Volume 3 I Issue 1 I March 2023

# JOURNAL OF MATERIALS INFORMATICS




## Synergizing ontologies and graph databases for highly flexible materials-to-device workflow representations

Max Dreger, Mohammad J. Eslamibidgoli, Michael H. Eikerling, Kourosh Malek*

Open Access
ISSN 2770-372X (Online)
www.jmijournal.com

# EDITORIAL BOARD

# GENERAL INFORMATION

## About the Journal

*Journal of Materials Informatics* (*JMI*), is an international peer-reviewed, open access, online journal. *JMI* aims to advance and accelerate the pace of materials discovery, design, and deployment by synergistically combining theory, experiment, computation, and artificial intelligence in a tightly integrated and smart manner. The journal synthesizes foundational physics, chemistry, mathematics, mechanics, computer science and engineering, and materials science and engineering with emerging experiments, computations, and real-world applications, for interdisciplinary researchers and those who are new to the field. Although materials informaticsis currently in the nascent stage, it is reshaping materials science and engineering in many ways at an exponential speed and its full power of synergy is still far from being realized.

## Information for Authors

Manuscripts should be prepared in accordance with Author Instructions.
Please check www.jmijournal.com/pages/view/author_instructions for details.
All manuscripts should be submitted online at https://oaemesas.com/login?JournalId=jmi.

## Copyright



## Permissions



## Disclaimer



## Publisher

OAE Publishing Inc.
245 E Main Street st112, Alhambra, CA 91801, USA
Website: www.oaepublish.com

## Contacts

E-mail: editorialoffice@jmijournal.com
Website: www.jmijournal.com

*Journal of Materials Informatics*

# CONTENTS

Volume 3 / Issue 1 / 2023 /

Hu *et al. J Mater Inf* 2023;3:1
**DOI:** 10.20517/jmi.2022.28

**Journal of Materials Informatics**

**Research Article**

**Open Access**

# Data-driven prediction of the glass-forming ability of modeled alloys by supervised machine learning

**Yuan-Chao Hu[1],*, Jiachuan Tian[2]**

[1]Department of Mechanical Engineering & Materials Science, Yale University, New Haven, CT 06520, USA.
[2]Meta Platforms, Inc., Menlo Park, CA 94025, USA.

***Correspondence to:** Dr. Yuan-Chao Hu, Department of Mechanical Engineering & Materials Science, Yale University, New Haven, CT 06520, USA. E-mail: ychu0213@gmail.com; ORCID: 0000-0001-9872-7854

**How to cite this article:** Hu YC, Tian J. Data-driven prediction of the glass-forming ability of modeled alloys by supervised machine learning. *J Mater Inf* 2023;3:1. http://dx.doi.org/10.20517/jmi.2022.28

**Received:** 27 Sep 2022 **First Decision:** 16 Nov 2022 **Revised:** 4 Jan 2023 **Accepted:** 1 Feb 2023 **Published:** 17 Feb 2023

**Academic Editor:** Xingjun Liu **Copy Editor:** Ke-Cui Yang **Production Editor:** Ke-Cui Yang

## Abstract

The ability of a matter to fall into a glassy state upon cooling differs greatly among metallic alloys. It is conventionally measured by the critical cooling rate $R_c$, below which crystallization inevitably happens. There are a lot of factors involved in determining $R_c$ for an alloy, including both elemental features and alloy properties. However, the underlying physical mechanism is still far from being well understood. Therefore, the design of new metallic glasses is mainly by time- and labor-consuming trial-and-error experiments. This considerably slows down the development process of metallic glasses. Nowadays, large-scale computer simulations have been playing a significant role in understanding glass formation. Although the atomic-scale features can be well captured, the simulations themselves are constrained to a limited timescale. To overcome these issues, we propose to explore the glass-forming ability of the modeled alloys from computer simulations by supervised machine learning. We aim to gain insights into the key features determining $R_c$ and found that the non-linear couplings of the geometrical and energetic factors are of great importance. An optimized machine learning model is then established to predict new glass formers with a timescale beyond the current simulation capability. This study will shed new light on both unveiling the glass formation mechanism and guiding new alloy design in practice.

**Keywords**: Metallic glasses, molecular dynamics simulations, glass-forming ability, machine learning, data mining

## 1. INTRODUCTION

Ever since the first discovery of an amorphous metal[1], also named metallic glass (MG) later, from Au-Si system, developing new MGs with exceptional glass-forming ability (GFA), i.e., low critical cooling rate $R_c$, has been one of the main goals in the field[2–5]. In turn, these new materials ensure the exploration of the physics, chemistry, and mechanics of glasses in experiments[6–8]. In the past several decades, thousands of new MGs with various GFAs have been synthesized successfully in labs all around the world. In addition, an increasing amount of fascinating knowledge has been acquired. This greatly enriches the glass family.

Starting from the periodic table, the principal elements for MGs are transition metals, sometimes with metalloids as minor additions. Empirically, more components usually make a better glass former. There are usually four to five elements in bulk MGs. This makes the glass formation problem rather complex to understand. First of all, the parameter space is huge with enormous elemental features and alloy properties[9]. These include but are not limited to composition, atom size ratio, cohesive energy, pairwise and many-body interactions, and their couplings. It is even impossible to sample the full space for a binary system by traditional methods. Secondly, with multiple components, there are many (metastable) phases involved during nucleation and growth of the equilibrium crystalline product[10]. These metastable phases can have very complex crystal structures and are hard to be captured by experimental observations. Thirdly, supercooled metallic liquids overall show a disordered state, but there are abundant types of local structures formed[11]. They are favored by either energy or entropy. Understanding the roles of these locally favored structures in glass transition and crystallization of supercooled melts has been becoming a hot topic[12]. Due to these complexities, we are still far from well understanding the crucial factors that govern MG formation.

In recent years, it is quite encouraging that advanced high throughput sputtering experimental technique has shown its capability of synthesizing a library of $\sim 1,000$ compositions at the same time[4,5]. By sputtering from multiple targets, a thin film of a system with continuous gradient composition is generated. These libraries are likely a good starting point for mining the GFA data by big data methods. However, because of the large species and compositional space, the experimental datasets can be rather sparse. The intercorrelation between different libraries is obscure to understand the data. This will make either building the physical model or predicting new materials challenging. The sophisticated design of the datasets (hence the experiments) is important for material prediction.

To create a large dataset of GFA with continuous controlled parameters change, we have carried out very large-scale molecular dynamics simulations to study the glass formation and crystallization process of binary alloys in recent years[13–17]. On the one hand, by carefully analyzing the crystallization kinetics of supercooled metallic liquids, the thermodynamic factor, interfacial energy, has been identified as the key to controlling the crystallization rate and thus the GFA[15]. At the microscopic scale, the competing ordering effect is crucial to determine the interfacial energy. In principle, the stronger the crystal-like preorder is frustrated by some locally favored structures with incompatible symmetries (such as icosahedra), the higher the interfacial energy will be[15,17]. Hence, the better GFA can be expected. The topological and especially the chemical properties of these local structures are very crucial in determining the interfacial energy. Furthermore, by tuning the local structures so as to decrease the wettability of the preorder at the liquid-crystal interface, the crystallization speed can be manipulated over several orders of magnitude[17]. It is more interesting to find that the preorder is very crucial not only in crystal nucleation but also in the crystal growth process. Accordingly, a critical modification has been proposed to the classical nucleation theory. On the other hand, by characterizing the $R_c$ (i.e., GFA) of binary alloys in a large parameter space, we have studied how different elemental features and alloy properties affect $R_c$[13,14,16]. These include the atomic size mismatch, cohesive energy, mixing energy, and "atomic symmetry". It is surprisingly found that local chemical ordering plays a deterministic role in $R_c$. In most of the previous studies, such as Cheng *et al.*[18] and Laws *et al.*[19], the metallic glasses are usually treated as hard-sphere-like systems where dense packing is critical. The local atomic packings, especially local

icosahedral order, have been considered as the most important factors in glass formation. However, in recent years, we performed systematical studies on the glass formation from model alloys [13–17] and found that the local chemical ordering can be more important in glass formation than previously thought. It can outperform local icosahedral orderings even when the size mismatch is considerable. Because of the multi-component nature, the atomic interactions are more complex, and the local chemical ordering can be more significant. The aging or cooling behavior is important in determining local chemical ordering by controlling atomic diffusion. Generally, the atomic rearrangements during structural relaxation towards the local equilibrium control local chemical ordering. Macroscopically, it will depend on the energetic parameters, atomic sizes, and composition. In addition, the competition among crystalline symmetry is much weaker than that between crystalline symmetry and crystal-incompatible symmetry. This explains why icosahedral clusters are usually found in metallic glass formers. These studies refresh the current understanding of the physical mechanisms of crystallization and glass formation and provide guidelines for experimental glass design. Nevertheless, we have never mined the data itself and built reliable models to predict new glasses. That falls into the efforts of the current work.

In this paper, we are going to utilize the supervised machine learning method to dig into the simulation dataset and try to build an optimized model to predict new binary glasses. Since the particle size ratio is helpful in grouping our dataset, we use the "out-of-group" strategy to make predictions. That is, we leave out a subgroup of samples with a specific particle size ratio and make predictions for them. Since these data are completely independent of the others and have not been seen by the training model, we can treat them as "new". More importantly, we aim to unveil the key features (factors) that determine the GFA of binary alloys. We find that non-linear coupling of the elemental features and alloy properties is critical in glass formation. In more detail, the GFA does not depend on the basic elemental features individually and additively; instead, it depends on the various non-linear couplings of them. The interactions of these basic elemental features to different polynomial degrees are more important in making good predictions. These interaction terms have never been identified previously and can serve as guidelines for future model development and experimental glass design. Therefore, the results will provide new insights into unravelling the physical mechanism of glass formation and help accelerate future material design.

## 2. METHODS

### 2.1. Molecular dynamics simulations

To generate a clean GFA dataset, we started from the simple binary models with Lennard-Jones potential:

$$V_{\alpha\beta}(r_{ij}) = 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{6}\right], \quad (1)$$

where $\alpha, \beta$ indicate which of the particles (A or B) are interacting and $r_{ij}$ is the separation between particles $i$ and $j$. All the simulations were performed with periodic boundary conditions in all directions. The cubic simulation box contains $N = 2000$ particles of equal mass $m$. We studied both monodisperse system ($\sigma_{AA} = \sigma_{BB} = \sigma_{AB}$) and additive bidisperse system ($\sigma_{AB} = (\sigma_{AA} + \sigma_{BB})/2$). Instead, we tuned the interparticle interaction strengths ($\epsilon_{AA}$, $\epsilon_{BB}$ and $\epsilon_{AB}$) widely. We keep $\epsilon_{BB} \leq \epsilon_{AA} = 1.0$ to differentiate the species. The sampling library is exemplified in Figure 1 as a two-dimensional function of $(\epsilon_{BB} - \epsilon_{AA})/(\epsilon_{BB} + \epsilon_{AA})$ and $2\epsilon_{AB}/(\epsilon_{BB} + \epsilon_{AA})$. These two variables take both the same species and inter-species interactions into consideration. We then simulated the systems with size mismatch within 5% so that they can mostly crystallize in the computational time scale. We also sample different compositions $f_B$ (the fraction of B particles in the total number) from 0.1 to 0.9 in an interval of 0.1. Because of the broad energetic preferences, we employed $NPT$ (constant number, constant pressure, constant temperature) ensemble with $P = 10$ to avoid cavitation in any system. To map the experiments, we quenched the high-temperature liquid to very low temperatures at a series of cooling rates $R$. In this way, the critical cooling rate can be quantified after characterizing the crystal

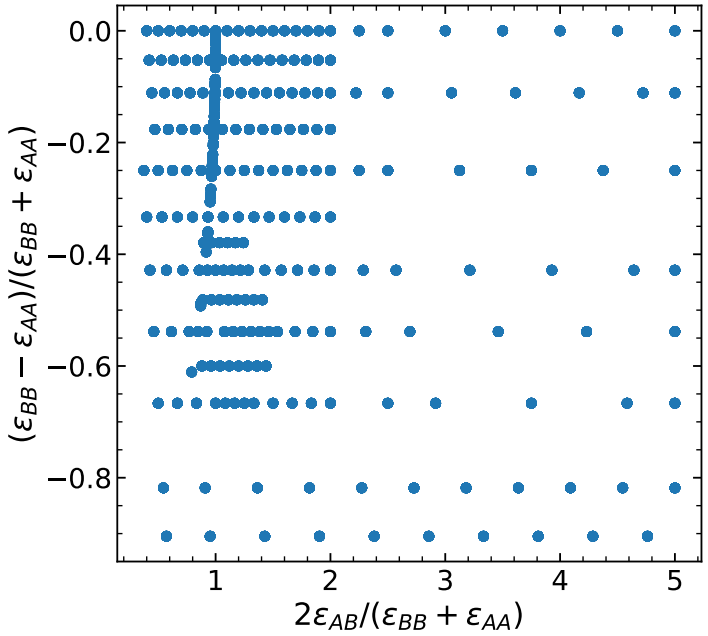


**Figure 1.** Sampling library of the energetic parameters for the pairwise Lennard-Jones binary systems.

fraction at each $R$. 10 independent simulations are generally performed for better statistics. Note that crystallizing a good glass former can take a very long time and even out the capability of the computational power. The basic units for energy, length, and mass are $\epsilon_{\mathrm{AA}}$, $\sigma_{\mathrm{AA}}$, and $m$, respectively. The pressure, temperature, and time scale are reported in reduced units of $\epsilon_{AA}/\sigma_{AA}^3$, $\epsilon_{AA}/k_B$, and $\sqrt{m\sigma_{AA}^2/\epsilon_{AA}}$, where $k_{\mathrm{B}}$ is the Boltzmann constant. The derived units for $R_{\mathrm{c}}$, $\rho$, and $\Delta H_{\mathrm{imix}}$ are $\sqrt{\epsilon_{\mathrm{AA}}^3/m\sigma_{AA}^2k_{\mathrm{B}}^2}$, $m/\sigma_{\mathrm{AA}}^3$, and $\epsilon_{\mathrm{AA}}/k_{\mathrm{B}}$, respectively. More details about the technical details are available in our previous works[13,14,16].

In addition to the above grid search of binary systems, we also performed extensive simulations to simulate many binary systems inspired by experiments[9]. The detailed information of these binary systems is provided in Table 1. Based on the experimental values of the elemental features, including particle size, cohesive energy and mass, we map them to the reduced units and ran the simulations. During the mapping, we also keep $\epsilon_{AA} \geq \epsilon_{BB}$. To capture the inter-species interactions, we follow the classical London's rule that $\epsilon_{AB} = \sqrt{\epsilon_{AA}\epsilon_{BB}}$ and set $\sigma_{AB} = (\sigma_{AA} + \sigma_{BB})/2$[13]. In this way, we can include more realistic models and explore a larger parameter space. To better display the data, we plot the data in Figure 2 in two dimensions with respect to $(\epsilon_{BB} - \epsilon_{AA})/(\epsilon_{BB} + \epsilon_{AA})$ and $\sigma_{BB}/\sigma_{AA}$. Obviously, the parameter ranges are quite broad for binary MGs and those with different GFA can overlap in the two-dimensional space. This indicates that the GFA problem is not single-parameter deterministic. The energetic parameters and geometrical one may couple in a higher order. We should emphasize that this set of simulations is not aiming to compare directly to experiments to model each specific system. Instead, we hope to explore a larger space with some sort of connection to experiments. Furthermore, an alloy with an element having both larger cohesive energy and particle size than the other has a higher probability of becoming glass. From Table 1, there are 40 samples out of the total (62) that fall into this group. If we include the mass comparison, this number decreases to 31 out of 62. This demonstrates the neutral effect of particle masses. Meanwhile, there is also a higher chance for glass formation when both elements are metals (39/62). These insights are helpful for future experimental glass design, but are still subject to the small number of binary MGs being developed.

**Table 1. Properties of binary systems explored by experiments**

| A | B | $\epsilon_A$ (eV/atom) | $r_A$ (Å) | $m_A$ (amu, g/mol) | $\epsilon_{BB}/\epsilon_{AA}$ | $r_B/r_A$ | $m_B/m_A$ | Glass type | Condition1 | Condition2 | Condition3 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ag | Ca | 2.95 | 1.44 | 107.87 | 0.6237 | 1.3681 | 0.372 | ribbon | 0 | 0 | 1 |
| Si | Ag | 4.63 | 1.1 | 28.085 | 0.6371 | 1.3091 | 3.841 | ribbon | 0 | 0 | 0 |
| Al | Ca | 3.39 | 1.4 | 26.982 | 0.5428 | 1.4071 | 1.485 | bulk | 0 | 0 | 1 |
| Cu | Al | 3.49 | 1.26 | 63.546 | 0.9713 | 1.1111 | 0.425 | ribbon | 0 | 0 | 1 |
| Zr | Al | 6.25 | 1.58 | 91.224 | 0.5424 | 0.8861 | 0.296 | film | 1 | 1 | 1 |
| Ge | Au | 3.85 | 1.14 | 72.63 | 0.9896 | 1.2544 | 2.712 | ribbon | 0 | 0 | 0 |
| Hf | Au | 6.44 | 1.58 | 178.49 | 0.5916 | 0.9051 | 1.104 | ribbon | 1 | 0 | 1 |
| Si | Au | 4.63 | 1.1 | 28.085 | 0.8229 | 1.3 | 7.013 | ribbon | 0 | 0 | 0 |
| Au | Sn | 3.81 | 1.43 | 196.97 | 0.8241 | 1.0839 | 0.603 | ribbon | 0 | 0 | 1 |
| Zr | Au | 6.25 | 1.58 | 91.224 | 0.6096 | 0.9051 | 2.159 | ribbon | 1 | 0 | 1 |
| B | Co | 5.81 | 0.88 | 10.81 | 0.7556 | 1.4205 | 5.452 | ribbon | 0 | 0 | 0 |
| B | Fe | 5.81 | 0.88 | 10.81 | 0.7367 | 1.4205 | 5.166 | ribbon | 0 | 0 | 0 |
| B | Ni | 5.81 | 0.88 | 10.81 | 0.7642 | 1.4318 | 5.430 | ribbon | 0 | 0 | 0 |
| Ti | Be | 4.85 | 1.42 | 47.867 | 0.6845 | 0.8169 | 0.188 | ribbon | 1 | 1 | 1 |
| Zr | Be | 6.25 | 1.58 | 91.224 | 0.5312 | 0.7342 | 0.099 | ribbon | 1 | 1 | 1 |
| Ca | Mg | 1.84 | 1.97 | 40.078 | 0.8207 | 0.8122 | 0.606 | ribbon | 1 | 1 | 1 |
| Ca | Zn | 1.84 | 1.97 | 40.078 | 0.7337 | 0.7056 | 1.631 | ribbon | 1 | 0 | 1 |
| Hf | Co | 6.44 | 1.58 | 178.49 | 0.6817 | 0.7911 | 0.330 | ribbon | 1 | 1 | 1 |
| Co | P | 4.39 | 1.25 | 58.933 | 0.7813 | 0.848 | 0.526 | ribbon | 1 | 1 | 0 |
| Ti | Co | 4.85 | 1.42 | 47.867 | 0.9052 | 0.8803 | 1.231 | bulk | 1 | 0 | 1 |
| Zr | Co | 6.25 | 1.58 | 91.224 | 0.7024 | 0.7911 | 0.646 | ribbon | 1 | 1 | 1 |
| Hf | Cu | 6.44 | 1.58 | 178.49 | 0.5419 | 0.7975 | 0.356 | bulk | 1 | 1 | 1 |
| Cu | Mg | 3.49 | 1.26 | 63.546 | 0.4327 | 1.2698 | 0.382 | ribbon | 0 | 0 | 1 |
| Ti | Cu | 4.85 | 1.42 | 47.867 | 0.7196 | 0.8873 | 1.328 | ribbon | 1 | 0 | 1 |
| Y | Cu | 4.37 | 1.8 | 88.906 | 0.7986 | 0.7 | 0.715 | ribbon | 1 | 1 | 1 |
| Zr | Cu | 6.25 | 1.58 | 91.224 | 0.5584 | 0.7975 | 0.697 | bulk | 1 | 1 | 1 |
| Hf | Fe | 6.44 | 1.58 | 178.49 | 0.6646 | 0.7911 | 0.313 | film | 1 | 1 | 1 |
| Fe | P | 4.28 | 1.25 | 55.845 | 0.8014 | 0.848 | 0.555 | ribbon | 1 | 1 | 0 |
| Si | Fe | 4.63 | 1.1 | 28.085 | 0.9244 | 1.1364 | 1.988 | film | 0 | 0 | 0 |
| Fe | Sn | 4.28 | 1.25 | 55.845 | 0.7336 | 1.24 | 2.126 | film | 0 | 0 | 1 |
| Zr | Fe | 6.25 | 1.58 | 91.224 | 0.6848 | 0.7911 | 0.612 | ribbon | 1 | 1 | 1 |
| Hf | Ge | 6.44 | 1.58 | 178.49 | 0.5978 | 0.7215 | 0.407 | ribbon | 1 | 1 | 0 |
| Pd | Ge | 3.89 | 1.4 | 106.42 | 0.9897 | 0.8143 | 0.682 | ribbon | 1 | 1 | 0 |
| Pt | Ge | 5.84 | 1.39 | 195.08 | 0.6592 | 0.8201 | 0.372 | ribbon | 1 | 1 | 0 |
| Zr | Ge | 6.25 | 1.58 | 91.224 | 0.616 | 0.7215 | 0.796 | ribbon | 1 | 1 | 0 |
| Hf | Ni | 6.44 | 1.58 | 178.49 | 0.6894 | 0.7975 | 0.329 | ribbon | 1 | 1 | 1 |
| Hf | Si | 6.44 | 1.58 | 178.49 | 0.7189 | 0.6962 | 0.157 | ribbon | 1 | 1 | 0 |
| Ni | Mg | 4.44 | 1.26 | 58.693 | 0.3401 | 1.2698 | 0.414 | ribbon | 0 | 0 | 1 |
| Y | Mg | 4.37 | 1.8 | 88.906 | 0.3455 | 0.8889 | 0.273 | ribbon | 1 | 1 | 1 |
| Mg | Zn | 1.51 | 1.6 | 24.305 | 0.894 | 0.8687 | 2.690 | ribbon | 1 | 0 | 1 |
| Zr | Mg | 6.25 | 1.58 | 91.224 | 0.2416 | 1.0127 | 0.266 | ribbon | 0 | 0 | 1 |
| Si | Mn | 4.63 | 1.1 | 28.085 | 0.6307 | 1.2 | 1.956 | ribbon | 0 | 0 | 0 |
| Zr | Mn | 6.25 | 1.58 | 91.224 | 0.4672 | 0.8354 | 0.602 | film | 1 | 1 | 1 |
| Mo | Zr | 6.82 | 1.39 | 95.95 | 0.9164 | 1.1367 | 0.951 | ribbon | 0 | 0 | 1 |
| Nb | Ni | 7.57 | 1.5 | 92.906 | 0.5865 | 0.84 | 0.632 | bulk | 1 | 1 | 1 |
| Nb | Si | 7.57 | 1.5 | 92.906 | 0.6116 | 0.7333 | 0.302 | ribbon | 1 | 1 | 0 |
| Ni | P | 4.44 | 1.26 | 58.693 | 0.7725 | 0.8413 | 0.528 | ribbon | 1 | 1 | 0 |
| Ni | Pd | 4.44 | 1.26 | 58.693 | 0.8761 | 1.1111 | 1.813 | film | 0 | 0 | 1 |
| Ti | Ni | 4.85 | 1.42 | 47.867 | 0.9155 | 0.8873 | 1.226 | film | 1 | 0 | 1 |
| Ta | Ni | 8.1 | 1.54 | 180.95 | 0.5481 | 0.8182 | 0.324 | ribbon | 1 | 1 | 1 |
| Ni | Y | 4.44 | 1.26 | 58.693 | 0.9842 | 1.4286 | 1.515 | film | 0 | 0 | 1 |
| Zr | Ni | 6.25 | 1.58 | 91.224 | 0.7104 | 0.7975 | 0.643 | ribbon | 1 | 1 | 1 |
| Pd | P | 3.89 | 1.4 | 106.42 | 0.8817 | 0.7571 | 0.291 | film | 1 | 1 | 0 |
| Pt | P | 5.84 | 1.39 | 195.08 | 0.5873 | 0.7626 | 0.159 | ribbon | 1 | 1 | 0 |
| Si | Pd | 4.63 | 1.1 | 28.085 | 0.8402 | 1.2727 | 3.789 | bulk | 0 | 0 | 0 |
| Zr | Pd | 6.25 | 1.58 | 91.224 | 0.6224 | 0.8861 | 1.167 | ribbon | 1 | 0 | 1 |
| Pt | Sb | 5.84 | 1.39 | 195.08 | 0.4709 | 1.1151 | 0.624 | ribbon | 0 | 0 | 1 |
| Pt | Si | 5.84 | 1.39 | 195.08 | 0.7928 | 0.7914 | 0.144 | ribbon | 1 | 1 | 0 |
| Pt | Ti | 5.84 | 1.39 | 195.08 | 0.8305 | 1.0216 | 0.245 | ribbon | 0 | 0 | 1 |
| Zr | Pt | 6.25 | 1.58 | 91.224 | 0.9344 | 0.8797 | 2.138 | ribbon | 1 | 0 | 1 |
| Ti | Si | 4.85 | 1.42 | 47.867 | 0.9546 | 0.7746 | 0.587 | ribbon | 1 | 1 | 0 |
| Zr | Si | 6.25 | 1.58 | 91.224 | 0.7408 | 0.6962 | 0.308 | ribbon | 1 | 1 | 0 |

Condition1: $\epsilon_{AA} \geq \epsilon_{BB}$ and $\sigma_{AA} \geq \sigma_{BB}$; Condition2: $\epsilon_{AA} \geq \epsilon_{BB}$ and $\sigma_{AA} \geq \sigma_{BB}$ and $m_A \geq m_B$ ; Condition3: both elements A and B are metals

## 2.2. Measuring glass-forming ability

After obtaining the quenched solids from the computer simulations, we employed bond orientational order parameters to characterize the local structures[11,20]. Taking any particle in the simulation box, the nearest

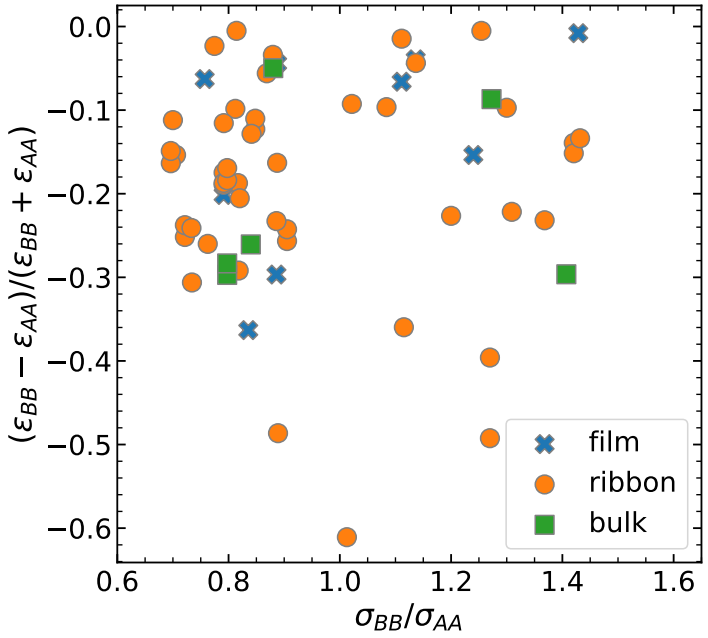


**Figure 2.** Data exploration of experimental binary alloys grouped by the glass type. The geometrical parameter and the energetic parameter are considered for comparison.

neighbors of each particle are obtained by radical Voronoi tessellation[21]. We calculate the bond orientational order parameter $q_{6m}(i)$ for each particle $i$:

$$q_{6m}(i) = \sum_{j=1}^{N_i} \frac{A_j}{A_{\text{tot}}^i} Y_{6m}(\theta(\mathbf{r}_{ij}), \phi(\mathbf{r}_{ij})), \tag{2}$$

where $N_i$ is the number of nearest Voronoi neighbors of particle $i$, $Y_{6m}(\theta(\mathbf{r}_{ij}, \phi(\mathbf{r}_{ij}))$ is the spherical harmonic function of degree 6 and order $m$, and $\theta$ and $\phi$ are the polar and azimuthal angles. The contribution from the spherical harmonics of each neighbor $j$ of particle $i$ is weighted by the fraction $A_j/A_{\text{tot}}^i$ of the area of the Voronoi face separating the two particles to the total area of all faces $A_{\text{tot}}^i$ of the polyhedron surrounding particle $i$. We determine the number of crystal-like atoms by calculating the correlations in the bond orientational order parameter:

$$s_6(i,j) = \frac{\sum_{m=-6}^{6} q_{6m}(i) q_{6m}^*(j)}{\sqrt{\sum_{m=-6}^{6} |q_{6m}(i)|^2 \sum_{m=-6}^{6} |q_{6m}(j)|^2}}, \tag{3}$$

where $q_{6m}^*(j)$ is the complex conjugate of $q_{6m}(j)$. If $s_6(i,j) > 0.7$, we treat the bond as crystal-like[22]. If the total number of crystal-like bonds for a given particle is larger than 10, the particle is considered to be in a crystalline environment. The sensitivity of the thresholds for $s_6(i,j)$ and the number of crystal-like bonds have been studied previously[22,23]. For each set of size ratios and energetic parameters, we calculate the fraction of crystalline particles $f_{\text{xtal}}$ as a function of the cooling rate $R$. Then we use the following function to model the rate-dependent $f_{\text{xtal}}$ and estimate the critical cooling rate $R_c$ when $f_{\text{xtal}} = 0.5$.

$$f_{\text{xtal}} = \frac{1}{2}\left(1 - \tanh\left[\log_{10}(R/R_c)^{-\kappa}\right]\right), \tag{4}$$

where $0 < \kappa < 1$ is the stretching exponent[13,14,16].

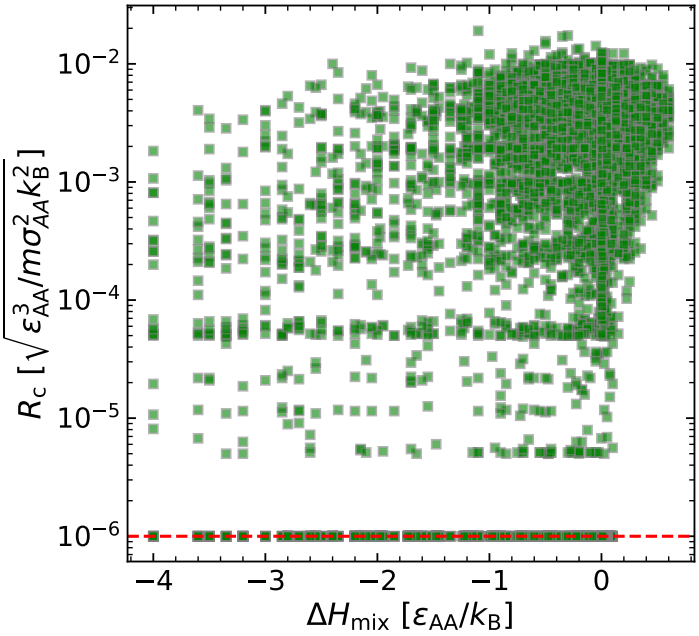


**Figure 3.** The relationship between the critical cooling rate $R_c$ and the heat of mixing $\Delta H_{mix}$. The samples with $R_c = 10^{-6}$ do not crystallize in the longest computational timescale. They are shown here for convenience.

## 3. RESULTS

We finish the above computer simulations by using several years' computational time by hundreds of CPU cores in parallel. By characterizing the local structures of each sample, we obtained a big dataset consisting of 7688 samples. Fortunately, there are only about 25% samples that have not crystallized with our lowest cooling rate. The reason why they cannot crystallize is that the computational running time is still limited and they are better glass formers than the others. They are not ultra-stable glasses but will crystallize at an extended simulation timescale. Therefore, we shall create a controllable high-quality dataset.

To explore the GFA dataset, we first try to gain insights with some empirical rules. As is known to all, the heat of mixing $\Delta H_{mix}$ has long been considered as a major parameter in glass formation. Negative heat of mixing becomes one of the central rules in glass formation criteria, including the famous Inoue's rules[24]. This is very important to make sure the multiple species will mix with each other. Otherwise, phase separation will happen, which will strongly deteriorate the GFA. In this sense, a critical question arises: is there any quantitative correlation between GFA and $\Delta H_{mix}$ or negative $\Delta H_{mix}$ is only a necessity for glass formation? To answer this question, we hence calculate the $\Delta H_{mix}$ for all the samples and show its relationship to the critical cooling rate $R_c$ in Figure 3. Interestingly, there is no quantitative correlation between them, even though a major of glass formers have negative heats of mixing. This finding has been corroborated by experimental data previously[13]. The ones with $\Delta H_{mix} > 0$ generally are poor glass formers ($R_c \approx 10^{-2.5}$).

Another common factor for glass formation is the density $\rho$. It is expected that a higher-density solid should have denser packing so that the atomic rearrangements are more difficult, which will impede crystallization. This relationship was observed in typical Cu-Zr binary systems[25]. Here we also check this behavior in our Lennard-Jones systems. To compare the results over all the studied systems, we take the density of the glassy solid fabricated by $R = 10^{-2}$. Figure 4 shows the scatter plot of $R_c$ as a function of $\rho$. Obviously, no quantitative relationship exists between them for the studied systems.

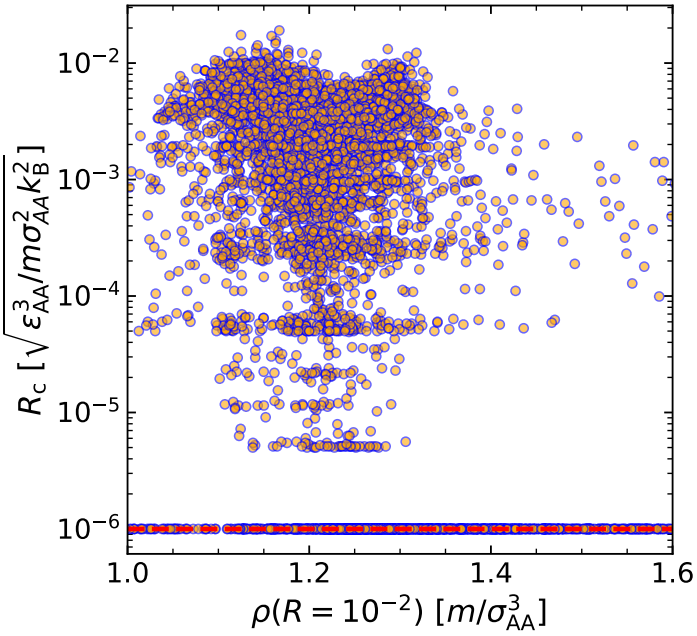


**Figure 4.** The relationship between the critical cooling rate $R_c$ and the low temperature glass density $\rho$. The samples with $R_c = 10^{-6}$ do not crystallize in the longest computational timescale. They are shown here for convenience.

These findings demonstrate that the GFA of an alloy is more complex than from some single parameters. We need to figure out the high-order correlations between the GFA and the elemental and alloy properties. Nowadays, there is no explicit function that can be used for this purpose. To better define the function and get the main factors, we should turn our attention to some advanced big data analysis techniques, for example, supervised machine learning. This methodology will enable us to explore such a kind of relationship without knowing the function *in priori*.

Machine learning has become an innovative tool to explore big datasets and make predictions based on the known features[26–28]. It has been applied in enormous numbers of fields, including materials science[29–33]. Meanwhile, many advanced theoretical models have been developed for different application cases. In this study, we are trying to explore the simulation GFA datasets and gain physical insights into glass formation. Usually, machine learning is likely a black box for users with high-dimensional inputs. The designed model with a specific algorithm will take care of the mathematical relationships from the input features to the labels. The nonlinearity involved is hard to explain in a physical manner. Here, we start with a simple model with a small number of features so as to capture all the details.

Considering our simple model systems, there are several independent variables. Namely, $\epsilon_{BB}/\epsilon_{AA}$, $\epsilon_{AB}/\epsilon_{AA}$, $\sigma_{BB}/\sigma_{AA}$, $f_B$. Thus, we are trying to build a minimal number of basic features for the machine learning model. Based on the previous understandings of the GFA, four fundamental features are thus considered: $\varepsilon_1 = (\epsilon_{BB}-\epsilon_{AA})/(\epsilon_{BB}+\epsilon_{AA})$ and $\varepsilon_2 = 2\epsilon_{AB}/(\epsilon_{BB}+\epsilon_{AA})$, $\varsigma = (\sigma_{BB}-\sigma_{AA})/\sigma_{AB}$, and $f_B$. These features mainly consider how different the two components are. Under the well-mixed condition, the more different the two species are, the stronger frustration towards crystallization will be. Hence, the GFA will be elevated.

With the critical cooling rates (in log scale, $\log_{10} R_c$) as the label, our machine learning problem is intrinsically a regression problem. We constructed the machine learning methods by employing the open-source Scikit-learn package[34]. Before building the model, the input features need cleaning and preprocessing. As a first step, we recall that about 25% of the samples cannot be crystallized in the simulation timescale. Therefore, they are not

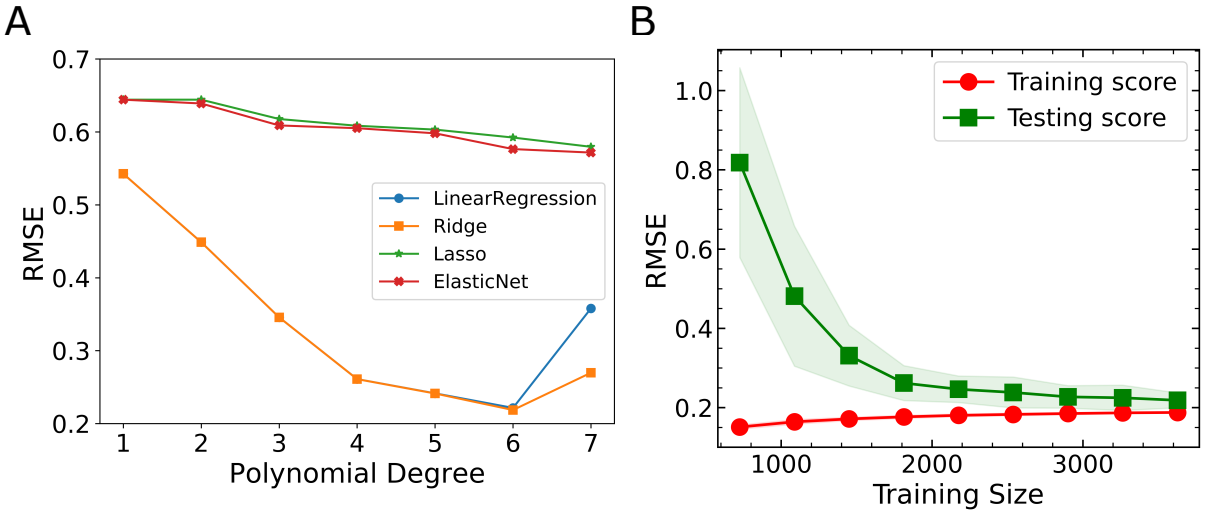


**Figure 5.** Machine learning model optimizations. (A) Cross-validation scores for various learning models with polynomial degrees up to 7. (B) The learning curve of the Ridge model. The training score and cross-validation score are compared with different training sizes. Both of them tend to saturate and merge at large training sizes.

good for the training, so they are removed from the dataset for the subsequent process. Then we standardize the features by removing the mean and scaling to unit variance. In this way, all the input data behaves like standard normal distribution. This data processing step is key to reducing the bias for many machine learning models, including the ones we will use below.

Since we already figure out from the above discussion that GFA is not a simple linear single parameter problem, we are trying to build higher-order correlations of these basic features. To this end, we build high-dimensional features from polynomial extrapolation. In detail, we generate polynomial and interaction features from the four basic features. The new feature matrix will thus consist of all polynomial combinations of the basic features with a degree less than or equal to the specified degree. In this way, we can capture not only the nonlinearity but also the feature interactions. The higher order is, the more input features will be. Meanwhile, the risk of overfitting will also increase. Starting from these polynomial features, we hope to train a linear model to map the features to the labels. We thus compare several linear models, including basic linear regression, and their derivatives with different regularizations. For example, Ridge regression includes the L2 regularization on the size of the coefficients, while Lasso regression imposes L1 regularization. By adding both L1 and L2-norm regularization, an ElasticNet model can be trained. To create a workflow, we build a pipeline from feature engineering, model construction and cross validation, covering different degrees of polynomials and linear algorithms. During the training, 10-fold cross validation is chosen for optimization. The root mean squared error (RMSE) between the real values and the predicted values is minimized. Figure 5A shows the comparison of the performance of different training models. For Lasso and ElasticNet, where feature selection is automatically involved by L1 regularization, the models always under-fit the training data and thus their RMSE is much higher. In addition, with an increasingly large number of features (from degree 1 to 7) fed to the training model, their performance is not much improved. These models are very aggressive in feature reduction and cannot pick up important high-degree features. This demonstrates their improbability in solving the current issue.

We then turn to the basic linear regression and Ridge regression models. They are behaving similarly, except that Ridge did a better job when the polynomial degree was 6. We first emphasize that the RMSE in Figure 5A is from the 10-fold cross-validations for the training model on the test sub-dataset. For machine learning models, with increasing model complexity, the bias will decrease while variance can greatly increase. There

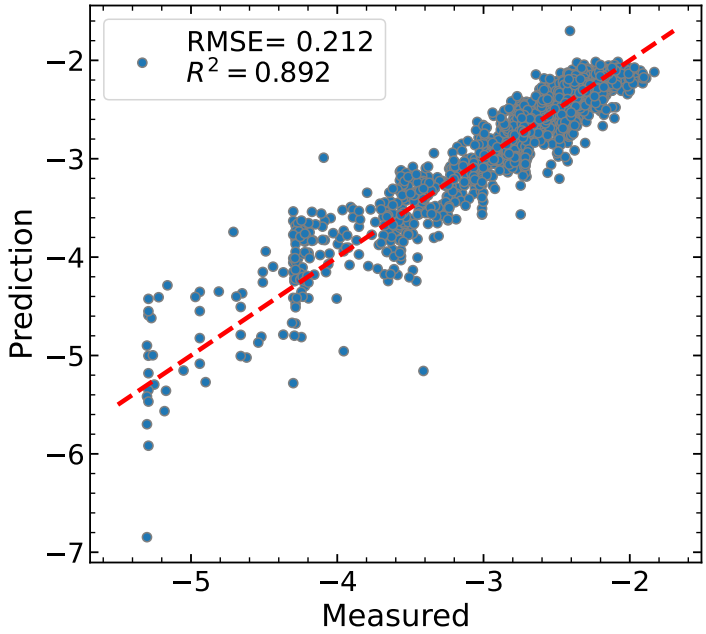


**Figure 6.** Comparison of the machine learning predicted $R_c$ with the simulation measurements. A linear fit (red dashed line) is included for illustration.

will be a variance-bias trade-off. With small polynomial degrees, the bias can be rather high, but the variance can be small. The model under-fits the training data and thus cannot capture the test data trends well. The performance can be improved by increasing the model complexity. The minimum at polynomial degree 6 indicates the best performance with a reasonable variance-bias trade-off. With a further increasing degree (> 6), the bias can be small, but the variance can be rather high. The model will be too complex so as to over-fit the training data but cannot well generalize to the test data. That is why RMSE shows a V-shape. Further increasing to degree 8 will greatly increase RMSE, especially for the linear regression model. Given the size of the dataset and consideration of the degree of freedom, any degree that is higher than 8 is not practical. Thus, we believe a model with polynomials at degree 6 is the global optimum.Therefore, we would expect that the optimized model is Ridge regression with a six-degree polynomial. There are hence around 210 features derived from the four basic features. Specifically, this algorithm is to estimate the coefficient set $\{\beta_0, \beta_1, \beta_2, ..., \beta_p\}$ that minimizes the loss function

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}\beta_j^2, \tag{5}$$

where the dataset has $n$ observations with $p$ predictors (features), $x_{ij}$ is the $j^{\text{th}}$ predictor of the $i^{\text{th}}$ feature, $y_i$ is the corresponding label, and $\lambda$ is the non-negative regularization strength. To account for the overfitting probability, we characterize the learning curve of the optimized model. Basically, a subset of the original data will be generated internally for training and the rest for testing. With 10-fold cross-validations, the model is trained with different training sizes, and its performance is plotted in Figure 5B. Remarkably, with increasing training size, the training score is only slightly worse, but the testing performance is dramatically improved. Both scores tend to saturate and merge at ~ 3000 training data. This excludes the overfitting risk in our machine-learning model.

