductility, two antagonistic properties. The workflow is showcased here by studying the properties of the as-grown graded film, which presents wide amorphous domains and contrasted mechanical properties. This experimental dataset then trains machine learning models to provide continuous predictions of the alloy properties over the entire composition space. We show that optimal alloy properties are expected close to the binary edges of the quinary.

© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

The daunting challenges of climate changes alongside the depletion of fossil fuel and mineral resources lead to a worldwide realization that an energy transition is inevitable. However, the

efficient production and storage of renewable or low carbon energy still requires innovations in materials science, which appears to be a major lever, but also a major bottleneck considering the amount of possible systems to explore and the constraints in time and resource availability. Several international projects [1–4] have thus operated a paradigm shift: acceleration of innovative material design requires moving away from slow, resource-intensive, experiment-heavy approaches, in favor of streamlined, high-throughput methodologies, integrated with computerized tool

\* Corresponding author.

E-mail address: [raphael.boichot@grenoble-inp.fr](mailto:raphael.boichot@grenoble-inp.fr) (R. Boichot).

chains, considering first and foremost the less critical mineral resources on Earth.

For a century, materials science has mainly focused on systems with one or two principal components. Until recently, the metallurgy of structural alloys has focused on studying vertices of the composition spaces, including aluminium, titanium, iron, nickel and copper-based systems. A real technology breakthrough may come from the little-explored highly multidimensional composition space, and thus recent advances in material science have been centered on multinaries, such as functional ceramics [5], polymers [6], metallic glasses [7,8] or High Entropy Alloys (HEA) [9,10].

## 2. The tools for a new approach in material design

The vastness of the strongly multidimensional spaces that must be explored makes the laborious “one-sample-at-a-time” historical approach inappropriate to beat this time-trial. From 3-component or ternary system, to 10-component or denary system,  $4.9 \times 10^4$  to  $1.7 \times 10^{12}$  possible compositions can be explored to find a global optimal composition – considering 1 %at variations and no influence of process parameters, microstructure, etc. Facing this “curse of dimensionality”, a more rational approach should be implemented for efficient experimental exploration and optimization of these multinaries. Additionally, a broader picture of the compositional dependence of properties is also desirable to identify recycling-tolerant alloys in those systems [11].

In this perspective, Boetcher *et al.*, initiate the bases of combinatorial thin film approach in 1955, followed by Kennedy *et al.*, in 1965 [12]. These methods were then formalized by Hanak *et al.*, [13] in 1970 as the “multi-sample concept”. This was the premise of the combinatorial approach, which consists in synthesizing and characterizing many compositions using high-throughput techniques, to build material libraries linking compositions, structures and properties of materials [12,14,15]. Many high-throughput techniques and combinatorial applications [16,17] have been developed since then, in particular for the exploration of multinary materials [18].

Now that combinatorial methodologies bring high-throughput experimental investigations at hand, efficient processing of the large amount of data they generate is critical. Coupling a combinatorial method to data mining tools is an example of a suitable strategy [19]. The broad availability of open access, coherent and good quality datasets (i.e. containing both outstanding and average results with the associated measurement error) is key to the successful development of combinatorial design.

The development of Machine Learning (ML) in the context of materials science has led to accelerated and enhanced material designs. ML models have been used on data gathered from literature alone [18], or as a guide to generate high-throughput experimental data [20–22], or from CALPHAD (CALculation of PHase Diagrams) data – on HEA amongst others [23]. These previous studies have shown a good ability to describe and predict the properties of the studied materials. Nonetheless, CALPHAD predictions face certain limitations in multinary systems [24], and literature-based data can be incomplete [25]. Combining high-throughput characterizations and alloy composition libraries to train ML models does away with both these restrictions, affording control over data distribution. The inherent statistical error associated to a model increases when data are noisy or sparse in a region of the explored space [26]. A uniform screening of the composition space would allow to have data as representative as possible – with the lower inherent bias possible – and thus is a necessary step to obtain reliable predictions over the entire composition space, while being able to estimate the model error as constant as possible. Mixture designs propose such a uniform screening. Many dif-

ferent mixture designs exist and the choice of the design is strongly linked to the chosen combinatorial fabrication process.

## 3. Rational method for the optimization of any N-element system

The key point of the method proposed in this paper is to create a synergy between high throughput approaches, mixture design and Machine Learning models, in order to optimize the composition of any multi-component material for chosen properties. High throughput experiments guarantees to build broad composition-structure-properties libraries, while mixture design guide the gathering of homogeneous data, representative of the composition space. This way, Machine Learning models will more likely be competent over the studied domain.

More precisely, in this paper, combinatorial experiments will be performed through the fabrication of compositionally graded samples, which permit great efficiency in terms of experimental resources. A single sample should thus ideally include as many of the points specified by the mixture design as possible. For linear gradients, it implies that those points are aligned in the composition space. Among different possible mixture designs, Simplex Designs proposed by Scheffé [27] propose such characteristics. The present study will assess a possible strategy using a Simplex-Centroid Design (SCD), in which the uniformity of screening is guaranteed by exploring all the equimolar or equimassic mixtures and pure elements through graded samples.

Gradients of composition fabricated with a combinatorial technique – as either thin film or bulk – can then be characterized using local techniques for their properties of interest [28]. These techniques will dictate the minimal probed volume or surface over which the composition can be approximated as constant. In general, a larger spatial extent of the gradient will enhance the compositional resolution and facilitate local characterizations. The collected datasets can then be used to train ML models, the choice of which will depend on the type and amount of data.

## 4. Application of the method on a quinary refractory high entropy alloy

Sometimes included in or associated with Complex Concentrated Alloys, HEA are alloys composed of five or more principal elements, in proportions ranging from 5 to 35 %at. They have represented a real challenge in terms of screening since their introduction with Cantor's alloy in 2004 [29], as their optimization requires studying the traditionally neglected center of a strongly multidimensional compositional space [30]. Later on, Gao *et al.*, [31] showed that the same kind of issue is encountered in developing high mixing entropy bulk metallic glasses. The complexity of both HEA and high entropy metallic glasses was also pointed out by Welk *et al.*, [32] and Andrew *et al.*, [33]. For this reason, these multinary systems have been studied through combinatorial approach in different papers [34–37], and using ML models [38]. HEAs are then an interesting test case to evaluate the proposed method.

Most of the interesting mechanical properties of HEAs were measured with Cantor-based alloys. These alloys present for example interesting wear resistance [39] or exceptional yield strength coupled to low density [40]. The introduction or substitution of more and more elements into Cantor-based formula in order to improve its creep or oxidation resistance leads to a very fuzzy frontier between the different classes of HEA. In this perspect, high throughput-based design strategies to accelerate composition selection were proposed [41].

This paper will focus on the exploration of a quinary system of refractory HEAs (RHEAs). RHEAs are studied both as bulk, as high temperature alloys, and as thin films [42], to improve wear resistance, irradiation resistance [43] or oxidation and corrosion resistance for instance. One of the main obstacles to their commercialization is their low ductility at room temperature. Based on the recurrence of refractory elements in RHEAs literature compositions [30], on their criticality [44], as well as the existing knowledge of their properties of interest [45–48], the studied system was chosen as Nb-Ti-Zr-Cr-Mo. The selection of these particular elements has no bearing on the method; any other multi-component alloy could have been used.

## 5. Material and methods

### 5.1. Screening an $N$ -element composition space

For any  $N$ -element composition space, the SCD recommends  $2^{N-1}$  composition points to explore. Geometrically, the domain to explore is an  $(N-1)$  simplex, thus a  $(N-1)$  dimension geometrical space. In such spaces, a finite number of 1D (one-dimensionnal) linear gradients cross three points of this mixture design each, and a finite number of 2D planar gradients encompass seven points of the design each, if these points have strictly complementary compositions (see [Supplementary Notes 1](#)). It is worth noting that fabricating 2D projection of 3D volumes and hypervolumes passing through several SCD points are possible. However, projecting hypergradients on a 2D sample can lead to overlapping of compositional domains and prevent coverage of all SCD points at once, which is problematic in the specific case of the proposed method (see [Supplementary Figure 19](#)).

By crossing a hypervolume with all possible linear and planar gradients, each SCD point is encountered several times. A reduced set of gradients is then sufficient to go at least once through each mixture composition that must be covered. This set can be selected by a brute-force algorithm, taking into account some user constraints (minimizing the number of sample or the cost of the experimental work, finely screening a region of interest etc. see [Supplementary Figure 3](#)). This allows precise scheduling and cost estimate for any experimental campaign designed with our method.

In the case of the quinary RHEA, 90 possible linear gradients and 65 possible planar gradients go through the 31 SCD compositions. The alloy chosen for this particular study will be screened by 1D linear gradients. Although thirteen gradients are sufficient for a complete screening of a quinary using SCD (see [Supplementary Notes 1](#)), a set of twenty gradients was preferred to introduce finer mapping of the center of the composition space as well as a more symmetrical distribution of the composition points (see [Supplementary Notes 2](#)).

### 5.2. Fabrication of graded samples

Synthesizing extended composition gradients using diffusion treatments is impractically slow with the refractory metals classically included in RHEAs, and requires delicate liquid processing of the base metals and careful sample preparation. Magnetron co-sputtering, on the contrary, is an easy way of mixing metals not prone to interdiffusion and allows making very broad gradients of composition. In order to illustrate the method, we chose to study RHEAs thin films sputtered in strictly identical conditions, so as not to introduce a supplementary process-related difficulty in the evaluation of the method. The choice of deposition method will however lead to specific microstructures. While magnetron sputtering lead to phases far from thermodynamic equilibrium, it can

be difficult to master grain size in bulk multiple diffusion for example. In the following, the properties were measured on the as-deposited thin films. The method can be adapted to reproduce bulk-like microstructures by heat-treating the film as appropriate, possibly above the maximum liquidus if a deposition mask is used [49]. Annealing of co-sputtered graded thin films of refractory metals has however been proven to be difficult due to composition-dependent surface oxidation and roughening [50].

The synthesis of broad 1D gradient alloys in thin films is advantageously based on equipment of limited sophistication. Gradients of composition were deposited by a two-cathode magnetron confocal co-sputtering in a Plassys MP500S chamber, using one DC (direct current) and one RF (radio frequency) cathode (factory setup of our sputtering machine). Epi-ready Sapphire <0001> was chosen as substrate for its pure elastic response in nanoindentation. Experimental parameters were optimized by a Design of Experiment to increase deposition rate while reducing the surface roughness in order to perform accurate nanoindentation tests on as grown-samples. Ideally, sample surface roughness must be several orders of magnitude less than the Berkovitch tip penetration depth. The samples present a mirror-like surface with very low roughness (See [Supplementary Notes 3](#)) for the following optimal conditions: plasma was ignited at 5 Pa, working pressure was then reduced down to 0.14 Pa with an argon inlet of 40 sccm (standard cubic centimetre per minute or multiple of  $7.44 \times 10^{-7} \text{ mol.s}^{-1}$ ), DC intensity was set to 190 mA (with a corresponding power of 65 to 70 W) and RF power was set at 150 W. Additionally, a bias was applied to the substrate (20 V) and the substrate holder temperature was set to 400 °C. A 5 h deposition in these conditions leads to a film thickness close to 5 μm all along the deposited gradients. Cathodes were crossed on test samples to ensure that DC or RF sputtering was not the key parameter leading to amorphous or crystalline phase on our samples.

The geometry of the cathodes was adjusted to maximize the spatial extent of deposited gradients and their composition variations. The chamber configuration that optimizes both spatial extent of the gradients and deposition rate corresponds to the cathodes aiming at mid-radius of the 100 mm substrate holder, cathode metal surface at 8 cm from the substrate while the sample holder is kept motionless. Typical gradients range from  $A_{0.2}B_{0.8}$  to  $A_{0.8}B_{0.2}$  over 100 mm, A and B being the compositions of the two cathode targets, carefully chosen to match the different sets of SCD point alignments. Being able to move the focal point of cathodes is key to control the gradient amplitude and thus to accurately characterize the subsequently discretized compositions. Sputtering targets (and SCD) were designed in mass fraction in order to maximize use of targets readily available on the market.

### 5.3. Properties characterization

Samples are discretized by 4 mm steps along the gradients to be systematically characterized. Property characterization of as-grown thin films focuses on finding a composition leading to an alloy both strong and ductile, an antagonistic combination whose optimization is critical for structural materials [51]. Composition, crystallinity and mechanical properties were assessed to build the alloy libraries, by EDX (energy-dispersive X-ray spectroscopy), EBSD (electron backscatter diffraction), XRD (X-ray diffraction) and nanoindentation.

Composition evaluation was performed with a field emission gun (FEG) ZEISS Gemini scanning electron microscope SEM 500 equipped with a silicon drift detector (SDD) (EDAX OCTANE ELITE 25 Si<sub>3</sub>N<sub>4</sub> windows- 60 mm<sup>2</sup>). EDX measurements were performed and analyzed with the EDAX TEAM software. The acquisition was done on a 10 μm<sup>2</sup> zone, during 30 s, at 20 kV, once per discrete

composition. All subsequent characterizations were performed in directly adjacent areas.

An EBSD map was obtained once for each discrete composition on a  $10 \mu\text{m}^2$  zone, with a 7 nm step, on a FEG ZEISS Gemini SEM 500, with a charge coupled device (CCD) detector (EDAX HIKARI SUPER) and OIM software. Image Quality (IQ) and Confidence Index (CI) were extracted in a.csv file by an automated treatment using MTEX Matlab toolbox [52]. EBSD CI and IQ were then associated to each composition.

Mechanical properties were acquired through five 700 nm deep indents per composition (5x1 matrix perpendicular to gradient axis, each separated by 20  $\mu\text{m}$ ), using a Berkovitch tip on a Nanoin-denter MTS-XP machine with continuous stiffness measurement at a constant strain rate  $\frac{d\varepsilon}{dt} = 0.05 \text{ s}^{-1}$ . Using the Oliver and Pharr [53] model, area of the tip was computed systematically from 16 measurements on fused silica before any characterization batch. Hardness, reduced modulus and Young's modulus were extracted from Oliver and Pharr model, by fixing the Poisson ratio to 0.3, for a contact depth of 400 nm (57% relative depth). Ductility was assessed using Galanov's model [54] on experimental and predicted modulus and hardnesses by computing the relative characteristic size of the elastic-plastic zone under the indenter  $\frac{h_c}{c}$ , the constrain factor C – linking yield strength and hardness – and the ductility characteristic  $\delta_H$  – ratio of plastic deformation and total deformation.

Systematic XRD measurements were made on the 5 axis Rigaku Smartlab in  $0\text{-}2\theta$  configuration with a copper anode (using the K- $\alpha$  wavelength, 1.5418 Å). Scanning in  $2\theta$  was performed in the range  $25^\circ\text{-}144^\circ$ . In order to probe a constant length of the graded sample (1 mm perpendicular to gradient axis), a variable divergent slit configuration has been used. This technique was less local than the other ones used here. It provides structural information that can be compared to EBSD results, albeit averaged over a larger probed volume.

#### 5.4. Dataset construction

For each characterization, relevant data should be added to the dataset. In order to improve dataset quality on nanoindentation replications, statistical processing is applied. Due to a surface detection failure or tip pollution, computed Young's modulus and hardness can be biased and bear no physical meaning: all compositions with both an H value under 2 GPa and an E value under 10 GPa are removed from the dataset. Then, for each composition, possible outliers are automatically identified and extracted from the threshold replications, using Dixon Q-test [55] on the minimum and on the maximum value of hardness and modulus (independently).

#### 5.5. Machine Learning models

From the discrete compositions-properties datasets obtained, ML allows a continuous prediction of properties over the entire composition space. Optimized models with an increasing complexity are tested: Multiple Regression (MR), Support Vector Machine (SVM), Random Forest (RF) and Neural Network (NN). Each model-dataset couples are evaluated. Hyperparameters optimization and model-dataset selection are detailed in [Supplementary Notes 6](#).

#### 5.6. Representation

Experimental results and predictions should be represented in a human-readable form. Representing a composition space with many spatial dimensions is delicate as most humans have a more limited grasp of spaces with more than three dimensions. Although

some visualizations have been proposed using projections [30,56], they deprive the user from seeing the entirety of the composition space at once, excluding notably the center of the multinary, which is expected to be particularly interesting for HEAs study. The hypertetrahedral quinary composition space will be represented here using the Delaunay triangulation [57,58] of five points normally distributed in a 3D geometrical space. In this representation, there is no overlapping of the different points of the SCD, in particular at the equimolar or equimassic mixture points. The entire composition domains can be visualized at once and explored through an interactive representation. Finally, linear and planar gradients can also be visualized without deformation (more details are given in [Supplementary Notes 4](#)).

The overall approach detailed in this method is summarized in Fig. 1: a) Linear or planar gradients are selected, in a quaternary system for instance, passing at least once through SCD points; b) Linear gradients being deposited as thin films by confocal magnetron sputtering; c) Characterization of linear gradients by EDX, nanoindentation, XRD and EBSD. Each 10 cm gradient is discretized into approximately 22 zones; d) From this raw experimental library, datasets are built; e) Four different Machine Learning model families are trained and the best one enables the identification of compositions yielding optimized properties.