Now we come to the most important step of machine learning, namely making predictions on unseen data. For our purpose, we leave out a subgroup of data with a specific $\sigma_{BB}/\sigma_{AA}$ before training (e.g., out-of-group prediction). This aims to avoid interpolation in the machine learning model and make sure the independent

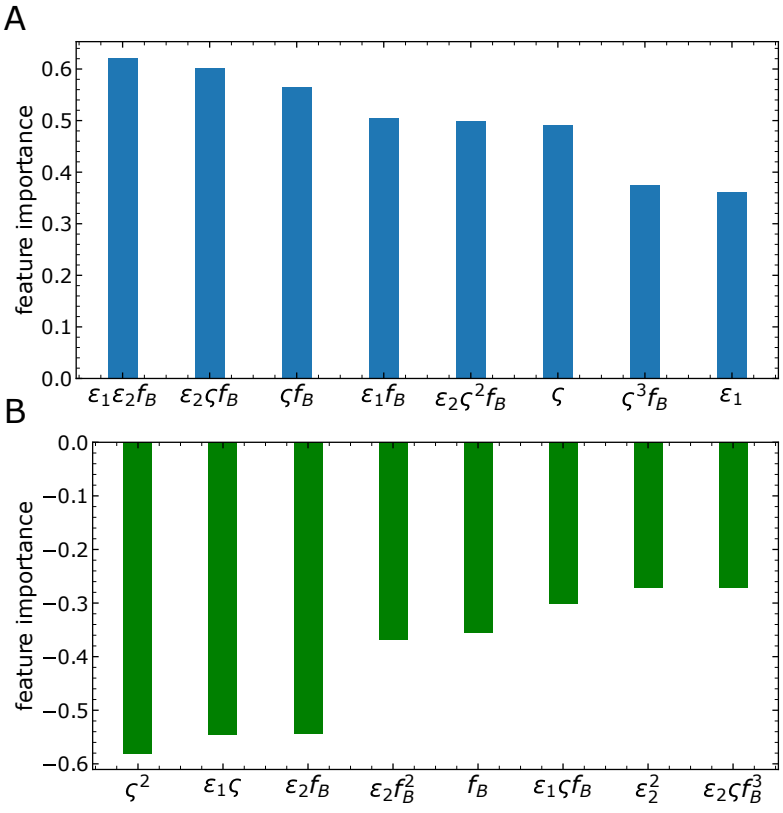


**Figure 7.** Feature importance analysis of the optimized machine learning model. (A) Positive correlations. (B) Negative correlations.

testing data has never been known by the training model. In this way, we show the predicted $R_c$ against the measured $R_c$ in Figure 6. It is quite encouraging that the RMSE is only around 0.212 and a direct linear fitting gives $R^2 \approx 0.9$. These demonstrate the reliability of our training model to predict new glass formers in the world of computer simulations. The larger fluctuation for better glass formers (lower $R_c$) is because of the greater difficulty in measuring the accurate $R_c$. Some of these binary alloys can require a longer time to crystallize beyond the current computational power. The machine learning model will be helpful in predicting new materials with enhanced GFA for the study of glass physics. For instance, the Kob-Andersen model has been the most popular model for glass study in the past three decades[35]. It was assumed as a very good glass former. However, recent studies show that with a larger simulation box and GPU acceleration, the Kob-Andersen model is vulnerable to crystallization[36,37]. It is actually a poor glass former. The non-additivity and non-classical energy mixing make the model difficult to understand. A simpler yet better model is of great interest for glass studies.

We note that there are a variety of machine learning algorithms available and many of them have been applied in materials development[33,38,39], such as linear models with regularization, tree-based models, and neural networks. For instance, the iterative random forest model has been studied widely in classifying glass formation and regressing glass properties[33,39]. These complex models already showed prediction power, but suffered from a higher risk of over-fitting. In addition, the interpretability will drastically decrease with increasing model complexity. Therefore, in this study, we choose to start from the simple linear model with non-linear combinations of basic features instead of complex, non-linear fancy models, such as artificial neural networks. We hope to better extract the important couplings of these basic features in glass formation. Interestingly, we found that the couplings of $\varepsilon_1$, $\varepsilon_2$, $\varsigma$ and $f_B$ are rather crucial in determining the GFA. In Figure 7, we plot the most important features in the linear model. We split them into two graphs based on whether it is positively or

negatively correlated with $R_c$. It is surprising to see that the energetic parameter ($\varepsilon_1\varepsilon_2 f_B$) is the most critical one, without the particle size information. This may provide a plausible explanation for why local chemical ordering is so important in glass formation [13,15,16]. This top feature is then followed by the couplings of the particle size and the energetics. These results clarified the critical features in glass formation, at least for binary systems. They may provide further insights for future experimental glass design. Note that having a binary alloy with exceptional GFA will be ideal for glass study and applications.

We note that bridging the important model features in the current study to those determined in experiments is interesting and important. From the work by Liu *et al.* [40], it was surprisingly found that the 'random' feature generation from elemental features without enough physical insights is insufficient in machine learning. Overfitting is readily there by feeding those high-dimensional features to a non-linear random forest algorithm. Instead, a model with only three features sophisticatedly derived from both elemental and alloy features provides some predictability. Similarly, in the work by You *et al.* [41], an artificial neural network model can classify crystalline versus amorphous phases in high-throughput fabrications by using a small number of elemental and alloy features, especially from excess electrical resistivity. The significance of these alloy properties unambiguously demonstrates the non-trivial couplings of elemental properties in metallic alloys. These studies convey critical messages. On the one hand, physics-driven features from elemental and alloy properties are significant. On the other hand, how the coupling of elemental properties to determine the alloy property is crucial in feature engineering. The current study is in line with these spirits: we first identified the four fundamental physics-driven features. They consider the energetic interactions, atomic sizes, and compositions, which are consistent with the experimental inputs. Furthermore, we identified their critical couplings [Figure 7] that may correlate with some alloy properties. How to directly map these fundamental model parameters in Figure 7 to experimentally measurable quantities, such as electrical resistivity and liquidus temperature, is interesting for future study.

## 4. SUMMARY

The glass-forming ability has been one of the central mysteries for MGs, unlike other families of glass. The critical cooling rates of MGs can differ by more than 10 orders of magnitude. This huge time gap has fascinated glass researchers to explore the underlying physics and to design new materials with desired properties. To accelerate MG development, we need to understand the physical mechanisms of glass formation and learn from the existing big data accumulated so far. In this study, without relying on collecting experimental data from the literature sea, we performed large-scale computer simulations over several years to generate a high-quality dataset. Based on the current understanding of these data, we build an optimized physics-based machine learning model with only four basic features. The model is able to make reliable predictions on new substances and provides insights into the most critical features. It is found that the non-linear couplings of the energetic parameters and geometric parameters are key for glass formation. This further demonstrates the complexity of the long-standing GFA issue. More interestingly, the most important factor for glass formation is found to be the coupling of energetic parameters and composition. This rationalizes the crucial role of local chemical ordering in glass formation and crystallization of metallic alloys, which has been overlooked in the past. A deeper understanding of the physics of GFA is desired in the future. Practically, generating and maintaining a high-quality data warehouse for the GFA with extended variables are important for future study. This may require the collaboration of the whole field.

Although here we focus on binary alloys for simulation convenience, the current study can be effectively extended to multi-component materials based on the acquired knowledge. With more components, there will be more independent variables. For example, there will be 12 independent variables for a ternary system. Even though this will greatly increase the sampling difficulty in molecular dynamics simulations, some optimized high-throughput computational strategies may be developed. It would be interesting to see whether

such a model persists or not. Extending the physical mechanism and model prediction of glass formation in single-component and binary systems to multi-component systems is an intrinsically important and intriguing direction for future work. Another interesting related topic would be machine learning study of the phase selection of high-entropy alloys. With mainly five elements of similar size, which is close to one set of our current simulations, high-entropy alloys usually form a finite number of simple crystals. This is an ideal case as a classification problem. The driver of the phase selection and local chemical ordering is the energetic competition. By using a similar simulation protocol or with an advanced patchy particle model, these issues can be well tackled by combining computer simulations and machine learning methods.

## DECLARATIONS

### Acknowledgments

Hu YC has been focusing on the computational study of the glass-forming ability of metallic glasses since he started his postdoc research with Prof. Corey O'Hern at Yale University in 2018. At Yale, he carried out most of the simulations to quantify the critical cooling rates of thousands of systems and studied the statistical physics of the glass-forming ability. Supported by a JSPS fellowship, Y.C.H. has also worked with Prof. Hajime Tanaka at the University of Tokyo to study the atomic-scale structural mechanism of glass formation and crystallization of binary alloys. Without these experiences and thoughts, this work would have never come to fruition. Y.C.H. is very grateful for all the support from all the members of the O'Hern lab and the Tanaka lab. Hu YC acknowledges the technical support from Yale Center for Research Computing. Hu YC thanks Y.C. Wu for carefully reviewing the manuscript before submission.

### Authors' contributions

Proposed and supervised the project, conducted the simulations, built the machine learning model, performed the analysis, and wrote the manuscript: Hu YC

Contributed to generalizing the machine learning codes and performing analyses to respond to the referees: Tian J

### Availability of data and materials

The dataset and the machine learning codes are available at the high-quality dataset and the machine learning package at https://github.com/yuanchaohu/ML_GFA_JMI or at https://github.com/jzt5132/ml_code.

### Financial support and sponsorship

None.

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.



## REFERENCES

1. Klement W, Willens RH, Duwez P. Non-crystalline structure in solidified gold–silicon alloys. *Nature* 1960;187:869–70. DOI
2. Lu ZP, Liu CT. A new glass-forming ability criterion for bulk metallic glasses. *Acta Mater* 2002;50:3501–12. DOI

3. Wang WH, Dong C, Shek CH. Bulk metallic glasses. *Mater Sci Eng: R Rep* 2004;44:45–89. DOI
4. Ding S, Liu Y, Li Y, et al. Combinatorial development of bulk metallic glasses. *Nat Mater* 2014;13:494–500. DOI
5. Li MX, Sun YT, Wang C, et al. Data-driven discovery of a universal indicator for metallic glass forming ability. *Nat Mater* 2022;21:165–72. DOI
6. Wang WH. Dynamic relaxations and relaxation-property relationships in metallic glasses. *Prog Mater Sci* 2019;106:100561. DOI
7. Qiao JC, Wang Q, Pelletier JM, et al. Structural heterogeneities and mechanical behavior of amorphous alloys. *Prog Mater Sci* 2019;104:250–329. DOI
8. Zhang LC, Jia Z, Lyu F, Liang SX, Lu J. A review of catalytic performance of metallic glasses in wastewater treatment: Recent progress and prospects. *Prog Mater Sci* 2019;105:100576. DOI
9. Li Y, Zhao S, Liu Y, Gong P, Schroers J. How many bulk metallic glasses are there? *ACS Comb Sci* 2017;19:687–93. DOI
10. Kurtuldu G, Shamlaye KF, Löffler JF. Metastable quasicrystal-induced nucleation in a bulk glass-forming liquid. *Proc Natl Acad Sci USA* 2018;115:6123–8. DOI
11. Tanaka H. Bond orientational order in liquids: Towards a unified description of water-like anomalies, liquid-liquid transition, glass transition, and crystallization. *Eur Phys J E* 2012;35:1–84. DOI
12. Xie Y, Sohn S, Wang M, et al. Supercluster-coupled crystal growth in metallic glass forming liquids. *Nat Commun* 2019;10:915. DOI
13. Hu YC, Schroers J, Shattuck MD, O'Hern CS. Tuning the glass-forming ability of metallic glasses through energetic frustration. *Phys Rev Mater* 2019;3:085602. DOI
14. Hu YC, Zhang K, Kube SA, et al. Glass formation in binary alloys with different atomic symmetries. *Phys Rev Mater* 2020;4:105602. DOI
15. Hu YC, Tanaka H. Physical origin of glass formation from multicomponent systems. *Sci Adv* 2020;6:eabd2928. DOI
16. Hu YC, Jin W, Schroers J, Shattuck MD, O'Hern CS. Glass-forming ability of binary Lennard-Jones systems. *Phys Rev Mater* 2022;6:075601. DOI
17. Hu YC, Tanaka H. Revealing the role of liquid preordering in crystallisation of supercooled liquids. *Nat Commun* 2022;13:4519. DOI
18. Cheng YQ, Ma E. Atomic-level structure and structure–property relationship in metallic glasses. *Prog Mater Sci* 2011;56:379–473. DOI
19. Laws KJ, Miracle DB, Ferry M. A predictive structural model for bulk metallic glasses. *Nat Commun* 2015 Dec;6:8123. DOI
20. Steinhardt PJ, Nelson DR, Ronchetti M. Bond-orientational order in liquids and glasses. *Phys Rev B* 1983;28:784. DOI
21. Rycroft CH, Grest GS, Landry JW, Bazant MZ. Analysis of granular flow in a pebble-bed nuclear reactor. *Phys Rev E* 2006;74:021306. DOI
22. Rein ten Wolde P, Ruiz-Montero MJ, Frenkel D. Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling. *J Chem Phys* 1996;104:9932–47. DOI
23. Russo J, Tanaka H. The microscopic pathway to crystallization in supercooled liquids. *Sci Rep* 2012;2:505. DOI
24. Takeuchi A, Inoue A. Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element. *Mater Trans* 2005;46:2817–29. DOI
25. Li Y, Guo Q, Kalb JA, Thompson CV. Matching glass-forming ability with the density of the amorphous phase. *Science* 2008;322:1816–9. DOI
26. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature* 2018;559:547–55. DOI
27. Tshitoyan V, Dagdelen J, Weston L, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* 2019;571:95. DOI
28. Friederich P, Häse F, Proppe J, Aspuru-Guzik A. Machine-learned potentials for next-generation matter simulations. *Nat Mater* 2021 Jun;20:750–61. DOI
29. Raccuglia P, Elbert KC, Adler PDF, et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* 2016;533:73–6. DOI
30. Hart GLW, Mueller T, Toher C, Curtarolo S. Machine learning for alloys. *Nat Rev Mater* 2021 Aug;6:730–55. DOI
31. Zhou ZQ, He QF, Liu XD, et al. Rational design of chemically complex metallic glasses by hybrid modeling guided machine learning. *npj Comput Mater* 2021 Aug;7:1–10. DOI
32. Sun YT, Bai HY, Li MZ, Wang WH. Machine learning approach for prediction and understanding of glass-forming ability. *J Phys Chem Lett* 2017 Jul;8:3434–39. DOI
33. Ren F, Ward L, Williams T, et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci Adv* 2018;4:eaaq1566. DOI
34. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 2011;12:2825–30. Available from: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https:/ [Last accessed on 10 Feb 2023]
35. Kob W, Andersen HC. Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture I: The van Hove correlation function. *Phys Rev E* 1995;51:4626–41. DOI
36. Pedersen UR, Schrøder TB, Dyre JC. Phase diagram of Kob-Andersen-type binary Lennard-Jones mixtures. *Phys Rev Lett* 2018;120:165501. DOI
37. Ingebrigtsen TS, Dyre JC, Schrøder TB, Royall CP. Crystallization instability in glass-forming mixtures. *Phys Rev X* 2019;9:031016. DOI
38. Iwasaki Y, Takeuchi I, Stanev V, et al. Machine-learning guided discovery of a new thermoelectric material. *Sci Rep* 2019;9:2751. DOI
39. Sarker S, Tang-Kong R, Schoeppner R, et al. Discovering exceptionally hard and wear-resistant metallic glasses by combining machine-

learning with high throughput experimentation. *Appl Phys Rev* 2022;9:011403. DOI

40. Liu G, Sohn S, Kube SA, et al. Machine learning versus human learning in predicting glass-forming ability of metallic glasses. *Acta Mater* 2023 Jan;243:118497. DOI
41. You D, Zhang H, Ganorkar S, et al. Electrical resistivity as a descriptor for classification of amorphous versus crystalline phases of alloys. *Acta Mater* 2022;231:117861. DOI

Dreger *et al. J Mater Inf* 2023;3:2
**DOI:** 10.20517/jmi.2023.01

**Journal of Materials Informatics**

**Research Article**

**Open Access**

Check for updates

# Synergizing ontologies and graph databases for highly flexible materials-to-device workflow representations

**Max Dreger[1,2], Mohammad J. Eslamibidgoli[1,2], Michael H. Eikerling[1,2,3], Kourosh Malek[1,2,*]**

[1]Theory and Computation of Energy Materials (IEK-13), Institute of Energy and Climate Research, Forschungszentrum Jülich GmbH, Jülich 52425, Germany.
[2]Centre for Advanced Simulation and Analytics (CASA), Simulation and Data Lab for Energy Materials, Forschungszentrum Jülich, Jülich 52425, Germany.
[3]Chair of Theory and Computation of Energy Materials, Faculty of Georesources and Materials Engineering, RWTH Aachen University, Aachen 52062, Germany.

[*]**Correspondence to:** Dr. Kourosh Malek, Theory and Computation of Energy Materials (IEK-13), Institute of Energy and Climate Research, Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, Jülich 52425, Germany. E-mail: k.malek@fz-juelich.de

**How to cite this article:** Dreger M, Eslamibidgoli MJ, Eikerling MH, Malek K. Synergizing ontologies and graph databases for highly flexible materials-to-device workflow representations. *J Mater Inf* 2023;3:2. https://dx.doi.org/10.20517/jmi.2023.01

**Received:** 12 Jan 2023 **First Decision:** 8 Feb 2023 **Revised:** 17 Feb 2023 **Accepted:** 27 Feb 2023 **Published:** 6 Mar 2023

**Academic Editors:** Xingjun Liu, Yi Liu **Copy Editor:** Ke-Cui Yang **Production Editor:** Ke-Cui Yang

## Abstract

The escalating adoption of high-throughput methods in applied materials science dramatically increases the amount of generated data and allows for the deployment and use of sophisticated data-driven methods. To exploit the full potential of these accelerated approaches, the generated data need to be managed, preserved and shared. The heterogeneity of such data calls for highly flexible models to represent the data from fabrication workflows, measurements and simulations. We propose the use of a native graph database to store the data instead of relying on rigid relational data models. To develop a flexible and extendable data model, we create an ontology that serves as the blueprint of the data model. The Python framework Django is used to enable seamless integration into the virtual materials intelligence platform VIMI. The Django framework relies on the Object Graph Mapper neomodel to create a mapping between database classes and Python objects. The model can store the whole bandwidth of the data from fabrication to simulation data. Implementing the database into a platform will encourage researchers to share data while profiting from rich and highly curated data to accelerate their research.

**Keywords:** FAIR, energy materials, fabrication workflow optimization, ontologies, graph databases

## INTRODUCTION

Accelerating the development of clean energy devices is pivotal for the energy transition. A significant proportion of development efforts in this realm is devoted to the complex materials, such as electrocatalysts, multifunctional electrodes and ionic and porous transport media. Integrating these materials into devices necessitates a symbiotic combination of their properties to achieve the target device functionality. The intertwined requirements define a need for energy materials to be comprehensively and thoroughly screened, characterized and fabricated. Consequently, this calls for the development and implementation of a holistic and seamless platform to manage and analyze the rapidly growing datasets along the materials-to-device development workflow[1].

High-throughput methods in materials research have rapidly evolved in recent years and are expected to greatly speed up the rate at which novel materials are developed[2-5]. The screening for new energy materials and the optimization of manufacturing processes are already being conducted using high-throughput computation (HTC) and high-throughput experimentation (HTE). HTC is enabled by the steep growth in computing power along with the robust and efficient implementation of physics-based models[6-7]. HTC paves the way for the automated large-scale screening of materials with the desired combinations of properties[8-13]. HTE allows many experiments to be conducted in parallel, thereby enabling fast materials screening[14,15]. The further acceleration of scientific research can be achieved by automating fabrication workflows. Material acceleration platforms (MAPs) seek to enable closed-loop development by performing HTE in a fully autonomous fashion[16,17]. Most MAPs are deployed to optimize a set of materials or workflow parameters with respect to predefined target properties[18-21]. A high degree of autonomy calls for a sophisticated computational backend where data from previous fabrication cycles must be extracted and used to design the next cycle on-the-fly. Bayesian optimization is the most commonly used method in closed-loop experimentation[18]. These techniques generate a large amount of data along the materials development pipeline, thereby necessitating the need for efficient data management strategies[22,23].

The standard approach to data management is to build rudimentary data infrastructures suited to the needs of a particular project. This method is effective for studies of limited scope, where the collected data can be used for incremental improvements on specific materials classes or target applications. Since the data lack standardization, it becomes, however, challenging to compare them from various sources or to reuse data for other purposes with similar scope. An effective data management system should adhere to the FAIR principles, i.e., the data should be findable, accessible, interoperable and reusable[24,25]. Data FAIRification improves the reproducibility of scientific results and makes data accessible to the whole research community. Thus, FAIR data management enlarges the pool of available highly curated data and allows the application of a wider variety of data-driven methods[26,27]. Standardized formats that originate from following the FAIR principles require less processing to make data machine-readable and AI-ready[27].

In materials research and other fields, database projects have emerged to collect and manage increasing amounts of research data. Data types represented by these projects tend to be specific, with examples including the Materials Project[28], a database containing materials simulation data, and the Cambridge Crystallographic Data Centre (CCDC)[29], which gathers crystallographic data on materials. Enabling the full potential of data-driven approaches for accelerated materials discovery requires databases that include not only simulation data but also suitable fabrication and characterization data.

A suitable database must be flexible to effectively represent heterogenous data that contain insights into fabrication processes, measurements and simulations in materials research. Furthermore, materials, components and devices need to be described on multiple spatial and temporal scales. A database that is capable of storing data with such bandwidth could be the foundation of platforms that can orchestrate and accelerate materials research. These platforms can assist in the screening of materials, experimental design, the optimization of workflows and the orchestration of devices within self-driven labs. Our efforts in this regard tie in with the recently developed VIMI platform and offer data management for simulation and fabrication data, providing data-driven analytics, accelerated characterization and computer-aided materials design to its users[1].

In this communication, we present a flexible data management approach for energy materials platforms to accelerate the search for advanced materials by exploiting the full potential of data-driven research. Our approach contains the following:

- An extension of the European Materials Modelling Ontology (EMMO) to create a standardized representation of the energy materials domain for mandating FAIR data generation.

- A new graph data model based on the classes manufacturing, measurement, matter and property, as well as the relations between them. The data model provides an intelligible, flexible and extendable representation of fabrication workflows, measurements and simulation data.

- Data storage within the native graph database neo4j for efficient access to its highly connected content.

- An encapsulation of the database in a Django framework to allow a straightforward integration into VIMI or other platforms.

- A mapping from objects within the database to Python classes *via* an Object Graph Mapper (OGM).

## METHODOLOGY

### Ontology

Ontologies are a formal representation of knowledge that connect various metadata and make them machine-readable[30]. As shown in Figure 1, an ontology employs classes, which can have properties stored as key-value pairs. The classes within an ontology are connected *via* relationships and rules and constraints can be specified. Ontologies are crucial for representing domain knowledge in various scientific fields and provide the basis for data and knowledge exchange among researchers within a specific domain. Ontologies are particularly useful in the context of FAIR data generation since they standardize the knowledge representation of a domain.

The interdisciplinary nature of materials science renders the standardization of information imperative for communication. The European Materials Modelling Council[31] implements EMMO[32], which is a versatile ontology for materials sciences. EMMO consists of three levels. The top level defines real-world objects and introduces "perspectives" to reflect their pluralistic nature (e.g., materials can be defined *via* their composition or function). The middle level contains specific perspectives that make EMMO applicable to various domains. Each of the perspectives represents a different branch that defines objects from a holistic, physicalistic, semiotic or mereotopological perspective.

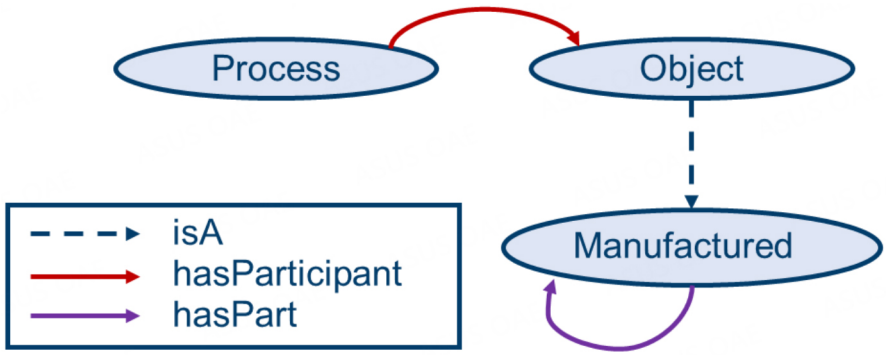


**Figure 1.** A minimalistic example of an ontology. Instances of the Process class are connected to Object instances *via* the has Participant relationship. The Manufactured class is a child class of the Object class, meaning that all properties and relationships can be inferred from Object to Manufactured. A Manufactured object is composed of other Manufactured parts. This mereological perspective is represented by the hasPart relationship.

The different perspectives enable EMMO to represent the fabrication, characterization and simulation of materials up to the device scale on a very general level, as illustrated in Figure 2. The bottom level contains ontologies of specific materials science domains. Other projects that require domain-specific ontologies can extend the bottom level of EMMO, while the higher levels of EMMO offer a ruleset that can be inferred from these extensions. Projects that use EMMO-based ontologies include NOMAD[33,34], CHAMEO[35] and BigMap[36,37]. Building on EMMO to introduce an ontology substantially simplifies its creation since the basics for most application domains are already contained in the EMMO. An EMMO-based ontology can therefore rely on a variety of existing relationships/classes and constraints. Extending the EMMO branches allows properties, relationships and constraints from the parent classes, e.g., Manufactured is a subclass of Object, to be inferred and it therefore also has the hasParticipant relationship to Process classes [Figures 1 and 2]. Creating an ontology relying on EMMO also improves interoperability with other EMMO-based ontologies, since they have the same structure and are following the same basic rule sets. In many aspects, ontologies are comparable to languages, centered around narrow domains, and like languages, the power of an ontology strongly relies on its level of adoption.

### Database

At the core of the data infrastructure, a database is required to store fabrication, simulation and measurement data in a FAIR format. Furthermore, the metadata need to be sufficient to ensure the reproducibility of each entry in the database. Relational databases, such as MySQL or Postgres, are the current industry standards[38,39]. These databases contain tables, each of them representing a class of real-world objects, e.g., materials or measurements. Each row within a table refers to a specific instance of these real-world objects, e.g., a specific material. Relationships between different objects are represented by joining tables with foreign keys. Relational databases excel when highly structured and sparsely connected data need to be stored. However, the usage of foreign keys makes processing these relationships slow and their table structure makes the data model rigid[40]. The differences in performance between relational databases and NoSQL databases have been investigated and benchmarked previously[41,42].

There is also a wide variety of NoSQL databases for the development of a more dynamic data model and these can be divided into document-oriented, key-value and graph databases. Document-oriented data storage uses documents in a specific encoding, such as XML, YAML, JSON or BSON. Each document is addressed *via* a unique key and can be related to other documents by joint keys. The difference to relational databases is the versatile structure of the documents, which allows for a more flexible data model[43]. The handling of relationships nevertheless remains inefficient, since documents are still joined *via* foreign keys,

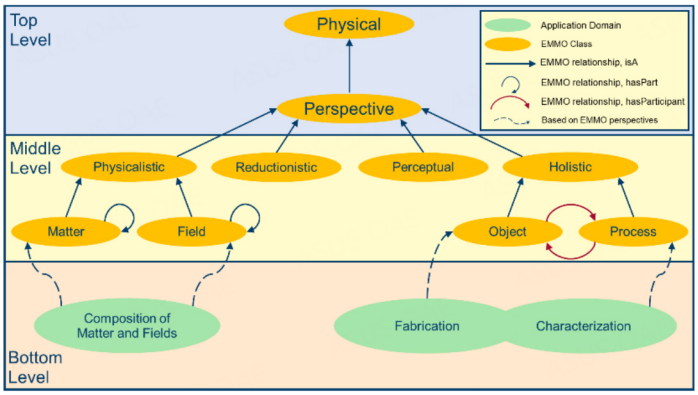


**Figure 2.** A minimalistic example of the overall EMMO structure. The green shapes on the bottom level represent application domains that can use and extend the respective EMMO classes in the middle level. The top level is strongly simplified since it contains further fundamental classes, relationships and constraints that the rest of the EMMO is based on.

just as in the relational data model[40]. Key-value databases store data as a single opaque collection of key-value pairs. This data structure also enables highly flexible data models, but its primitive structure needs to be extended for complex use cases[44,45].

In this work, we chose neo4j, a native graph database, for data storage. Native graph databases use graph theory to represent and store data. Graphs contain nodes that are connected *via* relationships[46]. They can be used to solve mathematical problems or to represent and store data using adjacency lists. Graphs can naturally represent complex domains inside the real world, allowing for data modelling without imposing layers of abstraction between the natural world and the associated data model. Directed graphs also contain valuable information since they can represent asymmetric relationships. The data structure of native graph databases allows for a highly efficient traversal of the graph along its relationships, which can be described as pointer hopping or dereferencing pointers. Since the adjacent nodes of each node are stored directly at the node itself, queries that require graph traversal *via* a path of relationships exhibit an O(1) complexity[40].

In neo4j, nodes and relationships can have properties stored as key-value pairs. Relationships in neo4j are always directed, which allows for the representation of asymmetric relations. These directed relationships provide more context, e.g., a simple workflow can be represented as several Process nodes, connected *via* followedBy relationships that need to be directed to fully represent the workflow. Relations between nodes of the same class are only unambiguous if they are intrinsically symmetric or if they are represented by directed relationships. Graph databases do not require a data schema since they are naturally additive, making them highly advantageous for storing heterogeneous data.

From a materials research perspective, fabrication workflows underline the heterogeneity of data in materials science research, since fabrication and characterization process data can contain a wide range of parameters, materials and subprocesses. These workflows and measurements can naturally be represented by graphs, since they are mostly sequences of subprocesses that have materials and parameters as inputs and manufactured materials or properties as outputs. A process can also be represented within a table, but it requires its structure to be predefined. If a process changes, its table must be altered or a new table must be created, leading either to massive, sparsely filled tables or many small tables representing variations of the same process. Graph databases store workflows as node sequences, one sequence for every stored workflow. Variations of these workflows do not require any changes to existing ones within the database since each workflow is stored as a separate sequence of nodes. Representing the workflows in Figure 3 within a graph

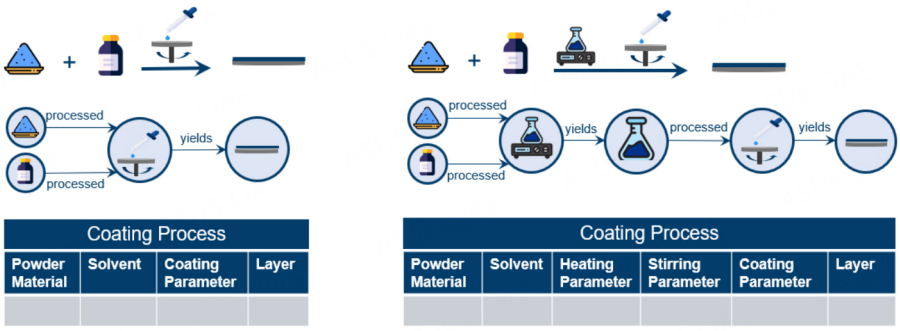


**Figure 3.** Schematic of fabrication workflows. The first row represents simplistic coating workflows. The workflows differ slightly since the left workflow contains an intermediate stirring and heating step. The fabrication workflows are represented as graphs and tables (bottom).

database would lead to two different sequences, which can both be accessed by queries that ask for workflows, leading from the given precursors to the corresponding product. The database is flexible in that regard since the data model does not predefine how a specific process is structured.

## SYNERGIZING ONTOLOGIES AND GRAPH DATABASES

### Ontology

Ontologies are naturally represented as graphs since they contain classes represented *via* relationships. The structure of an ontology makes it easily transferrable to a graph data model. The created ontology is an extension of the EMMO, which is already established as a highly sophisticated framework of classes, rules and constraints to represent materials science. The ontology with its central classes is shown in Figure 4.

The classes manufacturing and measurement both share the same parent class, Process. Matter represents all physical objects, from single atoms to manufactured components or devices. Property is a child class of PhysicalQuantity and is yielded by a Measurement process. Meta Data represents a collection of classes to make the scheme more understandable. It contains all classes that contain information stored as metadata to a specific process (e.g., measurement instruments, researchers or institutions, or experimental parameters). The Manufacturing, Measurement and Matter instances can be divided into components. The Matter instances are inputs to the Manufacturing and Measurement instances and outputs of the Manufacturing nodes. Processes can have metadata assigned to them and Measurement can have physical quantities as Property instances as outputs.

The ontology is tailored to the specific domain of energy materials by introducing domain-specific child classes. Most properties and relationships of these child classes can be inferred from the parent classes. Only specific properties and constraints need to be added to the class definition.

Currently, the EMMO extension focuses on fuel cell fabrication and characterization, but it can easily be extended to other technology domains. Classes of components, materials and fabrication procedures specific to the fuel cell domain were extracted from several well-cited fuel cell reviews. Therefore, the added taxonomy follows a widely accepted classification.

### Data model

The original data model developed in this work is inspired by the open provenance model in which every simulation, measurement or fabrication process acts as a function[47]. It takes an input, e.g., a dataset or a selection of materials, to a function, e.g., a simulation or a fabrication step, and yields an output, e.g., a

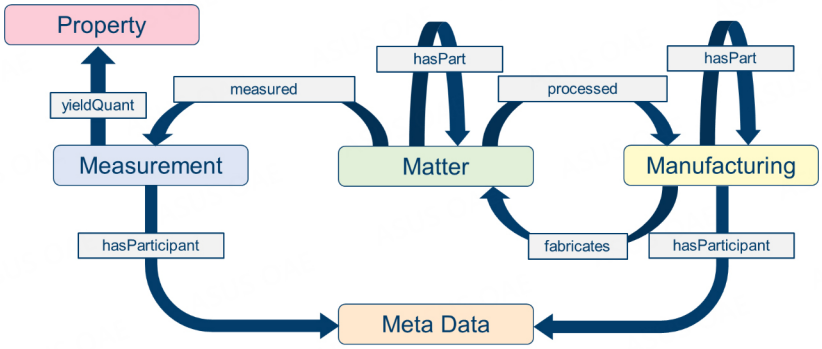


**Figure 4.** High-level schematic of proposed EMMO-based data model.

manufactured material or a measurement result. Due to the directed nature of the neo4j graph database, each process can naturally be represented by a tuple comprising an input node, a process node and an output node.

The overall data model follows the flowchart in Figure 4, thereby making it highly flexible and adaptable. To improve the findability of the data, we imported specific parts of the ontology into the database, namely, the classes Process, Matter and Quantity, including all their subclasses. These ontology branches represent the abstract concept of the real-world objects we aim to store in our database. The database therefore contains an ontology domain of the abstract concepts and a real-world domain that contains the actual data. Each node of the real-world domain is a specific instance of an ontology class and can be linked to a corresponding node of the ontology domain. The ontology nodes are used as labels for the nodes in the real-world domain. The connectivity of the ontology nodes creates not only a single label but also alternative labels. A real-world node that represents $H_2O$ as a solvent would be connected to the ontology node called PolarSolvent, which is a subclass of Solvent. Therefore, $H_2O$ is labeled with PolarSovent and Solvent can be retrieved as an alternative label [Figure 5]. These alternative labels greatly improve the findability of the data and the ontology domain can be easily extended to maintain the flexibility of the database.

Furthermore, for the holistic perspective on objects, their mereological description is crucial since materials, components and processes span multiple spatial and temporal scales. Device fabrication is a process that contains subprocesses and each of these can be split into a sequence of subprocesses [Figure 6]. The mereological representation of processes and matter allows for the representation of these objects down to an arbitrary degree of precision. This method of fractioning creates tree-structured graphs. Trees are specific graphs in which two nodes are always connected by exactly one path, making it a connected acyclic graph. Each node within a tree can have an arbitrary number of child nodes but must have only one parent node, except for the root node, which has no parent node. Each node can be treated as the root node of its subtree, thereby allowing recursion to traverse a tree.

The data model must represent experimental setups to enforce the reproducibility of the measurements and fabrication workflows. This ability will be crucial when the database is employed as part of a data infrastructure with an interface to automated labs. The scientific setup can be represented as a subgraph of connected nodes representing specific instruments/devices. Steps within fabrication workflows can then be mapped to the corresponding devices of the experimental setup, thereby allowing for precise process representations and enabling specific workflow optimization and troubleshooting. In the field of automation, challenging tasks include the orchestration of different devices and the exact positions of

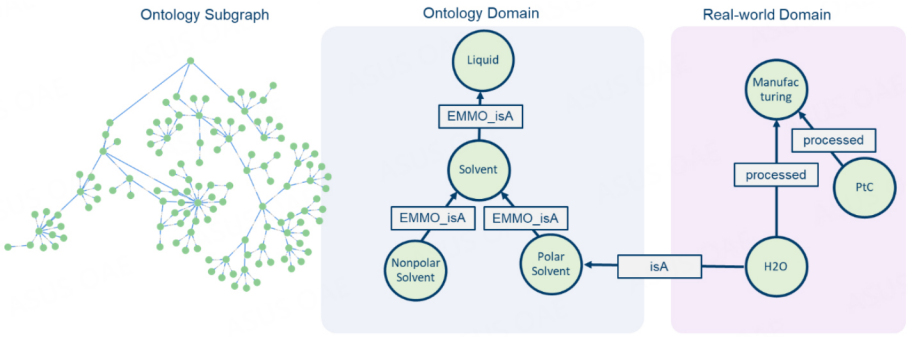


**Figure 5.** Screenshot of the imported Matter ontology within the database (left). Flowcharts of the ontology and real-world domain within the graph database (right) with the mapping between these domains via the isA relationship. Inheritance in the ontology domain is represented *via* the EMMO_isA relationship. The real-world data represented here shows $H_2O$ and platinum on carbon catalyst (PtC) processed by an arbitrary Manufacturing step.

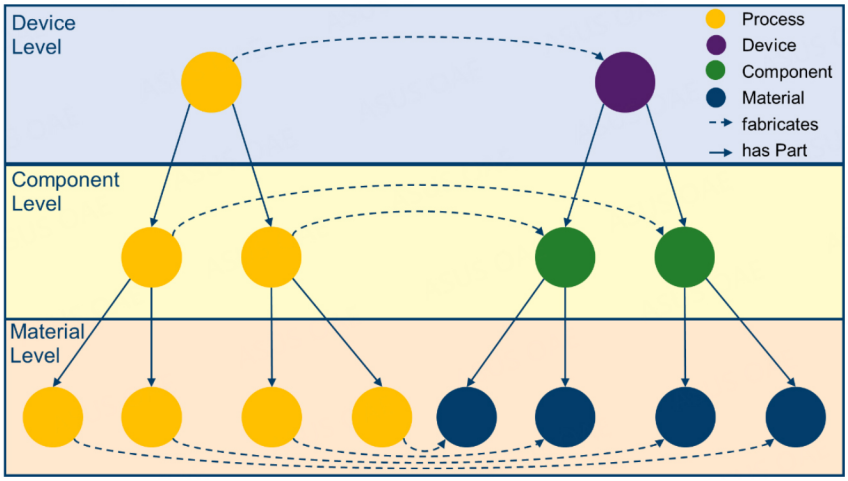


**Figure 6.** Schematic of the mereotopological structure of a process and its subprocesses (orange nodes) and a device, its components and their materials (purple, green and blue nodes, respectively).

samples and other moving or moveable parts of the setup at different times. This leads to the necessity to represent these workflows in extremely high granularity. The supplementary data generated by publications regarding automated labs show the level of detail that is needed to achieve automation[21,48,49]. Another important aspect of data management in automated labs is tracking the process itself in real-time[16]. Tracking and high-resolution representation could be intuitively accomplished using the proposed graph data model. Although a relational database could also satisfy these needs for a specific automated lab, it would be very complicated to use the same data model for a different setup.

The described data model is intuitive, meaning that the data model and the real-world domain are not separated by layers of abstraction that would add unnecessary complexity to the data model. It uses the middle-level classes and relationships of the EMMO ontology. Furthermore, using trees as data structures allows for the implementation of highly efficient queries along the hierarchical breakdown of objects into their components. Its performance concerning possible user requests must also be evaluated by a data model. The data model allows for intuitive querying for processing sequences or device compositions, as well as requests of fabrication workflows as sequences of input-function-output tuples.

### Object graph mapper

Researchers in materials science, as users of the database, should be able to upload and access data from fabrications, measurements and simulations. The data need to be retrievable in different formats, including in the form of Python objects. The usage of layers that connect the database with the platform it is embedded in, are a common practise. Platforms that rely on relational databases use object-relational mappers to map their database object and datatypes to a given programming language, e.g., the AiiDA platform for computational materials science employs an object-relational mapper to map from their SQL database[50].

We utilize the OGM neomodel to facilitate mapping between the objects and data types within a graph database and the Python programming language[51]. The neomodel library is also available as a Django plugin, as a robust Python framework to reduce the complexity of web application creation, making the OGM an ideal key component for an upcoming API implementation due to its Python compatible interface. The OGM allows for the introduction of domain-specific Python classes (e.g., Manufacturing Process). These classes contain properties and member functions, reflecting how they are defined within the ontology. The Python inheritance rules can create a hierarchy of classes. Neomodel enables mapping of the Python classes, including their properties and data types, to the classes within the database [Figure 7].

Neomodel can also be used to generate, modify and query in a high-level interface that is agnostic to database architectural details. Using neomodel enables query formulation following the Python syntax, thereby offering a Python-based interface to the database that makes interacting with it more intuitive due to the broad acceptance of the Python programming language in the scientific community.