## 6. Results

### 6.1. Experimental results

#### 6.1.1. Composition range

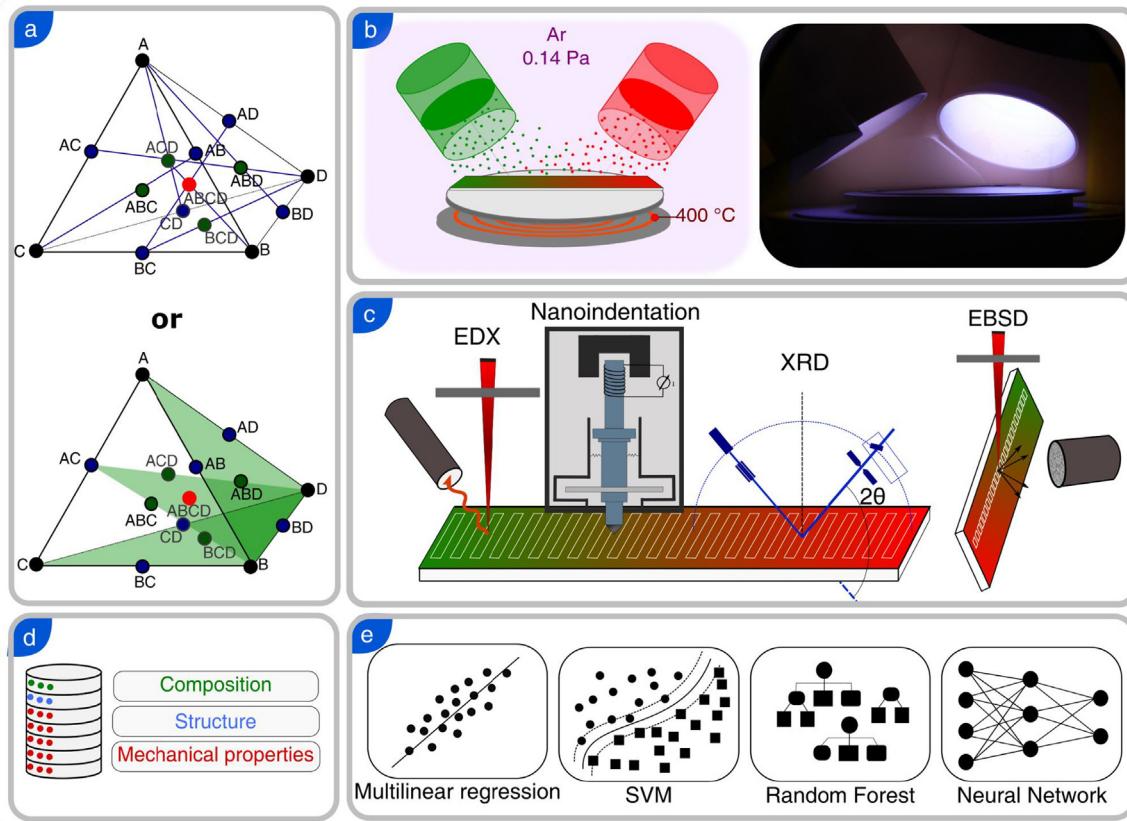
The fabricated gradients (Fig. 2) are mirror-like metallic thin films deposited on a 10 cm sapphire strip. Some of the gradients show a transition zone where roughness is different, which suggests that the microstructure may change along the gradient.

The composition variation over the span of each strip ranges between 50 %at to 90 %at of the associated target compositions, depending on the gradient and the elements considered. Mass compositions accurately follow the predicted gradients, with a fine screening of the center of the composition space. A shift towards the chromium vertex is systematically observed in chromium containing samples due to the higher sputtering rate of this element [59]. The cathode compositions used in this study were not adjusted to compensate for the difference in sputtering efficiency for each metal in each mixture. The sputtering parameters were kept constant too in order to avoid introducing a bias in the study, in addition to the existing ones.

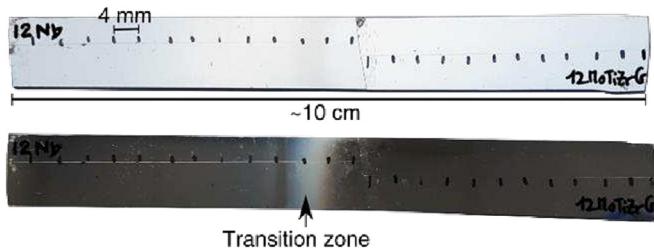
#### 6.1.2. Microstructure classes

From XRD diffraction patterns, two classes are extracted, depending on the appearance of diffraction peaks (presence of crystallized phases) or not (amorphous or nano-crystalline phases). The composition is then always either crystallized in single BCC phase, or in amorphous/nano-crystalline state. In order to compare XRD and EBSD, similar classes are extracted from backscatter diffraction (EBSD) maps. Confidence Index (CI) and Image Quality (IQ) values (Fig. 4 and [Supplementary Figure 20](#)) evolve with the thin film structure, from crystalline pure BCC to amorphous composition. A thresholding at 0.02 of the confidence index measured in EBSD was applied to separate crystalline and nano-crystalline/amorphous states.

EBSD and XRD classes are identical for 82% of the composition points. Variations are principally due to the class transition that is detected at different points of the gradients (difference of one or two steps). This may come from the thresholding imposed for CI and signal-to-noise ratio on XRD diffractograms in the crystalline-amorphous transition regions. However, other artifacts



**Fig. 1.** A) linear or planar gradients are selected, in a quaternary system for instance, passing at least once through scd points; b) linear gradients being deposited as thin films by confocal magnetron sputtering; c) characterization of linear gradients by edx, nanoindentation, xrd and ebsd. each 10 cm gradient is discretized into approximately 22 zones; d) From this raw experimental library, datasets are built; e) Four different Machine Learning model families are trained and the best one enables the identification of compositions yielding optimized properties.



**Fig. 2.** Photograph of a gradient Nb-TiZrCrMo under two different lighting angles.

may explain the fact that amorphous composition ranges are systematically broader with EBSD than by XRD. Fig. 5 (sample top view) illustrates the difference in identification of the two techniques originating from sampling artifacts. Indeed, due the small surface area covered by EBSD compared to XRD, a low density of crystalline inclusions dispersed into an overall amorphous sample will appear crystalline with XRD and likely amorphous by EBSD. It can also come from the difference in probed depth between the two techniques. During the deposition process, there is probably a balance between kinetic of deposition on surface and kinetic of crystallization in the bulk (see Supplementary Figure 22) due to the magnetron sputtering process that could, in turn, prevent the detection of crystalline phases for certain compositions using EBSD as the technique probes a depth of a few tens of nanometers only.

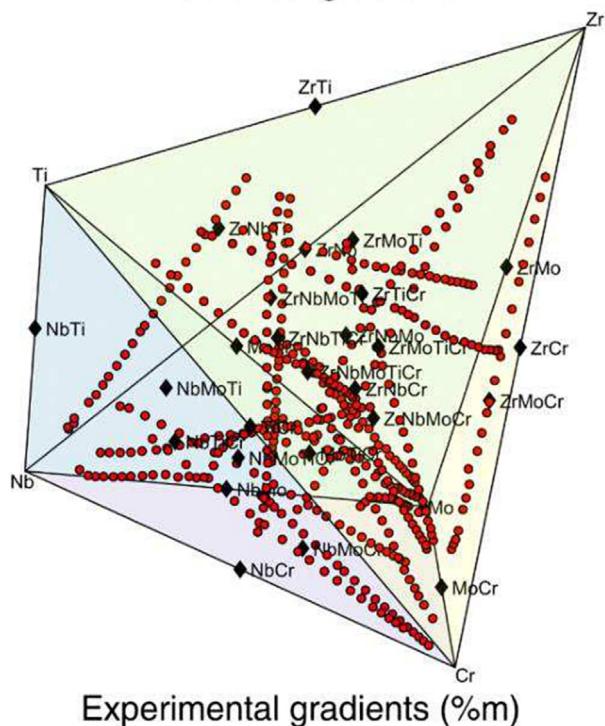
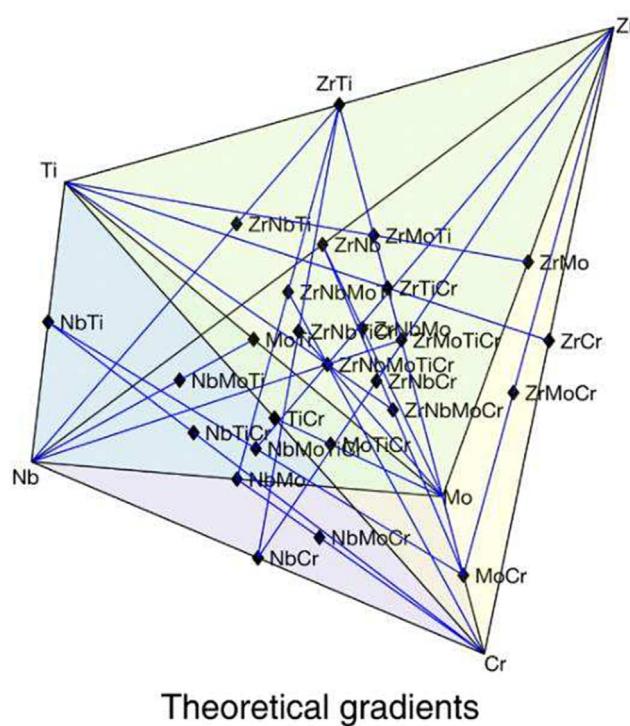
Fig. 6.a and 6b. show the result of the above classification based on EBSD and XRD experimental data, respectively.

### 6.1.3. Mechanical properties

Fig. 6 shows large variations of properties for all measured outputs (an interactive view is proposed in the Gitlab associated with this paper, see [Supplementary Notes 6](#), and a 2D 5-cell version is given in [Supplementary Figures 23 to 26](#)): a 250 GPa range for elastic modulus and a 20 GPa range for hardness are observed. Absolute values of hardness and elastic modulus should nevertheless be considered with caution as internal stresses in as-grown sputtered thin films can influence them. Error linked to the sample tilt during nanoindentation (negligible here) is evaluated in [Supplementary Notes 3](#).

Ductility index computed from experimental hardness and Young's modulus is also represented in Fig. 6.e. Again, wide variations of  $\delta_H$  are observed, evolving from 0.498 to 0.812, with a maximum near the niobium vertex. Ductility values predicted in the next section will be calculated from predicted hardness and Young's modulus and not from the data computed here.

Fig. 6.a and d. allow observing a possible correlation between microstructure and mechanical properties: the highest hardness values are observed on CrMo to Zr gradient, near CrMo binary, before a clear break. This abrupt change corresponds to a crystalline to amorphous transition on the EBSD phase representation, which is a hint of the well-known dependency of the mechanical properties from both microstructure and composition.

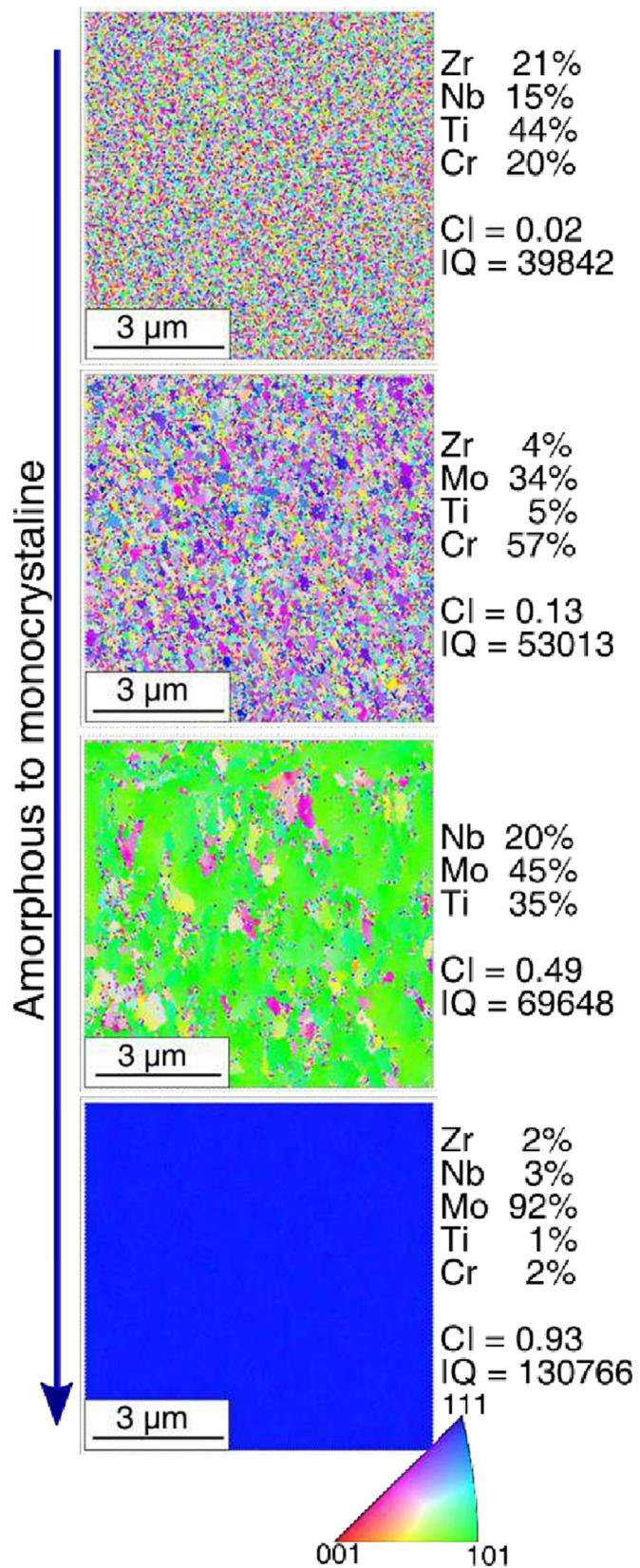


**Fig. 3.** Experimental mass composition gradients vs expected ones.

## 7. Properties predictions with neural networks

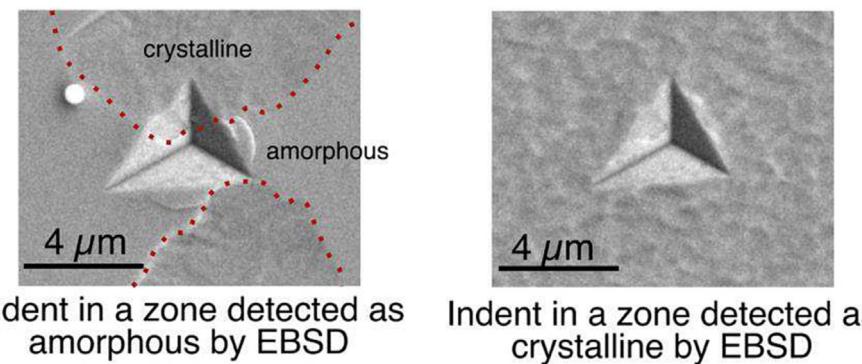
### 7.1. 20-gradient dataset

Four families of models were optimized, trained and evaluated. All the results are given in [Supplementary Notes 6](#). In this paper, we will focus on neural networks predictions, trained with EBSD and XRD classes and with mechanical properties dataset without outliers. These models are the most predictive and they do not



**Fig. 4.** EBSD maps of amorphous to preferentially oriented crystalline microstructure with associated CIs and IQs.

show any obvious artifacts. Properties are predicted over the entire composition, with a 2 %at step for each element.



**Fig. 5.** SEM micrographs of a nano-indent in an amorphous domain with crystallised inclusions and fully crystalline domain. Note the conchoidal fractures around the indent in amorphous zone.

XRD- and EBSD-predicted amorphous or nanocrystalline compositions hull are represented in Fig. 7.a. Neural Network predictions from EBSD mapping leads to a larger amorphous domain than the domain predicted from XRD, as expected. However, these domains overlap. In the following discussion, definition of amorphous compositions will be based on mathematical intersection of these two domains: compositions are considered as amorphous if both EBSD-based predictions and XRD-based predictions give them as such, and as crystalline if EBSD-prediction or XRD-prediction gives them as such.

Thus defined, phase prediction still does not return two clearly delineated structure domains. A closer look at the distribution of the compositions predicted classes (see Supplementary Figure 41) reveals three different regions: one in which all compositions are crystalline, one in which all compositions are amorphous, and a transition zone in which neighboring compositions are either. The hulls of all three domains are plotted in Fig. 7.b.

XRD dataset is overlaid on the predicted hulls on Fig. 7.b. which shows a good matching between prediction and data. However, the isolated circled (in red) experimental amorphous/nanocrystalline composition is outside of the corresponding hull. This shows that the model is not overfitting the data and is thus quite generalizable. It is interesting to note that the binary classification modeled by Machine Learning allows us to reconstruct a fairly clear transition zone as observed experimentally.

Predicted Young's modulus and hardness are used to compute the plasticity index  $\delta_H$  from Galanov's model. The three predicted mechanical properties are represented in Fig. 7.c. Young's modulus and hardness optimum are consistent with experimental data, with maximum values near Cr-Mo edge. The maximum of ductility corresponds to the minimum of hardness, near the niobium vertex, which was expected from experimental data. Finding a domain both ductile and strong consists in a compromise between these two antagonistic properties.