Neomodel allows for the creation, deletion and update of the nodes and relationships within the database. For example, it might therefore be used to update the single information of single nodes. Nevertheless, the OGM lacks efficiency for large-scale database operations. Each node/relationship that is added leads to separate requests to the database, resulting in high costs for the ingestion of large datasets. Thus, for large chunks of data, neomodel allows for custom-made cypher queries to conduct complex operations in an efficient manner. To ingest a fabrication workflow, for example, requires a predefined cypher query since such a workflow contains multiple nodes with a high degree of connectivity. Using an OGM has the advantage that complex database operations can be wrapped into Python functions that have well defined input and output interfaces. Neomodel creates an additional abstraction layer between database operations and the data management system itself. Cypher queries for complex database operations are thoroughly tested, wrapped into a Python function and are ready to be used in a Python environment, thereby hiding their underlying complexity.

### Database

The proposed data model is centered around the creation of node sequences. Labels are introduced to create sets of nodes that improve the structure within the database. These labels correspond to the Python classes created using neomodel and the classes defined within the ontology. Nodes can have multiple labels, e.g., a fuel cell node might have the labels Matter, Manufactured, Device and Electrochemical Device. The choice of indexing nodes of a specific label mainly depends on the nature of queries containing that label. Each label represents an entity containing attributes as key-value pairs necessary to identify that entity (e.g., unique identifiers or the SMILES representation of a molecule). Nodes that share one label can be indexed, thereby accelerating the node retrieval of a particular label. However, indexing has the drawback of slowing down the write efficiency and requires more storage capacity.

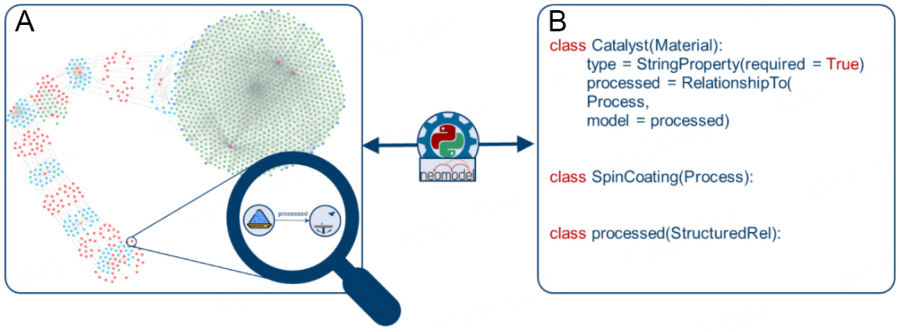


**Figure 7.** Nodes and relationships within the database (A) can be mapped to Python classes (B).

The property space of materials, components and devices and the parameter space of processes in materials science is high dimensional and constantly expanding. These physical quantities are stored within external Property and Parameter nodes to allow for the representation of all parameters and properties. The externalization of these nodes also allows for the efficient querying of nodes that share the same or similar properties and fabrication workflows that share the same or similar parameters. The database stores measurement results, fabrication parameters or properties in the form of scalars or a scalar array.

The highly efficient traversal of relationships is made possible by the data structure of neo4j, which has a complexity of O(1). To further increase query efficiencies, neo4j physically stores adjacent nodes close to each other and tailors the database architecture specifically to frequently used queries. The caching during queries also improves the efficiency of reading, writing and matching commands.

Large binary data types, such as image data from imaging techniques, will be stored within an external file server and are only referenced as metadata for properties derived from the images (e.g., size distribution). Referencing of the image files is carried out *via* UML links. Externalizing large data types, such as images and videos, is a good practice for data modelling. It keeps the image files connected to their corresponding data while queries do not have to handle clunky image data chunks, thereby improving the overall performance of the database. Furthermore, keeping binary data chunks within the database does not yield benefits since the database itself cannot query, index or compare binary data.

The data management system will be implemented into the VIMI platform[1], so that researchers and users from industry can upload their data *via* well-defined interfaces, e.g., the dragging and dropping of CSV files. Furthermore, cooperations with automated labs are in place for which custom tailored APIs will be created to streamline their generated data directly to the database.

### Data representation

To test the data model, a batch of data for the fabrication and characterization of fuel cells was stored. The batch spans data from the materials to the device including the data from various measurements on different length scales. The heterogeneity of the data and its high dimensionality makes them ideal for testing the proposed data model.

The data were ingested into the database *via* CSV files and represented following the proposed data model. Figure 8 shows the fabrication data and how they are stored in the graph database. Figure 8A shows a single fabrication workflow from the materials to the fuel cell device, and Figure 8B presents 25 fabrication workflows within the database. The screenshots were taken from the neo4j browser interface, which offers visual representations of the stored data.

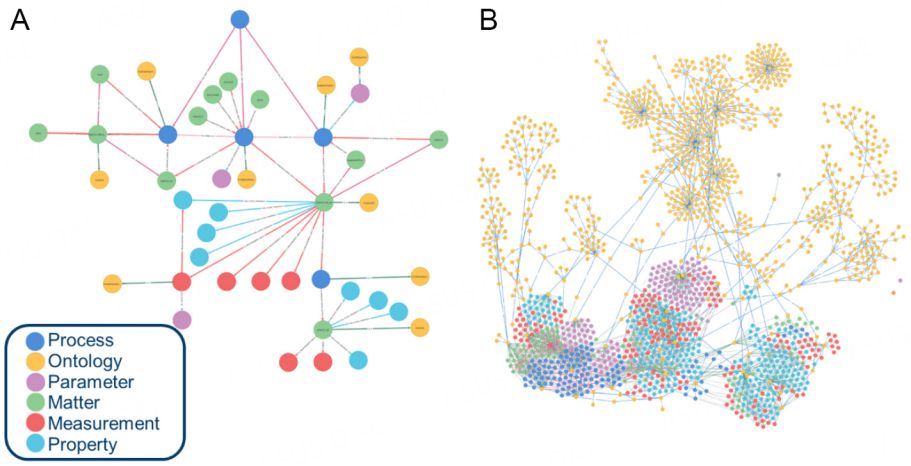


**Figure 8.** Representation of a single fabrication workflow from the starting materials to the fuel cell device (A) and 25 fabrication workflows and how they are linked to the ontology domain (yellow) within the database (B).

Ingesting sample data shows that the proposed data model is indeed able to represent complex fabrication workflows with a degree of detail. Even though the data model can represent these workflows, it is challenging to make them intuitively retrievable. The user of the data management system should be able to retrieve all variations of fabrication procedures for specific devices or components with a single command. This requires sophisticated queries scanning the database in an efficient manner to find the requested node sequences. The wider the bandwidth of the fabrication data stored within the database, the more challenging it will be to retrieve the requested data. These queries will be wrapped in Python functions and implemented into APIs to make the data accessible. We are currently cooperating with experimentalists from different parts of the energy materials domain to enrich our database, test our data model and especially improve and challenge our queries.

## CONCLUSIONS

We propose a new data model as the basis of a highly adaptable data infrastructure for the fabrication, measurements and simulations of energy materials. The data model is designed to represent workflows and processes at an arbitrary level of complexity. It can be modified to incorporate new materials, components or processes. The hydrogen technology domain, with an emphasis on fabrication and characterization, is represented in the data model by introducing an EMMO-based ontology. The data are stored in the native graph database neo4j and its structure allows for the efficient traversal of fabrication processes. To further increase efficiency, tree data structures can be used to represent the fabrication workflows in their subprocesses and the dissection of devices, components or materials into their constituent parts.

A use case for the proposed data management system is automated labs since they require automated data management. Current automated labs usually create data management systems tailored to their specific labs, thus, each automated lab requires a new data management system and a new data model. This creates additional overhead when these labs are set up, and it leads to small unconnected data lakes that lack standardization. The proposed data model is an answer to the growing number of automated labs and their need for data management since it can represent given workflow in an arbitrary level of detail.

For more advanced approaches to data-driven workflow optimization, it is essential to include data FAIRification in experimental workflows. In particular, the data generated in fabrication processes lack FAIR features due to their heterogeneous nature and the absent workflows for standardization. Storing data that are both FAIR and suitable for AI-driven models is possible by abandoning the relational data model and transit to the flexible, graph-based data model. Access to FAIR experimental data further pave the way for data-driven techniques.

The next phases of this project involve the integration of graph databases into the VIMI platform. This will make the database accessible to other users and the data infrastructure will be used to streamline data into the generation of training datasets and the creation of machine learning models. Since the data model is based on the concepts of the EMMO, it can represent characterization and fabrication of other domains in materials science. Its flexibility also allows for applications in other related research domains, such as batteries or solar cells.

## DECLARATIONS

### Acknowledgements

The authors gratefully acknowledge the cooperation of Prof. Jasna Jankovic (UCONN, USA) and Fabian Tipp (Forschungszentrum Jülich) for providing fabrication and simulation data. The authors also gratefully acknowledge the Gauss Centre for Supercomputing e.V. (http://www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

### Authors' contributions

Performed the research and drafted the manuscript: Dreger M
Revised and finalized the manuscript: Dreger M, Eslamibidgoli MJ, Eikerling MH, Malek K

### Availability of data and materials

The EMMO-based ontologies as well as the project itself are available at https://github.com/MaxDreger92/MatGraphAI.

### Financial support and sponsoring

The authors acknowledge the financial support from the Federal Ministry of Science and Education (BMBF) under the German-Canadian Materials Acceleration Centre (GC-MAC) grant number 01DM21001A and the financial support from the HiTEC graduate school for doctoral candidates at the Forschungszentrum Jülich.

### Conflict of interest

All authors declare that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

## REFERENCES


1. Malek A, Eslamibidgoli MJ, Mokhtari M, Wang Q, Eikerling MH, Malek K. Virtual materials intelligence for design and discovery of advanced electrocatalysts. *Chemphyschem* 2019;20:2946-55. DOI PubMed
2. Maier WF, Stöwe K, Sieg S. Combinatorial and high-throughput materials science. *Angew Chem Int Ed Engl* 2007;46:6016-67. DOI PubMed
3. Xiang XD, Sun X, Briceño G, et al. A combinatorial approach to materials discovery. *Science* 1995;268:1738-40. DOI
4. Koinuma H, Takeuchi I. Combinatorial solid-state chemistry of inorganic materials. *Nat Mater* 2004;3:429-38. DOI PubMed
5. Spivack JL, Cawse JN, Whisenhunt DW, et al. Combinatorial discovery of metal co-catalysts for the carbonylation of phenol. *Appl Catal A-Gen* 2003;254:5-25. DOI
6. Curtarolo S, Hart GL, Nardelli MB, Mingo N, Sanvito S, Levy O. The high-throughput highway to computational materials design. *Nat Mater* 2013;12:191-201. DOI PubMed
7. Martin R. Electronic structure: basic theory and practical methods. Cambridge University Press; 2004.
8. Ceder G, Chiang Y, Sadoway DR, Aydinol MK, Jang Y, Huang B. Identification of cathode materials for lithium batteries guided by first-principles calculations. *Nature* 1998;392:694-6. DOI
9. Jóhannesson GH, Bligaard T, Ruban AV, Skriver HL, Jacobsen KW, Nørskov JK. Combined electronic structure and evolutionary search approach to materials design. *Phys Rev Lett* 2002;88:255506. DOI PubMed
10. Stucke DP, Crespi VH. Predictions of new crystalline states for assemblies of nanoparticles: perovskite analogues and 3-D arrays of self-assembled nanowires. *Nano Lett* 2003;3:1183-6. DOI
11. Curtarolo S, Morgan D, Persson K, Rodgers J, Ceder G. Predicting crystal structures with data mining of quantum calculations. *Phys Rev Lett* 2003;91:135503. DOI PubMed
12. Morgan D, Ceder G, Curtarolo S. High-throughput and data mining with ab initio methods. *Meas Sci Technol* 2005;16:296-301. DOI
13. Curtarolo S, Morgan D, Ceder G. Accuracy of ab initio methods in predicting the crystal structures of metals: a review of 80 binary alloys. *Calphad* 2005;29:163-211. DOI
14. Boussie TR, Diamond GM, Goh C, et al. A fully integrated high-throughput screening methodology for the discovery of new polyolefin catalysts: discovery of a new class of high temperature single-site group (IV) copolymerization catalysts. *J Am Chem Soc* 2003;125:4306-17. DOI PubMed
15. Potyrailo RA, Chisholm BJ, Morris WG, et al. Development of combinatorial chemistry methods for coatings: high-throughput adhesion evaluation and scale-up of combinatorial leads. *J Comb Chem* 2003;5:472-8. DOI PubMed
16. Wagner J, Berger CG, Du X, Stubhan T, Hauch JA, Brabec CJ. The evolution of materials acceleration platforms: toward the laboratory of the future with AMANDA. *J Mater Sci* 2021;56:16422-46. DOI
17. Aspuru-Guzik A, Persson K. Materials acceleration platform: accelerating advanced energy materials discovery by integrating high-throughput methods and artificial intelligence. In. Mission Innovation; 2018.
18. Attia PM, Grover A, Jin N, et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* 2020;578:397-402. DOI PubMed
19. Tabor DP, Roch LM, Saikin SK, et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat Rev Mater* 2018;3:5-20. DOI
20. Burger B, Maffettone PM, Gusev VV, et al. A mobile robotic chemist. *Nature* 2020;583:237-41. DOI PubMed
21. MacLeod BP, Parlane FGL, Morrissey TD, et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Sci Adv* 2020;6:eaaz8867. DOI PubMed PMC
22. Jose R, Ramakrishna S. Materials 4.0: materials big data enabled materials discovery. *Appl Mater Today* 2018;10:127-32. DOI
23. Ghiringhelli LM, Vybiral J, Levchenko SV, Draxl C, Scheffler M. Big data of materials science: critical role of the descriptor. *Phys Rev Lett* 2015;114:105503. DOI PubMed
24. Stall S, Yarmey L, Cutcher-Gershenfeld J, et al. Make scientific data FAIR. *Nature* 2019;570:27-9. DOI
25. Draxl C, Scheffler M. NOMAD: The FAIR concept for big data-driven materials science. *MRS Bull* 2018;43:676-82. DOI
26. DeCost BL, Hattrick-Simpers JR, Trautt Z, Kusne AG, Campo E, Green ML. Scientific AI in materials science: a path to a sustainable and scalable paradigm. *Mach Learn Sci Technol* 2020;1:033001. DOI PubMed PMC
27. Scheffler M, Aeschlimann M, Albrecht M, et al. FAIR data enabling new horizons for materials research. *Nature* 2022;604:635-42. DOI PubMed
28. Jain A, Ong SP, Hautier G, et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater* 2013;1:011002. DOI
29. Allen FH, Bellard S, Brice MD, et al. The Cambridge crystallographic data centre: computer-based search, retrieval, analysis and display of information. *Acta Crystallogr B Struct Sci* 1979;35:2331-9. DOI
30. Guarino N, Oberle D, Staab S. What is an ontology? In: Staab S, Studer R, editors. Heidelberg: Springer; 2009. pp.1-17.
31. Adamovic N, Asinari P, Goldbeck G, et al. European materials modelling council. In Proceedings of the 4th World Congress on Integrated Computational Materials Engineering (ICME 2017): Springer; 2017, pp. 79-92. DOI
32. Council EMM. Report on workshop on interoperability in materials modelling; 2017.
33. Draxl C, Scheffler M. The NOMAD laboratory: from data sharing to artificial intelligence. *J Phys Mater* 2019;2:036001. DOI
34. Lenz-Himmer MO, Ghiringhelli L, Baldauf C, Scheffler M. Ontologies in computational materials science. *APS March Meeting*

*Abstracts* 2021;2021:X61-002.
35. Nostro P, Goldbeck G, Toti D. CHAMEO: an ontology for the harmonisation of materials characterisation methodologies. *Appl Ontol* 2022;17:401-21. DOI
36. Simon Clark FLB. Available from: https://github.com/BIG-MAP/BattINFO [Last accessed on 2 Mar 2023].
37. Deng Z, Kumar V, Bölle FT, et al. Towards autonomous high-throughput multiscale modelling of battery interfaces. *Energy Environ Sci* 2022;15:579-94. DOI
38. Maier D. The theory of relational databases. Rockville: Computer Science Press; 1983.
39. Halpin T, Morgan T. Information modeling and relational databases. Morgan Kaufmann; 2010.
40. Robinson I, Webber J, Eifrem E. Graph databases: new opportunities for connected data. O'Reilly Media, Inc.; 2015.
41. Holzschuher F, Peinl R. Performance of graph query languages: comparison of cypher, gremlin and native access in neo4j. Proceedings of the Joint EDBT/ICDT 2013 Workshops; 2013. pp. 195-204. DOI
42. Vicknair C, Macias M, Zhao Z, et al. A comparison of a graph database and a relational database: a data provenance perspective. Proceedings of the 48th annual Southeast regional conference; 2010. pp. 1-6. DOI
43. Banker K, Garrett D, Bakkum P, Verch S. MongoDB in action: covers MongoDB version 3.0. Simon and Schuster; 2016.
44. Seeger M, Ultra-Large-Sites S. Key-value stores: a practical overview. Available from: https://blog.marc-seeger.de/assets/papers/Ultra_Large_Sites_SS09-Seeger_Key_Value_Stores.pdf [Last accessed on 2 Mar 2023].
45. Idreos S, Callaghan M. Key-value storage engines. Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data; 2020. pp. 2667-72. DOI
46. Berge C. The theory of graphs. Courier Corporation; 2001.
47. Moreau L, Freire J, Futrelle J, Mcgrath RE, Myers J, Paulson P. The open provenance model: an overview. In: Freire J, Koop D, Moreau L, editors. Provenance and annotation of data and processes. Berlin: Springer Berlin Heidelberg; 2008. pp. 323-6. DOI
48. MacLeod BP, Parlane FGL, Rupnow CC, et al. A self-driving laboratory advances the Pareto front for material properties. *Nat Commun* 2022;13:995. DOI PubMed PMC
49. Rooney MB, Macleod BP, Oldford R, et al. A self-driving laboratory designed to accelerate the discovery of adhesive materials. *Digital Discovery* 2022;1:382-9. DOI
50. Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B. AiiDA: automated interactive infrastructure and database for computational science. *Comput Mater Sci* 2016;111:218-30. DOI
51. Edwards R. Neomodel documentation. Available from: https://neomodel.readthedocs.io/en/latest/ [Last accessed on 2 Mar 2023].

Wang *et al. J Mater Inf* 2023;3:3
**DOI:** 10.20517/jmi.2022.45

**Journal of Materials Informatics**

**Research Article**

**Open Access**

# Sulfur poisoning mechanism of LSCF cathode material in the presence of $SO_2$: a computational and experimental study

**Rui Wang[1], Lucas R. Parent[2], Yu Zhong[1,*]**

[1]Mechanical and Materials Engineering Department, Worcester Polytechnic Institute, Worcester, MA 01609, USA.
[2]Innovation Partnership Building, University of Connecticut, Storrs, CT 06269, USA.

[*]**Correspondence to:** Prof. Yu Zhong, Mechanical and Materials Engineering Department, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA. E-mail: yzhong@wpi.edu

**How to cite this article:** Wang R, Parent LR, Zhong Y. Sulfur poisoning mechanism of LSCF cathode material in the presence of $SO_2$: a computational and experimental study. *J Mater Inf* 2023;3:3. https://dx.doi.org/10.20517/jmi.2022.45

**Received:** 19 Dec 2022 **First Decision:** 30 Jan 2023 **Revised:** 15 Feb 2023 **Accepted:** 6 Mar 2023 **Published:** 9 Mar 2023

**Academic Editors:** Qian Li, William Yi Wang **Copy Editor:** Ke-Cui Yang **Production Editor:** Ke-Cui Yang

## Abstract

Aiming at the comprehensive understanding of the single sulfur poisoning effect and, eventually, the multiple impurities poisoning phenomena on the SOFC (Solid Oxide Fuel Cell) cathode materials, the sulfur poisoning effect on the $(La_{0.6}Sr_{0.4})_{0.95}Co_{0.2}Fe_{0.8}O_3$ (LSCF-6428) has been investigated in the presence of 10 ppm $SO_2$ at 800, 900, and 1,000 °C, respectively, with a combined computational and experimental approach. The good agreement between the CALPHAD (Computer Coupling of Phase Diagrams and Thermochemistry) simulations and the XRD (X-Ray Diffraction), SEM (Scanning Electron Microscopy), and TEM (Transmission Electron Microscopy) characterization results support the reliability of the CALPHAD approach in the SOFC field. Furthermore, comprehensive simulations were made to understand the impact of temperature, $P(SO_2)$, $P(O_2)$, and Sr concentration on the threshold of $SrSO_4$ stability. Results showed that the formation of $SrSO_4$ is thermodynamically favored at lower temperatures, higher $P(SO_2)$, higher $P(O_2)$, and higher Sr concentration. Finally, comparisons were also made between LSCF-6428 and LSM20 ($La_{0.8}Sr_{0.2}MnO_3$) using simulations, which confirmed that LSCF-6428 is a poor sulfur-tolerant cathode, in agreement with the literature.

**Keywords:** LSCF cathode, CALPHAD, sulfur poisoning, long-term degradation, accelerated testing

## INTRODUCTION

Mixed Ionic and Electronic Conducting (MIEC) (La, Sr)(Co, Fe)$O_3$ (LSCF) is one of the most promising Solid Oxide Fuel Cell (SOFC) cathodes due to its comprehensive electrochemical properties. It has been widely applied at intermediate conditions in the energy conversion from fossil fuels to electricity to tackle the penalty of $CO_2$ emission and poor efficiency from traditional oxy-fuel combustion[1,2]. However, SOFC cathode materials suffer from thermodynamic instability in their long-term durability in the presence of several impurities, of which $SO_2$ is known to be a concern[3,4]. Thus, developing a comprehensive understanding of the sulfur poisoning of LSCF cathodes in the presence of $SO_2$ impurity is necessary to address the long-term degradation of the SOFC cathode systems at operating conditions.

It is reported that a trace amount of $SO_2$ in air, as low as in part per million (ppm) level, can still cause a performance drop and shorten the service life of LSCF cathodes due to the formation of the detrimental secondary phases[5]. Liu *et al.*[6] first reported the sulfur poisoning of an LSCF cathode in the presence of 1 and 20 ppm of $SO_2$ at 800 °C for 1,000 h. Results suggested that the LSCF has a low tolerance to sulfur poisoning even with 1 ppm $SO_2$. Moreover, the degradation rate of LSCF was 8 times higher than that of LSM at 20 ppm $SO_2$ because of the formation of $SrSO_4$ in LSCF. In the same year, Wang *et al.*[7] reported the formation of $SrSO_4$, $La_2O_2SO_4$, and $CoFe_2O_4$ in 100 ppm $SO_2$ at 800 °C for 24 h, $SrSO_4$ and $CoFe_2O_4$ in 10 and 1 ppm $SO_2$ at 800 °C for 24 h, respectively. The cell was free from sulfur poisoning in 0.1 ppm $SO_2$ for 24 h and $SO_2$-free air for 100 h at 800 °C, respectively, indicating a thermodynamic threshold or a diffusion-controlled mechanism for the sulfur poisoning process. Wang *et al.*[8] reported the formation of $SrSO_4$ in the presence of 20mg/L $SO_2$ at 800 °C for 500 h and concluded that LSCF exhibited poorer durability than LSM against $SO_2$ poisoning. Later, Wang *et al.*[9] investigated the effect of Sr concentration on the sulfur poisoning of LSCF cathodes, namely $(La_{0.6}Sr_{0.4})(Co_{0.2}Fe_{0.8})O_3$ (LSCF-6428) and $(La_{0.8}Sr_{0.2})(Co_{0.2}Fe_{0.8})O_3$ (LSCF-8228), at 800 °C with 1 ppm $SO_2$. They concluded that the sulfur poisoning is much more severe for the less active LSCF-8228 cathode, even though the volume of $SrSO_4$ formed is much larger in LSCF-6428 due to it having more available electrochemically active sites. Moreover, Wang *et al.*[10] evaluated the $SO_2$ poisoning phenomena of LSCF-6428 cathodes with 0.1 ppm $SO_2$-air at 800 °C for 100 h by varying the flow rate (25, 50, and 90 mL $min^{-1}$). They concluded that $SrSO_4$ formed on the surface of the LSCF-6428 first due to the fast absorption of the $SO_2$ and then concentrated in the vicinity of the triple phase boundaries (TPBs) under a polarization condition. Subsequently, Wang and Jiang[11] reviewed the mechanism of $SO_2$ poisoning on both LSCF and LSM cathodes. They reported that the sulfur poisoning effect was more pronounced at lower temperatures for both LSCF and LSM cathodes. More importantly, it was most pronounced on the surface of the LSCF cathode and more evident on the interface of LSM/YSZ TPBs. Wang *et al.*[12] investigated the interrelation between sulfur poisoning and performance degradation of LSCF in the presence of 1 ppm $SO_2$ at 700 °C, 750 °C, and 800 °C, respectively. They found that the performance degradation was more severe at either 700 °C or 750 °C, even though no significant sulfur was detected in either of these two samples.

Up to now, there are two principal methods to investigate the relationship between the stabilities of the electrodes and the long-term degradation under SOFC applications, CALPHAD[13-17] and DFT (Density-Functional Theory)[18,19]. Both can provide a potential approach to designing highly efficient and stable electrode materials with excellent impurity tolerance for SOFCs. Darvish *et al.*[15] reported the impacts of temperatures, $SO_2$ partial pressure ($P(SO_2)$), oxygen partial pressure ($P(O_2)$), and Sr content on the sulfur poisoning phenomena of the LSCF cathode using the CALPHAD approach and electrochemical analysis. They concluded that the formation of $SrSO_4$ and spinel phase is thermodynamically in favor of higher $P(SO_2)$, higher $P(O_2)$, lower temperatures, and higher Sr content. Later, Wang *et al.*[5] reviewed the sulfur poisoning of SOFCs, especially the LSCF cathode, using a very detailed literature review with a corresponding chemical reaction sequence taking into account the parameters like $P(SO_2)$, temperature,

$P(O_2)$, and Sr content. The primary poisoning mechanism proposed was the formation of $SrSO_4$ from the reaction of $SO_2$ with SrO and $O_2$ to form $SrSO_4$. Subsequent reactions would start to form other sulfur-containing phases like $La_2O_2SO_4$ and $La_2(SO_4)_3$ through a reaction with the La components. However, much remains unknown as to how these poisoning reactions are affected by the structure and composition of the cathode material, which limits our ability to design cathodes with improved long-term stability in the presence of natural $SO_2$ impurities. Moreover, a comprehensive understanding of the relationship between the long-term degradation and single or eventually multiple gas impurities poisoning effects is still needed.

In this work, we have utilized a combined computational and experimental approach to understand the sulfur poisoning mechanism(s) and improve the long-term durability of LSCF cathodes in order to develop better-performing sulfur-tolerant cathode materials in the future. We first examined the agreement between the simulation approach and the experimental observations at various conditions, confirming the reliability of CALPHAD for poisoning simulations in these cathode systems. Using this approach, we then investigated the accelerated testing protocol, which has become the standard method for conducting sulfur poisoning experiments involving LSCF cathode materials. We found that our simulation approach is able to predict in which systems (cathode material and treatment environment) the accelerated testing protocols reflect actual operating conditions. Finally, in-depth simulations were done for LSCF and compared with those for LSM to help understand the LSCF cell system and suggest directions toward alternative cathode materials that have superior sulfur tolerance.

## EXPERIMENTAL PROCEDURE

The as-received $(La_{0.6}Sr_{0.4})_{0.95}Co_{0.2}Fe_{0.8}O_3$ (denoted as LSCF-6428) powders (FuelCellMaterials) were shaped into several pellets with thickness and diameter of around 1 and 13 mm, respectively. Afterward, the pellets were pre-sintered in the ambient air at 1,200 °C for 2h to facilitate handling and remove the binders. Later, the pre-sintered pellets were heat-treated at 800 °C, 900 °C, and 1,000 °C under $SO_2$-containing dry air (10ppm $SO_2$ balanced with dry air, Airgas) or dry Argon (10ppm $SO_2$ balanced with Argon, Airgas) in the tube furnace (OTF-1500X-III, MTI Corporation) for 2 days. The furnace was equipped with 3 independent temperature zones, which allowed the above experiments with the same atmospheric but different temperature conditions to be run in the same batch. Finally, the samples were collected after furnace cooling for further characterization.

The crystal structures of all the samples, including the commercial powder, pre-sintered and heat-treated pellets, were examined by X-ray Diffraction with a Cu tube (PANalytical EMPYREAN) in the range of 20-60°. Later, the microstructure and elemental distributions of the heat-treated pellets were investigated by Scanning Electron Microscopy (SEM) (JEOL JSM-7000F) coupled with Energy-dispersive X-ray Spectroscopy (EDS) (Oxford Instrument X-MAX$^N$) to characterize the secondary phases. Transmission electron microscope (TEM) samples, ca. 100 nm thick cross-sections mounted on Cu support grids, were prepared using a Thermo Fisher Helios 460F1 dual-beam focused ion beam (FIB-SEM) using a Ga ion beam and standard FIB lift-out procedures to minimize ion damage/implantation. The electron beam was used to deposit a carbon protective layer on the top surface of each lift-out lamella, and the voltage of the cutting ion beam was progressively lowered as each lift-out lamella was thinned down to ca. 100 nm and cleaned. Each FIB-prepared TEM sample was characterized by TEM imaging, diffraction, and spectroscopy (elemental mapping).

Bright Field TEM images and diffraction patterns were obtained for each FIB-prepared TEM sample using a Thermo Fisher Talos TEM equipped with a CETA-M camera, operating at 200 keV, gun lens = 4, spot size = 8, CLA(2) = 70 μm, and beam current: ca. 0.3 nA. An objective lens aperture (OLA) of 70 μm was

used for imaging, and a selected area diffraction (SAD) aperture of 40 μm was used for diffraction; in diffraction mode, the camera length was 1.1 m TEM images and diffraction patterns were acquired and analyzed using Thermo Fisher TIA software.

Scanning-TEM (STEM) images and STEM-EDS elemental maps were obtained for each FIB-prepared TEM sample using a Thermo Fisher Titan Themis TEM operating at 300 keV, gun lens = 3, spot size = 3, CLA(2) = 70 μm, camera length = 130 mm, and beam current: ca. 1.0 nA, and equipped with a Thermo Fisher Super-X Si drift detector (SSD) x-ray spectrometer (ca. 25-35 kcps during STEM-EDS map acquisition on the sample). STEM-EDS spectra and elemental maps were acquired and analyzed using Thermo Fisher Velox software, with input parameters set as: sample thickness: 100 nm, density correction: 6.0 g/cm$^3$, Brown-Powel Ionization cross-section model, 1 sigma Gaussian blur for pre-filter and for post-filter.

## COMPUTATIONAL PROCEDURE

The CALPHAD approach is considered the computational tool in the present study, which is highly dependent on the reliability of the thermodynamic databases that are usually assessed and constructed using existing literature data. With the corresponding database ready, thermodynamic simulations can thus be conducted using the minimization of the Gibbs free energy of the current system in the Thermo-Calc software. So far, the CALPHAD approach has already been considered and utilized extensively and proven to be a significantly effective tool in the field of SOFCs[13-16,20-24].

To perform the simulations in the current LSCF system, the thermodynamic database established by Zhang[25] has been employed in the present work, where detailed thermodynamic parameters can be found. Based on the compound energy formalism, the LSCF perovskite phase is modeled with three sublattices:

**Perovskite:** (La $^{3+}$, $Sr^{2+}$, Va)$_1$($Co^{2+}$, $Co^{3+}$, $Co^{4+}$, $Fe^{2+}$, $Fe^{3+}$, $Fe^{4+}$, Va)$_1$($O^{2-}$, Va)$_3$

where Va stands for Vacancy. There are, in total, 42 end-members in this perovskite phase, such as $(La^{3+})_1(Fe^{3+})_1(O^{2-})_3$, where most of them are not charge neutral. The charge neutrality will be considered and maintained when doing simulations in Thermo-calc. Thus, the Gibbs energy per mole molecule can be described via the CALPHAD formalism:

$$G_m = \sum_{i,j,k} y_i\, y_j y_k\, {}^{\circ}G^{Prv}_{i:j:k} + RT\left(\sum y_i ln y_i + \sum y_j ln y_j + 3\sum y_k ln y_k\right) + {}^{Ex}G^{Prv}_m$$

where $y_i$, $y_j$, and $y_k$ represent the site fraction of each species in A-, B-, and oxygen-site, respectively. ${}^{\circ}G^{Prv}_{i:j:k}$ is the molar Gibbs energy of the end-members in the system, while ${}^{Ex}G^{Prv}_m$ is the excess Gibbs energy taking into account the interaction parameters among the sublattices. Meanwhile, several oxide solution phases are predicted to form during the simulations, of which the Co/Fe-containing Halite and Spinel phases, as the most important secondary phases, are described as:

**Halite:** (Co $^{2+}$, $Fe^{2+}$, $Fe^{3+}$, Va)$_1$($O^{2-}$)$_1$

**Spinel:** (Co $^{2+}$, $Co^{3+}$, $Fe^{2+}$, $Fe^{3+}$)$_1$($Co^{2+}$, $Co^{3+}$, $Fe^{2+}$, $Fe^{3+}$, Va)$_2$($O^{2-}$)$_4$

Similar to the treatment in our previous works[15,17], the Gibbs free energy of the missing sulfur-containing secondary phases $La_2O_2S$[26], $La_2O_2SO_4$[27] and $La_2(SO_4)_3$[28] were added together with the SSUB5 database, which did not consider these phases in the database. With this, the final obtained LSCF + S database is able to perform thermodynamic simulations of the sulfur poisoning phenomena in LSCF cathode materials. More importantly, to make thermodynamic predictions, a fixed number of thermodynamic conditions are needed to obtain reliable results, which is based on the following concerns. Firstly, the simulation conditions should reflect the current experimental operating conditions as shown above, which include a temperature range from 800 up to 1,000 °C, a $P(O_2)$ range from Argon ($10^{-5}$ atm) up to ambient atmosphere (0.21 atm) and a fixed 10 ppm $P(SO_2)$ based on the concentration of the actual gas we received from Airgas. Only in this way can the simulation results be comparable to our experimental observations. In addition, simulations should also mimic the general sintering and operation conditions under the actual or accelerating testing circumstances. Here, $P(SO_2)$, ranging from ppb level (~100 ppb in the atmosphere) to ppm level (accelerated testing condition), with $P(O_2)$, from ambient air on the surface down to the reducing conditions on TPBs due to the polarization effect[29] is considered as the atmospheric conditions in the simulations. And the temperature is from 600 to 800 °C as the general operating temperature for IT-SOFCs (Intermediate-Solid Oxide Fuel Cells). Finally, the influence of the Sr content is considered in the current simulation to investigate the relationship between the Sr content and sulfur poisoning phenomenon, of which LSCF-8228, LSCF-7328 and LSCF-6428 are chosen. Finally, the sulfur poisoning results of the current LSCF cathode were cross-compared with the most widely-used LSM20 ($La_{0.8}Sr_{0.2}MnO_3$)[17] cathode in our previous literature to better understand the sulfur-tolerant cathode in the future.

## RESULTS AND DISCUSSIONS

The CALPHAD simulation approach introduced above was applied to the LSCF-6428 cathode in 10 ppm $SO_2$ conditions, shown in Figure 1 as a function of temperature, where Figure 1A is for dry air and Figure 1 is for Argon. It should be noted that the simulation is based on 0.1 mole molecule of the LSCF-6428 cathode. More importantly, the dry air condition is for the surface of the LSCF cathode (the cathode-gas interface), while the argon condition is for the TPBs (the cathode-electrolyte-gas interface) due to the polarization effect[14]. It can be seen that the sulfur-containing phase, $SrSO_4$, will be thermodynamically stable over the temperature range (600-1,000 °C) in both dry air and argon atmospheric conditions. However, at higher temperatures, the amount of $SrSO_4$ formed drops more severely in the argon condition than in dry air, indicating that the $SrSO_4$ secondary phase is more thermodynamically favorable in higher $P(O_2)$, which agrees well with the literature[5,12] and suggests that the formation of the sulfates necessitates the oxidation of $SO_2$. Furthermore, the stability of the sulfate phase decreases with increasing temperature in both dry air and Argon, suggesting that sulfate formation favors lower temperature conditions. In addition, a halite phase, instead of the spinel phase, shows up as the stable Co-Fe oxide phase at high temperatures in Argon [Figure 1B], but not in dry air, which is also intuitive due to the lower valance of the cations in reducing conditions. It is also worth mentioning that in both conditions, the structure of the spinel phase will change with increasing temperature from a Co-rich spinel (spinel #2) to a Fe-rich spinel (spinel #1) due to the miscibility gap of the spinel phase. The stability of two spinel structures is predicted thermodynamically but may be different from the actual experimental observations.

Results from the experimental characterization of LSCF-6428 samples heat-treated in both dry air and Argon were then cross-compared with the simulation results for empirical validation. The XRD spectra of the pre-sintered LSCF-6428 and the heat-treated samples are shown in Figure 2. It is apparent from the figure that some secondary phases do form during sulfur poisoning heat treatment experiments, which mainly correspond to the $SrSO_4$ phase. Meanwhile, a few isolated peaks cannot be identified due to their limited signals, which the simulations suggest could be related to the formation of other oxide phases. More

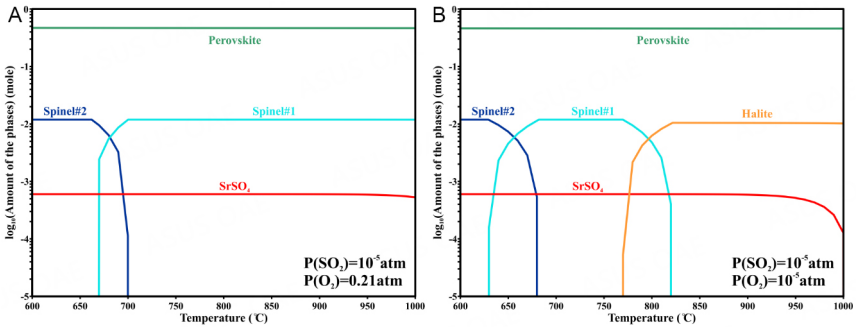


**Figure 1.** Computational results of the stability of LSCF-6428 under 10 ppm $SO_2$ in terms of temperature in a) dry air and b) argon.

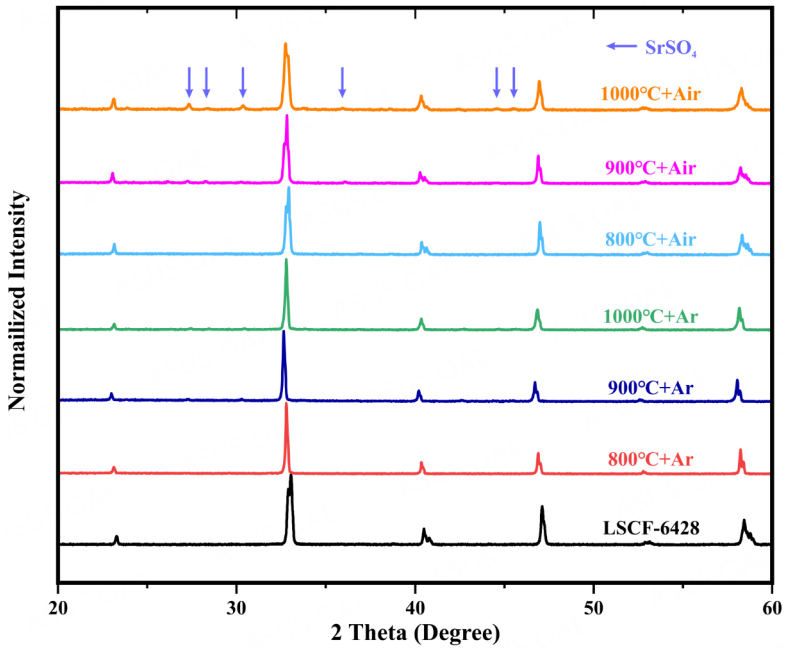


**Figure 2.** XRD spectra of the LSCF-6428 pellet (denoted as LSCF-6428) and heat-treated samples under the corresponding experimental conditions.

importantly, it can be seen from Figure 2 that the peaks of the $SrSO_4$ phase are more pronounced in the dry air condition and at higher temperatures, indicating that the process for the formation of the $SrSO_4$ phase is governed both thermodynamically and kinetically, as its formation is thermodynamically favored at higher oxygen partial pressure and lower temperature conditions.

To further identify and verify the phases formed during the LSCF-6428 heat-treatment experiments, electron microscopy analysis was used to image and spectroscopically measure the spatial elemental distribution of the samples heat-treated in 10 ppm $SO_2$ containing atmosphere for 2 days, shown in Figure 3, where Figure 3A is for dry air at 1,000 °C, Figure 3B is for dry air at 800 °C, and Figure 3C is for Argon at 1,000 °C, respectively. In 1,000 °C dry air [Figure 3A], there are two different grain morphologies besides the LSCF matrix based on the SEM-EDS mapping results: larger grains correspond to the Sr, S, and O-rich phases, while smaller ones are rich in Co, Fe, and O, which agrees well with the thermodynamic predictions from Figure 1. Similarly, as shown in Figure 3B, two different secondary phases were observed in the 800 °C dry air sample. The large but limited grains were a Co, Fe, and O-rich phase, while the tiny but well-

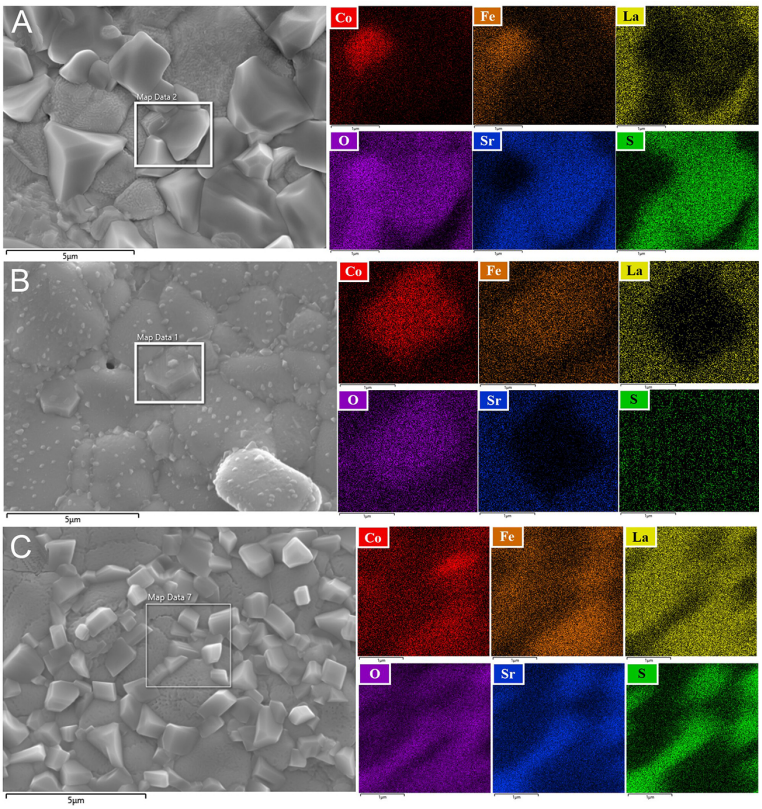


**Figure 3.** SEM micro-image and the corresponding EDS mapping results of the selected region of heat-treated LSCF-6428 for 2 days under 10 ppm $SO_2$ balanced with (A) dry air at 1,000 °C; (B) dry air at 800 °C, and (C) argon at 1,000 °C.

distributed grains along the larger grain boundaries and on the surfaces of the LSCF matrix cannot be characterized through SEM-EDS due to resolution/signal limitations. However, they were still believed to be the $SrSO_4$ phase based on the 1,000 °C dry air SEM-EDS and the predictions, where the $SrSO_4$ phase is thermodynamically more favorable at the lower temperature condition. The SEM results also indicate that the formation of the $SrSO_4$ is a kinetically limited process due to the apparent particle size difference (smaller $SrSO_4$ particles at a lower temperature), while the formation of the spinel phase is more likely governed by thermodynamics as there is no apparent grain size effect between the two temperature conditions. Finally, SEM-EDS results for the sample in Argon at 1,000 °C [Figure 3C] show that the average size of the $SrSO_4$ phase grained formed at 1,000 °C in Argon is between those formed under the conditions of 1,000 °C in dry air and 800 °C in dry air, indicating that the formation of the $SrSO_4$ phase is also governed by $P(O_2)$ when kinetics are not constrained by lower temperature. In general, the SEM-EDS results support the predictions from our thermodynamic simulations regarding the formation of the secondary phases of the LSCF cathode under $SO_2$ poisoning conditions. However, further electron microscopy characterization at higher resolution is also needed to draw determining conclusions about the exact identity of the secondary phases formed, especially the crystal structure of the oxides present.

To characterize the nano-scale features on the surface of the LSCF samples, particularly the S-rich and Co-rich phases, TEM imaging, diffraction, and spectroscopic elemental mapping (STEM-EDS) were performed on lift-out samples prepared by focused ion beam (FIB), shown in Figures 4 and 5 for the sample treated at 800 °C in dry air (same sample as Figure 3B), and in Figures 6 and 7 for the sample treated at 1,000 °C in Argon (same sample as Figure 3C). The regions of interest (ROIs) in Figure 4 and Figure 6 were chosen for

**Table 1. Quantification results from STEM-EDX analysis of LSCF treated at 800 °C in dry air from the ROIs in Figure 4**

| | ROI-4A (atomic %) | (% error) | ROI-4B (atomic %) | (% error) |
|---|---|---|---|---|
| **O** | **58.01** | 0.17 | **60.30** | 0.03 |
| **S** | **0.04** | 13.72 | **0.05** | 13.49 |
| **Fe** | **3.91** | 0.11 | **17.25** | 0.12 |
| **Co** | **37.95** | 0.04 | **3.71** | 0.30 |
| **Sr** | **0.04** | 8.22 | **7.96** | 0.19 |
| **La** | **0.05** | 2.51 | **10.72** | 0.10 |

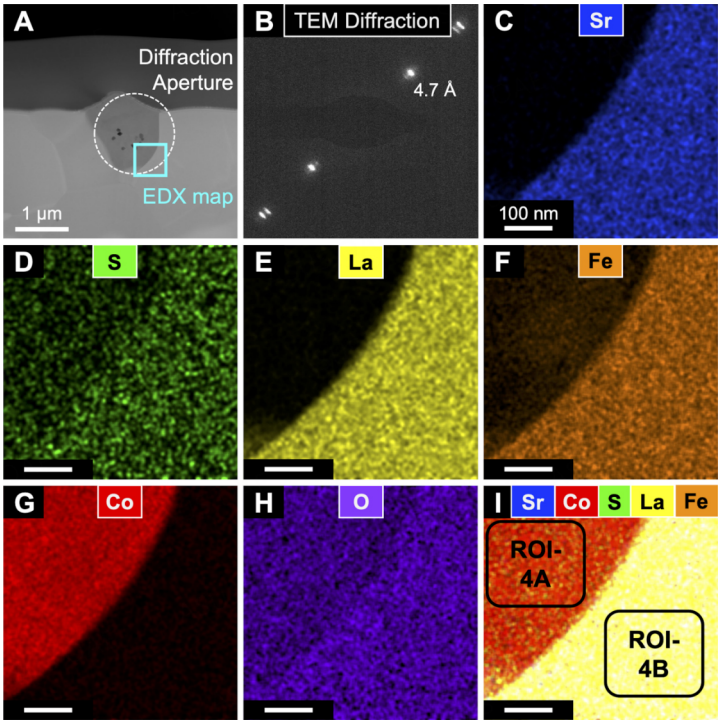


**Figure 4.** TEM characterization of a Co-rich region of the LSCF surface after treatment in dry air at 800 °C. (A) Dark field STEM image of Co-rich surface nanocrystal. White circle marks the location of the SAD aperture used to acquire a diffraction pattern of the nanocrystal, and the cyan box marks the location of the STEM EDS elemental map (C-I); (B) TEM diffraction pattern of nanocrystal; d-spacing values match those of $Co_3O_4$ for $d_{111}$; (C-H) STEM-EDX elemental maps for Sr, S, La, Fe, Co, and O, displayed as relative atomic composition for each element; (I) combined elemental map for atomic fractions of Sr, S, La, Co, and Fe. ROI-4A and ROI-4B mark the regions where atomic composition was quantified using the STEM-EDS data, shown in Table 1.

containing large Co-rich phase crystals, and the ROIs in Figure 5 and Figure 7 were chosen for containing several tiny S-rich phase crystals. For LCSF at 800 °C in the air [Figure 4], the surface crystal (ROI-4A) only contains Cr and O (and trace Fe), at ca 3:4 ([Co/Fe]:O) atomic ratio by STEM-EDS analysis [Table 1]. Based on TEM diffraction of the same surface crystal [Figure 4B], the measured d-spacing values are consistent with $Co_3O_4$, spinel #1, which fits the EDS composition results and agrees well with our modeling predictions. While the small surface crystal (ROI-5A) in Figure 5 only contains Sr, S, and O, at ca 1:1:2.5 atomic ratio by STEM-EDS analysis [Table 2]. Based on the TEM diffraction of the surface crystal, shown in Figure 5B, the crystal's measured d-spacing values are consistent with $SrSO_4$. Based on the STEM-EDX and TEM diffraction analysis of the LSCF sample treated in dry air at 800 °C, it can be concluded that the Co/O rich phase is primarily composed of the spinel #1 phase ($[Co/Fe]_3O_4$), while the small S-rich surface grains are $SrSO_4$, in agreement with the simulation predictions [Figure 1A].

**Table 2. Quantification results from STEM-EDX analysis of LSCF treated at 800 °C in dry air from the ROIs in Figure 5**

| | ROI-5A (atomic %) | (% error) | ROI-5B (atomic %) | (% error) |
|---|---|---|---|---|
| **O** | **50.31** | 1.49 | **46.14** | 0.34 |
| **S** | **22.98** | 1.15 | **0.04** | 10.15 |
| **Fe** | **0.88** | 7.58 | **23.17** | 0.05 |
| **Co** | **0.20** | 26.91 | **5.12** | 0.09 |
| **Sr** | **24.96** | 1.12 | **11.17** | 0.16 |
| **La** | **0.67** | 5.38 | **14.35** | 0.09 |

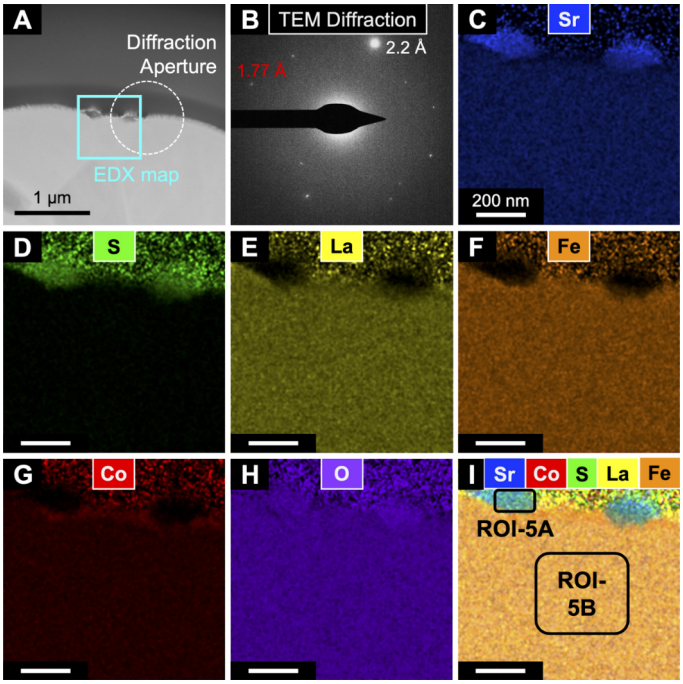


**Figure 5.** TEM characterization of a Sulfur-rich region of the LSCF surface after treatment in dry air at 800 °C. (A) Dark field STEM image of S-rich surface nanocrystals. White circle marks the location of the SAD aperture used to acquire a diffraction pattern of the nanocrystal, and the cyan box marks the location of the STEM EDS elemental map (C-I); (B) TEM diffraction pattern of nanocrystal; d-spacing values match those of $SrSO_4$ for $d_{220}$ and $d_{303}$; (C-H) STEM-EDX elemental maps for Sr, S, La, Fe, Co, and O, displayed as relative atomic composition for each element; (I) combined elemental map for atomic fractions of Sr, S, La, Co, and Fe. ROI-5A and ROI-5B mark the regions where atomic composition was quantified using the STEM-EDS data, shown in Table 2.

For the LSCF sample treated in Argon at 1,000 °C, the near-surface crystal (ROI-6A) in Figure 6 only contains Cr and O (and trace Fe), at ca 1:1 ([Co/Fe]:O) atomic ratio by STEM-EDS analysis [Table 3]. Based on TEM diffraction of the same crystal [Figure 6B], the measured d-spacing values are consistent with $Co_1O_1$, halite, which agrees with our modeling predictions. The surface crystal (ROI-7B) in Figure 7 only contains Sr, S, and O, at ca 1:1:4 atomic ratio by STEM-EDS analysis [Table 3], and based on TEM diffraction of the surface crystal [Figure 7B], the crystal's measured d-spacing values are consistent with $SrSO_4$. From experimental characterization using the STEM-EDX and TEM diffraction, it can be concluded that the LSCF sample treated in Argon at 1,000 °C has a Co/O rich phase that is primarily composed of the halite phase ($[Co/Fe]_1O_1$), while the small S-rich surface grains are $SrSO_4$, also in agreement with our simulation predictions [Figure 1B].

**Table 3. Quantification results from STEM-EDX analysis of LSCF treated at 1,000 °C in Argon from the ROIs in Figure 6**

| | ROI-6A (atomic %) | (% error) | ROI-6B (atomic %) | (% error) |
|---|---|---|---|---|
| O | 51.36 | 3.28 | 60.31 | 1.71 |
| S | 0.12 | 0.02 | 0.02 | 0.00 |
| Fe | 0.27 | 0.04 | 17.35 | 1.93 |
| Co | 48.06 | 3.32 | 3.45 | 0.45 |
| Sr | 0.04 | 0.01 | 7.72 | 0.94 |
| La | 0.15 | 0.02 | 11.15 | 1.10 |

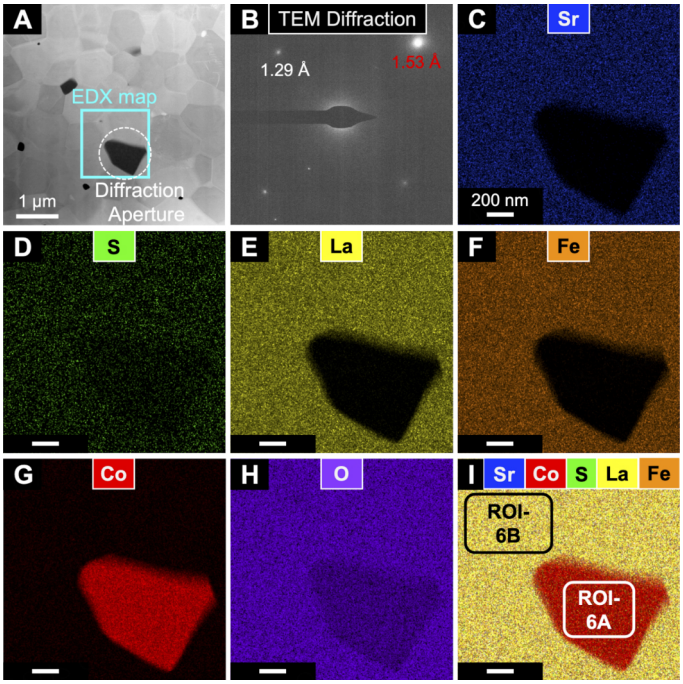


**Figure 6.** TEM characterization of a Co-rich region of the LSCF surface after treatment in Argon at 1,000 °C. (A) Dark field STEM image of Co-rich nanocrystal. White circle marks the location of the SAD aperture used to acquire a diffraction pattern of the nanocrystal, and the cyan box marks the location of the STEM EDS elemental map (C-I); (B) TEM diffraction pattern of nanocrystal; d-spacing values match those of $Co_1O_1$ for $d_{022}$ and $d_{133}$; (C-H) STEM-EDX elemental maps for Sr, S, La, Fe, Co, and O, displayed as relative atomic composition for each element; (I) combined elemental map for atomic fractions of Sr, S, La, Co, and Fe. ROI-6A and ROI-6B mark regions where atomic composition was quantified using the STEM-EDS data, shown in Table 3.

So far, the combined approach of CALPHAD simulation and XRD, SEM, and TEM characterization has provided a deeper insight into the sulfur poisoning phenomena of LSCF cathode material in the presence of $SO_2$ under various treatment temperatures and gas environments. Furthermore, the good agreement between the thermodynamic predictions and the experimental observations for LSCF highlights the utility of computational tools to study materials degradation phenomena in SOFC systems. Thus, further simulations were conducted to address more specific details, such as the influence of the LSCF cathode's Sr-composition on its sulfur poisoning, the validity of the $SO_2$ accelerated testing method for LSCF cathodes, and the long-term degradation mechanism(s) during the sulfur poisoning of LSCF cathodes.

**Table 4. Quantification results from STEM-EDX analysis of LSCF treated at 1,000 °C in Argon from the ROIs in Figure 7**

| | ROI-7A (atomic %) | (% error) | ROI-7B (atomic %) | (% error) | ROI-7c (atomic %) | (% error) |
|---|---|---|---|---|---|---|
| O | 59.57 | 0.29 | 66.28 | 0.21 | 51.14 | 0.43 |
| S | 0.04 | 12.67 | 16.13 | 0.10 | 0.11 | 18.78 |
| Fe | 19.88 | 0.08 | 0.18 | 2.97 | 3.91 | 0.61 |
| Co | 1.90 | 0.29 | 0.06 | 7.56 | 41.41 | 0.10 |
| Sr | 4.70 | 0.37 | 17.29 | 0.18 | 0.55 | 5.78 |
| La | 13.90 | 0.04 | 0.06 | 3.85 | 2.88 | 0.44 |

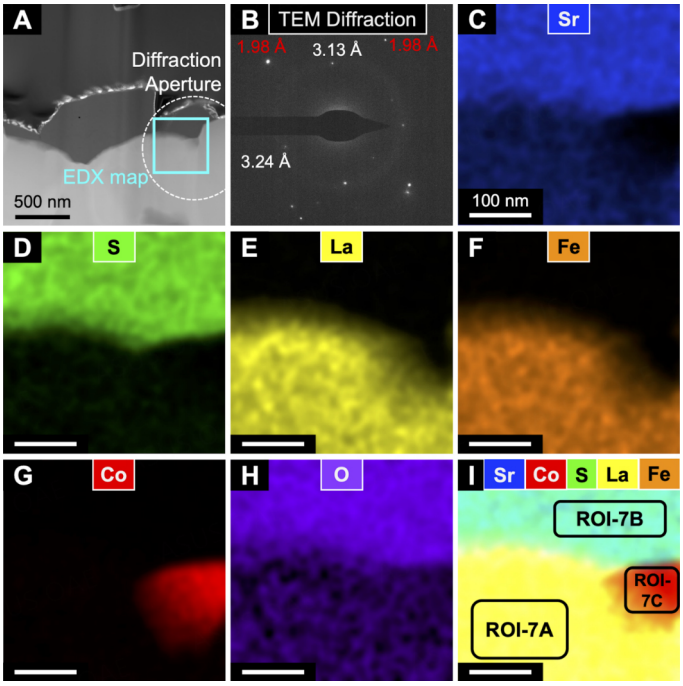


**Figure 7.** TEM characterization of a Sulfur-rich region of the LSCF surface after treatment in Argon at 1,000 °C. (A) Dark field STEM image of S-rich surface nanocrystals. White circle marks the location of the SAD aperture used to acquire a diffraction pattern of the nanocrystal, and the cyan box marks the location of the STEM EDS elemental map (C-I); (B) TEM diffraction pattern of nanocrystal; d-spacing values match those of $SrSO_4$ for $d_{102}$, $d_{210}$, and $d_{401}$; (C-H) STEM-EDX elemental maps for Sr, S, La, Fe, Co, and O, displayed as relative atomic composition for each element; (I) combined elemental map for atomic fractions of Sr, S, La, Co, and Fe. ROI-7A, ROI-7B, and ROI-7C mark the regions where atomic composition was quantified using the STEM-EDS data, shown in Table 4.

To test the effect of Sr concentration on the sulfur poisoning phenomena of LSCF cathode materials, three commonly used compositions, LSCF-6428, LSCF-7328, and LSCF-8228, listed from highest to lowest Sr content, respectively, were chosen as candidates for investigation using simulations. The thresholds for the formation of $SrSO_4$ with these three candidates are plotted in Figure 8 as a function of $P(SO_2)$ and temperature, where Figure 8A is for treatment in dry air while Figure 8B is for treatment in Argon. The $SrSO_4$ stable region infers that the $SrSO_4$ phase is thermodynamically stable under given conditions, while the $SrSO_4$ free region suggests the chemical instability of the $SrSO_4$ phase. Consequently, the optimal operating condition for each composition would be in the $SrSO_4$ free region to prevent the formation of the detrimental $SrSO_4$ secondary phase. The threshold diagrams show that the $SrSO_4$ phase is thermodynamically more favorable at lower temperatures, higher $P(SO_2)$, higher $P(O_2)$, and for higher Sr composition electrodes. This agrees well with our experimental observations for LSCF-6428, shown in Figure 3. It also matches well with prior studies by Wang *et al.*[7,11], where the formation of $SrSO_4$ was

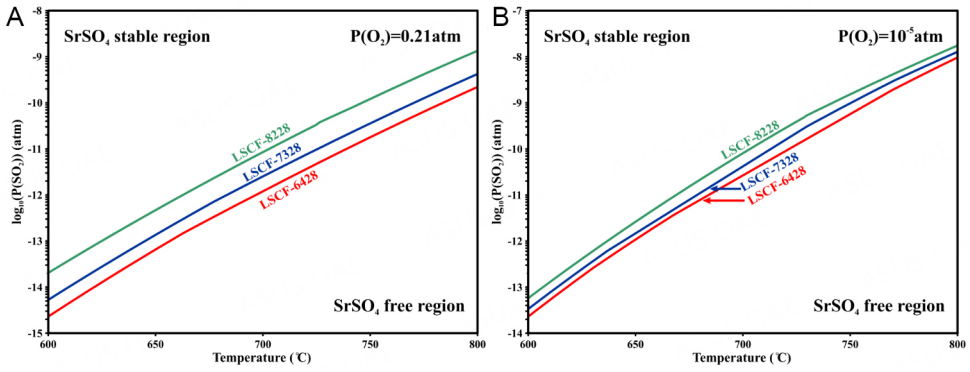


**Figure 8.** $SrSO_4$ threshold diagrams in terms of the $P(SO_2)$ and temperatures for LSCF-6428 (red), LSCF-7328 (blue) and LSCF-8228 (green) under (A) dry air and (B) Argon.

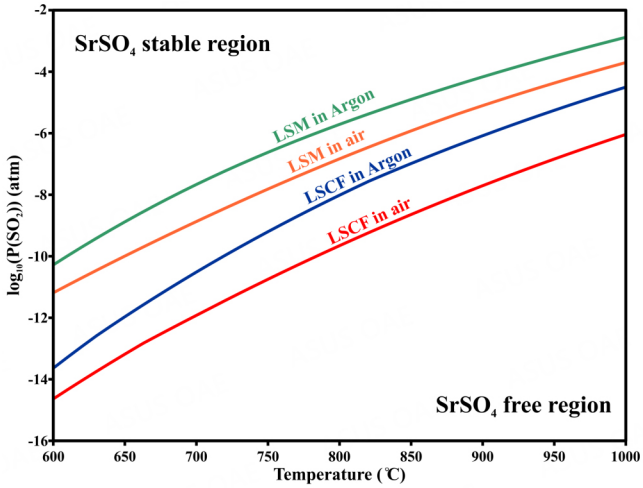


**Figure 9.** $SrSO_4$ threshold diagrams in terms of the $P(SO_2)$ and temperatures LSCF-6428 and LSM20 under air and argon conditions.

observed in the presence of 1, 10, and 100 ppm $SO_2$ but not in 0.1ppm $SO_2$, and where sulfur poisoning was more pronounced at lower temperatures. In addition, Wang *et al.*[9] investigated the effect of Sr concentration on the sulfur poisoning of LSCF by mainly using LSCF-6428 and LSCF-8228. Their results showed that the thickness of $SrSO_4$ formed in LSCF-6428 was almost four times thicker than that in LSCF-8228, which has the lower relative Sr composition of the two LSCF materials.

Under typical $P(SO_2)$ atmospheric conditions, the only sulfur-containing secondary phase observed to form in LSCF cathodes is $SrSO_4$[6-12,30,31]. It is widely accepted that the formation of $SrSO_4$ is responsible for the long-term degradation of the LSCF cathode materials[5]. Based on the simulation results in Figure 8, the threshold for $SrSO_4$ formation in LSCF-6428 at 800 °C is around $10^{-10}$ atm in air and $10^{-8}$ atm under reducing conditions. The typical ($SO_2$) content in the ambient air is at the level of $10^{-7}$ atm, indicating that ambient $SO_2$ in the air is sufficient to drive long-term degradation of the LSCF-6428 cathode. Therefore, sulfur poisoning results from accelerated testing, which is done at elevated $SO_2$ levels, should mimic the natural sulfur poisoning that occurs during cathode operation at standard testing conditions with atmospheric $SO_2$, as in both cases, there is excess $SO_2$. Thus, accelerated testing will not only reduce the total time for running poisoning experiments but is also capable of reproducing the sulfur poisoning phenomena that occur in the presence of ambient $SO_2$ during standard testing conditions. In systems where the CALPHAD simulations

predict that ambient $SO_2$ will be insufficient to drive $SrSO_4$ stability, the use of accelerated testing can potentially generate erroneous experimental results regarding $SrSO_4$ poisoning related to standard testing conditions[17]. Using this simulation approach, the reliability of accelerated testing can be scrutinized for different cathode materials under different environmental conditions.

The threshold diagrams of the LSCF cathode suggest that LSCF has poor sulfur tolerance under operating conditions, and finding an alternative sulfur-resistant cathode material thus remains of great importance. To better understand the sulfur poisoning of LSCF cathode in the context of other potential cathode materials, CALPHAD threshold diagrams were created for LSM20[17] and compared to LSCF-6428 in terms of $P(SO_2)$ and temperature under air and argon atmospheric conditions, as shown in Figure 9. Simulations predict that LSM20 has superior sulfur resistance to LSCF-6428 both in the air and argon conditions. Specifically, the threshold of $SrSO_4$ stability for LSM20 is around $10^{-7}$ atm at 800 °C in air, while it is $10^{-10}$ atm for LSCF-6428. Unlike LSCF-6428, LSM20 could be free from sulfur poisoning at ambient $SO_2$. This sulfur resistance behavior was also experimentally observed by Liu *et al.*[6], where LSM20 was found to be free from sulfur poisoning even following exposure to 1ppm of $SO_2$ for 1,000 h, and the degradation rate at 20 ppm $SO_2$ for LSCF-6428 was 8 times higher than that of LSM20. These computational and experimental results confirmed that the LSCF cathode is a poor sulfur-tolerant material in comparison with LSM. One potential solution for the poor sulfur resistance of LSCF is to place an additional sulfur filter before contacting the LSCF cathode to purge the sulfur concentration to below the threshold concentration. Another solution would be to utilize a core-shell structure[32], having LSM as the sulfur-protecting shell and LSCF cathode as the core.

## CONCLUSIONS

In this work, we have evaluated the sulfur poisoning behavior of LSCF cathode materials in the presence of $SO_2$ with a combined computational and experimental approach and tested the reliability of the simulation method for predicting sulfur poisoning behavior under different environmental conditions. The results from CALPHAD simulations regarding the stability of the secondary phases were validated using the experimental characterization, XRD, SEM, and TEM, of the same cathode materials following operations at the simulation conditions. Further simulation predictions were made to better understand the effects that other determining factors, such as temperature, $P(O_2)$, $P(SO_2)$, and cathode Sr-composition, have on the formation of $SrSO_4$ and the overall sulfur poisoning behavior. We find that the formation of $SrSO_4$ on LSCF cathodes (sulfur poisoning) is more thermodynamically favorable at lower temperatures, higher $P(SO_2)$, higher $P(O_2)$, and higher Sr composition. Finally, comparisons were made between LSCF-6428 and LSM20 cathode materials, which confirmed that LSCF-6428 has a much lower sulfur tolerance than LSM20, in agreement with recent literature. The CALPHAD simulation approach used here can be extended to other potential cathode systems to theoretically test their tolerance to poisoning at different environmental conditions, which has the potential to accelerate the experimental development of novel poisoning-resistant cathode materials.

## DECLARATIONS

### Authors' Contributions

Made substantial contributions to conception and design of this research, data analysis, writing the draft and editing: Wang R

Performed data analysis, figures preparation, writing-review and editing: Parent LR

Made substantial contributions to conception and design of this research, data analysis, funding acquisition, writing and editing: Zhong Y

### Availability of data and materials
Not Applicable.

### Financial support and sponsorship
This material is based upon work supported by the Department of Energy under Award Number (DE-FE0031652). The authors would like to thank the support and guidance from the DOE National Energy Technology Laboratory program managers, Jason Montgomery, Venkat Venkataraman and Andrew O'Connell.

### Conflicts of interest
All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate
Not applicable.

### Consent for publication
Not applicable.



## REFERENCES
1. Skinner SJ. Recent advances in Perovskite-type materials for solid oxide fuel cell cathodes. *Int J Inorg Mater* 2001;3:113-21. DOI
2. Stambouli A, Traversa E. Solid oxide fuel cells (SOFCs): a review of an environmentally clean and efficient source of energy. *Renew Sustain Energy Rev* 2002;6:433-55. DOI
3. Schuler AJ, Wuillemin Z, Hessler-wyser A, Van Herle J. Sulfur as pollutant species on the cathode side of a SOFC system. *ECS Trans* 2009;25:2845-52. DOI
4. Xiong Y, Yamaji K, Horita T, et al. Sulfur poisoning of SOFC cathodes. *J Electrochem Soc* 2009;156:B588-92. DOI
5. Wang F, Kishimoto H, Ishiyama T, et al. A review of sulfur poisoning of solid oxide fuel cell cathode materials for solid oxide fuel cells. *J Power Sources* 2020;478:228763. DOI
6. Liu RR, Taniguchi S, Shiratori Y, Ito K, Sasaki K. Influence of $SO_2$ on the long-term durability of SOFC cathodes. *ECS Trans* 2011;35:2255-60. DOI
7. Wang F, Yamaji K, Cho D, et al. Sulfur poisoning on $La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_3$ cathode for SOFCs. *J Electrochem Soc* 2011;158:B1391-7. DOI
8. Wang DJ, Jing L. Effect of $SO_2$ on performance of solid oxide fuel cell cathodes. Available from: https://www.semanticscholar.org/paper/Effect-of-SO2-on-Performance-of-Solid-Oxide-Fuel-De-jun-Jing/82c3339827f4bedfe62b0270ecab17a1aac3d0ca#citing-papers [Last accessed on 9 Mar 2023].
9. Wang F, Yamaji K, Cho D, et al. Effect of strontium concentration on sulfur poisoning of LSCF cathodes. *Solid State Ionics* 2012;225:157-60. DOI
10. Wang F, Yamaji K, Cho D, et al. Evaluation of sulfur dioxide poisoning for LSCF cathodes. *Fuel Cells* 2013;13:520-5. DOI
11. Wang C, Jiang SP. Mechanism of $SO_2$ poisoning on the electrochemical activity of LSCF and LSM electrodes. *ECS Trans* 2015;68:1023-9. DOI
12. Wang F, Kishimoto H, Develos-bagarinao K, Yamaji K, Horita T, Yokokawa H. Interrelation between sulfur poisoning and performance degradation of LSCF cathode for SOFCs. *J Electrochem Soc* 2016;163:F899-904. DOI
13. Darvish S, Asadikiya M, Hu B, Singh P, Zhong Y. Thermodynamic prediction of the effect of $CO_2$ to the stability of $(La_{0.8}Sr_{0.2})_{0.98}MnO_{3\pm\delta}$ system. *Int J Hydrogen Energy* 2016;41:10239-48. DOI
14. Darvish S, Gopalan S, Zhong Y. Thermodynamic stability maps for the $La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3\pm\delta}$-$CO_2$-$O_2$ system for application in solid oxide fuel cells. *J Power Sources* 2016;336:351-9. DOI
15. Darvish S, Wang CC, Jiang SP, Zhong Y. Thermodynamic stability mapping and electrochemical study of $La_{1-x}Sr_xCo_{0.2}Fe_{0.8}O_{3\pm\delta}$ (x = 0.2-0.4) as a cathode of solid oxide fuel cells in the presence of $SO_2$. *Electrochim Acta* 2018;287:68-77. DOI
16. Darvish S, Hu B, Singh P, Zhong Y. Thermodynamic and experimental evaluation of $La_{1-x}Sr_xMnO_{3\pm\delta}$ cathode in presence of Cr-containing humidified air. *JOM* 2019;71:3814-24. DOI
17. Wang R, Parent LR, Gopalan S, Zhong Y. Experimental and computational investigations on the $SO_2$ poisoning of $(La_{0.8}Sr_{0.2})_{0.95}MnO_3$ cathode materials. *Adv Powder Mater* 2023;2:100062. DOI

18. Walker E, Ammal SC, Suthirakun S, Chen F, Terejanu GA, Heyden A. Mechanism of sulfur poisoning of $Sr_2Fe_{1.5}Mo_{0.5}O_{6-\delta}$ perovskite anode under solid oxide fuel cell conditions. *J Phys Chem C* 2014;118:23545-52. DOI
19. Su M, Huan D, Hu X, Zhu K, Peng R, Xia C. Understanding the favorable $CO_2$ tolerance of Ca-doped $LaFeO_3$ perovskite cathode for solid oxide fuel cells. *J Power Sources* 2022;521:230907. DOI
20. Ta N, Chen M, Zhang L, et al. Numerical simulation of kinetic demixing and decomposition in a $LaCoO_{3-\delta}$ oxygen membrane under an oxygen potential gradient. *J Membr Sci* 2018;548:526-39. DOI
21. Wang CC, Darvish S, Chen K, et al. Combined Cr and S poisoning of $La_{0.8}Sr_{0.2}MnO_{3-\delta}$ (LSM) cathode of solid oxide fuel cells. *Electrochim Acta* 2019;312:202-12. DOI
22. Xu H, Cheng K, Chen M, Zhang L, Brodersen K, Du Y. Interdiffusion between gadolinia doped ceria and yttria stabilized zirconia in solid oxide fuel cells: experimental investigation and kinetic modeling. *J Power Sources* 2019;441:227152. DOI
23. Sabarou H, Wang R, Zhong Y. The origin of the phase separation in $(La_{0.8}Sr_{0.2})_{0.95}(Cr_xFe_{1-x})O_{3\pm\delta}$ perovskites for oxygen transport membranes applications. *Solid State Ion* 2020;349:115293. DOI
24. Cheng K, Xu H, Zhang L, et al. Computational engineering of the oxygen electrode-electrolyte interface in solid oxide fuel cells. *NPJ Comput Mater* 2021;7:119. DOI
25. Zhang W, Barfod R. Investigation of degradation mechanisms of LSCF based SOFC cathodes-by CALPHAD modeling and experiments. Available from: https://orbit.dtu.dk/en/publications/investigation-of-degradation-mechanisms-of-lscf-based-sofc-cathod [Last accessed on 9 Mar 2023].
26. Kumar RV, Kay DAR. Thermodynamics of the Ca-S-O, Mg-S-O, and La-S-O systems at high temperatures. *Metall Trans B* 1985;16B:287-94. DOI
27. Dwivedi RK, Kay DAR. Thermodynamics of the oxidation of rare earth oxysulfides at high temperatures. *Metall Trans B* 1984;15B:523-8. DOI
28. Kellogg HH. A critical review of sulfation equilibira. Available from: https://archive.org/details/sim_american-institute-of-mining-metallurgical-petroleum_1964-12_230_7/page/n131/mode/2up [Last accessed on 9 Mar 2023].
29. Levy C, Zhong Y, Morel C, Marlin S. Thermodynamic stabilities of $La_2Zr_2O_7$ and $SrZrO_3$ in SOFC and their relationship with LSM synthesis processes. *J Electrochem Soc* 2010;157:B1597-601. DOI
30. Gao J, Li L, Yin Z, Zhang J, Lu S, Tan X. Poisoning effect of $SO_2$ on the oxygen permeation behavior of $La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$ perovskite hollow fiber membranes. *J Membr Sci* 2014;455:341-8. DOI
31. Liu RR, Wang DJ, Jing L. Effect of $SO_2$ on the performance of LSCF cathode. *Adv Mater Res* 2014;902:41-4. DOI
32. Gopalan S, Levitas B. Core-shell heterostructures as functional materials for solid oxide fuel cell (SOFC) electrodes. Available from: https://www.osti.gov/biblio/1872369/ [Last accessed on 9 Mar 2023].

Mooraj *et al. J Mater Inf* 2023;3:4
**DOI:** 10.20517/jmi.2022.41

**Journal of Materials Informatics**

**Review**

**Open Access**



# A review on high-throughput development of high-entropy alloys by combinatorial methods

**Shahryar Mooraj, Wen Chen***

Department of Mechanical and Industrial Engineering, University of Massachusetts Amherst, Amherst, MA 01003, USA.

***Correspondence to:** Prof. Wen Chen, Department of Mechanical and Industrial Engineering, University of Massachusetts Amherst, 160 Governors Drive, Amherst, MA 01003, USA. E-mail: wenchen@umass.edu

**How to cite this article:** Mooraj S, Chen W. A review on high-throughput development of high-entropy alloys by combinatorial methods. *J Mater Inf* 2023;3:4. https://dx.doi.org/10.20517/jmi.2022.41

**Received:** 8 Dec 2022 **First Decision:** 13 Jan 2023 **Revised:** 6 Feb 2023 **Accepted:** 7 Mar 2023 **Published:** 17 Mar 2023

**Academic Editors:** Tong-Yi Zhang, Xingjun Liu, Yong Yang **Copy Editor:** Ke-Cui Yang **Production Editor:** Ke-Cui Yang

## Abstract

High-entropy alloys (HEAs) are an emerging class of alloys with multi-principal elements that greatly expands the compositional space for advanced alloy design. Besides chemistry, processing history can also affect the phase and microstructure formation in HEAs. The number of possible alloy compositions and processing paths gives rise to enormous material design space, which makes it challenging to explore by traditional trial-and-error approaches. This review highlights the progress in combinatorial high-throughput studies towards rapid prediction, manufacturing, and characterization of promising HEA compositions. This review begins with an introduction to HEAs and their unique properties. Then, this review describes high-throughput computational methods such as machine learning that can predict desired alloy compositions from hundreds or even thousands of candidates. The next section presents advances in combinatorial synthesis of material libraries by additive manufacturing for efficient development of high-performance HEAs at bulk scale. The final section discusses the high-throughput characterization techniques used to accelerate the material property measurements for systematic understanding of the composition-processing-structure-property relationships in combinatorial HEA libraries.

**Keywords:** High-entropy alloys, machine learning, combinatorial studies, high throughput, additive manufacturing, alloy design

## INTRODUCTION AND MOTIVATION

Throughout history, metallurgists have altered the properties and compositions of alloys to achieve higher-performance materials. Traditional alloy design strategies involved microalloying trace elements into a primary base element, resulting in the discovery of many valuable alloys such as Cu-based bronze, Fe-based steels, and Ni-based superalloys. Over time the increasing demand for high-performance materials has led to increasingly complex alloys[1]. This trend has peaked in the past 20 years with the introduction of multi-principal element alloys or high-entropy alloys (HEAs)[2]. Unlike traditional alloys, HEAs do not contain a single primary element; instead, multiple elements in the alloy are mixed in relatively similar (almost equiatomic) concentrations. Cantor and Yeh first popularized this new alloying strategy concept in 2004 when they independently published works describing the manufacture and design philosophy of this new class of alloys[3,4]. Since the publication of these two works, the field of HEAs has exploded as such a new alloy design paradigm opens up a vast compositional space that was previously unexplored[5]. Although some fundamental questions such as phase selection and diffusion kinetics in HEAs remain elusive, many HEAs have shown high strength[6-8], large ductility[9], exceptional hardness and wear resistance[10-12], and superior corrosion resistance[13].

Despite the great potential that HEAs present for researchers, some crucial challenges must be overcome to increase their viability for future applications. While HEAs open up an uncharted multicomponent compositional space for material design, the vast compositional space makes it impractical to explore via traditional metallurgical techniques[14]. Additionally, the cost of HEAs can vary wildly due to the variety of possible elemental combinations. Some alloy systems only contain cheap transition metals (Fe, Ni, Cr)[15] that may be easy to scale, while other systems contain refractory elements (W, Nb, Ta)[16], which can significantly raise the cost of material. Finally, processing history significantly affects the microstructure and material properties even for a given nominal alloy composition. Many processing conditions including temperature, cooling rate, mechanical deformation, and irradiation can play a significant role in the formation of constituent phases and microstructures in HEAs[17-20]. Hence, processing imposes an additional and orthogonal dimension that multiplies with the huge compositional dimension and makes it more difficult to efficiently identify high-performance alloys using conventional alloy development strategies[21-23]. Thus, it is paramount for researchers to utilize efficient workflow to minimize the cost and experimental trials to study HEAs.

Over the past decade, many high-throughput material development techniques have emerged to tackle the combinatorial nature of HEAs. These techniques include magnetron sputtering, diffusion multiples, and additive manufacturing. Magnetron sputtering uses a magnetically confined plasma to accelerate positively charged ions toward a target material, leading to the sputtering of the target atoms onto a substrate to form a thin film with a thickness ranging from a few nanometers to a few microns[24]. A combinatorial materials library can be built by sputtering multiple elemental targets onto a single substrate[24-29]. The diffusion multiples method involves arranging different metals such that they are physically touching. Then this configuration is heated to an elevated temperature that enables atomic diffusion across the interfaces between the different metals. This process leads to a compositional gradient near the interface that serves as a compositional library[30-34]. Despite the large compositional space that diffusion multiples and magnetron sputtering can achieve, these approaches encounter some difficult issues. Both techniques involve samples of microscopic length scales, and thus, the microstructures and material properties observed from these libraries may not be representative of these materials at bulk scales. In addition, magnetron sputtering involves extremely high cooling rates on the order of $10^{10}$ K/s, which are substantially higher than those involved in routine metal manufacturing[35,36]. As such, the phases and microstructures in sputtered thin films are almost exclusively polymorphic or even amorphous and thus do not represent the microstructures of bulk materials for most practical applications.

Additive manufacturing (AM), also called 3D printing, is a technology to make objects from 3D digital data, usually layer upon layer, as opposed to subtractive manufacturing technologies[37]. There exist several types of AM systems that can be used to produce metal alloys: powder bed systems, powder feed systems, thermal spray systems, and wire feed systems. In the case of a powder bed system, the feedstock powders are spread over a flatbed, and a 2D pattern is selectively melted over the bed using either a laser or electron beam as a heat source[38,39]. Powder feed systems flow powders through a delivery nozzle using a carrier gas and then melt the powders onto the substrate as it impacts the substrate using heat from a laser[40,41]. Thermal spray 3D printing (TS3DP) systems spray heated powders at high velocities onto a substrate leading to bonding between powder particles as they impact the substrate surface. This allows parts to be built layer by layer without the large heating and cooling rates of laser-based techniques[42]. Finally, wire feed systems use metal wires as feedstock and can use either electric- or plasma-based welding arcs to melt the wire and build a part layer by layer[43-46]. AM of multiple elemental feedstock powders or wires offers the capability to build large compositional libraries at bulk length scales. Furthermore, careful control of the printing parameters during AM, such as laser power and scan speed, allows for tailoring the cooling rates and resulting solidification microstructures to expand the material development space.

In order to rapidly discover new HEAs with desirable properties, researchers need to utilize an efficient workflow to leverage the strengths of various design and characterization techniques. Figure 1 illustrates a typical protocol for high-throughput development of HEAs. First, the elements of interest are selected based on their fundamental properties and interactions, which are fed into a high-throughput computational method like machine learning, molecular dynamics, CALculation of PHAse Diagram (CALPHAD), or first-principles calculations. These computational methods can then predict the bulk materials' phase formation, microstructure, and properties for initial screening of potential compositions of interest. Subsequently, high-throughput manufacturing can be used to fabricate the vast material library and high-throughput materials characterization enables rapid measurements of the material properties. This review focuses on high-throughput computational techniques, synthesis methods, and characterization studies that produce and analyze alloys with reasonable cooling rates at *bulk* scale. First, this review explores the high-throughput computational methods that can easily identify the potential compositions that show promising properties for structural or functional applications. Then, it discusses the high-throughput manufacturing of bulk compositional libraries encompassing a wide range of potentially interesting alloys by AM. The final section of this review describes some high-throughput characterization techniques to accelerate screening of multicomponent metal alloys. This combination of high-throughput methods offers a guideline for researchers to discover new alloys rapidly and efficiently.

## OVERVIEW OF HEAS

### Definition of HEAs

There currently exist two well-accepted definitions of HEAs. The first one, referred to as the "compositional definition", states that HEAs are alloys with multiple principal elements (at least 5) where each principal element makes up 5 at. % to 35 at. % of the overall composition[4,48]. The most commonly studied HEA is the Cantor alloy system which contains equiatomic CoCrFeNiMn, a prime example of this definition[2,3]. Figure 2A illustrates this high-entropy region within a ternary phase diagram, with the center of the phase diagram covered by the high-entropy region[49]. It should be noted that the edges of the phase diagram in Figure 2A may contain two or more elements to match the composition definition. Additionally, minor elements can be added to a base HEA system to tune its properties further[50].