A Pareto front is used to this end, in the central amorphous or nanocrystalline domain, and in the crystalline domain. Each front is represented in Fig. 8 and the Pareto optimal compositions are represented in the composition space in Fig. 9.

In both phase domains, the Pareto optimal compositions are near the crystalline/amorphous hull. The same observations stand for other model-dataset couples (Supplementary Figure 42), however all Pareto optimal compositions are not exactly the same from one model to another. Moreover, for all kinds of model, the error can be plotted depending on data repartition (see Supplementary Notes 5 for detailed computation). It appears that the error increases while approaching binary edges, which is expected as they were not screened in details on purpose. Predictivity of the model will thus be evaluated on validation sets, chosen along bina-

ries where prediction of properties are the worse. Here, three binaries were chosen to evaluate the model predictivity on the border: Nb-Ti, Zr-Mo and Cr-Nb. This choice is detailed in Supplementary Notes 7.

#### 7.1.1. Validation of neural networks for hardness and ductility index prediction.

Hardness, ductility index and XRD phase class of initial and validation datasets are plotted in Fig. 10 and show a good consistency with preceding experimental data.

Predictions of neural networks trained with the initial 20-gradient dataset can be compared to experimental results of the validation datasets (Supplementary Figure 17). While Nb-Cr properties are quite accurately predicted, prediction on Zr-Mo et Nb-Ti properties are less accurate. It shows that the method requires to be iterated in order to increase the model accuracy on the seeming zones of interest.

#### 7.1.2. Method iterations in the amorphous/nanocrystalline domain

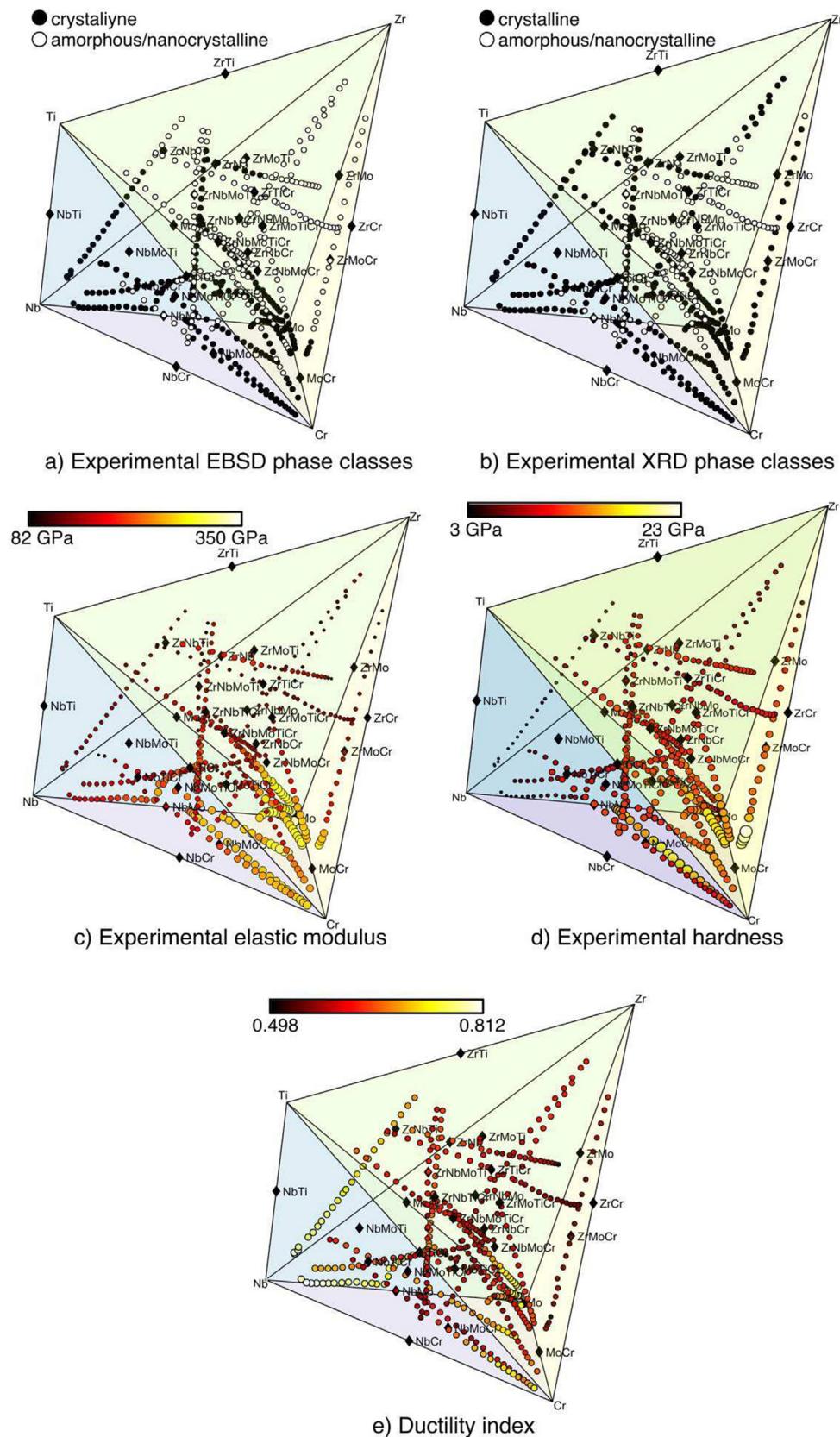
Each binary sample is added one by one to the original 20-gradient dataset, after outliers were removed by a Dixon test. The neural networks models are trained again with these new datasets, new predictions are performed and Pareto optimal compositions are identified. The evolution of the Pareto compositions is represented in Fig. 11 (for amorphous zone).

It shows that the optimal compositions identified from modeled properties are modified by adding new data in the dataset, for neural network models but also for the rest of the models (see Supplementary Figure 43). A converged optimum based on Pareto front is identified throughout the iterations:  $Zr_{2-4}Nb_{0-4}Mo_{0-4}Ti_{18-22}Cr_{72-74}$  is for instance a good compromise for a hardness of 11 to 13 GPa and a ductility index of 0.63 to 0.65. It can be observed that the pareto optimal zones converge with added experimental data.