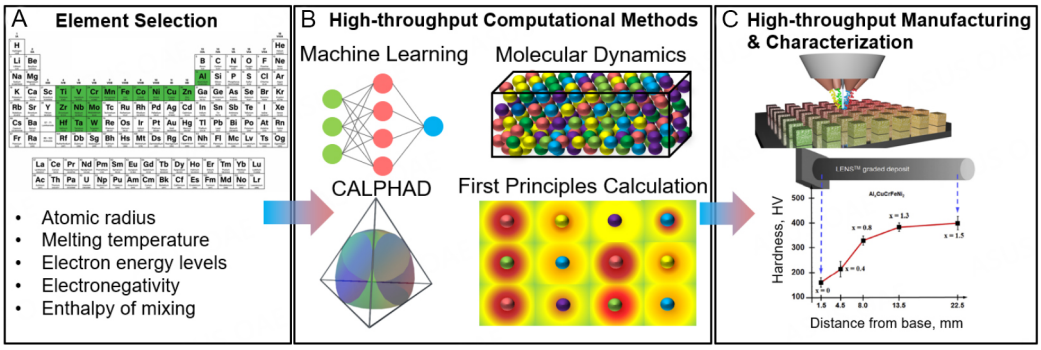


**Figure 1.** Schematic illustration of a typical protocol for high-throughput development of HEAs. (A) Selection of elements for a prospective alloy system; (B) high-throughput computational methods are used to select a range of promising compositions that can be explored experimentally; (C) high-throughput manufacturing and characterization of the promising compositions selected via computation to determine the target composition. The lower illustration in (C) is quoted with permission from Borkar *et al.*[47], copyright 2016, Elsevier. HEA: high-entropy alloy.

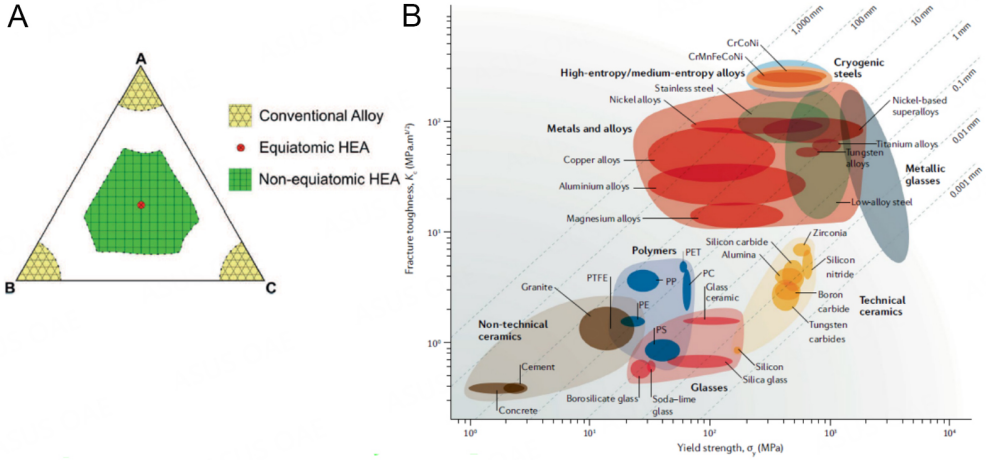


**Figure 2.** (A) Schematic illustration of the composition space of conventional alloys, equiatomic HEAs, and non-equiatomic HEAs. This figure is quoted with permission from Li *et al.*[49]; (B) ashby chart of the yield strength *vs.* fracture toughness of many material groups showing high/medium entropy alloys have excellent damage tolerance, adapted from George *et al.*[5], copyright 2019, Springer Nature. HEA: High-entropy alloy.

The second widely accepted definition is based on the mixing entropy of an alloy system, assuming an ideal random solution state. The mixing entropy is calculated as $\Delta S_{mix} = -R * \sum_{i=1}^{n} x_i \, ln(x_i)$ where $R$ is the ideal gas constant, $n$ is the number of principal elements, and $x_i$ is the atomic fraction of the i$^{th}$ element[48]. Yeh *et al.* separated the alloy design space into three regimes where a low entropy alloy has $\Delta S_{mix} < 0.69R$, a medium entropy alloy has $0.69R < \Delta S_{mix} < 1.61R$ and a HEA has $1.61R < \Delta S_{mix}$[51]. It should be noted that the definition of mixing entropy above includes the assumption that the random solution state is defined as the liquid state or a high-temperature solid solution state such that the atoms have enough energy to maintain completely random configurations[51]. However, it has been pointed out by Miracle *et al.* that the above threshold for HEAs would exclude certain non-equiatomic alloys with five principal elements, as calculations show that such alloys exhibit $\Delta S_{mix} < 1.61R$[48]. Miracle *et al.* also pointed out that others have suggested using a threshold of $1.36R < \Delta S_{mix}$ which would include the alloys that were excluded by the threshold proposed by Yeh *et al.*, making the entropy definition more consistent with the principal element definition[48].

## Four core effects

Despite their relatively short history, HEAs have already shown great potential for practical applications. Their properties are already competitive with and even exceed those of state-of-the-art materials. This potential is highlighted in Figure 2B, which illustrates the exceptional combination of high toughness and yield strength of HEAs compared to traditional structural materials[5]. The origin of these outstanding properties is often attributed to four core effects associated with HEAs: the high entropy effect, severe lattice distortion, sluggish diffusion, and cocktail effect[48]. Figure 3 presents a schematic illustration of the four core effects associated with HEAs. Each of these effects contributes to the unique properties observed in HEAs, and these contributions will be discussed below.

### *High entropy effect*

Traditional alloying strategies suggest that alloys with multi-principal elements form multi-phase, brittle intermetallic systems[51]. However, many works on HEAs show they could achieve metastable and stable single-phase solid solutions[53-55]. Even in HEAs that show multiple phases, the number of phases is much lower than the maximum number predicted by the Gibbs phase rule[56-58]. These results suggest that the high mixing entropy leads to increased mutual solubility of elements in HEA systems. The effect of high mixing entropy is described by the equation for Gibbs free energy of formation, which implies that phases with high entropy will have a lower Gibbs free energy and thus be more stable[59]. Thus, the high mixing entropy aids in stabilizing single-phase solid solutions as long as this contribution overcomes the enthalpy of formation of possible intermetallic phases, especially at elevated temperatures. Additionally, this relationship also implies that the contribution of the mixing entropy to the Gibbs free energy decreases at lower temperatures and suggests that HEAs in the form of solid solutions at high temperatures may become metastable and decompose at low temperatures. For example, Stepanov *et al.* showed that Cantor alloy exhibits a typical single-phase face-centered cubic (FCC) structure upon quenching; however, it can decompose with the precipitation of a secondary Cr-rich $\sigma$-phase after prolonged annealing at 600 °C[60]. This representative finding again underscores the importance of processing history in the phase selection of HEAs, which will be discussed in later sections.

### *Sever lattice distortion*

In HEA systems, various atoms with different atomic sizes lead to varying bond configurations and local lattice energies. These bond configurations create a high lattice distortion within the crystal structure[61,62]. The severe lattice distortion has been experimentally confirmed in many HEA systems via X-ray diffraction (XRD), neutron diffraction, and TEM[62-66]. Such severe lattice distortion leads to more diffuse scattering through the lattice and causes the broadening of diffraction peaks with a decrease in the peak intensities compared to traditional dilute alloy systems. The increase in lattice distortion also impedes the motion of dislocations through the matrix, which leads to solid solution strengthening. Traditional solid solution strengthening models typically involve the contributions of solute atoms to a matrix of solvent atoms. Still, these models are challenging to apply to HEAs as the solvent and solute atoms cannot be clearly distinguished[59]. To that end, new solid solution hardening models have been developed by accounting for the lattice and shear modulus distortion in the local environment near each constituent atom[61,67]. The lattice distortion within HEA systems has also been shown to correlate strongly with the stability of various phases. For example, single-phase solid solutions tend to be more stable in systems with low lattice distortions. In contrast, intermetallic and multi-phase structures are more likely to form in systems with high lattice distortions. This effect can sometimes outweigh the effect of high configurational entropy in phase selection[55,68,69].

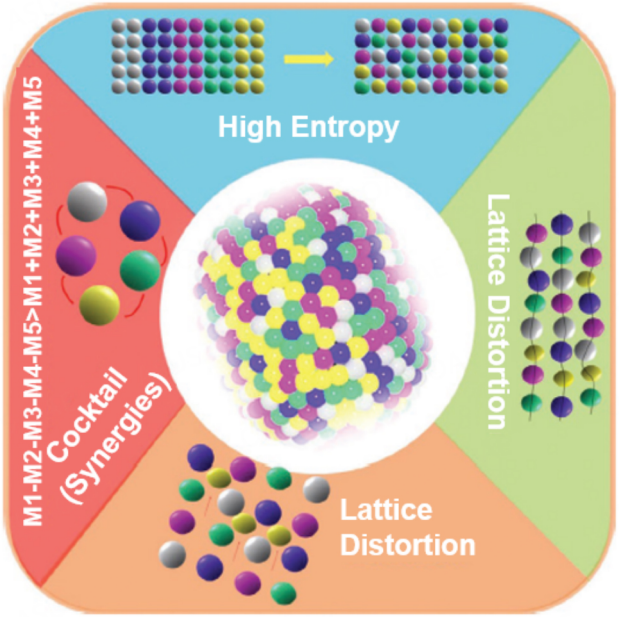


**Figure 3.** Schematic illustration of properties of HEAs. This figure is quoted with permission from Li *et al.*[52], copyright 2021, John Wiley and Sons. HEA: High-entropy alloy.

*Sluggish diffusion*

Diffusion through HEAs can be much slower than diffusion in conventional alloys. Many researchers have investigated the elemental diffusion in HEA systems and have found that the diffusivities are often much lower than those in binary or dilute alloy systems[70-72]. This sluggish diffusion can improve the stability of solid solution phases as harmful intermetallic phases can be largely suppressed. Intermetallic phases can only form under non-polymorphic solidification conditions, which require long-range diffusion. Additionally, metastable solid solutions form under polymorphic crystallization conditions, which only require topological atomic rearrangements on the atomic length scale[23]. Thus, the sluggish diffusion in HEAs suppresses the long-range diffusion that would lead to the formation of brittle intermetallic phases and instead promotes polymorphic crystallization to form solid solutions. Additionally, the coarsening of grains can be inhibited due to sluggish diffusion, leading to improved thermal stability and thermomechanical performance at elevated temperatures[73-75].

*Cocktail effect*

Dr. Ranganathan first proposed the cocktail effect to describe the synergistic nature of compositionally complex alloys[76]. This effect describes the unexpected properties observed in HEAs, bulk metallic glasses, and super-elastic and super-plastic metals (also called "gum" metals)[48]. Unlike the other effects described earlier, the cocktail effect does not predict the expected properties of HEAs. Still, it serves as a reminder that certain elemental combinations can achieve synergistic effects that are not predicted from the base constituent elements.

## HIGH-THROUGHPUT COMPUTATIONAL METHODS TO DESIGN HEAS

As previously mentioned, the compositional space opened by the concept of HEAs is vast. This design space is too large to explore through traditional trial-and-error means. Thus, it is of significant interest to identify promising compositions and phases via high-throughput computational methods[77]. These computational methods include machine learning, first-principles calculations, molecular dynamics, and CALPHAD. The

field of high-throughput computational studies is extremely wide and covers too many topics to discuss succinctly. As such, the discussion of computational methods is limited to studies focused on phase formation and mechanical properties of HEAs to illustrate the potential advantages and disadvantages of the previously mentioned methods.

### Machine learning

Machine learning (ML) is a powerful computational tool to rapidly explore vast design space through statistical methods[78]. These methods include artificial neural networks (ANN), support vector machines (SVM), and decision trees, which can often be used to quantitatively predict material properties such as hardness[79] or to predict qualitative factors such as the expected phases of a given alloy composition[80]. Over the past decade, as computational power has continued to increase, there has been an explosion in the topics of machine learning and big data[81]. Machine learning methods have an extremely high potential to handle large databases due to their statistical nature. This section includes examples from literature of various ML techniques and methods that are representative of the state-of-the-art results achieved in the field.

ML techniques are capable of predicting the structure and properties of various alloys in reasonably short periods. However, this predictive capability is largely dependent on the size and quality of the training data, a thorough consideration of appropriate input variables (also known as feature engineering), and the choice of ML model[82]. Typically robust databases of training data only exist for materials that have been well studied, such as the Ni-Ti-Hf shape memory alloy (SMA) systems[83]. For example, Liu *et al.* developed Gaussian process regression (GPR) models to estimate thermal parameters related to the martensite and austenite finish temperatures in a Ni-Ti-Hf alloy system to design a SMA[83]. The predicted parameters were described as $\bar{T} = (A_f + M_f)/2$, and $\Delta T = A_f - M_f$, where $A_f$ and $M_f$ are the austenite finish and martensite finish temperatures, respectively. The value of $\bar{T}$ represents the average of the austenite finish and martensite finish temperatures and thus illustrates the temperature region where an SMA is expected to transform. Tuning this range can be useful in aerospace applications where autonomous actuation can be induced due to the temperature difference of the surroundings at take-off (typically 275 K) and cruising (usually 215 K)[83]. On the other hand, $\Delta T$ represents the total temperature range of the austenite finish and martensite finish temperatures, indicating the hysteresis during the transformation. A low $\Delta T$ can lead to more efficient actuation when the martensitic and austenitic phase transformations are activated.

As previously mentioned, an essential aspect of building an ML model is the determination of the input variables that will most accurately predict the output variables. Typically, adding more input variables can improve the model's accuracy, as variables that do not correlate strongly with the output variables will have to be emphasized less through training sets. However, using too many input variables increases the dimensionality of the model, making it computationally expensive to execute. Additionally, the solution space formed by many input variables can often contain local minima that require many iterations to escape. For this reason, Liu *et al.* initially started with 48 input variables based on the relevance of those variables to the physical processes involved in martensitic and austenitic phase transformations[83]. These chosen features included fundamental atomic properties (e.g., atomic radius, atomic number, relative atomic mass, etc.), thermal properties (e.g., melting point, boiling point, the heat of fusion, thermal conductivity, etc.), overall alloy compositions, electronic configurations, and process conditions (e.g., solution temperature, aging temperature, etc.). This variable space was refined via mutual information (MI) and Pearson correlation (PC). MI indicates the dependency of the output variable on the input variables, which ensures that only the most impactful variables are used. In contrast, the PC between the two variables illustrates their correlations. Input variables strongly correlated to each other produce redundant

information and can thus be disregarded. Using this method, Liu *et al.* built a model that explored a 4500-point compositional space in which seven compositions were identified that exhibit $230 < \bar{T} < 260$ and the lowest $\Delta T$ values. These results illustrate ML models' capabilities to selectively tune a material system's properties using robust training datasets from previous studies and carefully selected input variables.

Most ML techniques are black-box ML with limited interpretability, which can hinder the development of chemical insight into the origin of preferable properties. Recently, a new method to implement ML described as a ML-based alloy design system (MADS) has been developed to predict alloys and maximize the hardness within an Al-Co-Cr-Cu-Fe-Ni-V alloy system[79]. This method is schematically illustrated in Figure 4A and consists of four steps. First, a database containing alloy compositions within the selected system and their measured hardness is established. Then a set of 142 features to model the hardness is established and refined to remove all except the five most crucial factors. This refinement step is important to reduce the computational cost and redundancy of information within the ML model. A model utilizing the most critical parameters is constructed and then executed to optimize the composition toward maximum hardness. Finally, the designed alloy compositions are experimentally fabricated and tested to verify the predicted properties. The five features selected were the average deviation of the atomic weight, the average deviation of the period column in the periodic table, the average deviation of the specific volume, the valance electron concentration, and the mean melting point for the alloy. After exploring the presented alloy system, the optimized composition was determined to be $Co_{18}Cr_7Fe_{35}Ni_5V_{35}$ which was predicted to have a hardness of 1,002 HV and was experimentally verified to show a hardness of 1,148 HV, showing the prediction is in good agreement with the experimental value. This new HEA exhibits about 25% greater hardness than the maximum hardness in the original training dataset. The hardness improvement illustrates ML methods' ability to take previous experimental data and extrapolate it to discover new compositions with better properties than previously achieved.

Artificial neural networks (ANNs) are common ML methods that use a layered architecture of input, hidden, and output nodes trained to predict useful material properties such as phase formation, hardness, and yield strength. The input layer consists of multiple nodes which hold values of the parameters that are known either a priori or from databases. Then each node in the 1$^{st}$ hidden layer is calculated by a weighted sum of the nodes from the input layer. Nodes in subsequent hidden layers are calculated by a weighted sum of the nodes from the previous layer. Finally, the output layer consists of the target/output parameters calculated from a weighted sum of the nodes from the hidden layer immediately preceding the output layer. These parameters can include the predicted properties of the studied alloys, such as the hardness of a material, the elastic properties, phase prediction classifier etc. The weights for every calculated sum are initialized as a best guess and then adjusted to minimize the error between the predicted and experimental values for training data. Once the error is minimized, the adjusted weights can then be used in conjunction with input data for new alloy systems outside of the training set to predict properties of interest[81,84,85].

Notably, Nassar *et al.* used two different ANNs (NN1 and NN2) to predict the phase formation with 37 possible elements in the alloy composition[86]. NN1 had only composition data as its inputs and, thus, only 37 input nodes. NN2 used composition data and some calculated thermodynamic properties of each composition, such as the entropy of mixing, enthalpy of mixing, valence electron concentration (VEC), atomic radius difference, and Pauling electronegativity difference. The output node values were a binary of 0 or 1, where 1 indicates the formation of a single-phase solid solution (SS) or a solid solution plus intermetallic (SS + IM). At the same time, 0 predicts a primary IM phase or IM + amorphous phase structure. After training, the neural networks could accurately predict the type of microstructure given an arbitrary composition with 92% and 90% accuracy for NN1 and NN2, respectively. The improved accuracy

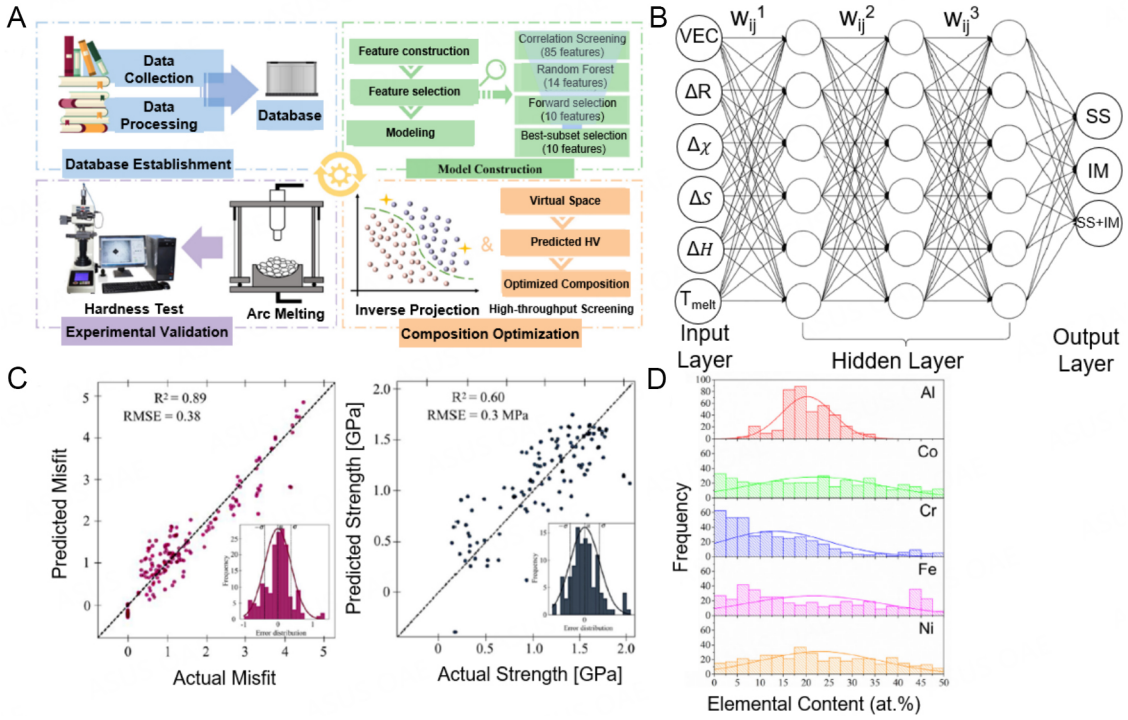


**Figure 4.** (A) Schematic diagram of machine learning-based approach to design new HEAs. This figure is quoted with permission from Yang *et al.*[79], copyright 2022, Elsevier; (B) schematic illustration of artificial neural network method, adapted from Risal *et al.*[87], copyright 2021, Elsevier; (C) actual versus predicted misfit and yield strength for 10-fold cross-validation of machine learning models, insets show the error distribution around the mean. This figure is quoted with permission from Vazquez[90], copyright 2022, Elsevier; (D) elemental content distribution of predicted eutectic HEAs, adapted from Wu *et al.*[6], copyright 2020, Elsevier. HEA: High-entropy alloy.

of NN1 is surprising, given that it only used the elemental composition as input, while NN2 included features related to thermodynamic properties.

Another work that shows consistent results with NN2 is that of Risal *et al.*, where 598 alloy compositions extracted from the literature were used as the training set, and the input parameters included the VEC, melting temperature of the alloy, enthalpy of fusion and variance of atomic radius[87]. The basic structure of the neural network used in their work is illustrated in Figure 4B. Interestingly, they achieved a prediction accuracy of 90.66%, slightly lower than that of NN1 and almost the same as NN2 in Nassar *et al.*'s work[86]. This result can be rationalized by the fact that NN typically only elucidates the correlation between parameters and thus may not always reveal the underlying physical connection between the input and output variables. Many examples exist in the literature on NNs, providing valuable predictions for HEAs' microstructure type and material properties. However, further study is needed to understand the mechanisms that lead to these valuable properties.

A common criticism of ML models is that they often lack interpretability despite their high predictive accuracy[88,89]. Sure-independence screening and sparsifying operator (SISSO) is an example of an ML method that can produce easy-to-understand relationships between the input and output variables. SISSO can output these relationships as analytical equations such that the dependence of the output variables on each input variable can be easily understood. Vazquez *et al.* recently used SISSO to predict the mechanical properties of alloys within a Ta-W-Nb-Mo-V refractory HEA (RHEA) system[90]. This method functions very differently from other ML algorithms as most methods attempt to filter the possible valuable features to

build an input space that is computationally efficient to analyze, as shown in previous examples. In the case of SISSO, the features are compiled as mathematical functions (descriptors) by applying mathematical operators to arbitrary groupings of the features. The descriptor space is then narrowed using sure independence screening (SIS) to identify descriptors that most strongly correlate to the target properties. Then, the sparsifying operator (SO) produces a linear model of the descriptors that best predicts the target property[91]. In this way, SISSO can produce models which converge even if the initial feature space is larger than the data set. Additionally, Vazquez *et al.* point out that SISSO is computationally inexpensive compared to typical ab-initio calculation methods like DFT[90]. Figure 4C shows the prediction of misfit volume and yield strength *vs.* the actual values calculated by DFT. The accuracy of the prediction of the misfit volume suggests that SISSO can reliably predict the mechanical properties of RHEA systems while remaining computationally much cheaper than DFT calculations. While the yield strength prediction overall shows a very low root mean squared error (RMSE), the R2 value is quite large, which arises due to limited experimental data and a lack of documentation of the processing conditions related to many compositions in the yield strength database. This result highlights the need for larger, more robust, and more detailed databases of experimental HEA data to improve the training quality of future ML models.

As previously mentioned, ML models can predict phase formation using solely composition information. This concept is taken even further by Wu *et al.*, who used a NN to study the effect of each element in a HEA system on the phase to predict the primary phase fraction after casting[6]. With this technique, they could design near-eutectic compositions within the Al-Co-Cr-Fe-Ni system[6]. The database[6] to train the model was prepared using experimental data from the literature, and CALPHAD calculations were performed using the nickel-based superalloy database TTNI8. Wu *et al.* chose to only use the elemental compositions as the input nodes and the primary phase fraction as the output node. The primary phase fraction was defined as 0 for eutectic compositions. In contrast, hyper- and hypo-eutectic compositions showed a positive value when FCC was predicted as the primary phase and a negative value when body-centered cubic (BCC) was predicted as the primary. After training and executing the NN, the authors identified 400 near-eutectic compositions and correlated them with the atomic fraction of each element. This plot is shown in Figure 4D, where it can be seen that the majority of the near-eutectic compositions fall into the region when Al content (at. %) is between 15% and 20% and the Cr content is below 25%. The other elements do not seem to significantly affect the formation of eutectic structures, which suggests that the Al and Cr contents are most crucial for eutectic structure formation in this alloy system. Thus, the NN was first used to predict the amount of Al that needed to be added to an equiatomic CoFeNi alloy to form a eutectic microstructure and how much Cr could be added to maintain that microstructure. Finally, the ratios of the other elements were further adjusted to predict a near-eutectic microstructure. The best composition based on the criteria of stable eutectic microstructure was $Ni_{32}Co_{30}Fe_{10}Cr_{10}Al_{18}$. This work presents the potential of ML models to refine a huge design space containing thousands of unique compositions down to a single optimized composition that can then be experimentally studied in detail.

While ML techniques such as those discussed in this section can readily analyze extremely large data sets, their accuracy depends heavily on the robustness and comprehensive nature of experimentally verified training sets[77]. There is currently a severe lack of such high-fidelity datasets to accurately train ML models to ensure ML can accurately predict the properties of future alloy systems[92]. In the meantime, as these databases expand, the scientific community is also implementing other computational methods that do not rely so heavily on previous results to predict future alloying behavior. These methods include first-principles calculations, molecular dynamics (MD), and CALPHAD calculations and will be discussed in the following sections.

### First-principles calculations

First-principles calculations (also called ab-initio) are computational methods that rely purely on fundamental quantum physical laws without additional assumptions[93]. The literature on first-principles methods is vast and presents a comprehensive overview beyond the scope of this work. This section will provide a sufficiently broad outline of the general concepts, advantages and disadvantages of these techniques pertinent to combinatorial studies of HEAs[93]. The main strength of these methods is that they do not require previous empirical observations of the predicted properties, and thus very little prior work is needed to implement them[94]. The most common practical implementation of ab-initio calculations is density functional theory (DFT) in the Khon-Sham formalism[94,95]. This method maps the quantum-mechanical many-electron Schrodinger equation onto an effective one-electron problem using electron density as a key variable. This mapping also requires the use of the exchange-correlation functional of the electron density which is not known for most systems and must be approximated either with the local density approximation (LDA)[96,97] or the generalized gradient approximation (GGA)[98-100]. Once these fundamental functions are calculated, the overall energy of the system can be calculated and used to determine the energy of formation for the possible phases of the system. This result can help researchers determine the stability of different phases to determine which phases are likely to form. It should be noted that in its initial state, DFT is a ground state theory and thus only provides the ground state energy at 0 K for a given configuration of atoms[94]. These results can be combined with thermodynamic concepts and statistical sampling techniques to bridge the gap between 0 K to a finite temperature[94].

Despite the strong predictive power of ab-initio calculations, they often suffer from high computational costs, which can significantly decrease the ability of researchers to explore the vast design space that is necessary to build accurate property maps for HEAs. To overcome this challenge, many researchers either combine first-principle calculations with more high-throughput methods like ML[101] or use new algorithms and models to improve the computational efficiency to the point where first-principles calculations can be used to explore hundreds to thousands of compositions in relatively short periods. Examples of such works will be discussed in this section.

One approach that is considered highly promising toward high fidelity and high throughput computations of HEAs is based on the small set of ordered structures (SSOS) containing several atoms[92]. This method works well to predict properties of equiatomic configurations of HEAs but loses computational efficiency when employed for non-equiatomic compositions. Sorkin *et al.* implemented a preselected set of small ordered structures (PSSOS) approach to address the issue of computational efficiency and used it to estimate the stability of BCC and FCC phases within the Al-Co-Cr-Fe-Ni system[92]. Traditionally the SSOS method uses a set of small, ordered structures (SOS) to model a HEA with a given composition. First, symmetry-unique SOS are constructed using non-conventional, non-primitive unit cells of cubic lattices. Each SOS has a unique pair correlation function. The complete set of possible SOS solutions is constructed and optimized using DFT. Then a small subset of SOS is selected by matching the pair correlation function of the target composition as a linear combination of the pair correlation functions of the selected SOS. This small set of SOS constitutes the solution of the SSOS. Screening the entire SOS solution space is impractical when studying HEAs, so the authors restricted their SOS space to those containing 5, 6, or 7 atoms. They selected the most frequent SOS structures in the solution set to further reduce the SOS space and only optimized those using DFT. This selection decreases the original SOS from over 50,000 sets to 1,500.

Through the above-mentioned process, the authors can predict the formation energy and density of the alloy system's BCC and FCC phases of 8,801 compositions. This result is exemplified in Figure 5A, which shows a plot of the formation energy and density of the BCC phase with varying Al and Cr compositions. Here the marker color represents the Ni content, and the marker size represents the Co content. It can be

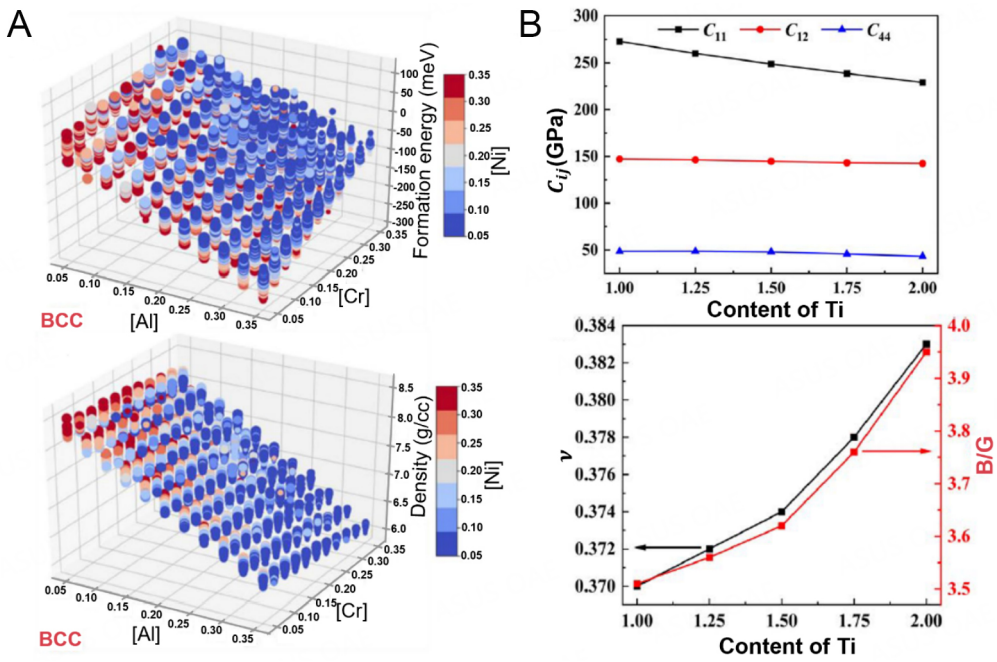


**Figure 5.** (A) Predicted formation energy and density per atom of BCC lattice structure in AlCoCrFeNi system calculated via the SSOS method as a function of Al and Cr content, the color denotes the Ni content, and marker size denotes the Co content. This figure is quoted with permission from Sorkin *et al.*[92]; (B) plot of predicted elastic constants ($C_{11}$, $C_{12}$, $C_{44}$), Poisson ratio, and Bulk modulus to shear modulus ratio calculated with the VCA model as a function of Ti in the $Ti_x$VNbMo system. This figure is quotedwith permission from Chen[104]. BCC: Body-centered cubic; SSOS: small set of ordered structures.

seen that the addition of Al leads to a substantial decrease in the formation energy of the BCC phase. After calculating the same parameters for the FCC phase, the authors found that the difference in the formation energies of BCC and FCC ($\Delta E_{BCC \rightarrow FCC} = E_{BCC} - E_{FCC}$) goes from positive to negative as the Al content increases, which is consistent with DFT calculations of the system. These results illustrate that the PSSOS method provides a new opportunity to achieve similar accuracy predictions of phase formations as DFT but with much cheaper computational costs, making this method highly suitable for high-throughput exploration of HEA space.

Virtual crystal approximation (VCA) serves as a computationally efficient alternative to more complex first-principles methods like special quasirandom structure (SQS) and similar local atomic environment (SLAE). Normally, DFT methods must use approximations to study highly disordered systems[102]. The approximation is carried out by constructing a supercell that contains multiple disordered configurations with artificially imposed boundary conditions[102]. However, such calculations require large supercells that are computationally taxing to utilize in DFT calculations. VCA deals with this issue using a pseudo-potential that averages the properties of each atom in different positions in the lattice cell. Ramer and Rappe previously investigated multiple methods to produce the averaged pseudo-potential such as averaging the pseudo-potentials for each atom within the lattice and averaging 'all-electron results'[103]. It was found that the averaging of 'all-electron results' provided the most accurate result when compared to experiments. This method involved averaging the Coulombic potentials and charge densities of the constituent atoms and then using these values to generate wavefunctions that are self-consistent solutions to the Kohn-Sham equation[103].

The computational efficiency of VCA makes it uniquely suited to explore HEA systems, as shown by Chen *et al.*, who used the VCA method to explore the effect of Ti within the $Ti_xVNbMo$ system[104]. VCA only requires the construction and analysis of a primitive cell, while other DFT methods require the use of a supercell as previously discussed, making them much more difficult to calculate. Since VCA has previously achieved reliable results for studying RHEA systems, Chen *et al.* proposed that it is reasonable to implement it to analyze the mechanical properties of this RHEA. Figure 5B illustrates the effect of Ti content on the lattice elastic constants and elastic properties [Poisson ratio and ratio of the bulk modulus to the shear modulus (B/G)]. Looking at the lattice elastic constants, it is clear that $C_{12} > C_{44}$ for all compositions, and the Cauchy pressure ($C_p = C_{12} - C_{44}$) is positive for all the compositions. This suggests that the nature of the bonding for all these compositions remains metallic. The Born-Huang mechanical stability criterion is also met ($C_{11} - C_{12} > 0$, $C_{11} + 2C_{12} > 0$ and $C_{44} > 0$), which indicates that the BCC crystal structure remains stable for these compositions. The Poisson ratio and B/G ratio seem to both increase with increasing Ti. Both of these values have been suggested to correlate well with the ductility of a material, implying that higher Ti content improves ductility. Chen *et al.* also indicated that Young's modulus (and hence yield strength) decreases with increasing Ti. To verify the accuracy of these results, the authors compared the properties of the equiatomic composition TiVNbMo to experimental values from literature and found a reasonable consistency.

New fist-principles methods such as Lederer-Toher-Vecchio-Curtarolo (LTVC) have been established over the last five years to provide novel approaches towards calculating solid solution phase stability in HEAs in order to guide future alloy discovery[105]. This method incorporates energy calculations into a mean-field statistical mechanics model, which uses order parameters to predict the transition temperature of a HEA system into a solid solution phase. The authors lay out the development of their protocol in 3 stages: (i) The automatic flow for material discovery (AFLOW)[106] repositories are used to train cluster expansion (CE) models[107] within the Alloy Theoretic Automated Toolkit (ATAT)[108] and estimate zero temperature energy configuration of atomic configurations, which are derivative structures from either FCC or BCC lattices, on which HEAs show solid solution formability; (ii) Then, the estimated atomic configurations are entered into a mean field statistical mechanical model called the generalized quasi-chemical approximation (GQCA)[109]; (iii) Finally, an order parameter is proposed by calculating the evolution of the probability of finding certain ordered configurations of atoms within the lattice.

To test this new method, the authors verified its accuracy by comparing its predictions to Monte Carlo simulations and experimental data for binary alloys. They also compared CALPHAD predictions via Thermo-calc for ternary alloys and experimental data from the literature. Once the method was considered reliable and accurate, the authors used it to predict the solid solution formation in many different alloys. They compared their predictions to the well-known empirical rules that usually inform the design of HEAs to form solid solutions. Figure 6A and B show the plots of the electronegativity and atomic size differences, as well as the VEC and atomic size differences. The large scatter of the green data points suggests that the LTVC method can be used to predict the formation of solid solutions beyond what is typically expected by the usual empirical rules. This study suggests that LTVC shows excellent potential to efficiently explore a large compositional space and discover new alloys that would not be considered under previous knowledge.

As HEAs have been increasingly studied over the past decade, certain empirical rules have been established that correlate well with the observed properties[110]. In the past, it has been suggested that these empirical rules can provide a guideline surrounding the design of HEAs with desirable properties, such as the formation of single-phase solid solutions[111]. However, certain empirical rules, such as the VEC threshold for the stability of FCC and BCC solid solutions, have failed to maintain predictive accuracy over the

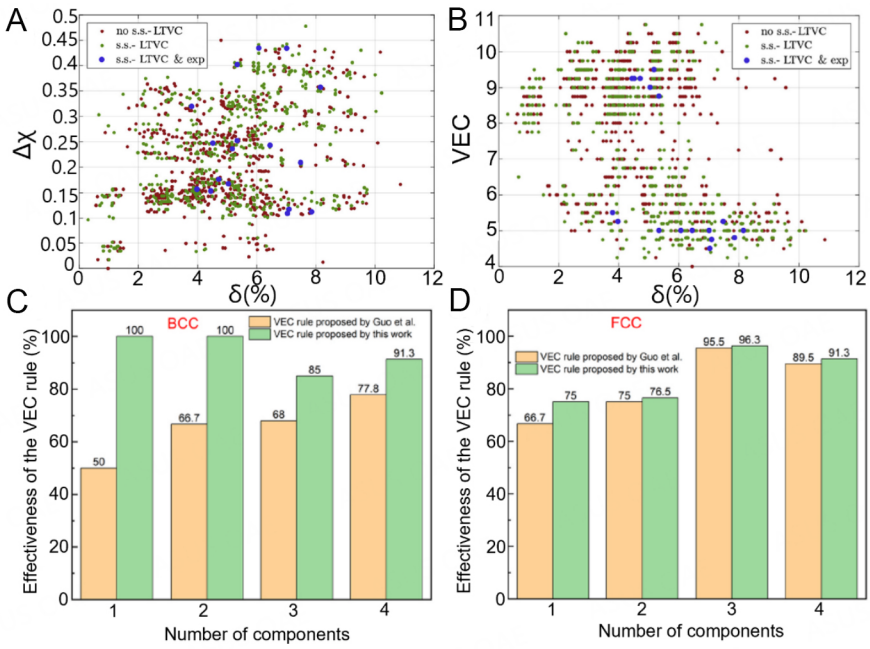


**Figure 6.** LTVC model predictions for quaternary and quinary alloys (green - predicted SS, blue - predicted SS and verified by experiments) plotted as a function of (A) the electronegativity difference and the atomic size difference; (B) the VEC and the atomic size difference. These figures are quoted with permission from Lederer *et al.*[105], copyright 2018, Elsevier. The effectiveness of the VEC rule previously used and the proposed rule from the work of Yang *et al.* in the Al-Co-Cr-Fe-Ni system[112], copyright 2022, Elsevier; (C) BCC structure (D) FCC structure. These figures are quoted with permission from Yang *et al.*[112], copyright 2022, Elsevier. BCC: body-centered cubic; FCC: face-centered cubic; LTVC: Lederer-Toher-Vecchio-Curtarolo; VEC: valence electron concentration; SS: solid solution.

years[112]. Yang *et al.* revisited this rule using special quasirandom structures (SQS) to investigate 180 compositions within the Al-Co-Cr-Fe-Ni alloy system[112]. The predictions of the phase selection between FCC and BCC were then compared to previous predictions made using the VEC of the alloy system. In their findings, Yang *et al.* argued that the threshold of VEC < 6.87 for the stability of the BCC phase could not accurately predict the phase formation in this system. The results presented by Yang *et al.* suggest that the FCC phase is stable when VEC < 8 and VEC < 5, while the BCC phase is most stable when 5 < VEC < 6.87. This finding reflects more accurately the trends illustrated in their work and is consistent with experimental data. The comparison for the accuracy of the two VEC rules is presented in Figure 6C and D, which shows that the new VEC rule has a superior prediction accuracy for both the BCC and FCC phases when the number of elemental alloy components increases from 1 to 4. It should be noted that this new rule also works well for quinary alloys but does not work well in predicting the dual-phase region. Further study is needed to produce more robust empirical rules that allow for simple rule-of-thumb predictions.

## Molecular dynamics

Molecular dynamics (MD) simulations represent a powerful tool to explore and predict material properties of potentially useful materials before significant investments in experimental characterization are made. In MD simulations, the researchers typically define an MD box that outlines the boundary conditions of the system as well as the initial positions and velocities of each atom[113]. The ambient conditions of the simulated system must also be defined, such as temperature and pressure. Once it is initialized, the system is allowed to reach thermal equilibrium[113]. Then the microscopic trajectory of each atom is determined by Newton's equations of motion depending on the potential energy functions utilized in defining the system[113,114]. The accuracy of simulated atom trajectory via MD simulations makes these methods well-

suited to study the nucleation and evolution of defects such as vacancies, dislocations, grain boundaries, and twinning[115-117].

In the past, MD simulations have also been used to explore phase transformation, mechanical behavior, nucleation and crystallization processes within HEAs[118-120]. MD simulations can study much larger systems with faster computation times than ab-initio calculations because they use classical Newtonian mechanics versus the quantum mechanical interactions on which ab-initio methods are typically based. They can also accurately simulate non-equilibrium systems due to the rapid time scales over which a simulation is conducted[113]. Despite these impressive advantages, a known weakness of MD simulations is that their accuracies depend heavily on the accuracy of the potential energy functions used to define them. However, these potential energy functions must first be measured by experimentation or calculated via ab-initio methods, which can limit the applicability of MD simulations to novel systems that have not been studied before[113]. This section presents works that take advantage of the strengths of MD simulations to explore large composition and application spaces with relatively low computation times.

As previously discussed, many computational methods can be used to investigate and predict material properties, such as yield strength, hardness, and phase formation. However, MD simulation has the added benefit of allowing researchers to investigate deformation mechanisms within an alloy via simulation of atomic motion under various ambient and loading conditions[118]. This ability is especially important as it is very difficult and laborious to observe plastic deformation processes under experimental[118,121]. Pan *et al.* applied atomic-scale tensile MD simulations to a $Fe_{80-x}Mn_xCo_{10}Cr_{10}$ alloy system to investigate transformation-induced plasticity (TRIP) and twinning-induced plasticity (TWIP) mechanisms in this system[121]. In this work, the atomic fraction of Mn, strain rate, and grain size were all adjusted to investigate each variable's effect on the system's deformation mechanisms[121]. Figure 7A shows a schematic illustration of the model where green dots represent the FCC phase and white dots denote grain boundaries. The FCC transforms into BCC and HCP during deformation, and this transformation was found to be most prevalent when x = 40. The addition of Mn also reduced the stacking fault energy, which facilitated twinning during deformation, leading to improved strain hardening. Interestingly the transformations and twinning mechanisms were suppressed for smaller nano-grain sizes, which Pan *et al.* attributed to the transformation from the intragranular evolution mechanism at larger grain sizes to the intergranular evolution mechanism at smaller grain sizes. This study shows the potential of MD simulations to explore compositional space and to provide a detailed analysis of deformation mechanisms before significant investments in experimental characterization.

MD simulations can be used to investigate the relationship between the stacking fault energy and strengthening mechanisms within an alloy system. Understanding this relationship can then provide guidelines for designing new HEAs with tailored properties and deformation mechanisms suited to specific applications[122]. Jarlov *et al.* performed MD simulations using the Large-scale Atomic/Molecularly Massively Parallel Simulator (LAMMPS) to investigate the effect of the chemical composition in the Co-Cr-Fe-Ni alloy system on the generalized stacking fault energy (GSFE)[122]. The authors used this method to explore the system's strengthening and deformation mechanisms during tensile tests. Figure 7B shows the simulated cell, and the planes marked as I, II, and III indicate the planes displaced during the tensile simulation. Based on the simulations, it was found that increasing Ni and Co contents led to an increase in the energy required to introduce stacking faults and deformation, while increasing Cr and Fe contents led to a decrease in the energy required to introduce these defects. When carrying out tensile simulations of the various compositions, it was found that the yield strength correlated linearly with the energy required to introduce intrinsic stacking faults. Thus, the strongest composition was identified as $(CoCrNi)_{90}Fe_{10}$,

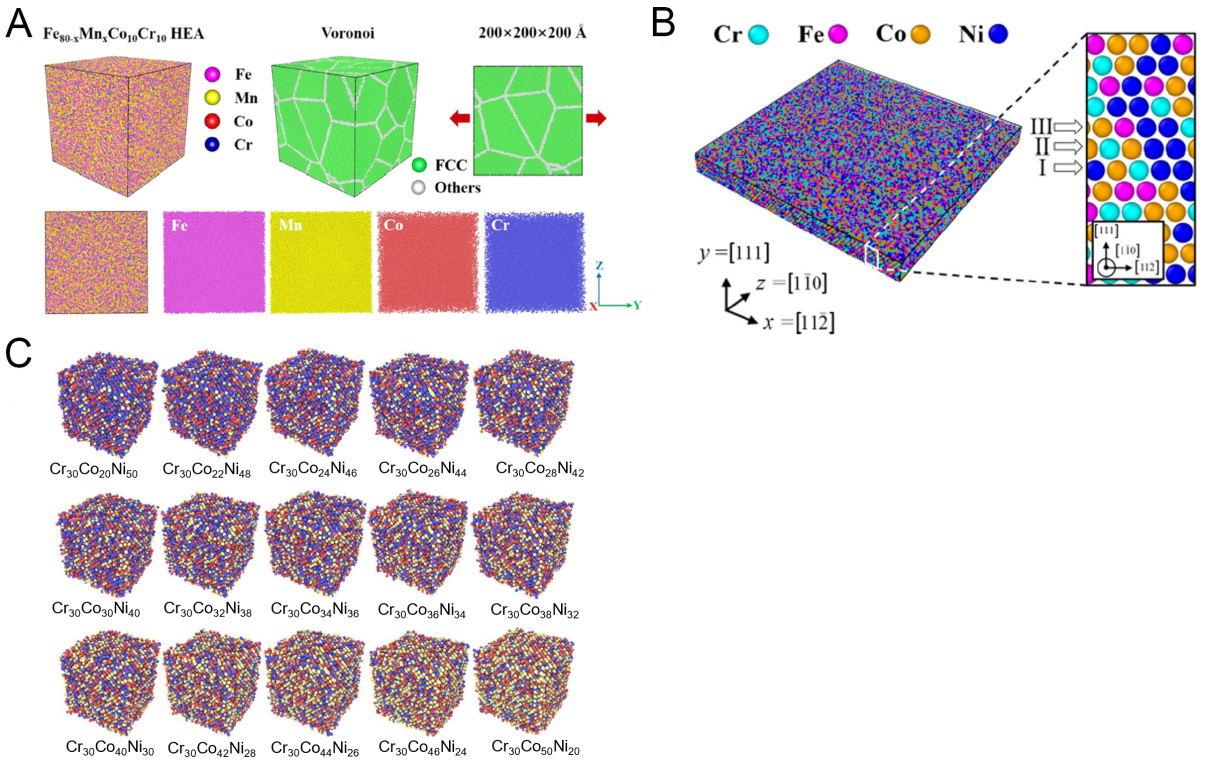


**Figure 7.** (A) Constructed molecular dynamics model of $Fe_{80-x}Mn_xCo_{10}Cr_{10}$ HEAs with FCC lattice structure and distribution of elements in the model. This figure is quoted from Pan *et al.*[121], copyright 2022, Elsevier; (B) simulation cell used in LAMMPS to calculate generalized stacking fault energy of the Co-Cr-Fe-Ni system. This figure is quoted with permission from Jarlov *et al.*[122], copyright 2022, Elsevier; (C) elemental distribution in the Co-Cr-Ni MEA system with different compositions produced by MD simulation. This figure is quoted with permission from Li *et al.*[123], copyright 2021, Elsevier. FCC: Face-centered cubic; HEA: high-entropy alloy; MD: molecular dynamics.

achieving the highest number of deformation twins. This result illustrates the power of MD simulations to optimize the alloy composition based on yield strength and tailoring of the simulated deformation mechanism.

While MD simulations can be powerful tools to predict the material properties of alloys, the high computational cost associated with these simulations makes it difficult to rapidly produce large datasets for high-throughput studies[123]. On the other hand, ML techniques are known for their potential to quickly and efficiently process and output huge amounts of data and thus offer a means to overcome the low data output of MD simulations. Li *et al.* combined high throughput MD simulation with ML to leverage both techniques' strengths to explore an extensive data set and provide accurate and detailed information on the material properties[123]. MD simulations can produce highly accurate predictions of yield strength, but the data produced by these simulations have high dimensional input-low dimensional output characteristics. These properties make it challenging to produce mathematical models to predict the correlation between input factors and yield strength. On the other hand, ML techniques can produce enormous amounts of data. Still, their accuracy requires a large and robust set of training data that experimentation cannot do. Thus, Li *et al.* utilized high-throughput MD simulation to produce an extensive training data set to train an ANN that can almost fully explore the composition space of the Co-Cr-Ni medium entropy alloy (MEA) system[123]. Figure 7C shows examples of different MD simulation models prepared for this study. The predictions made by the ANN were shown to be highly accurate. This work highlights the potential for high throughput MD simulations used in tandem with ML techniques to produce vast amounts of highly

accurate data that can efficiently identify optimal compositions within a large composition space, thereby overcoming the inherent weakness in each technique by leveraging the strengths of the other. Applying this method to other combinations of computational techniques may offer researchers new opportunities to expand the computational speed, size, and accuracy of future computational studies, which can accelerate alloy discovery far beyond the current state-of-the-art results.

The use of combining MD simulations with ML techniques was also explored by Zhang *et al.* to explore the non-equiatomic compositions within the Fe-Co-Cr-Ni-Mn alloy system[124]. In this case, the deformation of 100 compositions with a single-crystal structure was simulated in three different crystallographic directions, [100], [110], and [111]. The simulated stress-strain responses of these compositions are shown in Figure 8A. Three different ML techniques were then used to predict further the yield stress of non-equiatomic compositions within the alloy system. Unlike other ML tasks, the authors of this work used ML techniques to carry out binary classification of "Good" and "Weak" yield strength rather than quantitative prediction of yield strength. The advantage of this method is that ML programs trained with simulations of single crystals can be used to find optimized compositions that show promise as polycrystalline structures. Typically, polycrystalline models are much larger than single-crystal ones, which can make them more computationally expensive[125]. By leveraging the ability of ML classification techniques and the computational efficiency of high-throughput MD simulations of single crystals, the authors can produce highly efficient means to rapidly identify candidates for optimized compositions of HEA space. This technique was used again by Zhang *et al.* to carry out similar classification predictions for the Cu-Fe-Cr-Co-Ni alloy system[124]. It was again shown to be highly accurate and efficient at pointing out candidates with optimal yield strength[126]. This approach significantly refines the potential compositional space that experimentation needs to explore.

While MD simulations are useful in exploring the compositional space of a system, they can also be used to study material performance within other design dimensions, such as application temperature. Jian *et al.* used MD simulations to study the effect of aluminum concentration, temperature, and strain rate in amorphous $Al_xCoCrFeNi$ HEAs to study their potential as low-density structural materials[127]. Figure 8B shows the stress-strain curves of two of the three simulated compositions ranging from 300 K to 1,200 K. For all three compositions, the yield strength and Young's modulus both strongly depended on the temperature rather than the Al content. The temperature dependence of the yield strength originated from the high migration ability of atoms at higher temperatures, especially at 1,200 K, which was above the simulated glass transition temperature of about 1,100 K. The authors also varied the strain rate from $10^8$-$10^{11}$/s and found that the yield strength and Young's modulus increased with increasing the strain rate. The authors explained that a higher strain rate leads to a larger free volume but that at high strain rates, the times required for free volume rearrangement and atomic diffusion increase greatly. This relationship between free volume and atomic diffusion causes the effective free volume conducive to atomic migration to decrease. Thus, the atomic motion is impeded, which leads to increased strength. This study highlights the flexibility of MD simulations to explore compositional space and various ambient and application conditions that can provide a more holistic understanding of material performance.

### CALPHAD calculations

Phase diagrams are geometric representations of alloy systems under thermal equilibrium and typically denote the boundaries of composition and temperature where phase transformations are expected to occur[14]. These diagrams form the basis for studying solidification, crystal growth, and solid-solid phase transformations. Since the 1970s, the calculation of phase diagrams has become an integral part of alloy design, specifically through CALPHAD technology[14]. The technique relies on the minimization of the total Gibbs free energy of the system using the temperature, pressure, overall composition, and Gibbs energy function stored in databases[128].

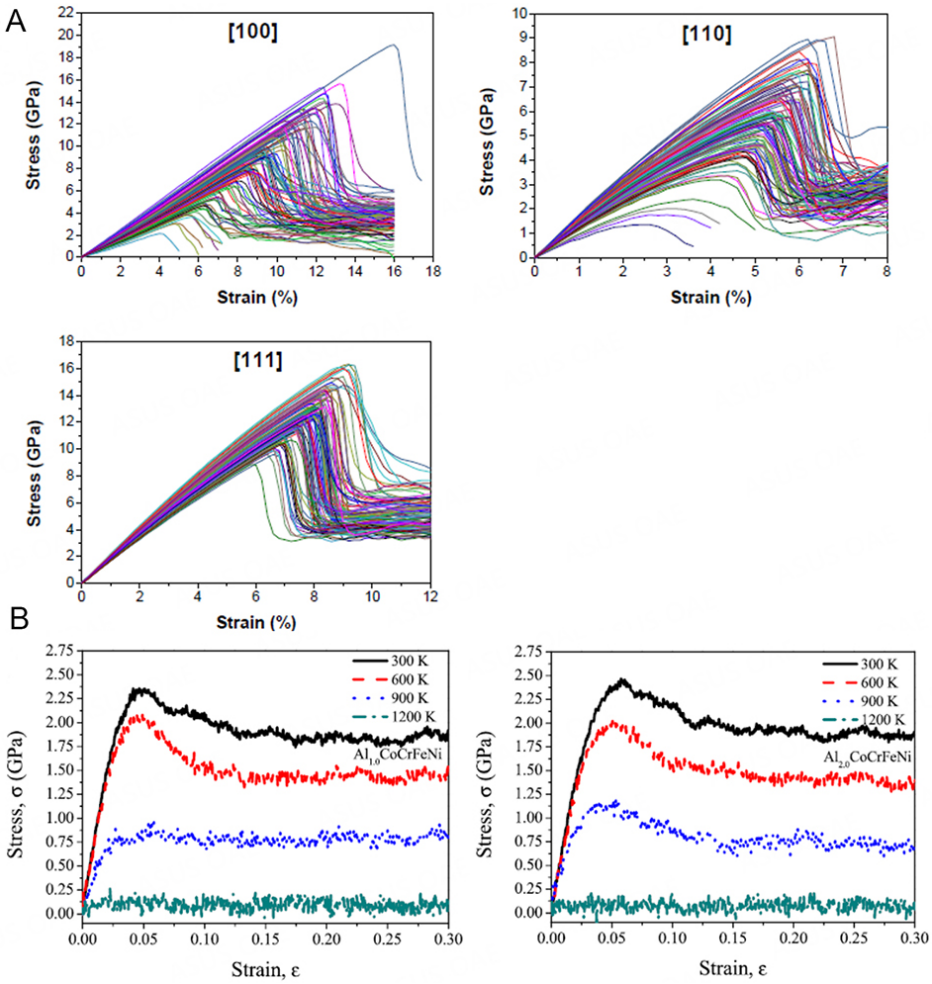


**Figure 8.** (A) MD simulated stress-strain response of single-crystal Fe-Co-Cr-Ni HEA system with different compositions (all elements are adjusted from 5 at. % to 35 at. %) loaded in different directions. This figure is quoted with permission from Zhang *et al.*[124], copyright 2021, Elsevier; (B) MD simulated stress-strain response of amorphous $Al_xCoCrFeNi$ (x = 1.0 and x = 2.0) HEAs at different temperatures. This figure is quoted with permission from Jiang *et al.*[127], copyright 2022, Elsevier. HEA: High-entropy alloy; MD: molecular dynamics.

The selection of the appropriate database is crucial for accurate calculations as the database should at least cover all the constituent binary and ternary sub-systems to provide accurate phase predictions for complicated alloy systems[129]. It should be noted that a current bottleneck in the field is the lack of comprehensive thermodynamic databases which cover large compositional and temperature spaces. Future experimental works are needed to help fill this gap. Recently, even first principle calculations have shown promise to build such databases with less effort than required for experimental characterization[129]. The current section provides examples of works that take advantage of the computational efficiency of CALPHAD methods to rapidly explore huge compositional spaces, which can reduce the large compositional spaces to ones that can be feasibly explored by experimentation.

One of the pioneering works to tackle the issue of combinatorial high-throughput studies using CALPHAD is carried out by Senkov *et al.*[130]. In this study, the authors used 9 different CALPHAD databases to

calculate every equiatomic alloy containing 3-6 elements out of 26 elements[130]. This calculation resulted in screening 130,000 different alloy compositions to predict the phases at both their melting temperatures and 600 °C. Interestingly, Senkov *et al.* found that the proportion of alloys with solid solution (SS) microstructures decreased as the number of components increased, as seen in Figure 9A[130]. This contradicts the general notion that increasing the number of elements would increase the configurational entropy and thus promote SS formation. In order to investigate the cause of this discrepancy, Senkov *et al.* calculated the entropy of mixing ($\Delta S_{mix}$) and enthalpy of mixing ($\Delta H_{mix}$) for each composition which describes the Gibbs free energy for SS phases. They also calculated the entropy of formation ($\Delta S_f$) and enthalpy of formation ($\Delta H_f$) for the intermetallic (IM) phases in each composition using CALPHAD. Then they used the entropy and enthalpy change of the different predicted phases to calculate the minimized Gibbs free energy to obtain quantitative predictions of the phase formation within each composition and compare them to reported phases in the experimental literature.

Through the previous analysis, they explained that the configurational entropy increases with ln(N), where N is the number of elements, while the possible binary interactions increase with (N/2)·(N-1). Thus, the number of binary interactions increases much faster than the configurational entropy, which increases the likelihood that an IM phase with a highly negative enthalpy of formation exists within a HEA system. Thus, the Gibbs free energy of possible IM phases decreases more rapidly than that of solid solution solutions as the number of elements increases. This work highlights the ability of large computational datasets to allow us to re-evaluate our fundamental assumptions of alloy design by providing large statistical datasets that reveal trends that may not be obvious from experimental testing.

Although many HEAs have been reported to form SS phases at lower temperatures, these are often metastable due to the inherent sluggish diffusion in HEAs. The fast computational speed of CALPHAD methods allows researchers to rapidly screen the composition phase for compositions that maintain a SS as the stable equilibrium phase even at low temperatures. Such methods have been utilized to predict the stable phases of 3 million compositions in 4 different alloy systems of AlCrMnNbTiV, AlCrMoNbTiV, AlCrFeTiV and AlCrMnMoTi[131]. This process was enabled by running approximately 100 calculations in parallel on single CPU cores in a computing cluster. This study aimed to identify various compositions that form single-phase solid solutions (SPSS) at low temperatures and then design compositions that are likely to exhibit good oxidation resistance. By incrementally adjusting the contents of various elements, the authors were able to investigate the effect of each element on the stability of SPSS. The alloy systems shown to have the most significant number of SPSS compositions were the AlCrMnNbTiV and AlCrMoNbTiV systems. Figure 9B shows the 2D projection of the compositional space explored in the AlCrMoNbTiV, where each red dot represents a composition with a predicted single-phase BCC microstructure. It was found that placing constraints to limit the Al and Cr contents can improve SPSS formation, as seen in the top 2 rows of Figure 9B but lowers the oxidation resistance. Thus, the optimal compositions found near the center of the high SPSS formation region and still maintaining a high oxidation resistance were $Al_{25}Cr_7Mn_{25}Nb_1Ti_1V_{41}$ or $Al_{21}Cr_7Mn_{21}Nb_1Ti_9V_{41}$. This study highlights the ability of high throughput CALPHAD methods to reduce a massive design space of over 3 million compositions down to a handful of promising candidates that can feasibly be explored even using conventional manufacturing methods.

The equiatomic Cantor alloy (CoCrFeMnNi) has been studied extensively in the past, including its deformation mechanism, phase formation, and mechanical properties at varying temperatures[132-134]. However, the non-equiatomic compositions have not been explored as deeply[135]. Assuming that a 1 at. % increment in any element's atomic fraction constitutes a new alloy then the compositional space for a generic 5-element alloy system covers an excess of $10^6$ unique compositions. Thus, CALPHAD or ML

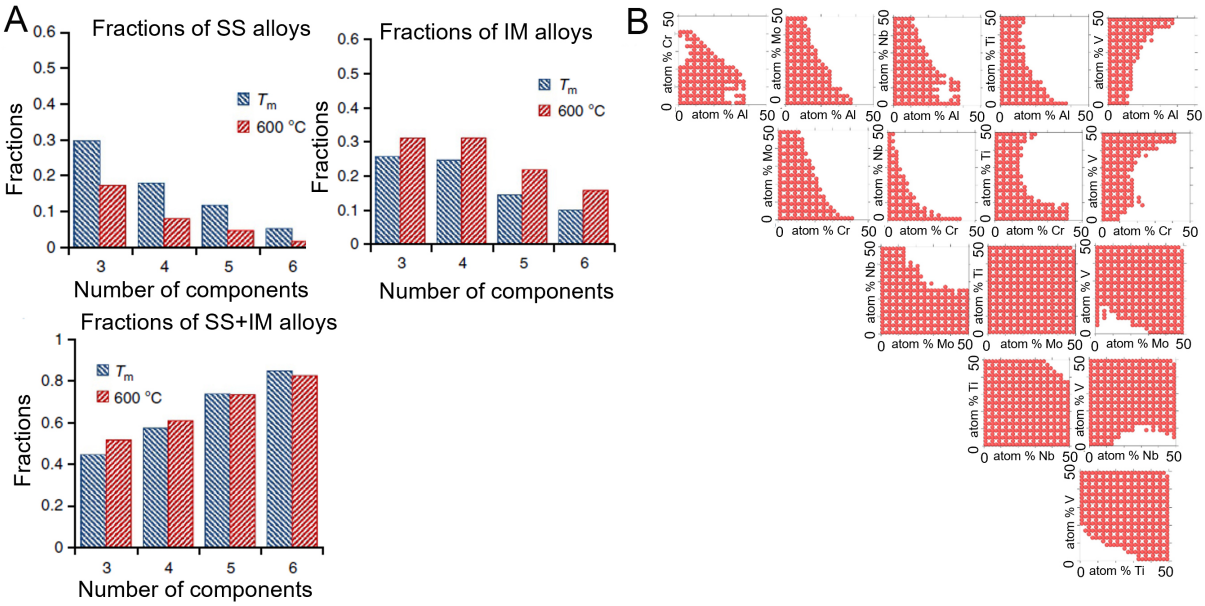


**Figure 9.** (A) Fractions of CALPHAD predicted single-phase solid solution, intermetallics, and solid solution and intermetallic equimolar alloys in 3 to 6 component alloy systems at the melting temperature ($T_m$) and at 600 °C. This figure is quoted with permission from Senkov *et al.*[130]; (B) two-dimensional projection of $Al_aCr_bMo_cNb_dTi_eV_{1-a-b-c-d-e}$ phase diagram from CALPHAD showing compositions within two-dimensional space where a BCC solid solution phase forms at 800K. This figure is quoted with permission from Klaver *et al.*[131]. IM: Intermetallic; SS: solid solution.

models are the best methods to screen through the massive composition space. However, the experimental databases on this system lack size and detail, and thus a ML approach cannot be adequately trained. For this reason, Conway *et al.* used high throughput CALPHAD methods to design composition within the Cantor alloy system (Co-Cr-Fe-Ni-Mn) that possesses a combination of high SPSS stability, good mechanical properties, and low material cost[135]. The high-throughput screening analyzed 1.78 million compositions where the elemental contents were gradually incremented by 1-2 at. % interval step. The phase fractions were calculated every 50 K between 500 K to 2,500 K to screen for compositions that produced thermally stable SPSS. Further constraints were applied to ensure every element was present in at least 10 at. %, and the Co and Ni contents were limited to 15 and 20 at. % to reduce the cost of the alloys. Twinning-induced plasticity (TWIP) and solid solution hardening (SSH) were fundamental strengthening mechanisms within this system. Thus, the authors used parameters within the TC-HEA database for their CALPHAD calculations of the SSH values and stacking fault energies (SFEs) for the screened compositions. Figure 10A shows the SFE and SSH plots in a quaternary diagram where the Co content was assumed constant at 10 at. %. The red circle illustrates the composition chosen by the authors ($Co_{10}Cr_{12}Fe_{43}Mn_{18}Ni_{17}$), while the red stars indicate the optimal composition using only the SFE or SSH as the guiding parameter. The composition explored showed only slightly lower yield strength than the equiatomic Cantor alloy at room temperature but showed high strength and ductility at elevated temperatures and exhibited a 40% reduction in cost compared to the equiatomic Cantor alloy. Based on these results, future thermodynamic screening for alloys can incorporate the strengthening mechanisms and material cost into complex alloy design.

The process-structure-property-performance (PSPP) relationship is the central paradigm in materials science. The fundamental goal of many materials scientists is to use computation, theory and experimentation to establish causal trends between the individual elements of PSPP to systematically achieve better material performance. To that end, Abu-Odeh *et al.* contextualized alloy design as an inverse phase stability problem (IPSP)[136]. IPSP is defined as the need to identify the set of thermodynamic

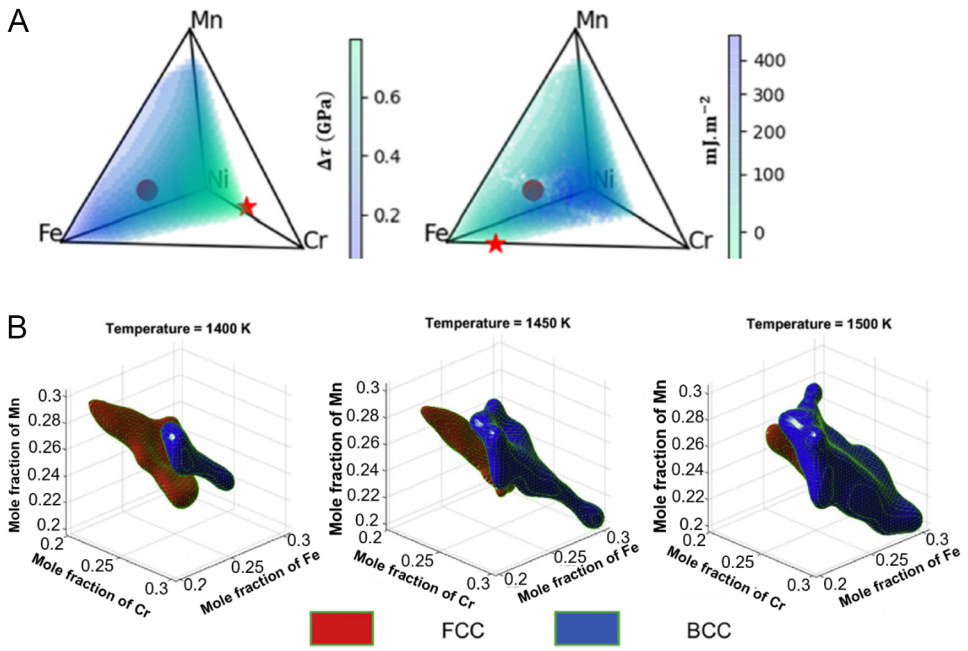


**Figure 10.** (A) Quaternary phase diagrams at fixed 10 at. % Co illustrating explored composition space. The red circle points out the composition that is experimentally tested, and the red stars indicate the target compositions to maximize hardness (left) and minimize stacking fault energy (right). This figure is quoted with permission from Conway *et al.*[135]; (B) CSA predicted single-phase solid solution compositional spaces for FCC and BCC at 1,400 K, 1,450 K, and 1,500 K. This figure is quoted with permission from Abu-Odeh *et al.*[136], copyright 2018, Elsevier. BCC: Body-centered cubic; CSA: constraint satisfaction algorithm; FCC: face-centered cubic.

conditions that lead to the stabilization of desirable phases which produce high-performance materials. One such example is to provide the coordinates composition and temperature space that result in SPSS for HEAs. The approach Abu-Odeh *et al.* took to tackle this problem is described as a constraint satisfaction algorithm (CSA) which involves the use of ML protocols executed in tandem with CALPHAD calculations to satisfy specific material property criteria/constraints.

This method enables efficient exploration of a large composition region to identify regions of arbitrarily complex phase constitution characteristics. This approach has the potential to design alloy compositions of any phase fraction rather than just focusing on the discovery of SPSS, as previously shown in other works. Abu-Odeh *et al.* applied their framework to the Cantor alloy (Co-Cr-Fe-Ni-Mn) system, where they explored the regions of SPSS stability for both FCC and BCC phases. Figure 10B visually represents the change in FCC and BCC stability with increasing temperature for a ternary sub-section of the compositions explored. After confirming the outcomes of the SPSS regions in the quinary compositions of the system, the approach was expanded to search for precipitation hardening compositions in the Al-CoCrFeNi system by identifying composition regions that include minor secondary phases. It was expressed that the secondary phase would only be considered if it did not form via spinodal decomposition, as this would not lead to any significant precipitation hardening. With this technique, the authors could identify composition spaces most likely to exhibit precipitation-hardening behavior. They highlighted that providing more detailed constraints can further refine the predicted composition space to provide a target region that can be practically explored via experimental methods.

### Comparison of computational methods

The previous categories of computational methods all serve important functions in the process of predicting and narrowing the huge compositional space of HEAs. To ensure efficient usage of computational resources

and time, it is crucial for researchers to understand which method is most useful for their individual applications. This section offers a direct comparison between their individual strengths and weaknesses based on the previously discussed studies.

ML techniques have numerous advantages that researchers can use to explore many compositions at once. Firstly, the computational efficiency of most ML methods allows some studies to screen up to $10^5$ compositions in a reasonable time frame[90,137]. The faster computation speed of ML than other computational techniques makes it particularly suited to exploring large composition spaces. Additionally, the ability to select various combinations of features as input variables, such as composition, atomic radius, valence electron concentration etc., gives ML a significant advantage in versatility allowing for pattern recognition between features that would normally not be possible with the human mind[82,138,139]. Despite the impressive capabilities, ML techniques rely heavily on large robust datasets to generate useful models[81,90,139]. Unfortunately, the relatively young age of the field of HEAs and the vast composition space means that the relative size of the currently available datasets is still quite small[48]. The small datasets limit the compositional regions where accurate ML models can be trained and applied[81,82]. The other common criticism of ML models is their lack of interpretability. ML models essentially act as a computational black box, meaning that even when they provide accurate predictions, the underlying physics is obscured by the complicated statistical calculations that are performed, making it difficult to build useful intuition from such models[78,79,87,140].

In contrast to the ML models, both first-principles and MD simulation methods rely primarily on well-known quantum mechanical and classical laws instead of statistical models[114,117,138]. This ensures that a strong fundamental understanding of the predicted properties can be extracted from such models. The reliance on fundamental physics also reduces the need for large training datasets as the required datasets are often already contained in readily available databases[125,141]. MD simulations also have the added benefit of illustrating the dynamic evolution of microstructures during an experiment, thereby providing atomic scale information on the phase transformation and deformation of materials during usage, which cannot be achieved using any other computational technique[127,142]. However, both first-principles and MD simulation methods are much more computationally expensive than ML and CALPHAD methods[138]. Thus, first-principles and MD techniques cannot explore as many compositions as ML and CALPHAD methods as seen in Table 1, where first-principles and MD can screen up to $10^4$ and $10^3$ compositions, respectively.

Recent studies have attempted to overcome this flaw by combining first-principles calculations with ML to produce models that are computationally efficient and highly accurate and provide physical insight into chemical segregation and phase formation[143-145]. Leong *et al.* used a cluster expansion (CE) model, which expands the configurational energy of an alloy structure in terms of various atomic clusters[144]. This model was trained using data obtained through first-principles calculations. Once the configurational energy is calculated for the clusters in the test set, the authors calculate the probability of the nearest neighbor (NN) atomic pairing between the different atomic species in a Mo-V-Nb-Ti-Zr alloy system to predict the Warren-Cowley short-range order (SRO) parameters[144,146]. This SRO allows the authors to highlight the tendency of Zr to segregate and cluster leading to the formation of intermetallic phases below 1,400 K and single-phase solutions above 1,400 K.

Finally, CALPHAD methods are both computationally efficient and have sufficiently large databases to produce accurate predictions for many HEA compositions[147,148]. In fact, CALPHAD methods are able to screen more compositions than any of the other computational methods (up to $10^6$ compositions) in a reasonable time span[135]. Despite this large computational efficiency, CALPHAD methods can only provide

**Table 1. Comparison of pros, cons, and capabilities of various computational methods**

| Computational method | Predicted Properties | Pros | Cons | Number of screened compositions | References |
|---|---|---|---|---|---|
| Machine learning | Elastic constants<br>Phase formation<br>Phase transformation temperature<br>Hardness<br>Tensile strength<br>Compressive strength | High computational efficiency<br>Versatility in predictive features | Requires large training sets<br>Lack of physical interpretability<br>Only gives statistical understanding | $10^5$ | [48,78,79,81,82,87,90,137-139] |
| First-principles | Elastic constants<br>Phase formation<br>Phase transformation temperature | Low input information needed<br>Provides fundamental understanding<br>Atomic scale detail | Computationally expensive<br>Time-consuming | $10^4$ | |
| Molecular dynamics | Elastic constants<br>Phase formation<br>Phase transformation temperature<br>Hardness<br>Tensile strength<br>Compressive strength | Provides fundamental understanding<br>Atomic scale detail<br>Dynamically simulate microstructure evolution | Computationally expensive<br>Time-consuming<br>Cannot provide macroscopic results | $10^3$ | [114,115,125,127,141,142] |
| CALPHAD | Phase formation<br>Phase transformation temperature | High computational efficiency<br>High accuracy<br>Easily interpretable | Only predicts equilibrium conditions<br>No kinetic information | $10^6$ | |

information about equilibrium phase formation and transformation temperatures which may not be representative of manufacturing or application conditions. This limitation is especially important for HEAs,c where sluggish diffusion limits the kinetics within the system, which can often lead to the formation of metastable phases that may not be expected under equilibrium conditions.

## COMBINATORIAL ADDITIVE MANUFACTURING TO EXPLORE LARGE COMPOSITIONAL SPACE

After narrowing a target composition space using computational methods, the remaining candidate compositions are still too numerous to reasonably explore via traditional metallurgical techniques. Thus, high-throughput manufacturing techniques are needed to rapidly produce samples that cover the candidate composition region. Previous studies have utilized magnetron sputtering and diffusion multiples to produce combinatorial libraries[28,149-151]. However, as previously discussed, these techniques produce samples at micro- or nano-scale, which may not be representative of bulk materials.

Additionally, the cooling rates experienced during magnetron sputtering are orders of magnitude greater than the cooling rates in traditional manufacturing settings[35,36]. Thus, there is a need for a manufacturing technique that can produce vast compositional libraries at a bulk length scale with practically relevant cooling rates. Laser additive manufacturing (LAM) has shown great promise towards that end. Previously LAM has been used to produce alloys with improved properties compared to their conventionally manufactured counterparts[36,152-159]. Two main types of LAM are used in combinatorial studies, i.e., laser directed energy deposition (DED), also known as laser engineered net shaping (LENS), and laser powder-bed fusion (L-PBF)[160]. The DED process utilizes a carrier gas that allows the powder to flow continuously while shielding it from oxidation during deposition. A laser source simultaneously heats the material upon contact with the printing substrate or previous layer[37]. In the case of L-PBF, a flatbed of powder is deposited on a substrate. A laser is then used to melt the particles in a pattern determined by design software to form a part layer by layer[161].

### Functionally graded materials

LAM can produce compositionally graded materials, making it a powerful tool for rapid combinatorial material exploration[160]. DED is the more common method used to produce graded materials, as the multiple nozzles can be coaxially aligned with the laser, ensuring that each nozzle's flow rate can be individually adjusted to spatially control the deposited alloy composition[162,163]. By dynamically changing the flow rate during AM, a compositional gradient can be formed, allowing for the exploration of a large compositional space within the same sample [164,165]. This method is schematically illustrated in Figure 11A, where a compositionally graded wall is produced starting from pure Cantor alloy at the base and increasing the content of refractory metals with increasing the build height[166]. Pegues *et al.* used this method to add Nb, Ta, and Ti6Al4V to CoCrFeMnNi to produce 3 different materials libraries[166]. They explored the effect of these additions on the microstructure and mechanical properties of the resulting alloys. Utilizing this method allowed them to efficiently explore a large compositional space without producing large samples. Micro-hardness tests on all three libraries revealed that the addition of the refractory elements resulted in increased hardness, likely due to the formation of secondary intermetallic phases such as a Laves phase[166].

While DED is the most common method to produce gradient materials compositionally, Wen *et al.* used L-PBF to produce a gradient material, as shown in Figure 11B[167]. Normally, L-PBF is considered undesirable for compositionally graded materials as the powder composition cannot be systematically controlled once loaded in a chamber. Wen *et al.* addressed this issue by adding a partition within the powder hopper so that two different powders could be loaded together. Then a mixer is placed below the hopper that mixes powders along the width of the mixer. This mixture forms a compositional gradient in the laser scan plane when the powder bed is deposited. Thus, a horizontal compositional gradient forms rather than the typical vertical gradients achieved in DED combinatorial studies[168-170]. Wen *et al.* used CoCrFe medium entropy alloy and Inconel 718 as the feedstock powders to prove this new technique's concept. At the pure CoCrFe end of the alloy gradient, a pure FCC phase structure was formed and as the Ni content increased due to the addition of Inconel 718, a secondary HCP phase was formed. The HCP phase content increased with increasing Ni content. The decrease in hardness occurred with increasing Ni-content, which is likely due to the larger sub-grain size observed near the Inconel 718 end.

Li *et al.* used DED to explore the effects of compositional and cooling rate changes on the microstructure and mechanical properties of the Al-Co-Cr-Fe-Ni alloy system [171]. First, they produced a pure CoCrFeNi substrate via casting. Then, they deposited varying amounts of Al on the substrate using a LENS system which formed different compositions of $Al_xCoCrFeNi$ along the substrate surface, ranging from x = 0.51 to x = 1.25, as shown in Figure 11C. The laser was also used for remelting straight lines parallel to the compositional gradient with different laser powers and scan speeds which induced different cooling rates in the compositional library. Three compositions from the library were also chosen to produce casting counterparts to achieve cooling rates far below what is achievable through DED. This method allowed the cooling rate to be varied from 25-6,400 K/s. Their findings showed that the lowest Al-containing compositions exhibited a dual-phase FCC + BCC structure which transitioned to a pure BCC/B2 at near equiatomic compositions. Additionally, compositions with low Al content showed a primary FCC phase with a cellular microstructure. The cellular microstructure followed a power law of the form $\lambda = A \cdot \dot{T}^{-1/3}$ where $\lambda$ is the cell size, $A$ is a fitted parameter, and $\dot{T}$ is the cooling rate. The microstructure refinement resulted in hardening following the Hall-Petch relationship. This work illustrates the potential for laser-based AM methods to rapidly and simultaneously explore the effects of composition and cooling rate on the phase evolution and mechanical properties of HEAs.

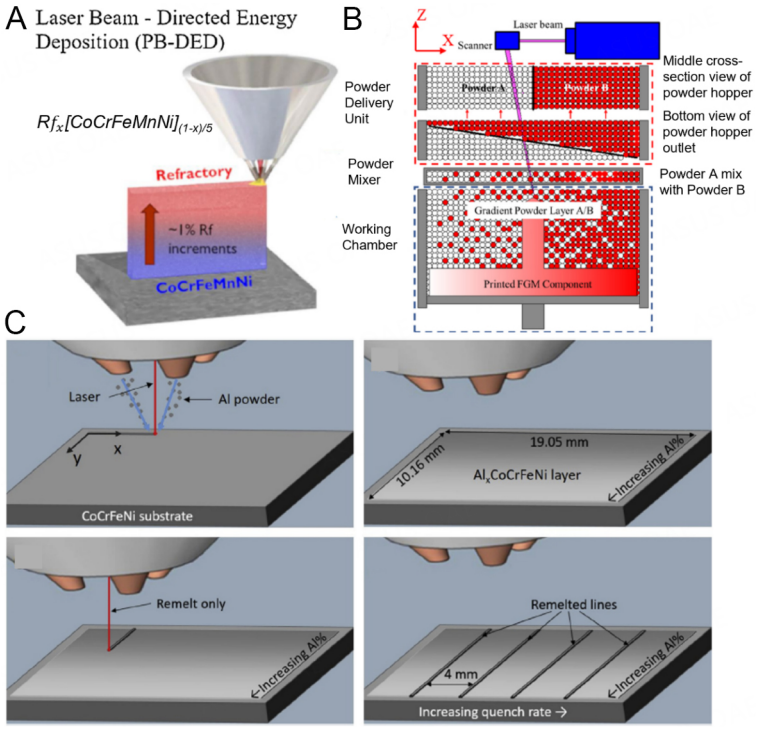


**Figure 11.** (A) Schematic illustration of high-throughput manufacturing of HEAs in a graded material under laser-directed energy deposition (L-DED). This figure is quoted with permission from Pegues *et al.*[166], copyright 2021, Elsevier; (B) schematic illustration of manufacturing a graded material under laser powder bed fusion (L-PBF) conditions. This figure is quoted with permission from Wen *et al.*[167], copyright 2021, Elsevier; (C) graded material library produced via L-DED in $Al_xCoCrFeNi$. The graded HEA library is remelted to investigate the effects of composition and cooling rate. This figure is quoted with permission from Li *et al.*[171], copyright 2020, Elsevier. HEA: High-entropy alloy.

Teh *et al.* used DED to produce compositionally graded pillars within the Co-Fe-Ni alloy system[172]. By adjusting the content of each element along the build direction, the phase fraction of FCC *vs.* BCC was varied from pure BCC at the base of the pillar to dual phase FCC + BCC to pure FCC at the top. The hardness also varied with build height due to changes in composition and grain size caused by increasing the Ni concentration. They characterized the functional properties of each composition in addition to the mechanical properties by measuring the saturation magnetization, coercivity, and electrical resistivity. After analyzing the combination of properties, the authors presented a radar chart comparing some promising compositions to pure Fe, as shown in Figure 12A.

Gwalani *et al.* varied the V content in an $AlMoCrFeV_x$ (from x = 0 to x = 1) HEA system [Figure 12B][173]. The addition of V led to solid solution hardening, increasing the hardness monotonically from 485 HV at x = 0 to 581 HV at x = 1. The microstructure remained purely BCC for all compositions and remained stable after annealing at 1,100 °C for 30 min. The grain size was also negligibly changed, which indicated high thermal stability. Zhao *et al.* blended Ti and CoCrFeNi powders in various compositions[174]. They then layered the different compositions within a powder supply bin to build a compositionally graded pillar by increasing the Ti content along the build direction[174]. All compositions showed an FCC structure primarily with minor BCC, Laves, and phases that contain Ti. Figure 12C shows a hardness map based on the results taken from the printed graded structure. As the secondary phase volume fractions increased, the hardness increased, and analysis of the various strengthening mechanisms suggested that the inclusion of the secondary phases was the main cause of the increase in strength. However, high Ti content layers also showed significant cracking. Thus, the authors concluded that 10 at. % was the maximum threshold of Ti content to produce crack-free samples and parts in CoCrFeNiTix HEA system.

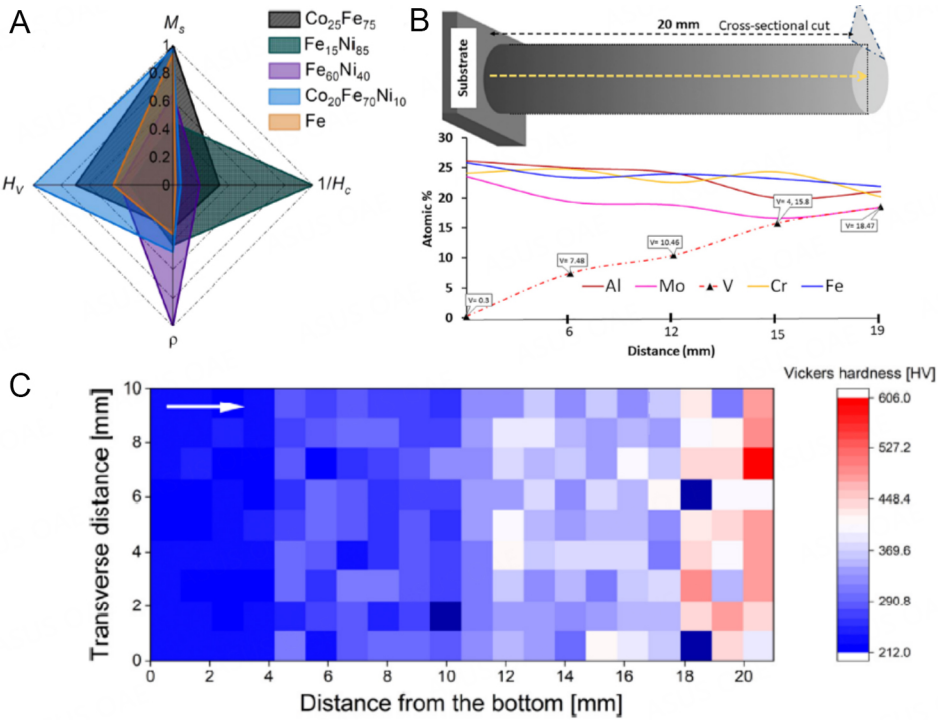


**Figure 12.** (A) Radar chart comparing properties of various promising compositions discovered from graded material library of the Co-Fe-Ni system. This figure is quoted with permission from Teh *et al.*[172], copyright 2022, Elsevier; (B) change in V concentration along the build direction for $AlMoV_xCrFe$ alloy. This figure is quoted with permission from Gwalani *et al.*[173], copyright 2019, Elsevier; (C) vickers hardness map of the side surface of graded $CoCrFeNiTi_x$ HEA system. This figure is quoted with permission from Zhao *et al.*[174], copyright 2021, Elsevier. HEA: High-entropy alloy.