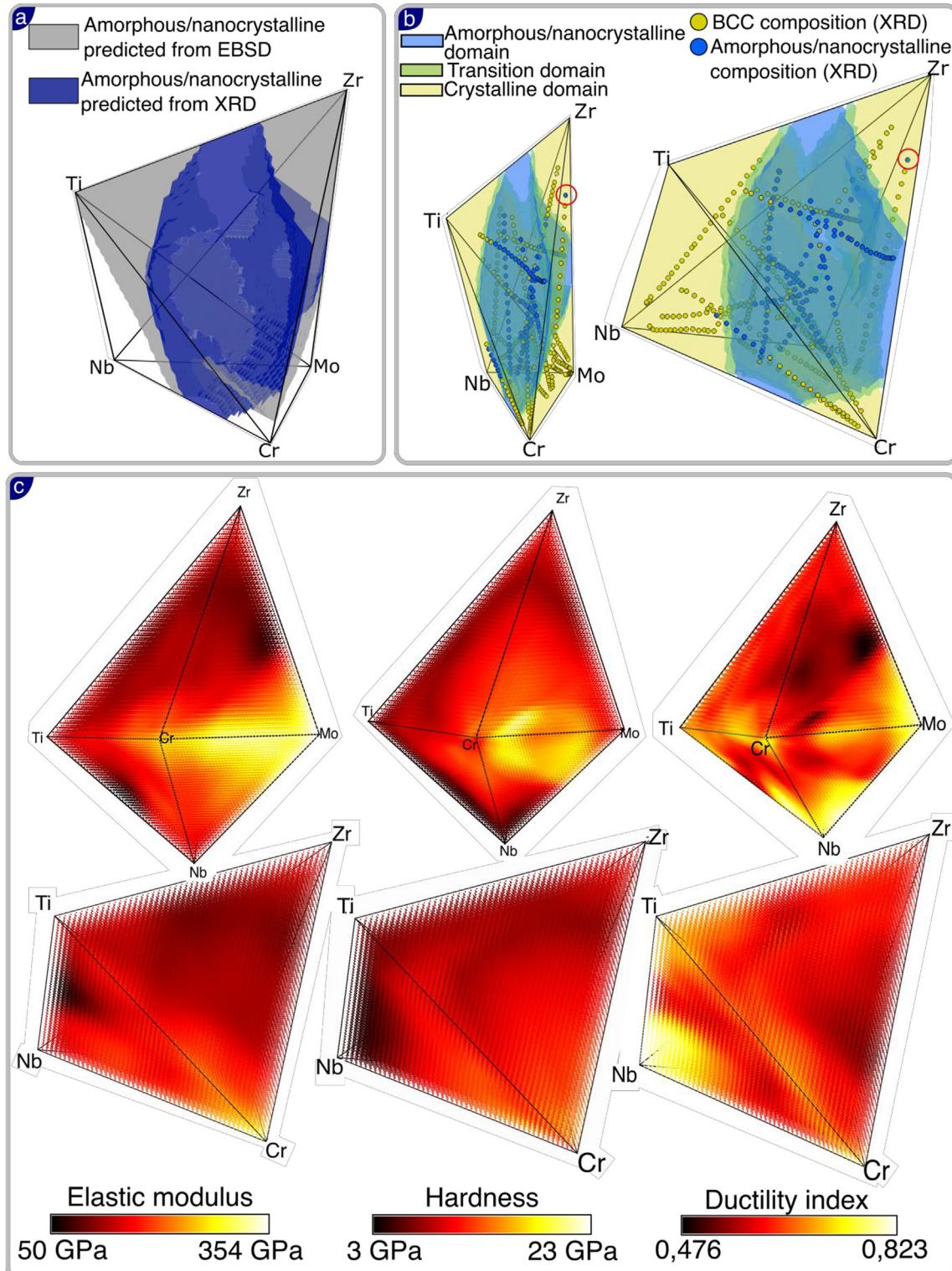
## 8. Discussion and conclusions

The method proposed here allowed to identify clearly delineated domains of Pareto optimal compositions. We chose to study in particular the central amorphous/nanocrystalline domain. It appears that optimal compositions, balancing ductility index and hardness, remain on the border of this domain even after introduction of new experimental validation sets. This leads to different hypotheses:

- The dependence of mechanical properties on composition is dominated by their dependence on the microstructure of the alloy. Typically, the amorphous/crystalline transition region likely corresponds to a nanocrystalline domain, in which the



**Fig. 6.** Experimental phase classes extracted from EBSD (a) and XRD (b), experimental values of elastic modulus (c) and hardness (d) and ductility index calculated from experimental data (e).

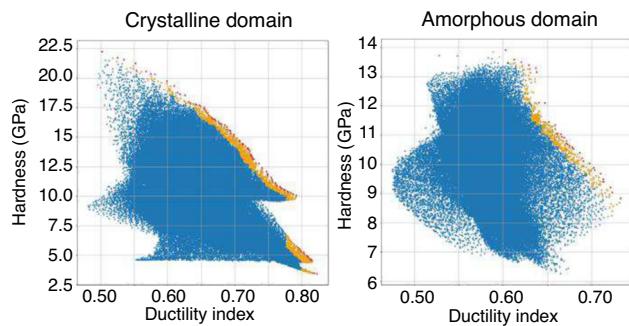


**Fig. 7.** Amorphous hulls predicted from XRD and EBSD data (a), final phase domain predicted from both XRD and EBSD, compared to XRD experimental results (b), predicted elastic modulus, hardness and ductility index (c).

microstructure enhances the ductility/hardness compromise. Additional TEM work could help ascertain this effect but is out of the scope of the present study.

- Even though the experimental effort was focused on screening the center of the composition space as finely as possible, where HEA and BGM systems are suspected to offer disruptive proper-

ties, optimal compositions were systematically lying close to the edges of the composition space. This observation casts substantial doubt on the cocktail effect being a systematic core effect of HEA or BGM systems. The present results did not reveal any evidence of said effect in the Nb-Ti-Zr-Cr-Mo system. Mul-



**Fig. 8.** Pareto front in crystalline and amorphous domains.

tilinear regression (see [supplementary Figures 27 and 28](#)) even indicates a detrimental effect of interactions involving more than 3 elements.

It is encouraging to note that alloy compositions close to the Pareto front identified in the present work are studied in the literature for biomedical applications [60,61]. It is however difficult to directly compare characterization results given the markedly different processing conditions and the equally different resulting microstructures. It does suggest that additional efforts should be committed to refine the screening of this system for load-bearing biocompatible alloys, possibly using the present method over a narrower portion of the composition space.

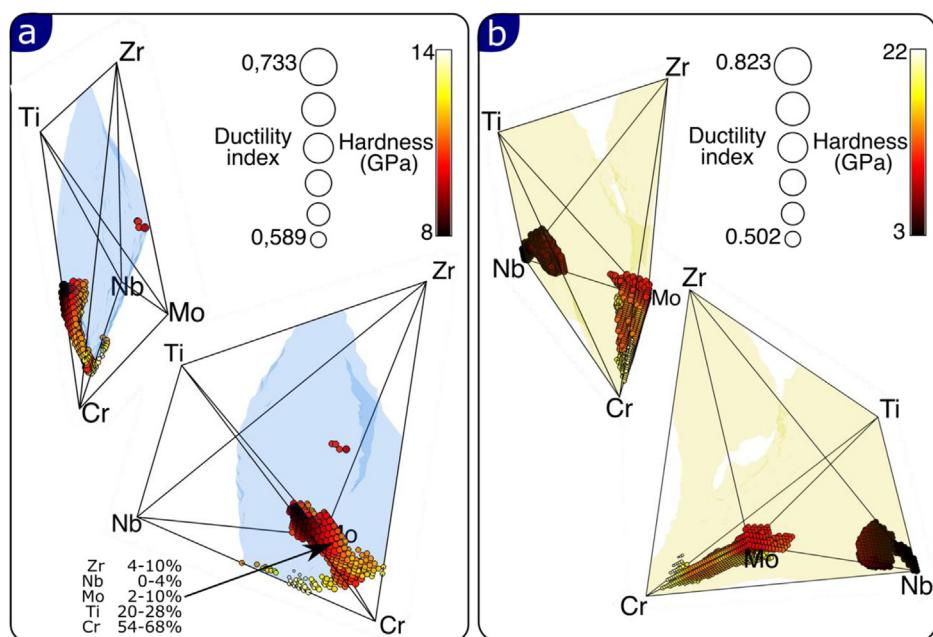
This study shows that a real asset lies into the synergistic combination of SCD, high-throughput experiments and ML in rapidly mapping highly multidimensional composition space. However, considering ML as a mere standalone tool for data mining of disparate pre-existing experimental data is insufficient to reveal its full potential in the ICME (integrated computational materials engineering) approach. Instead, it was shown here that ML models give a more favorable outcome with training datasets tailored to their operating principles. This need for ML-oriented data can be fulfilled with SCD-based high-throughput experiments, which result in the rapid and automated acquisition of evenly distributed

data and their associated error. It is worth stressing the importance of making such data open source, so that it can readily benefit all ulterior ML studies. The dataset proposed here is meant to integrate with other materials science databases. It is intended to be supplemented with further elements, alloys and processing parameters.

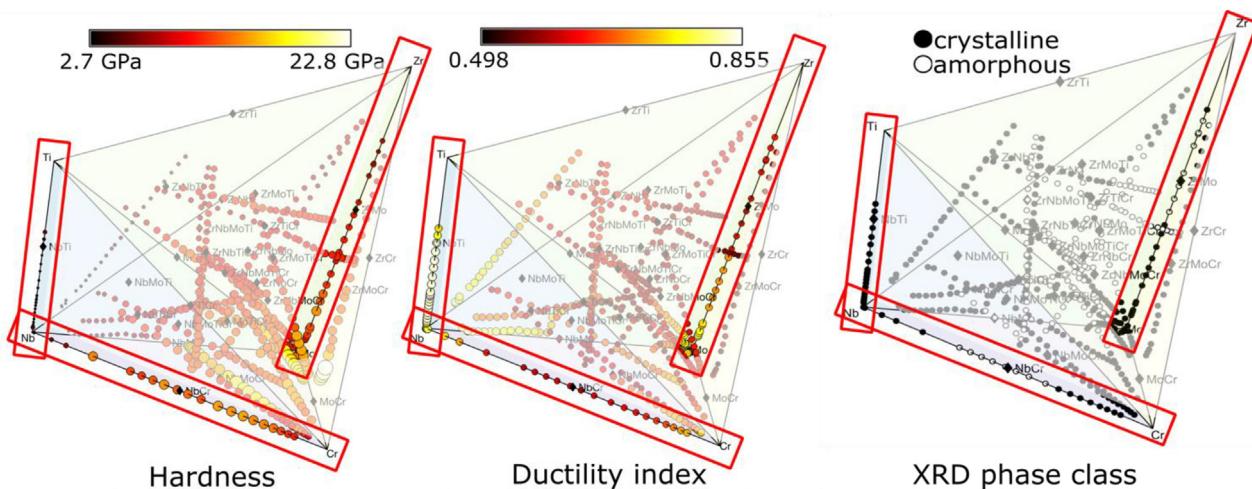
Following these guidelines, it was possible, with twenty-three thin film gradients only, to isolate a set of converged Pareto-optimal compositions based on their hardness/ductility compromise and to mitigate the cocktail-effect hypothesis in the case of as-grown RHEA thin films – i.e. out of equilibrium. The proposed representation of the composition space using Delaunay triangulation enables visualization of these results with enhanced legibility, which is a real advance for multinary studies.