### Bulk materials library

While graded materials provide a convenient means to explore multiple compositions within a single sample, they cannot give a full picture of material performance due to the possible mixing between layers during LAM, which can be difficult to control. In this sense, bulk materials libraries can produce individual samples to be studied in further detail while maintaining a high-throughput approach if rapid characterization techniques can be applied. One such example is illustrated in Figure 13A, where Yu *et al.* used DED to produce a library of bulk samples using elemental Al powder and $Cu_{50}Zr_{50}$ powder in separate powder hoppers, as seen in Figure 13A[175]. This work aimed to find the optimal compositions and processing conditions to produce bulk metallic glass composites (BMGC). BMGCs are materials formed by adding a crystalline phase into a glassy amorphous matrix[176].

The crystalline phase helps hinder the propagation of shear bands and dissipate fracture energy, which can significantly improve the room temperature ductility of BMGCs compared to monolithic bulk metallic glasses[155]. In their work, BMGC samples were deposited and then remelted to produce initially deposited melt pools with similar dimensions and different cooling rates. An 11 × 11 sample library was produced where the Al content was adjusted from 0 at. % to 10 at. % along the x-direction while the laser power was varied along the y-direction from 150 W to 400 W. Finite element modeling (FEM) was used to estimate the cooling rates and XRD analysis was used to confirm the phase constitution. After identifying phases with both amorphous and crystalline phases, the authors defined a uniformity coefficient to estimate the sample that would be expected to show the highest ductility. This criterion arose because ductility is closely related

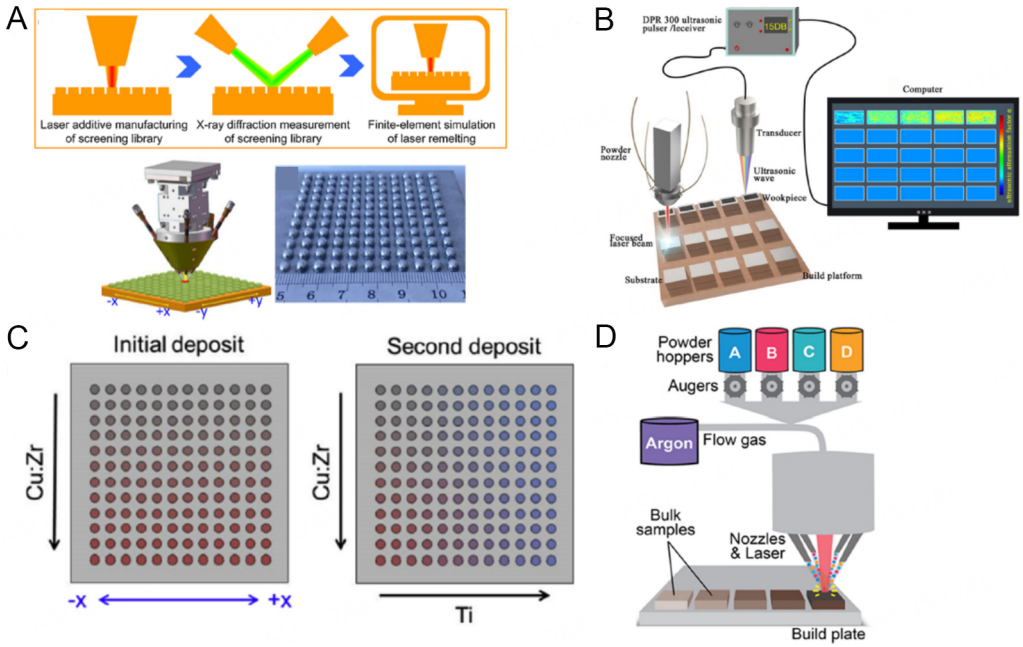


**Figure13.**(A) Schematic of high-throughput fabrication and screening of combinatorial materials library and images of the printed library of Cu-Zr-Al alloy system. This figure is quoted with permission from Yu *et al.*[175], copyright 2021, Elsevier; (B) schematic of combinatorial material library fabrication and ultrasonic screening to rapidly estimate effective processing parameters of $Zr_{51}Ti_5Ni_{10}Cu_{25}Al_9$ bulk metallic glass (BMG), adapted from Zhai *et al.*[177]; (C) schematic illustration of materials library produced with discrete dots of varying compositions. This figure is quoted with permission from Tsai *et al.*[178], copyright 2016, Elsevier; (D) schematic of DED processing of Fe-Ni-Cr-Mo bulk materials library. This figure is quoted with permission from Islam *et al.*[180], copyright 2021, AIP Publishing. DED: Directed energy deposition.

to the spatial distribution and uniformity of the crystalline dendrites within the glassy matrix. The optimal composition contained 4 at. % Al and used a remelting power of 175 W.

The rapid cooling rates induced by laser-based AM techniques can encourage the formation of amorphous structures in additively manufactured alloys and hence offer a unique opportunity to study bulk metallic glass (BMG) formation[177]. Zhai *et al.* fabricated a library of one composition with varying processing conditions to rapidly determine the optimal conditions to produce a defect-free BMG[177]. The composition used was $Zr_{51}Ti_5Ni_{10}Cu_{25}Al_9$. Using ultrasonic wave attenuation, they were able to rapidly determine the presence of defects, pores, or crystalline grain boundaries that may affect the performance of the BMG. This technique is schematically illustrated in Figure 13B. A laser power of 1,300 W and 600 mm/min was determined to provide the highest fraction of amorphous material while remaining defect-free. This work highlights the use of DED combined with ultrasonic wave attenuation to provide a non-destructive and easy way to rapidly investigate and verify the glass-forming ability of many compositions immediately after they are printed.

It is currently very difficult to predict the glass-forming ability (GFA) of an alloy composition. Thus, the current exploration of BMGs requires a high-throughput investigation similar to that of HEAs. For this reason, Tsai *et al.* deposited a combinatorial library of Cu-Zr-Ti to identify the composition with optimal GFA[178]. Figure 13C shows a schematic illustration of the construction of this library[178]. The library was built by depositing discrete hemispherical samples with varying Cu:Zr ratios between each row of samples. After the initial deposition, a layer of Ti was deposited with various feed rates and simultaneously melted onto the library. Each sample was remelted 2 more times to ensure the elements were fully melted and incorporated into each sample.

Once the library was prepared, each sample was remelted with different laser powers of 200 W, 240 W, and 280 W. Since a higher laser power leads to a lower cooling rate, compositions that maintain glassy microstructure with higher laser power should have higher GFA. Differential interference contrast (DIC) imaging under optical microscopy was used to screen for amorphous materials as samples with amorphous structures show a smooth liquid-like topography under DIC. At the same time, crystalline microstructures appear rough[179]. A total of 144 discrete samples were investigated, and 92 were identified as amorphous for the lowest power. Based on the previously mentioned criteria, the composition with the highest GFA was $Cu_{51.7}Zr_{36.7}Ti_{11.6}$, as it showed a high fraction of amorphous microstructure and was located in the center of the region of compositions that exhibit an amorphous microstructure after remelting at 280 W. It was also pointed out that this method could be extended to alloy systems with even more components by using pre-alloyed powders. Thus, the procedure laid out by Tsai *et al.* illustrates a means to rapidly identify BMGs with excellent GFA within a given alloy system. In addition to the optimal composition, combinatorial studies can also be used to rapidly determine optimal printing parameters for a given alloy system. Islam *et al.* carried out such a study on 25 different compositions in the Fe-Ni-Cr-Mo alloy system to and define a normalized dimensionless parameter based on the energy input density from the laser and the material properties of the constituent atoms[180]. A schematic illustration of their experimental method is illustrated in Figure 13D.

Eutectic HEAs (EHEAs) combine design concepts from both HEAs and eutectic alloys and show great potential for structural applications due to their impressive combination of strength and ductility[156,181]. This combination of properties arises from a hard and soft phase which help provide strength and ductility, respectively. However, further optimization is possible through minor composition adjustments to achieve near-eutectic HEAs. Joseph *et al.* produced a library of bulk $Al_xCoCrFeNi_{2.1}$ samples using the DED method to analyze the effect of Al-content on the microstructure and mechanical properties of alloys with near-eutectic compositions[182]. Figure 14A presents the XRD peak patterns of the compositions explored and shows an increase in the B2 phase with increasing the Al content. Additionally, cast samples with the compositions of each phase were prepared. These allowed the authors to investigate samples with single-phase microstructures that were either purely FCC or purely B2 phase. After analyses of the phase fractions and compressive properties of each composition, it was found that the alloys' yield strength followed a rule of mixtures based on the yield strength of the individual phases. This work highlights the ability of DED to provide large sample sets that allow for rapid characterization of multiple compositions that can elucidate strengthening trends within a system to achieve an optimal composition.

When testing the radiation damage resistance of a material, it is imperative to use bulk samples as the damage layer thickness is typically on the order of microns, and the compositional gradient struggles to maintain chemical homogeneity over large length scales. Additionally, thin-film-based materials typically form nano-grain microstructures, which artificially increase the radiation damage resistance of a material, making the results misleading compared to application conditions. Moorehead *et al.* printed a compositional library of Cr-Fe-Mn-Ni alloys to assess their irradiation properties[183]. It was found that Cr-rich compositions showed an increase in BCC phase fraction, while Fe and Ni-rich compositions showed higher FCC content, and Cr and Ni tended to segregate together preferentially. This trend can be seen in Figure 14B, where the compositions with a higher Cr content show more severe segregation. Nanoindentation was utilized as a high-throughput means to measure the effect of ion irradiation on the hardness of each composition. Radiation-induced hardening was found in all compositions with FCC, BCC, and FCC + BCC phases. The increase in hardness was consistently shown to be 1-1.5 GPa, with BCC-rich

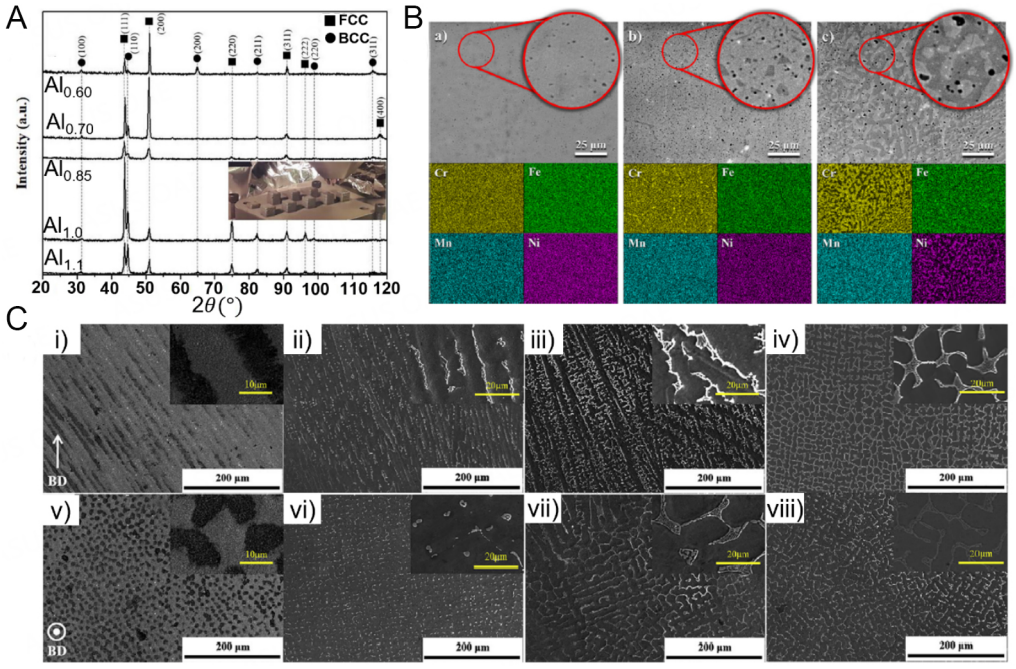


**Figure 14.** (A) X-ray diffraction profiles of $Al_xCoCrFeNi_{2.1}$ showing the change in FCC and BCC phase fractions with changing Al content. The inset shows an image of material library. This figure is quoted with permission from Joseph *et al.*[182], copyright 2020, Elsevier; (B) SEM images and EDS maps of $Cr_{19}Fe_{31}Mn_{11}Ni_{39}$, $Cr_{28}Fe_{23}Mn_{20}Ni_{29}$, and $Cr_{42}Fe_{17}Mn_{20}Ni_{21}$, from left to right. This figure is quoted with permission from Moorehead *et al.*[183], copyright 2021, Elsevier; (C) SEM images of the as-deposited $CoCrFeNiNb_x$ parallel to the build direction: (i) $CoCrFeNiNb_0$; (ii) $CoCrFeNiNb_{0.1}$; (iii) $CoCrFeNiNb_{0.15}$; (iv) $CoCrFeNiNb_{0.2}$; and perpendicular to the build direction: (v) $CoCrFeNiNb_0$; (vi) $CoCrFeNiNb_{0.1}$; (vii) $CoCrFeNiNb_{0.15}$; and (viii) $CoCrFeNiNb_{0.2}$. This figure is quoted with permission from Zhou *et al.*[9], copyright 2019, Elsevier. BCC: Body-centered cubic; FCC: face-centered cubic.

compositions showing the least increase in hardness. Additionally, Moorehead *et al.* found that compositions with a high Mn content may have a large amount of Mn depletion after homogenization due to the depressed melting point of high Mn-content alloys. Thus, the authors laid out a guideline to keep the Mn-content below 25 at. %. Finally, the time saved using the high throughput AM approach is highlighted compared to the traditional metallurgical approach of melting and casting. The authors state that traditional melting and casting could take up to 1-2 hours per composition compared to the 10 min per composition required by the DED method.

Zhou *et al.* utilized DED to rapidly produce samples with compositions $CoCrFeNiNb_x$ (referred to here as $Nb_x$)[9]. The authors investigated the mechanical properties of each alloy composition and its correlation to the phase and microstructures present. Figure 14C shows SEM images of 4 compositions ($Nb_0$, $Nb_{0.1}$, $Nb_{0.15}$, $Nb_{0.2}$) prepared by DED. The top row shows images taken parallel to the building direction, while the bottom row shows images taken perpendicular to the building direction. The authors concluded that the addition of Nb to the CoCrFeNi system led to a transition from a columnar to an equiaxed structure due to the formation of a secondary Laves phase in addition to the primary FCC phase. The Laves phase also caused an increase in yield strength in the $Nb_{0.2}$ composition more than three times that of the Nb-free composition while maintaining a ductility above 10%.

## HIGH-THROUGHPUT CHARACTERIZATION TECHNIQUES

While high-throughput computational methods can narrow down the alloy design space and high-throughput manufacturing methods enable rapid fabrication of samples within the design space, high-

throughput experimental methods are needed to characterize the manufactured materials' properties to experimentally verify which set of composition and processing conditions ultimately leads to the target performance. This section focuses primarily on the different high-throughput methods which can rapidly characterize important material properties such as hardness, strength, ductility, phase and composition, magnetic hysteresis, saturation magnetization, and corrosion resistance.

### Mechanical property characterization

Mechanical properties such as strength and ductility are crucial to assess the performance of a material for structural applications[184]. A few important criteria exist for a mechanical test considered to be useful for screening HEAs. First, the sample size and microstructure must represent bulk-like conditions[185]. Second, the test should include a dominant tensile component, as real application conditions often include some tensile stresses[185]. Microhardness and nanoindentation tests are the most common approach towards high-throughput screening of structural materials. These methods can quickly and accurately estimate bulk yield strength[185]. Nanoindentation can also offer broad insights into the post-yielding attributes through analysis of the stress-strain curves it produces[186]. The local nature of these two techniques also makes them highly useful in graded materials libraries where many compositions and microstructures can be manufactured in a single sample for rapid screening.

Here two example studies are provided that use microhardness testing to investigate the effect of composition on hardness. Jiang *et al.* produced various compositions of $CoFeNi_xVMo_y$ alloys to test the effects of composition and microstructure on the hardness[187]. This work concluded that an increase in the Mo content led to increased precipitation of the $CoMo_2Ni$-type intermetallic phase. An increase in Ni content increased the FCC solid solution phase and decreased the hardness. Figure 15A depicts the hardness of the various compositions showing that the peak hardness was reached at the composition equiatomic CoFeNiVMo. Pegues *et al.* also used micro-hardness indentation to build a hardness map of a graded $Ta_x$ CoCrFeMnNi sample, as shown in Figure 15B[166]. This map allowed them to rapidly determine the effect of Ta addition on the hardness of this Cantor alloy-based system. Higher Ta contents encouraged the formation of TaNi-rich intermetallic in the interdendritic region, which caused significant increases in hardness.

Although micro-indentation methods can provide reasonable data for screening materials, the most reliable method to investigate material properties is a lab-scale tension test with samples that conform to either the ASTM E8 standard or another equivalent internationally recognized standard. Here the authors of this review present one example from the literature and their own unpublished data to illustrate the typical results that can be achieved in a combinatorial HEA library. Ma *et al.* added Nb to the AlCoCrFeNi system, which led to the formation of Laves phase that increased the strength of the alloy while decreasing the ductility[188]. Tuning the Nb content allowed them to tune the compressive properties [Figure 15C]. This result indicates that the addition of intermetallic forming elements can be used to achieve a wide array of properties that can be optimized for application-specific uses. Following this design philosophy, the authors of this review recently used DED to produce $CoCrFeNiTi_x$ alloys to achieve a composition with improved mechanical properties. The tensile stress-strain curves of our investigated $CoCrFeNiTi_x$ HEAs are shown in Figure 15D. These results show that adding Ti to the base quaternary alloy increases the yield strength while decreasing the ductility until $x = 0.2$. Beyond this threshold, the yield strength and the ductility of the alloy drop simultaneously. It has been well established that the introduction of Ti into the CoCrFeNi system leads to the formation of brittle intermetallic phases that causes decreased ductility and increased hardness and yield strength[189-191]. The drop in yield strength from $x = 0.2$ to $x = 0.25$ is likely a result of defects that occurred during printing due to a higher fraction of brittle phases.

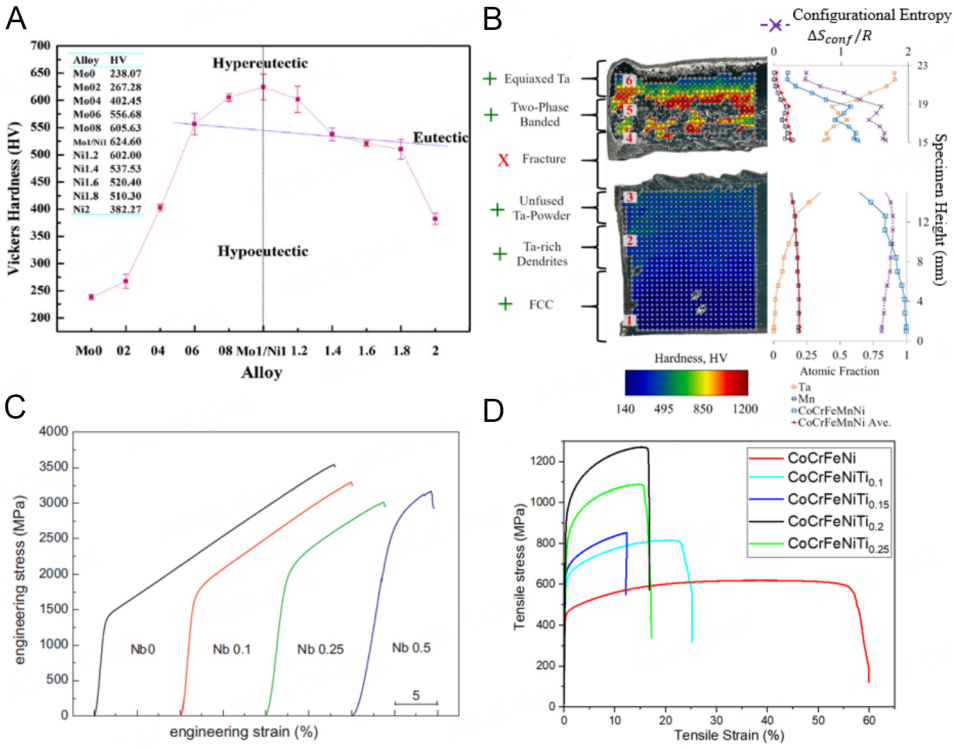


**Figure 15.** (A) Effects of Ni and Mo on Vickers hardness of $CoFeNi_xVmo_y$. This figure is quoted with permission from Jiang *et al.* [187], copyright 2015, Elsevier; (B) $Ta_xCoCrFeMnNi$ hardness map with associated elemental distribution. This figure is quoted with permission from Pegues *et al.*[166], copyright 2021, Elsevier; (C) compressive stress-strain curves of $CoCrFeNiNb_x$ samples. This figure is quoted with permission from Ma *et al.*[188], copyright 2012, Elsevier; (D) tensile stress-strain curves of $CoCrFeNiTi_x$ HEAs printed in the authors' lab. HEA: High-entropy alloy.

In order to fulfill the rapid testing needs of high-throughput experiments, a high degree of automation must be integrated into the characterization process to decouple experimental progress from the number of hours available to human researchers. To that end, Huang *et al.* developed a high-throughput tensile testing platform to automate the tensile testing procedure and increase the rate and which specimens can be characterized[192]. This method uses a large grip that is held onto the bottom of many dog-bone samples, and this grip is attached to a motorized table that moves the grip laterally into position. A top grip is aligned in the direction of travel, allowing for automated testing of many samples in a small-time frame. In the case of their work, Huang *et al.* tested many samples of 316L stainless steel printed through a combinatorial study of different printing conditions by L-PBF. This platform may also show great potential for combinatorial studies related to the compositions of HEAs by automating tensile testing of compositional libraries produced by laser-based AM.

### Phase and composition analysis

XRD is a common tool used to analyze the phases present in a material. The following paragraph illustrates some examples of typical data extracted from XRD analysis in combinatorial studies. Chen *et al.* studied the phase evolution in $(AlCoCrFeNi)_{100-x}Ni_x$ and $(CoCrCuFeNi)_{100-x}Mo_x$ HEAs[193]. XRD analysis shows that when x is between 0 and 4 at. %, both alloys exhibited single-phase solid solution structure where the $(AlCoCrFeNi)_{100-x}Ni_x$ alloy shows a BCC structure and the $(CoCrCuFeNi)_{100-x}Mo_x$ alloy shows an FCC structure. As the Ni content increases, a dual-phase FCC/BCC structure forms, and the FCC phase fraction increases. On the other hand, when the Mo content increases beyond 4 at. %, the FCC/BCC structure also forms, and the BCC phase fraction increases with the Mo content. The XRD patterns of the Ni- and Mo-

doped samples are shown in Figure 16A and B. Moorehead *et al.* used DED to produce a library of Mo-Nb-Ta-W HEAs, which were then examined via XRD[194]. The XRD patterns are shown in Figure 16C. In this case, all the compositions tested showed a full BCC phase.

However, energy dispersive spectroscopy (EDS) analysis showed that Nb was segregated to the interdendritic region. In order to tune the phase structure in a $CoCrFeNiNb_x$ alloy system, our group used DED by placing pre-alloyed CoCrFeNi powder in one powder feeder and pure Nb powder in another. Then, the feed rates from each feeder were adjusted to produce a graded material with increasing Nb along the build direction. The $CoCrFeNiNb_x$ compositions were selected as going from $x = 0$ to $x = 1$, where x was increased by about 0.1 for every 1 mm increase in height. Synchrotron XRD (SXRD) at the Cornell High Energy Synchrotron Source was then performed along the build direction to analyze the phase composition in each region. The beam size was maintained as 0.5 × 0.5 mm such that the measured phase compositions correspond accurately to the designed compositions. As the Nb content increased, an FCC/Laves dual-phase structure formed with the Laves phase volume fraction increased from 0% to 57% as the Nb content increased from 0 to 20 at. %. Figure 16D shows the SXRD results taken for each composition.

In addition to XRD analysis, EBSD can offer a means to probe phases at higher spatial resolution (around 200 nm for EBSD *vs.* about 1 mm for XRD), which may be especially important for gradient compositional libraries where the compositional change can be quite drastic over small length scales[185]. EBSD also has the advantage of being equipped onto SEM facilities. Thus, EDS analysis can often be carried out in parallel such that phase and composition can be resolved almost simultaneously. For experiments that include many phases and samples, there is a need to automate the phase analysis process to make the process more efficient[195,29]. Many groups have used high-throughput SXRD to rapidly identify phases in combinatorial material libraries[196]. The majority of these studies used thin films produced by magnetron sputtering[29,195,197,198]. However, there are almost no studies of large bulk materials that used similar high-throughput methods for phase structure analysis.

It should be noted that a bottleneck for the previously mentioned methods is the human intervention needed during data analysis. This analysis can require impractical time commitments when the number of compositions reaches hundreds or thousands. Thus, the development of automated systems for analyzing XRD, EBSD, and EDS data is crucial to ensure that experimental results of high-throughput experiments can be achieved in a timely manner. Machine learning has recently shown impressive results in this field by correctly indexing phases within an EBSD pattern without requiring human input to guess at the present phases[199]. Although this process has not been attempted for HEAs, it shows great promise to be applied to new materials. Extending this practice further to carry out more in-depth analysis, such as Rietveld refinement for XRD patterns, will greatly accelerate the development of future HEAs.

### Magnetic property measurement

Tang *et al.* added Ho to a $FeCoNi(CuAl)_{0.8}$ alloy to investigate the effect of the addition of rare earth (RE) element on the magnetic properties of this system[200]. The initial parent alloy showed a fully FCC structure, and adding Ho led to the formation of a secondary BCC phase. Increasing the atomic fraction of Ho led to an increase in the volume fraction of the BCC phase fraction and no change in the lattice parameter of the FCC phase. As shown in Figure 17A, increasing the Ho content led to lower energy losses via eddy currents and hysteresis until $x = 0.05$. Once $x = 0.07$, the energy losses increased substantially. The decrease in hysteresis loss is due to lower Cu segregation at the BCC-FCC phase boundaries with increasing Ho. This segregation leads to a lower magnetic domain pining effect, decreasing hysteresis losses[200]. At $x = 0.07$, the Cu and Ho tend to segregate heavily to phase boundaries, leading to higher hysteresis losses. The BCC

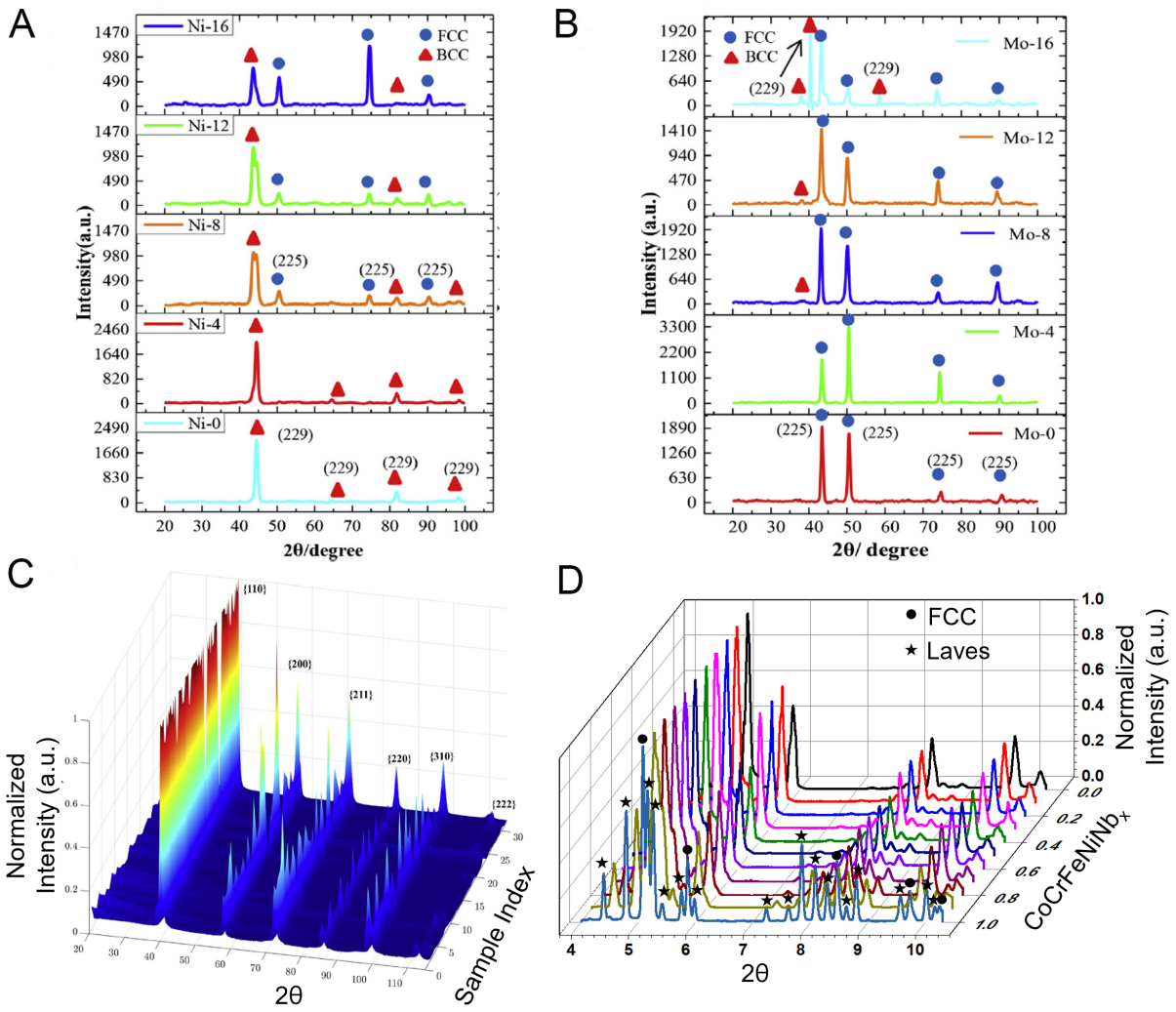


**Figure 16.** (A) XRD patterns of $(AlCoCrFeNi)_{100-x}Ni_x$. This figure is quoted with permission from Chen *et al.*[193], copyright 2018, Elsevier; (B) XRD patterns of $(CoCrCuFeNi)_{100-x}Mo_x$. This figure is quoted with permission from Chen *et al.*[193], copyright 2018, Elsevier; (C) XRD patterns of 31 samples fabricated by L-DED in the Mo-Ta-Nb-W alloy system. This figure is quoted with permission from Moorehead *et al.*[194]; (D) XRD patterns of $CoCrFeNiNb_x$ printed in the authors' lab.

phase shows thinner magnetic stripe domains than the FCC phase, which leads to lower eddy losses as the BCC phase fraction increases.

Zhang *et al.* studied the effect of composition and phase fraction in a FeCoNi(CuAl)x alloy system on the magnetic and mechanical properties[201]. The general trend from Figure 17B shows that the saturation magnetization ($M_s$) decreases with increasing Cu and Al content and shows a slight increase from x = 0.8 to x = 0.9. Fe, Co, and Ni are all ferromagnetic elements, while Cu and Al are not, so the authors rationalize the decrease in $M_s$ originating from the increase in non-ferromagnetic components. Borkar *et al.* also studied the effect of the Co/Cr ratio on the microstructure and magnetic properties of $AlCo_xCr_{1-x}FeNi$[202]. Increasing the Co/Cr ratio leads to increased $M_s$ as the magnetization depends heavily on the composition, as shown in Figure 17C.

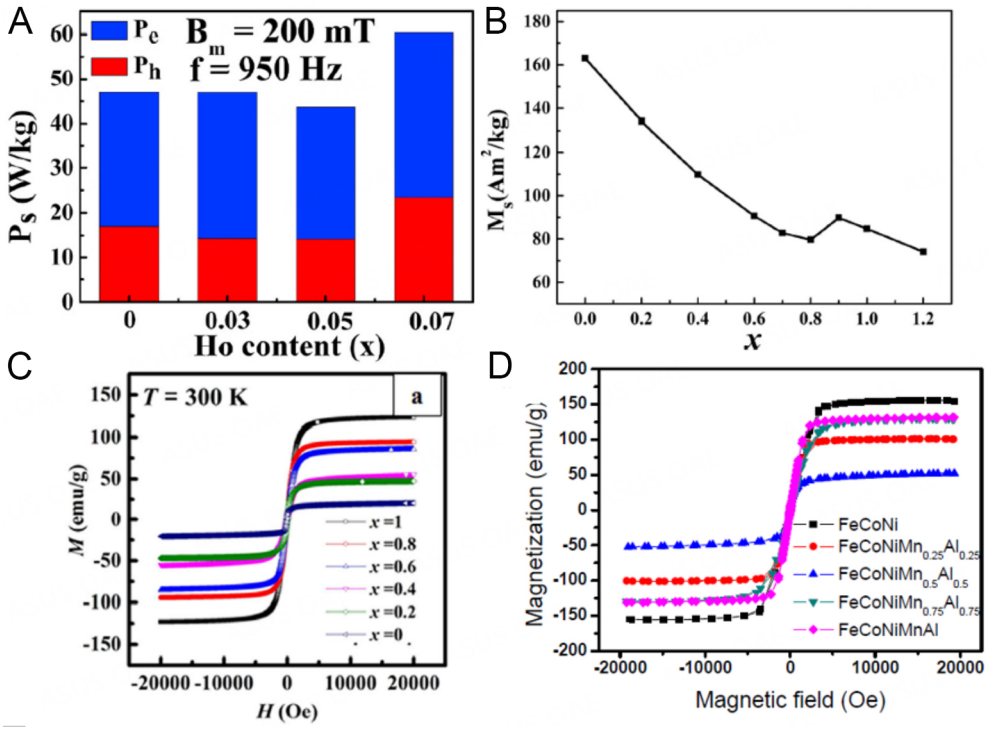


**Figure 17.** (A) Effect of Ho addition to $FeCoNi(CuAl)_{0.8}$ on magnetic hysteresis response of this alloy measured at room temperature. This figure is quoted with permission from Tang *et al.*[200], copyright 2021, Elsevier; (B) saturation magnetization of $FeCoNi(CuAl)_x$ alloys as CuAl is added. This figure is quoted with permission from Zhang *et al.*[201], copyright 2017, Elsevier; (C) magnetization curves of $NiAlFeCo_xCr_{1-x}$. This figure is quoted with permission from Borkar *et al.*[202], copyright 2017, John Wiley and Sons; (D) magnetization curves of $FeCoNi(MnAl)_x$. This figure is quoted with permission from Li *et al.*[203], copyright 2017, Elsevier.

Li *et al.* adjusted the composition of a $FeCoNi(MnAl)_x$ to study the effect of composition on the magnetic properties of this material[203]. The results presented in Figure 17D show that the saturation magnetization decreases as the Cu and Al content increases to x = 0.5 and then increases as the Cu, and Al content increases further. The FeCoNi alloy shows a fully FCC structure, while the $FeCoNi(MnAl)_{0.5}$ and $FeCoNi(MnAl)_{0.75}$ show an FCC + BCC dual-phase structure, and the FeCoNi(MnAl) composition shows a nearly fully BCC structure. Also, the lattice parameter of the FCC phase increases as Mn and Al are added, and the lattice parameter of the BCC phase decreases. The authors explain that the magnetization of the BCC phase decreases with decreasing lattice parameters and the magnetization of the FCC phase decreases with increasing lattice parameters[203]. Thus, the $FeCoNi(MnAl)_{0.5}$ composition shows the lowest magnetic performance because both phases show their lowest performance at that composition.

### Corrosion resistance

Corrosion resistance is a crucial property when selecting materials for real-life applications. The degradation of materials due to corrosion leads to over $500 billion in repair and maintenance costs in the US alone[204]. The corrosion process is highly complicated and includes various mechanisms that depend on the type and concentration of the corrosive electrolyte, the composition and microstructure of the chosen material, the ambient temperature, and the time spent in service[185]. Due to the complexity of corrosion phenomena, there is currently no unifying computational model to predict corrosion resistance, and there are limited empirical models for certain systems. This challenge forces researchers to rely on experimental results to screen materials for corrosion resistance. Thus, high-throughput corrosion resistance methods are extremely important to characterize and screen HEAs for future applications. Typical high-throughput screening methods utilize multi-electrode arrays placed in a common electrolyte to allow multiple materials

to be tested simultaneously[205,206]. Rapid characterization via wire resistance to indicate a reduction in the cross-sectional area has shown great potential to accelerate the analysis of corrosion resistance[207]. Additionally, optical characterization of the color change of a corroded substance can be correlated well with traditional polarization curves. It can thus provide an easy screening method without the need for in-depth analysis[208,209]. The remainder of this section will summarize the results of combinatorial studies on HEAs taken from the literature that illustrate the improvements in corrosion resistance achieved so far.

It is well known that Cr is a useful element to improve corrosion resistance, and this result has also been verified for many HEA systems[210]. Thus, a Cr-containing HEA system with elements that show a low thermal neutron absorption cross-section is a prime candidate for nuclear applications[210]. Xiang *et al.* studied the effect of Cr addition in a $Mo_{0.5}$VNbTi system on the microstructure and properties, including corrosion resistance[210]. The corrosion resistance tests for the $Mo_{0.5}VNbTiCr_x$ HEAs (x = 0, 0.25, 0.5, 0.75, 1.0, 1.5, and 2.0, denoted as $Cr_0$, $Cr_{0.25}$, $Cr_{0.5}$, $Cr_{0.75}$, $Cr_{1.0}$, $Cr_{1.5}$, and $Cr_{2.0}$, respectively) was carried out in superheated steam at 400 °C at 10.3 MPa in a static autoclave which is in accordance with ASTM G2/G2M guidelines. The weight gain per unit surface area was measured after 10, 20, 30, 40, 50, 60, and 70 days. The weight gain for each sample was compared to that of Zr-4 alloy (Zr-1.41Sn-0.21Fe-0.10Cr), which is commonly used as fuel rod cladding in nuclear reactors due to its excellent corrosion resistance[210,211]. The results are depicted in Figure 18A, where the weight gain decreases with adding Cr, except for the $Cr_{1.5}$ alloy, which may be due to defects. All the compositions outperform the Zr-4 alloy showing much lower weight gain. This HEA system shows significantly improved corrosion resistance compared to a state-of-the-art alloy used in current applications, illustrating the potential of HEAs to achieve incredible improvements in corrosion resistance properties.

Q235 steel is a common structural steel used in many applications, but it often requires a coating to be used in corrosive environments[212]. A HEA coating with good corrosion resistance can offer protection to the Q235 steel without degradation to its mechanical properties[212]. Qiu *et al.* explored the effect of composition on corrosion resistance in the $Al_2CrFeCo_xCuNiTi$[212]. The addition of Co to the system leads to a more positive corrosion potential which implies a higher corrosion resistance especially compared to the control Q235 steel, as seen in Figure 18B. The authors mention that severe elemental segregation can lead to the formation of micro-potentials during potentiodynamic polarization tests which can lead to micro-corrosion and accelerate the corrosion process. The authors also note that the equiaxed grain structure observed in the HEAs also leads to improved corrosion resistance. Qiu *et al.* also conducted a similar study on $Al_2CrFeCoCuNi_xTi$[213]. Their results are illustrated in Figure 18C. It was also found that the corrosion resistance increased with increasing Ni content and then decreased. The increase in corrosion resistance is because the Ni element has a high corrosion resistance which contributes to the improved corrosion resistance of the alloy. However, as the Ni content increases further beyond $Ni_{1.0}$, the elemental segregation increases greatly, which leads to the formation of micro-potentials that accelerate the corrosion process.

Ti-Zr-based HEAs have been suggested as potential biomedical implant materials[214]. These implants invariably undergo friction wear over long periods, and the complex chemical environment of the human body also leads to corrosion. Thus, a comprehensive understanding of the corrosion and wear resistance of such HEAs is needed to assess their viability for use as implants. Hua *et al.* studied the corrosion resistance of $Ti_xZrNbTaMo$ HEAs and compared it to Ti-6Al-4V, which has long been favored for biomedical applications[215]. The corrosion resistance on these alloys is depicted through the potentiodynamic polarization curves in Figure 18D. The $Ti_{0.5}ZrNbTaMo$ composition shows the highest corrosion resistance[214]. SEM image analysis of all the tested compositions showed that no pitting occurred after the potentiodynamic polarization tests, which points to the high corrosion resistance of these alloys. Further

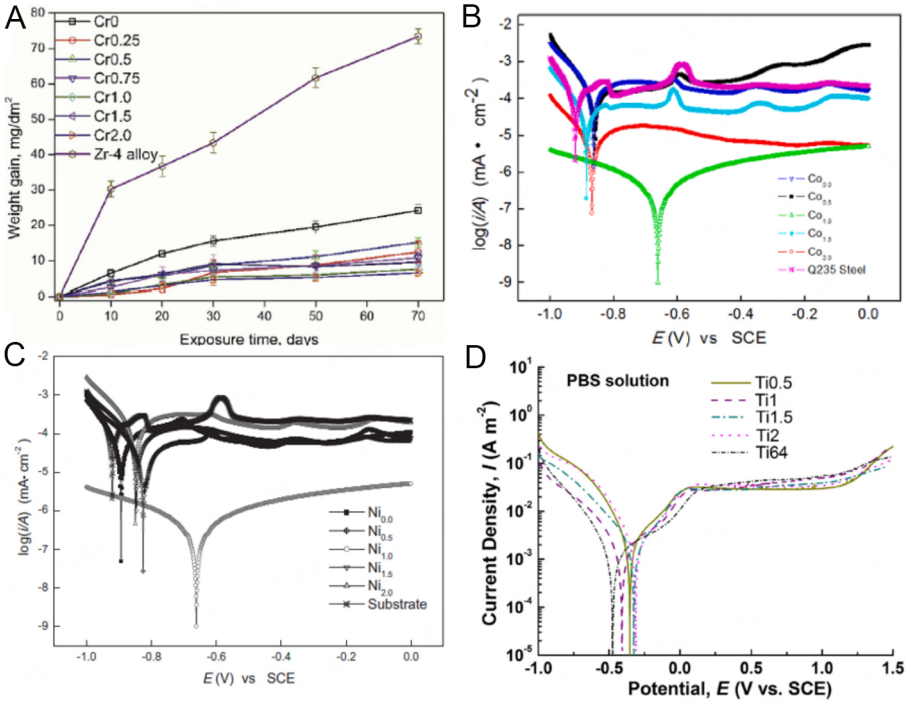


**Figure 18.** (A) Weight gain of $Mo_{0.5}VNbTiCr_x$ after corrosion test in superheated steam at 400 °C at 10.3 MPa pressure for 70 days, Zr-4 alloy is provided for comparison. This figure is quoted with permission from Xiang *et al.*[210], copyright 2020, Elsevier; (B) potentiodynamic polarization curves of $Al_2CrFeCo_xCuNiTi$ HEA compared to Q235 steel. This figure is quoted with permission from Qiu *et al.*[212], copyright 2019, Elsevier; (C) potentiodynamic polarization curves of $Al_2CrFeCoCuTiNi_x$ HEAs and Q235 steel substrate. This figure is quoted with permission from Qiu *et al.*[213], copyright 2013, Elsevier; (D) potentiodynamic polarization curves of $Ti_xZrNbTaMo$ HEAs and Ti6Al4V. This figure is quoted with permission from Hua *et al.*[214], copyright 2021, Elsevier.

surface analysis via XPS showed that the surface film of the HEAs is mainly composed of the $Ti^{4+}$, $Zr^{4+}$, $Nb^{5+}$, $Ta^{5+}$, $Mo^{4+}$, and $Mo^{6+}$ oxides, which indicates the formation of a passivation layer that protected the alloys from severe corrosion.