The development of the method exposed herein is timely given the recent interest for the center of multinaries, expected to bring a fresh wave of discoveries in material science. It highlights two points of attention regarding this trend, while providing ways to address them:

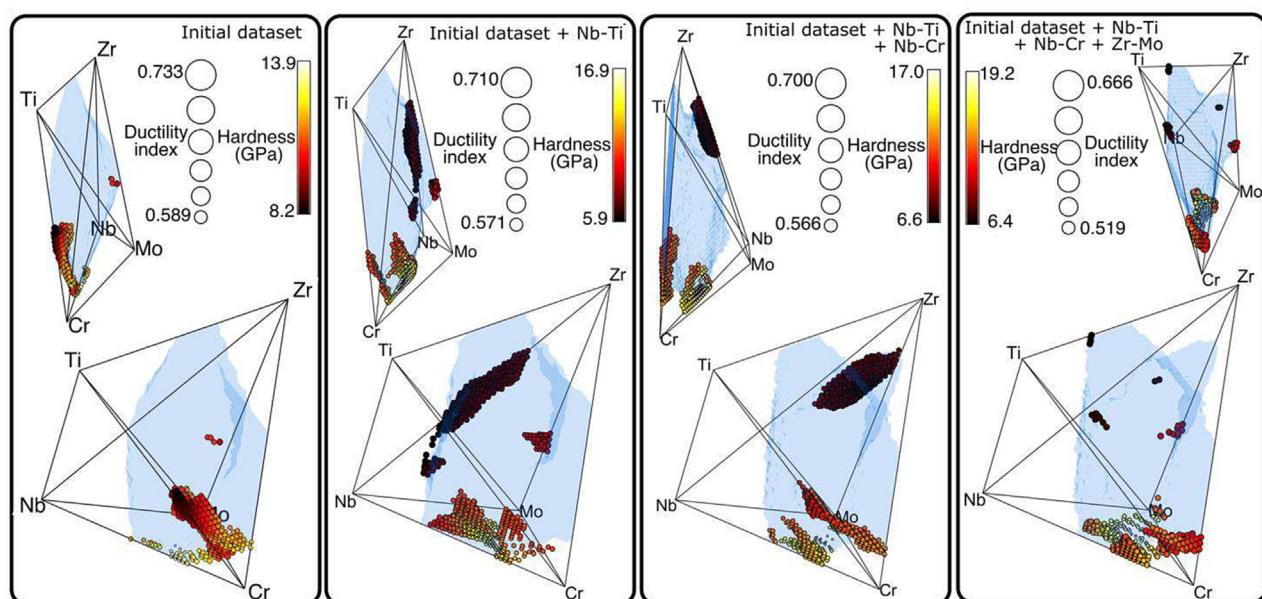
- The first is related to the location in the composition space of the property optima. It directly illustrated here that, depending on the target properties, the optima can lie away from the center of the composition space as well as from equimolar points. Thus, it seems that increasing the complexity of concentrated alloys will show diminishing returns. Using a rational approach to optimize alloys of moderate complexity could prove more favorable. The method proposed here can easily be transposed for ternary and quaternary systems. Furthermore, the data produced could also allow focusing on simpler subsystems from a complex dataset, which constitutes a supplementary asset.
- The second is, on a similar note, connected to the recyclability of complex alloys. The efficiency gain offered by high-performance materials is only desirable if they can sustainably be used and reused. It was shown that achieving maximum performance might require fine compositional tuning, which, in turn, would mean reduced recyclability. However, with a complete picture



**Fig. 9.** Pareto-optimal compositions obtained in amorphous (a) and crystalline (b) domains from Neural Network predictions.



**Fig. 10.** Validation dataset experimental XRD phase class, hardness and ductility index compared to initial dataset properties.



**Fig. 11.** Pareto-optimal compositions in amorphous zone predicted by Neural Network models trained with initial dataset without outliers, iteratively completed with binaries.

of the composition space, it is also possible to identify alloys whose properties remain constant in spite of compositional deviation and are thus recycling-ready.

### Declaration of Competing Interest

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

### Acknowledgements

We acknowledge the IDRIS and GRICAD for computational resources, and the network MIAI@Grenoble Alpes (French ANR-19-P3IA-0003) for fruitful discussions, in particular with Rémi Molinier and Emilie Devijver. The authors also thank the members of the Consortium des Moyens Technologiques Communs (CMT), Rachel Martin and Frédéric Charlot as well as the Centre of Excellence of Multifunctional Architectured Materials (CEMAM, French ANR-10-LABX-44-01) for characterization support, and Francis Baillet for his help on Mixture Designs. The authors have no funding to declare.

### 9. Data availability

Datasets and Machine Learning trained models are available in a Zenodo repository: [Accelerated exploration of multinary systems | Zenodo](#).

All Python and Matlab modules are available in the GitLab <https://gricad-gitlab.univ-grenoble-alpes.fr/garele/accelerated-exploration-of-multinary.git>.

A Ph.D. thesis covering more details on this study and complementary analysis is also available for reading <https://cnrs.hal.science/tel-04043261v1>.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.matdes.2023.112055>.

## References

- [1] Materials Genome Initiative for Global Competitiveness. [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/materials\\_genome\\_initiative-final.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf) (2011).
- [2] Materials Genome Initiative Strategic Plan. 36 (2021).
- [3] Jarvis, D. et al. Metallurgy Europe - a renaissance programme for 2012–2022: science position paper. (2012).
- [4] K. Alberi et al., The 2019 materials by design roadmap, *J. Phys. Appl. Phys.* 52 (2018) 013001.
- [5] E. Danielson, J.H. Golden, E.W. McFarland, C.M. Reaves, W.H. Weinberg, X.D. Wu, A combinatorial approach to the discovery and optimization of luminescent materials, *Nature* 389 (6654) (1997) 944–948.
- [6] J.C. Meredith, A.P. Smith, A. Karim, E.J. Amis, Combinatorial materials science for polymer thin-film dewetting, *Macromolecules* 33 (26) (2000) 9747–9756.
- [7] S. Ding et al., Combinatorial development of bulk metallic glasses, *Nat. Mater.* 13 (2014) 494–500.
- [8] Y.P. Deng, Y.F. Guan, J.D. Fowlkes, S.Q. Wen, F.X. Liu, G.M. Pharr, P.K. Liaw, C.T. Liu, P.D. Rack, A combinatorial thin film sputtering approach for synthesizing and characterizing ternary ZrCuAl metallic glasses, *Intermetallics* 15 (9) (2007) 1208–1216.
- [9] Akbari, A. Combinatorial Screening Approach in Developing Non-Equiatomic High Entropy Alloys. (University of Kentucky Libraries, 2018). doi:10.13023/ETD.2018.407.
- [10] Z. Li, A. Ludwig, A. Savan, H. Springer, D. Raabe, Combinatorial metallurgical synthesis and processing of high-entropy alloys, *J. Mater. Res.* 33 (2018) 3156–3169.
- [11] D. Raabe, C.C. Tasan, E.A. Olivetti, Strategies for improving the sustainability of structural metals, *Nature* 575 (7781) (2019) 64–74.
- [12] J. Zhao, Combinatorial approaches as effective tools in the study of phase diagrams and composition–structure–property relationships, *Prog. Mater. Sci.* 51 (5) (2006) 557–631.
- [13] J.J. Hanak, The “multiple-sample concept” in materials research: synthesis, compositional analysis and testing of entire multicomponent systems, *J. Mater. Sci.* 5 (1970) 964–971.
- [14] E.J. Amis, X.-D. Xiang, J.-C. Zhao, Combinatorial materials science: what's new since edison?, *MRS Bull.* 27 (2002) 295–300.
- [15] X.-D. Xiang, X. Sun, G. Briceño, Y. Lou, K.-A. Wang, H. Chang, W.G. Wallace-Freedman, S.-W. Chen, P.G. Schultz, A combinatorial approach to materials discovery, *Science* 268 (5218) (1995) 1738–1740.
- [16] T. Gebhardt, D. Music, T. Takahashi, J.M. Schneider, Combinatorial thin film materials science: from alloy discovery and optimization to alloy design, *Thin Solid Films* 520 (17) (2012) 5491–5499.
- [17] M.L. Green et al., Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies, *Appl. Phys. Rev.* 4 (2017) 011105.
- [18] A. Ludwig, Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods, *Npj Comput. Mater.* 5 (2019) 70.
- [19] A. Deschamps, F. Tancret, I.-E. Benrabah, F. De Geuser, H.P. Van Landeghem, Combinatorial approaches for the design of metallic alloys, *Comptes Rendus Phys.* 19 (2018) 737–754.
- [20] Z. Zhou et al., Machine learning guided appraisal and exploration of phase design for high entropy alloys, *Npj Comput. Mater.* 5 (2019) 128.
- [21] S. Sarker et al., Discovering exceptionally hard and wear-resistant metallic glasses by combining machine-learning with high throughput experimentation, *Appl. Phys. Rev.* 9 (2022) 011403.
- [22] F. Ren et al., Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments, *Sci. Adv.* 4 (2018) eaao1566.
- [23] J. Qi, A.M. Cheung, S.J. Poon, High entropy alloys mined from binary phase diagrams, *Sci. Rep.* 9 (2019) 15501.
- [24] S. Gorsse, O. Senkov, About the reliability of CALPHAD predictions in multicomponent systems, *Entropy* 20 (2018) 899.
- [25] S. Gorsse, M.H. Nguyen, O.N. Senkov, D.B. Miracle, Database on the mechanical properties of high entropy alloys and complex concentrated alloys, *Data Brief* 21 (2018) 2664–2678.
- [26] H.K.D.H. Bhadeshia, Neural networks in materials science, *ISIJ Int.* 39 (10) (1999) 966–979.
- [27] H. Scheffé, The simplex-centroid design for experiments with mixtures, *J. R. Stat. Soc. Ser. B Methodol.* 25 (2) (1963) 235–251.
- [28] J.-C. Zhao, The diffusion-multiple approach to designing alloys, *Annu. Rev. Mater. Res.* 35 (2005) 51–73.
- [29] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Microstructural development in equiatomic multicomponent alloys, *Mater. Sci. Eng. A* 375–377 (2004) 213–218.
- [30] D.B. Miracle, O.N. Senkov, A critical review of high entropy alloys and related concepts, *Acta Mater.* 122 (2017) 448–511.
- [31] X.Q. Gao et al., High mixing entropy bulk metallic glasses, *J. Non-Cryst. Solids* 357 (2011) 3557–3560.
- [32] B.A. Welk, M.A. Gibson, H.L. Fraser, A combinatorial approach to the investigation of metal systems that form both bulk metallic glasses and high entropy alloys, *JOM* 68 (3) (2016) 1021–1026.
- [33] High entropy alloy and bulk metallic glass coatings, *J. Therm. Spray Technol.* 31 (2022) 920–922.
- [34] A. Kauffmann et al., Combinatorial exploration of the high entropy alloy system Co–Cr–Fe–Mn–Ni, *Surf. Coat. Technol.* 325 (2017) 174–180.
- [35] G.C. Gruber et al., Synthesis and structure of refractory high entropy alloy thin films based on the MoNbTaW system, *Surf. Coat. Technol.* 439 (2022) 128446.
- [36] A. Marshal et al., Combinatorial evaluation of phase formation and magnetic properties of FeMnCoCrAl high entropy alloy thin film library, *Sci. Rep.* 9 (2019) 7864.
- [37] M.A. Tunes, V.M. Vishnyakov, Microstructural origins of the high mechanical damage tolerance of NbTaMoW refractory high-entropy alloy thin films, *Mater. Des.* 170 (2019) 107692.
- [38] W. Huang, P. Martin, H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, *Acta Mater.* 169 (2019) 225–236.
- [39] S. González et al., Wear resistant CoCrFeMnNi<sub>0.8</sub>V high entropy alloy with multi length-scale hierarchical microstructure, *Mater. Lett.* 331 (2023) 133504.
- [40] Y. Li et al., A low-density high-entropy dual-phase alloy with hierarchical structure and exceptional specific yield strength, *Sci. China Mater.* 66 (2023) 780–792.
- [41] F. Liu et al., High-throughput method-accelerated design of ni-based superalloys, *Adv. Funct. Mater.* 32 (2022) 2109367.
- [42] Y.-Y. Chen, S.-B. Hung, C.-J. Wang, W.-C. Wei, J.-W. Lee, High temperature electrical properties and oxidation resistance of V-Nb-Mo-Ta-W high entropy alloy thin films, *Surf. Coat. Technol.* 375 (2019) 854–863.
- [43] Y. Zong, N. Hashimoto, H. Oka, Study on irradiation effects of refractory bcc high-entropy alloy, *Nucl. Mater. Energy* 31 (2022) 101158.
- [44] E. Commission, Directorate General for Internal Market, Industry, Entrepreneurship and SMEs., Deloitte Sustainability., British Geological Survey, Bureau de Recherches Géologiques et Minières, & Toegepast natuurwetenschappelijk onderzoek. Study on the review of the list of critical raw materials: final report, (Publications Office, 2017).
- [45] Z.D. Han, N. Chen, S.F. Zhao, L.W. Fan, G.N. Yang, Y. Shao, K.F. Yao, Effect of Ti additions on mechanical properties of NbMoTaW and VNbMoTaW refractory high entropy alloys, *Intermetallics* 84 (2017) 153–157.
- [46] S.-B. Hung, C.-J. Wang, Y.-Y. Chen, J.-W. Lee, C.-L. Li, Thermal and corrosion properties of V-Nb-Mo-Ta-W and V-Nb-Mo-Ta-W-Cr-B high entropy alloy coatings, *Surf. Coat. Technol.* 375 (2019) 802–809.
- [47] O.N. Senkov, S.I. Rao, T.M. Butler, K.J. Chaput, Ductile Nb alloys with reduced density and cost, *J. Alloys Compd.* 808 (2019) 151685.
- [48] C.-C. Juan, M.-H. Tsai, C.-W. Tsai, C.-M. Lin, W.-R. Wang, C.-C. Yang, S.-K. Chen, S.-J. Lin, J.-W. Yeh, Enhanced mechanical properties of HfMoTaTiZr and HfMoNbTaTiZr refractory high-entropy alloys, *Intermetallics* 62 (2015) 76–83.
- [49] Y. Li et al., Combinatorial strategies for synthesis and characterization of alloy microstructures over large compositional ranges, *ACS Comb. Sci.* 18 (2016) 630–637.
- [50] Q. Xing, J. Ma, Y. Zhang, Phase thermal stability and mechanical properties analyses of (Cr, Fe, V)–(Ta, W) multiple-based elemental system using a compositional gradient film, *Int. J. Miner. Metall. Mater.* 27 (2020) 1379–1387.
- [51] O.N. Senkov, D.B. Miracle, K.J. Chaput, J.-P. Couzinie, Development and exploration of refractory high entropy alloys—a review, *J. Mater. Res.* 33 (2018) 3092–3128.
- [52] F. Bachmann, R. Hielscher, H. Schaeben, Texture analysis with MTEX – free and open source software toolbox, *Solid State Phenom.* 160 (2010) 63–68.
- [53] W.C. Oliver, G.M. Pharr, Measurement of hardness and elastic modulus by instrumented indentation: advances in understanding and refinements to methodology, *J Mater Res* 19 (1) (2004) 3–20.
- [54] B. Galanov, S. Ivanov, V. Kartuzov, Improved core model of the indentation for the experimental determination of mechanical properties of elastic-plastic materials and its application, *Mech. Mater.* 150 (2020) 103545.
- [55] W.J. Dixon, Processing data for outliers, *Biometrics* 9 (1) (1953) 74.
- [56] L. Banko et al., Unravelling composition–activity–stability trends in high entropy alloy electrocatalysts by using a data-guided combinatorial synthesis strategy and computational modeling, *Adv. Energy Mater.* 12 (2022) 2103312.
- [57] Lemaire, C. Triangulation de Delaunay et arbres multidimensionnels. (Ecole Nationale Supérieure des Mines de Saint-Etienne ; Université Jean Monnet - Saint-Etienne, 1997).
- [58] Delaunay, B. Sur la sphère vide. A la mémoire de Georges Voronoï. 9.
- [59] J.-J. Besson, Dépôts par pulvérisation cathodique. 27 (1985).
- [60] Murayama, Y., Okubo, A. & Kimura, H. Mechanical Properties of Ti–Cr System Alloys Prepared by Powder Metallurgy. 4.
- [61] A.B. Elshalakany et al., Microstructure and mechanical properties of Ti–Mo–Zr–Cr biomedical alloys by powder metallurgy, *J. Mater. Eng. Perform.* 26 (2017) 1262–1271.