## CONCLUSIONS AND FUTURE OUTLOOK

HEAs present abundant opportunities to search for new materials with properties and performance that can exceed traditional dilute alloys. While the potential for this new class of materials is promising, the vast composition and microstructure space is too large to explore efficiently via traditional metallurgical techniques based on trial-and-error approaches. This review article highlights important advances in combinatorial studies that either present high-throughput methods to rapidly filter out undesirable materials or provide insights into general rules of thumb to allow researchers to design high-performance materials more efficiently.

The ultimate goal is to ensure that researchers spend more time understanding how to design and manufacture high-performance HEAs for industrial applications and less time on repetitive sample preparation and characterization methods. Implementing efficient high-throughput methods can minimize the time spent studying sub-optimal alloy compositions, which maximizes the resources spent on improving the most promising alloys. First, this review explores the high-throughput computational techniques that can down-select the design space before experimental characterization is even attempted. Then, it presents works that use additive manufacturing as a solution to produce large combinatorial libraries of bulk sample materials at length scales comparable to those expected during service and

applications. Finally, high-throughput material characterization is highlighted for rapid understanding of the relationships between composition, microstructure, and material properties. This review article serves as a guideline for developing workflows that can efficiently discover new high-performance HEAs. To this end, several research frontiers in the field are put forward:

1. Machine learning (ML) techniques can provide predictions of massive design space, but there currently exists a shortage of robust training sets for HEA compositions. Further investment in high-throughput computational techniques that can produce these robust databases, such as CALPHAD, first-principles calculations, and molecular dynamics simulation, is needed. Once these databases are sufficiently established, ML techniques can provide highly reliable predictions of the phase constitution for unknown compositions. They can even predict bulk properties such as yield strength and density.

2. Additive manufacturing provides a means to rapidly produce bulk samples of varying compositions and microstructures. However, AM materials are prone to defects that can significantly deteriorate performance. Further studies, including in-situ characterization during 3D printing, are needed to better characterize the small-scale physics, in-situ alloying chemistry, and macroscale defect formation to reduce the work needed in preliminary optimization.

3. Data collection and analysis of material characterization techniques need to be further automated to enable high-throughput characterization of enormous materials libraries without significant time investments from researchers. Such techniques as phase, composition and microstructure characterization may need to be carried out in parallel to maximize the efficient use of equipment with overlapping functionalities, such as SEM with EBSD capabilities. Additionally, data processing automation is critically needed to rapidly characterize the vast number of compositions that are explored in high-throughput experiments.

## DECLARATIONS

### Acknowledgments

This work is based upon research conducted at the Center for High Energy X-ray Sciences (CHESS), which is supported by the National Science Foundation under award (DMR-1829070). The authors are grateful to Katharine Shanks at CHESS for her support in data acquisition and analysis at the ID3A beamline.

### Authors' contributions

Writing: Mooraj S
Manuscript supervision and editing: Chen W

### Availability of Data and Materials

Not applicable.

### Financial support and sponsorship

Chen W acknowledges the support from National Science Foundation (DMR-2004429) and UMass Amherst Faculty Startup Fund.

### Conflicts of interest

All authors declare that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Copyright



## REFERENCES


1. Oh HS, Kim SJ, Odbadrakh K, et al. Engineering atomic-level complexity in high-entropy and complex concentrated alloys. *Nat Commun* 2019;10:2090. DOI PubMed PMC
2. Cantor B. Multicomponent and high entropy alloys. *Entropy* 2014;16:4749. DOI
3. Cantor B, Chang I, Knight P, Vincent A. Microstructural development in equiatomic multicomponent alloys. *Mater Sci Eng A* 2004;375-377:213-8. DOI
4. Yeh J, Chen S, Lin S, et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv Eng Mater* 2004;6:299-303. DOI
5. George EP, Raabe D, Ritchie RO. High-entropy alloys. *Nat Rev Mater* 2019;4:515-34. DOI
6. Wu Q, Wang Z, Hu X, et al. Uncovering the eutectics design by machine learning in the Al-Co-Cr-Fe-Ni high entropy system. *Acta Mater* 2020;182:278-86. DOI
7. Chanda B, Verma A, Das J. Nano-/ultrafine eutectic in CoCrFeNi(Nb/Ta) high-entropy alloys. *Trans Indian Inst Met* 2018;71:2717-23. DOI
8. Zhang H, Liu P, Hou J, Qiao J, Wu Y. Prediction of strength and ductility in partially recrystallized $CoCrFeNiTi_{0.2}$ high-entropy alloy. *Entropy* 2019;21:389. DOI PubMed PMC
9. Zhou K, Li J, Wang L, Yang H, Wang Z, Wang J. Direct laser deposited bulk $CoCrFeNiNb_x$ high entropy alloys. *Intermetallics* 2019;114:106592. DOI
10. Poletti MG, Fiore G, Gili F, Mangherini D, Battezzati L. Development of a new high entropy alloy for wear resistance: $FeCoCrNiW_{0.3}$ and $FeCoCrNiW_{0.3}$ + 5 at.% of C. *Mater Design* 2017;115:247-54. DOI
11. Xiao J, Tan H, Chen J, Martini A, Zhang C. Effect of carbon content on microstructure, hardness and wear resistance of $CoCrFeMnNiC_x$ high-entropy alloys. *J Alloys Compd* 2020;847:156533. DOI
12. Cui Y, Shen J, Manladan SM, Geng K, Hu S. Wear resistance of $FeCoCrNiMnAl_x$ high-entropy alloy coatings at high temperature. *Appl Surf Sci* 2020;512:145736. DOI
13. Shi Y, Yang B, Liaw P. Corrosion-resistant high-entropy alloys: a review. *Metals* 2017;7:43. DOI
14. Li R, Xie L, Wang WY, Liaw PK, Zhang Y. High-throughput calculations for high-entropy alloys: a brief review. *Front Mater* 2020;7:290. DOI
15. Yin Y, Chen Z, Mo N, et al. High-temperature age-hardening of a novel cost-effective $Fe_{45}Ni_{25}Cr_{25}Mo_5$ high entropy alloy. *Mater Sci Eng A* 2020;788:139580. DOI
16. Ma Y, Wu S, Jia Y, et al. Hexagonal closed-packed precipitation enhancement in a NbTiHfZr refractory high-entropy alloy. *Metals* 2019;9:485. DOI
17. Liu Z, Zhao D, Wang P, et al. Additive manufacturing of metals: microstructure evolution and multistage control. *J Mater Sci Technol* 2022;100:224-36. DOI
18. Fu C, Li J, Bai J, et al. Effect of helium bubbles on irradiation hardening of additive manufacturing 316L stainless steel under high temperature He ions irradiation. *J Nucl Mater* 2021;550:152948. DOI
19. Farshidianfar MH, Khajepour A, Gerlich A. Real-time control of microstructure in laser additive manufacturing. *Int J Adv Manuf Technol* 2016;82:1173-86. DOI
20. Shamsaei N, Yadollahi A, Bian L, Thompson SM. An overview of Direct Laser Deposition for additive manufacturing; Part II: mechanical behavior, process parameter optimization and control. *Addit Manuf* 2015;8:12-35. DOI
21. Derimow N, Clark T, Abbaschian R. Solidification processing and cooling rate effects on hexagonal $Co_{22}Cr_{18}Cu_{20}Mn_{16}Ti_{24}$ high-entropy alloys. *Mater Chem Phys* 2020;240:122188. DOI
22. Xu X, Guo S, Nieh T, Liu C, Hirata A, Chen M. Effects of mixing enthalpy and cooling rate on phase formation of $Al_xCoCrCuFeNi$ high-entropy alloys. *Materialia* 2019;6:100292. DOI
23. Kube SA, Schroers J. Metastability in high entropy alloys. *Scr Mater* 2020;186:392-400. DOI
24. Braeckman B, Boydens F, Hidalgo H, et al. High entropy alloy thin films deposited by magnetron sputtering of powder targets. *Thin Solid Films* 2015;580:71-6. DOI
25. Al Hasan NM, Hou H, Sarkar S, et al. Combinatorial synthesis and high-throughput characterization of microstructure and phase transformation in Ni-Ti-Cu-V quaternary thin-film library. *Engineering* 2020;6:637-43. DOI
26. Liu X, Zou P, Song L, et al. Combinatorial high-throughput methods for designing hydrogen evolution reaction catalysts. *ACS Catal*

2022;12:3789-96. DOI

27. Ludwig A. Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *NPJ Comput Mater* 2019;5:70. DOI
28. Shi Y, Yang B, Rack PD, Guo S, Liaw PK, Zhao Y. High-throughput synthesis and corrosion behavior of sputter-deposited nanocrystalline Al (CoCrFeNi)100- combinatorial high-entropy alloys. *Mater Design* 2020;195:109018. DOI
29. Kube SA, Sohn S, Uhl D, Datye A, Mehta A, Schroers J. Phase selection motifs in high entropy alloys revealed through combinatorial methods: large atomic size difference favors BCC over FCC. *Acta Mater* 2019;166:677-86. DOI
30. Keil T, Utt D, Bruder E, Stukowski A, Albe K, Durst K. Solid solution hardening in CrMnFeCoNi-based high entropy alloy systems studied by a combinatorial approach. *J Mater Res* 2021;36:2558-70. DOI
31. Geuser FD. High-throughput in-situ characterization and modeling of precipitation kinetics in compositionally graded alloys. *Acta Mater* 2015;101:1-9. DOI
32. Zhang X, Xiang Y. Combinatorial approaches for high-throughput characterization of mechanical properties. *J Materiomics* 2017;3:209-20. DOI
33. Wang Z, Zhang L, Li W, et al. A high-throughput approach to explore the multi-component alloy space: a case study of nickel-based superalloys. *J Alloys Compd* 2021;858:158100. DOI
34. Zhu C, Li C, Wu D, et al. A titanium alloys design method based on high-throughput experiments and machine learning. *J Mater Res Technol* 2021;11:2336-53. DOI
35. Liu YH, Fujita T, Aji DP, Matsuura M, Chen MW. Structural origins of Johari-Goldstein relaxation in a metallic glass. *Nat Commun* 2014;5:3238. DOI PubMed
36. Li MX, Zhao SF, Lu Z, et al. High-temperature bulk metallic glasses developed by combinatorial methods. *Nature* 2019;569:99-103. DOI PubMed
37. Frazier WE. Metal additive manufacturing: a review. *J Materi Eng Perform* 2014;23:1917-28. DOI
38. Ngo TD, Kashani A, Imbalzano G, Nguyen KT, Hui D. Additive manufacturing (3D printing): a review of materials, methods, applications and challenges. *Compos B Eng* 2018;143:172-96. DOI
39. Snow Z, Nassar AR, Reutzel EW. Invited review article: review of the formation and impact of flaws in powder bed fusion additive manufacturing. *Addit Manuf* 2020;36:101457. DOI
40. Clare A, Mishra R, Merklein M, et al. Alloy design and adaptation for additive manufacture. *J Mater Process Technol* 2022;299:117358. DOI
41. Bandyopadhyay A, Traxel KD. Invited review article: metal-additive manufacturing - Modeling strategies for application-optimized designs. *Addit Manuf* 2018;22:758-74. DOI PubMed PMC
42. Zhang C, Ouyang D, Pauly S, Liu L. 3D printing of bulk metallic glasses. *Mater Sci Eng R Rep* 2021;145:100625. DOI
43. Silva LJ, Souza DM, de Araújo DB, Reis RP, Scotti A. Concept and validation of an active cooling technique to mitigate heat accumulation in WAAM. *Int J Adv Manuf Technol* 2020;107:2513-23. DOI
44. Dhinakaran V, Ajith J, Fathima Yasin Fahmidha A, Jagadeesha T, Sathish T, Stalin B. Wire arc additive manufacturing (WAAM) process of nickel based superalloys - a review. *Mater Today* 2020;21:920-5. DOI
45. Kozamernik N, Bračun D, Klobčar D. WAAM system with interpass temperature control and forced cooling for near-net-shape printing of small metal components. *Int J Adv Manuf Technol* 2020;110:1955-68. DOI
46. Hou P, Mooraj S, Champagne VK, et al. Effect of build height on temperature evolution and thermally induced residual stresses in plasma arc additively manufactured stainless steel. *Metall Mater Trans A* 2022;53:627-39. DOI
47. Borkar T, Gwalani B, Choudhuri D, et al. A combinatorial assessment of $Al_xCrCuFeNi_2$ ($0 < x < 1.5$) complex concentrated alloys: microstructure, microhardness, and magnetic properties. *Acta Mater* 2016;116:63-76. DOI
48. Miracle D, Senkov O. A critical review of high entropy alloys and related concepts. *Acta Mater* 2017;122:448-511. DOI
49. Li Z, Raabe D. Strong and ductile non-equiatomic high-entropy alloys: design, processing, microstructure, and mechanical properties. *JOM* 2017;69:2099-106. DOI PubMed PMC
50. Choi W, Jung S, Jo YH, Lee S, Lee B. Design of new face-centered cubic high entropy alloys by thermodynamic calculation. *Met Mater Int* 2017;23:839-47. DOI
51. Yeh JW. Recent progress in high-entropy alloys. Available from: https://www.researchgate.net/profile/Jien-Wei-Yeh/publication/245440481_Recent_progress_in_high-entropy_alloys/links/02e7e52456c6fbaec9000000/Recent-progress-in-high-entropy-alloys.pdf [Last accessed on 16 Mar 2023].
52. Li H, Lai J, Li Z, Wang L. Multi-sites electrocatalysis in high-entropy alloys. *Adv Funct Mater* 2021;31:2106715. DOI
53. Marshal A, Pradeep K, Music D, Zaefferer S, De P, Schneider J. Combinatorial synthesis of high entropy alloys: introduction of a novel, single phase, body-centered-cubic FeMnCoCrAl solid solution. *J Alloys Compd* 2017;691:683-9. DOI
54. Yao H, Qiao J, Hawk J, Zhou H, Chen M, Gao M. Mechanical properties of refractory high-entropy alloys: experiments and modeling. *J Alloys Compd* 2017;696:1139-50. DOI
55. Zhang Y, Zhou Y, Lin J, Chen G, Liaw P. Solid-solution phase formation rules for multi-component alloys. *Adv Eng Mater* 2008;10:534-8. DOI
56. Bao N, Zuo J, Du Z, Yang M, Jiang G, Zhang L. Computational characterization of the structural and mechanical properties of $Al_x$ $CoCrFeNiTi_{1-x}$ high entropy alloys. *Mater Res Express* 2019;6:096519. DOI
57. Dong Y, Chen QS, Lu YP, Zhang PC, Li TJ. Effect of aging temperature on microstructure and hardness of $CoCrFeNiTi_{0.5}$ high

entropy alloy. *Mater Sci Forum* 2014;789:48-53. DOI
58. Guo S, Ng C, Liu C. Anomalous solidification microstructures in Co-free $Al_xCrCuFeNi_2$ high-entropy alloys. *J Alloys Compd* 2013;557:77-81. DOI
59. Chen J, Zhou X, Wang W, et al. A review on fundamental of high entropy alloys with promising high-temperature properties. *J Alloys Compd* 2018;760:15-30. DOI
60. Stepanov N, Shaysultanov D, Ozerov M, Zherebtsov S, Salishchev G. Second phase formation in the CoCrFeNiMn high entropy alloy after recrystallization annealing. *Mater Lett* 2016;185:1-4. DOI
61. Toda-caraballo I, Rivera-díaz-del-castillo PE. Modelling solid solution hardening in high entropy alloys. *Acta Mater* 2015;85:14-23. DOI
62. He Q, Yang Y. On lattice distortion in high entropy alloys. *Front Mater* 2018;5:42. DOI
63. Lee C, Chou Y, Kim G, et al. Lattice-distortion-enhanced yield strength in a refractory high-entropy alloy. *Adv Mater* 2020;32:e2004029. DOI PubMed
64. Lee C, Song G, Gao MC, et al. Lattice distortion in a strong and ductile refractory high-entropy alloy. *Acta Mater* 2018;160:158-72. DOI
65. Dirras G, Lilensten L, Djemia P, et al. Elastic and plastic properties of as-cast equimolar TiHfZrTaNb high-entropy alloy. *Mater Sci Eng A* 2016;654:30-8. DOI
66. Owen L, Pickering E, Playford H, Stone H, Tucker M, Jones N. An assessment of the lattice strain in the CrMnFeCoNi high-entropy alloy. *Acta Mater* 2017;122:11-8. DOI
67. Senkov O, Scott J, Senkova S, Miracle D, Woodward C. Microstructure and room temperature properties of a high-entropy TaNbHfZrTi alloy. *J Alloys Compd* 2011;509:6043-8. DOI
68. Guo S. Phase selection rules for cast high entropy alloys: an overview. *Mater Sci Technol* 2015;31:1223-30. DOI
69. Zhang Y, Lu ZP, Ma SG, et al. Guidelines in predicting phase formation of high-entropy alloys. *MRS Commun* 2014;4:57-62. DOI
70. Kottke J, Laurent-brocq M, Fareed A, et al. Tracer diffusion in the Ni-CoCrFeMn system: transition from a dilute solid solution to a high entropy alloy. *Scr Mater* 2019;159:94-8. DOI
71. Mehta A, Sohn Y. Investigation of sluggish diffusion in FCC $Al_{0.25}$CoCrFeNi high-entropy alloy. *Mate Res Lett* 2021;9:239-46. DOI
72. Dąbrowa J, Danielewski M. State-of-the-art diffusion studies in the high entropy alloys. *Metals* 2020;10:347. DOI
73. Sathiaraj G, Ahmed M, Bhattacharjee P. Microstructure and texture of heavily cold-rolled and annealed fcc equiatomic medium to high entropy alloys. *J Alloys Compd* 2016;664:109-19. DOI
74. Bhattacharjee P, Sathiaraj G, Zaid M, et al. Microstructure and texture evolution during annealing of equiatomic CoCrFeMnNi high-entropy alloy. *J Alloys Compd* 2014;587:544-52. DOI
75. Sathiaraj G, Bhattacharjee P. Effect of starting grain size on the evolution of microstructure and texture during thermo-mechanical processing of CoCrFeMnNi high entropy alloy. *J Alloys Compd* 2015;647:82-96. DOI
76. Ranganathan S. Alloyed pleasures: multimetallic cocktails. Available from: http://eprints.iisc.ac.in/6189/1/Alloyed_pleasures.pdf [Last accessed on 16 Mar 2023].
77. Qiao L, Liu Y, Zhu J. A focused review on machine learning aided high-throughput methods in high entropy alloy. *J Alloys Compd* 2021;877:160295. DOI
78. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature* 2018;559:547-55. DOI PubMed
79. Yang C, Ren C, Jia Y, Wang G, Li M, Lu W. A machine learning-based alloy design system to facilitate the rational design of high entropy alloys with enhanced hardness. *Acta Mater* 2022;222:117431. DOI
80. Krishna YV, Jaiswal UK, Rahul RM. Machine learning approach to predict new multiphase high entropy alloys. *Scr Mater* 2021;197:113804. DOI
81. Zhou T, Song Z, Sundmacher K. Big data creates new opportunities for materials research: a review on methods and applications of machine learning for materials design. *Engineering* 2019;5:1017-26. DOI
82. Liu X, Xu P, Zhao J, Lu W, Li M, Wang G. Material machine learning for alloys: Applications, challenges and perspectives. *J Alloys Compd* 2022;921:165984. DOI
83. Liu S, Kappes BB, Amin-ahmadi B, Benafan O, Zhang X, Stebner AP. Physics-informed machine learning for composition - process - property design: shape memory alloy demonstration. *Appl Mater Today* 2021;22:100898. DOI
84. Yi W, Liu G, Lu Z, Gao J, Zhang L. Efficient alloy design of Sr-modified A356 alloys driven by computational thermodynamics and machine learning. *J Mater Sci Technol* 2022;112:277-90. DOI
85. White AD. Deep learning for molecules and materials. *LiveCoMS* 2022:3. DOI
86. Nassar A, Mullis A. Rapid screening of high-entropy alloys using neural networks and constituent elements. *Comput Mater Sci* 2021;199:110755. DOI
87. Risal S, Zhu W, Guillen P, Sun L. Improving phase prediction accuracy for high entropy alloys with machine learning. *Comput Mater Sci* 2021;192:110389. DOI
88. Montavon G, Samek W, Müller K. Methods for interpreting and understanding deep neural networks. *Digit Signal Process* 2018;73:1-15. DOI
89. Zhang Y, Wen C, Wang C, et al. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater* 2020;185:528-39. DOI

90. Vazquez G, Singh P, Sauceda D, et al. Efficient machine-learning model for fast assessment of elastic properties of high-entropy alloys. *Acta Mater* 2022;232:117924. DOI
91. Purcell TAR, Scheffler M, Carbogno C, Ghiringhelli LM. SISSO++: A C++ implementation of the sure-independence screening and sparsifying operator approach. *J Open Res Softw* 2022;7:3960. DOI
92. Sorkin V, Yu ZG, Chen S, Tan TL, Aitken ZH, Zhang YW. A first-principles-based high fidelity, high throughput approach for the design of high entropy alloys. *Sci Rep* 2022;12:11894. DOI PubMed PMC
93. Hautier G, Jain A, Ong SP. From the computer to the laboratory: materials discovery and design using first-principles calculations. *J Mater Sci* 2012;47:7317-40. DOI
94. Ikeda Y, Grabowski B, Körmann F. Ab initio phase stabilities and mechanical properties of multicomponent alloys: a comprehensive review for high entropy alloys and compositionally complex alloys. *Mater Charact* 2019;147:464-511. DOI
95. Kohn W, Sham LJ. Self-consistent equations including exchange and correlation effects. *Phys Rev* 1965;140:A1133-8. DOI
96. Ceperley DM, Alder BJ. Ground state of the electron gas by a stochastic method. *Phys Rev Lett* 1980;45:566-9. DOI
97. Perdew JP. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys Rev B Condens Matter* 1986;33:8822-4. DOI
98. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. *Phys Rev Lett* 1996;77:3865-8. DOI PubMed
99. Perdew JP, Chevary JA, Vosko SH, et al. Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. *Phys Rev B Condens Matter* 1992;46:6671-87. DOI
100. Kim G, Diao H, Lee C, et al. First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation. *Acta Mater* 2019;181:124-38. DOI
101. Rittiruam M, Noppakhun J, Setasuban S, et al. High-throughput materials screening algorithm based on first-principles density functional theory and artificial neural network for high-entropy alloys. *Sci Rep* 2022;12:16653. DOI PubMed PMC
102. Bellaiche L, Vanderbilt D. Virtual crystal approximation revisited: application to dielectric and piezoelectric properties of perovskites. *Phys Rev B* 2000;61:7877-82. DOI
103. Ramer N, Rappe A. Application of a new virtual crystal approach for the study of disordered perovskites. *J Phys Chem Solids* 2000;61:315-20. DOI
104. Chen L, Hao X, Wang Y, Zhang X, Liu H. First-principles calculation of the effect of Ti content on the structure and properties of TiVNbMo refractory high-entropy alloy. *Mater Res Express* 2020;7:106516. DOI
105. Lederer Y, Toher C, Vecchio KS, Curtarolo S. The search for high entropy alloys: a high-throughput ab-initio approach. *Acta Mater* 2018;159:364-83. DOI
106. Curtarolo S, Setyawan W, Hart GL, et al. AFLOW: an automatic framework for high-throughput materials discovery. *Comput Mater Sci* 2012;58:218-26. DOI
107. Sanchez J, Ducastelle F, Gratias D. Generalized cluster description of multicomponent systems. *Physica A* 1984;128:334-50. DOI
108. de Walle A, Asta M, Ceder G. The alloy theoretic automated toolkit: a user guide. *Calphad* 2002;26:539-53. DOI
109. Berding MA, Sher A. Electronic quasichemical formalism: application to arsenic deactivation in silicon. *Phys Rev B* 1998;58:3853-64. DOI
110. Jiang L, Lu Y, Jiang H, et al. Formation rules of single phase solid solution in high entropy alloys. *Mater Sci Technol* 2015. DOI
111. Guo S, Ng C, Lu J, Liu CT. Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys. *J Appl Phys* 2011;109:103505. DOI
112. Yang S, Liu G, Zhong Y. Revisit the VEC criterion in high entropy alloys (HEAs) with high-throughput ab initio calculations: a case study with Al-Co-Cr-Fe-Ni system. *J Alloys Compd* 2022;916:165477. DOI
113. Zhou K & Liu B. Molecular dynamics simulation: fundamentals and applications. Academic Press; 2022. DOI
114. Car R, de Angelis F, Giannozzi P, Marzari N. First-principles molecular dynamics. In: Yip S, editor. Handbook of Materials Modeling. Dordrecht: Springer; 2005. pp. 59-76. DOI
115. Tang Y, Li D. Nano-tribological behavior of high-entropy alloys CrMnFeCoNi and CrFeCoNi under different conditions: a molecular dynamics study. *Wear* 2021;476:203583. DOI
116. Yin S, Zuo Y, Abu-Odeh A, et al. Atomistic simulations of dislocation mobility in refractory high-entropy alloys and the effect of chemical short-range order. *Nat Commun* 2021;12:4873. DOI PubMed PMC
117. Fan Y, Wang W, Hao Z, Zhan C. Work hardening mechanism based on molecular dynamics simulation in cutting Ni-Fe-Cr series of Ni-based alloy. *J Alloys Compd* 2020;819:153331. DOI
118. Li J, Fang Q, Liu B, Liu Y, Liu Y. Mechanical behaviors of AlCrFeCuNi high-entropy alloys under uniaxial tension via molecular dynamics simulation. *RSC Adv* 2016;6:76409-19. DOI
119. Trong DN, Long VC, Ţălu Ş. Effects of number of atoms and doping concentration on the structure, phase transition, and crystallization process of $Fe_{1-x-y}Ni_xCo_y$ alloy: a molecular dynamic study. *Appl Sci* 2022;12:8473. DOI
120. Xie L, Brault P, Thomann A, Yang X, Zhang Y, Shang G. Molecular dynamics simulation of Al-Co-Cr-Cu-Fe-Ni high entropy alloy thin film growth. *Intermetallics* 2016;68:78-86. DOI
121. Pan Z, Fu Y, Wei Y, Yan X, Luo H, Li X. Deformation mechanisms of TRIP-TWIP medium-entropy alloys via molecular dynamics simulations. *Int J Mech Sci* 2022;219:107098. DOI
122. Jarlöv A, Ji W, Zhu Z, et al. Molecular dynamics study on the strengthening mechanisms of Cr-Fe-Co-Ni high-entropy alloys based on the generalized stacking fault energy. *J Alloys Compd* 2022;905:164137. DOI

123. Li J, Xie B, Fang Q, Liu B, Liu Y, Liaw PK. High-throughput simulation combined machine learning search for optimum elemental composition in medium entropy alloy. *J Mater Sci Technol* 2021;68:70-5. DOI
124. Zhang L, Qian K, Huang J, Liu M, Shibuta Y. Molecular dynamics simulation and machine learning of mechanical response in non-equiatomic FeCrNiCoMn high-entropy alloy. *J Mater Res Technol* 2021;13:2043-54. DOI
125. Morrissey LS, Nakhla S. Considerations when calculating the mechanical properties of single crystals and bulk polycrystals from molecular dynamics simulations. *Mol Simul* 2020;46:1433-42. DOI
126. Zhang L, Qian K, Schuller BW, Shibuta Y. Prediction on mechanical properties of non-equiatomic high-entropy alloy by atomistic simulation and machine learning. *Metals* 2021;11:922. DOI
127. Jiang J, Sun W, Luo N. Molecular dynamics study of microscopic deformation mechanism and tensile properties in $Al_x$CoCrFeNi amorphous high-entropy alloys. *Mater Today Commun* 2022;31:103861. DOI
128. Guruvidyathri K, Hari Kumar KC, Yeh JW, Murty BS. Topologically close-packed phase formation in high entropy alloys: a review of calphad and experimental results. *JOM* 2017;69:2113-24. DOI
129. Gao MC. Design of high-entropy alloys. In: Gao MC, Yeh J, Liaw PK, Zhang Y, editors. High-entropy alloys. Cham: Springer International Publishing; 2016. pp. 369-98. DOI
130. Senkov ON, Miller JD, Miracle DB, Woodward C. Accelerated exploration of multi-principal element alloys with solid solution phases. *Nat Commun* 2015;6:6529. DOI PubMed PMC
131. Klaver TPC, Simonovic D, Sluiter MHF. Brute force composition scanning with a CALPHAD database to find low temperature body centered cubic high entropy alloys. *Entropy* 2018;20:911. DOI PubMed PMC
132. Thurston KV, Gludovatz B, Hohenwarter A, Laplanche G, George EP, Ritchie RO. Effect of temperature on the fatigue-crack growth behavior of the high-entropy alloy CrMnFeCoNi. *Intermetallics* 2017;88:65-72. DOI
133. Li YJ, Savan A, Ludwig A. Atomic scale understanding of phase stability and decomposition of a nanocrystalline CrMnFeCoNi Cantor alloy. *Appl Phys Lett* 2021;119:201910. DOI
134. Zeng Z, Xiang M, Zhang D, et al. Mechanical properties of Cantor alloys driven by additional elements: a review. *J Mater Res Technol* 2021;15:1920-34. DOI
135. Conway PL, Klaver T, Steggo J, Ghassemali E. High entropy alloys towards industrial applications: high-throughput screening and experimental investigation. *Mater Sci Eng A* 2022;830:142297. DOI
136. Abu-odeh A, Galvan E, Kirk T, et al. Efficient exploration of the high entropy alloy composition-phase space. *Acta Mater* 2018;152:41-57. DOI
137. Zhao DQ, Pan SP, Zhang Y, Liaw PK, Qiao JW. Structure prediction in high-entropy alloys with machine learning. *Appl Phys Lett* 2021;118:231904. DOI
138. Schleder GR, Padilha ACM, Acosta CM, Costa M, Fazzio A. From DFT to machine learning: recent approaches to materials science-a review. *J Phys Mater* 2019;2:032001. DOI
139. Zhou Z, Zhou Y, He Q, Ding Z, Li F, Yang Y. Machine learning guided appraisal and exploration of phase design for high entropy alloys. *NPJ Comput Mater* 2019:5. DOI
140. Davydov AV, Kattner UR. Predicting synthesizability. *J Phys D Appl Phys* 2019;52:013001. DOI PubMed PMC
141. Jiang J, Chen P, Qiu J, et al. Microstructural evolution and mechanical properties of $Al_x$CoCrFeNi high-entropy alloys under uniaxial tension: a molecular dynamics simulations study. *Mater Today Commun* 2021;28:102525. DOI
142. Wang L, Liu W, Zhu B, et al. Influences of strain rate, Al concentration and grain heterogeneity on mechanical behavior of $CoNiFeAl_xCu_{1-x}$ high-entropy alloys: a molecular dynamics simulation. *J Mater Res Technol* 2021;14:2071-84. DOI
143. Leong Z, Tan TL. Robust cluster expansion of multicomponent systems using structured sparsity. *Phys Rev B* 2019:100. DOI
144. Leong Z, Ramamurty U, Tan TL. Microstructural and compositional design principles for Mo-V-Nb-Ti-Zr multi-principal element alloys: a high-throughput first-principles study. *Acta Mater* 2021;213:116958. DOI
145. Fernández-caballero A, Wróbel JS, Mummery PM, Nguyen-manh D. Short-range order in high entropy alloys: theoretical formulation and application to Mo-Nb-Ta-V-W system. *J Phase Equilib Diffus* 2017;38:391-403. DOI
146. Fontaine D. The number of independent pair-correlation functions in multicomponent systems. *J Appl Crystallogr* 1971;4:15-9. DOI
147. Kattner UR. The calphad method and its role in material and process development. *Tecnol Metal Mater Min* 2016;13:3-15. DOI PubMed PMC
148. Zeng Y, Man M, Bai K, Zhang Y. Revealing high-fidelity phase selection rules for high entropy alloys: a combined CALPHAD and machine learning study. *Mater Design* 2021;202:109532. DOI
149. Wu D, Tian Y, Zhang L, et al. Optimal design of high-strength Ti-Al-V-Zr alloys through a combinatorial approach. *Materials* 2018;11:1603. DOI PubMed PMC
150. Gumbmann E, De Geuser F, Deschamps A, Lefebvre W, Robaut F, Sigli C. A combinatorial approach for studying the effect of Mg concentration on precipitation in an Al-Cu-Li alloy. *Scr Mater* 2016;110:44-7. DOI
151. Li Y, Jensen KE, Liu Y, et al. Combinatorial strategies for synthesis and characterization of alloy microstructures over large compositional ranges. *ACS Comb Sci* 2016;18:630-7. DOI PubMed
152. Tang M, Pistorius PC, Narra S, Beuth JL. Rapid solidification: selective laser melting of $AlSi_{10}Mg$. *JOM* 2016;68:960-6. DOI
153. Aboutaleb AM, Mahtabi MJ, Tschopp MA, Bian L. Multi-objective accelerated process optimization of mechanical properties in laser-based additive manufacturing: case study on Selective Laser Melting (SLM) Ti-6Al-4V. *J Manuf Process* 2019;38:432-44. DOI
154. Jung HY, Peter NJ, Gärtner E, Dehm G, Uhlenwinkel V, Jägle EA. Bulk nanostructured AlCoCrFeMnNi chemically complex alloy

synthesized by laser-powder bed fusion. *Addit Manuf* 2020;35:101337. DOI

155. Lu Y, Su S, Zhang S, et al. Controllable additive manufacturing of gradient bulk metallic glass composite with high strength and tensile ductility. *Acta Mater* 2021;206:116632. DOI
156. Ren J, Zhang Y, Zhao D, et al. Strong yet ductile nanolamellar high-entropy alloys by additive manufacturing. *Nature* 2022;608:62-8. DOI PubMed
157. Zhang S, Hou P, Mooraj S, Chen W. Printability of $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10.0}Be_{22.5}$ metallic glass on steel by laser additive manufacturing: a single-track study. *Surf Coat Technol* 2021;428:127882. DOI
158. Wang YM, Voisin T, McKeown JT, et al. Additively manufactured hierarchical stainless steels with high strength and ductility. *Nat Mater* 2018;17:63-71. DOI PubMed
159. Chen W, Voisin T, Zhang Y, et al. Microscale residual stresses in additively manufactured stainless steel. *Nat Commun* 2019;10:4338. DOI
160. Li Z, Ludwig A, Savan A, Springer H, Raabe D. Combinatorial metallurgical synthesis and processing of high-entropy alloys. *J Mater Res* 2018;33:3156-69. DOI
161. Santa-aho S, Kiviluoma M, Jokiaho T, et al. Additive manufactured 316L stainless-steel samples: microstructure, residual stress and corrosion characteristics after post-processing. *Metals* 2021;11:182. DOI
162. Zhang C, Chen F, Huang Z, et al. Additive manufacturing of functionally graded materials: a review. *Mater Sci Eng A* 2019;764:138209. DOI
163. del Val J, Arias-gonzález F, Barro O, et al. Functionally graded 3D structures produced by laser cladding. *Procedia Manuf* 2017;13:169-76. DOI
164. Gwalani B, Gangireddy S, Shukla S, et al. Compositionally graded high entropy alloy with a strong front and ductile back. *Mater Today Commun* 2019;20:100602. DOI
165. Li L, Wang J, Lin P, Liu H. Microstructure and mechanical properties of functionally graded $TiC_p$/Ti6Al4V composite fabricated by laser melting deposition. *Ceram Int* 2017;43:16638-51. DOI
166. Pegues JW, Melia MA, Puckett R, Whetten SR, Argibay N, Kustas AB. Exploring additive manufacturing as a high-throughput screening tool for multiphase high entropy alloys. *Addit Manuf* 2021;37:101598. DOI
167. Wen Y, Zhang B, Narayan RL, et al. Laser powder bed fusion of compositionally graded CoCrMo-Inconel 718. *Addit Manuf* 2021;40:101926. DOI
168. Shishkovsky I, Kakovkina N, Scherbakov V. Rapid TMC laser prototyping: Compositional and phase-structural sustainability via combinatorial design of titanium-based gradient alloy reinforced by nano-sized TiC or $TiB_2$ ceramics. *SPIE* 2018;10523:172-7. DOI
169. Traxel KD, Bandyopadhyay A. Influence of in situ ceramic reinforcement towards tailoring titanium matrix composites using laser-based additive manufacturing. *Addit Manuf* 2020;31:101004. DOI PubMed PMC
170. Gong X, Yabansu YC, Collins PC, Kalidindi SR. Evaluation of Ti-Mn alloys for additive manufacturing using high-throughput experimental assays and gaussian process regression. *Materials* 2020;13:4641. DOI PubMed PMC
171. Li M, Flores KM. Laser processing as a high-throughput method to investigate microstructure-processing-property relationships in multiprincipal element alloys. *J Alloys Compd* 2020;825:154025. DOI
172. Teh WH, Chaudhary V, Chen S, et al. High throughput multi-property evaluation of additively manufactured Co-Fe-Ni materials libraries. *Addit Manuf* 2022;58:102983. DOI
173. Gwalani B, Soni V, Waseem OA, Mantri SA, Banerjee R. Laser additive manufacturing of compositionally graded AlCrFeMoVx (x = 0 to 1) high-entropy alloy system. *Opt Laser Technol* 2019;113:330-7. DOI
174. Zhao Y, Lau KB, Teh WH, et al. Compositionally graded CoCrFeNiTi high-entropy alloys manufactured by laser powder bed fusion: a combinatorial assessment. *J Alloys Compd* 2021;883:160825. DOI
175. Yu Z, Zheng W, Li Z, et al. Accelerated exploration of TRIP metallic glass composite by laser additive manufacturing. *J Mater Sci Technol* 2021;78:68-73. DOI
176. Wu Y, Wang H, Liu X, et al. Designing bulk metallic glass composites with enhanced formability and plasticity. *J Mater Sci Technol* 2014;30:566-75. DOI
177. Zhai L, Lu Y, Zhao X, Wang L, Lu X. High-throughput screening of laser additive manufactured metallic glass via ultrasonic wave. *Sci Rep* 2019;9:17660. DOI PubMed PMC
178. Tsai P, Flores KM. High-throughput discovery and characterization of multicomponent bulk metallic glass alloys. *Acta Mater* 2016;120:426-34. DOI
179. Tsai P, Flores KM. A combinatorial strategy for metallic glass design via laser deposition. *Intermetallics* 2014;55:162-6. DOI
180. Islam Z, Nelaturu P, Thoma DJ. A dimensionless number for high-throughput design of multi-principal element alloys in directed energy deposition. *Appl Phys Lett* 2021;119:231901. DOI
181. Zhang W, Liu L, Peng S, et al. The tensile property and notch sensitivity of $AlCoCrFeNi_{2.1}$ high entropy alloy with a novel "steel-frame" eutectic microstructure. *J Alloys Compd* 2021;863:158747. DOI
182. Joseph J, Imran M, Hodgson P, Barnett M, Fabijanic D. Towards the large-scale production and strength prediction of near-eutectic $Al_xCoCrFeNi_{2.1}$ alloys by additive manufacturing. *Manuf Lett* 2020;25:16-20. DOI
183. Moorehead M, Nelaturu P, Elbakhshwan M, et al. High-throughput ion irradiation of additively manufactured compositionally complex alloys. *J Nucl Mater* 2021;547:152782. DOI
184. Miracle D, Majumdar B, Wertz K, Gorsse S. New strategies and tests to accelerate discovery and development of multi-principal

element structural alloys. *Scr Mater* 2017;127:195-200. DOI

185. Miracle DB, Li M, Zhang Z, Mishra R, Flores KM. Emerging capabilities for the high-throughput characterization of structural materials. *Annu Rev Mater Res* 2021;51:131-64. DOI
186. Pathak S, Kalidindi SR. Spherical nanoindentation stress-strain curves. *Mater Sci Eng R Rep* 2015;91:1-36. DOI
187. Jiang L, Cao Z, Jie J, et al. Effect of Mo and Ni elements on microstructure evolution and mechanical properties of the $CoFeNi_xVMo_y$ high entropy alloys. *J Alloys Compd* 2015;649:585-90. DOI
188. Ma S, Zhang Y. Effect of Nb addition on the microstructure and properties of AlCoCrFeNi high-entropy alloy. *Mater Sci Eng A* 2012;532:480-6. DOI
189. Wang X, Liu Q, Huang Y, Xie L, Xu Q, Zhao T. Effect of Ti content on the microstructure and corrosion resistance of $CoCrFeNiTi_x$ high entropy alloys prepared by laser cladding. *Materials* 2020;13:2209. DOI PubMed PMC
190. Shun T, Chang L, Shiu M. Microstructures and mechanical properties of multiprincipal component $CoCrFeNiTi_x$ alloys. *Mater Sci Eng A* 2012;556:170-4. DOI
191. Tong Y, Chen D, Han B, et al. Outstanding tensile properties of a precipitation-strengthened $FeCoNiCrTi_{0.2}$ high-entropy alloy at room and cryogenic temperatures. *Acta Mater* 2019;165:228-40. DOI
192. Huang K, Kain C, Diaz-vallejo N, Sohn Y, Zhou L. High throughput mechanical testing platform and application in metal additive manufacturing and process optimization. *J Manuf Process* 2021;66:494-505. DOI
193. Chen R, Qin G, Zheng H, et al. Composition design of high entropy alloys using the valence electron concentration to balance strength and ductility. *Acta Mater* 2018;144:129-37. DOI
194. Moorehead M, Bertsch K, Niezgoda M, et al. High-throughput synthesis of Mo-Nb-Ta-W high-entropy alloys via additive manufacturing. *Mater Design* 2020;187:108358. DOI
195. Ren F, Pandolfi R, Van Campen D, Hexemer A, Mehta A. On-the-fly data assessment for high-throughput X-ray diffraction measurements. *ACS Comb Sci* 2017;19:377-85. DOI PubMed
196. Long CJ, Bunker D, Li X, Karen VL, Takeuchi I. Rapid identification of structural phases in combinatorial thin-film libraries using X-ray diffraction and non-negative matrix factorization. *Rev Sci Instrum* 2009;80:103902. DOI PubMed
197. Datye A, Alexander Kube S, Verma D, Schroers J, Schwarz UD. Accelerated discovery and mechanical property characterization of bioresorbable amorphous alloys in the Mg-Zn-Ca and the Fe-Mg-Zn systems using high-throughput methods. *J Mater Chem B* 2019;7:5392-400. DOI PubMed
198. Zhao L, Jiang L, Yang L, et al. High throughput synthesis enabled exploration of CoCrFeNi-based high entropy alloys. *J Mater Sci Technol* 2022;110:269-82. DOI
199. Kaufmann K, Zhu C, Rosengarten AS, Maryanovsky D, Wang H, Vecchio KS. Phase mapping in EBSD using convolutional neural networks. *Microsc Microanal* 2020;26:458-68. DOI PubMed
200. Tang Y, Sun S, Lv M, et al. Effect of Ho addition on AC soft magnetic property, microstructure and magnetic domain of $FeCoNi(CuAl)_{0.8}Ho_x$ (x = 0-0.07) high-entropy alloys. *Intermetallics* 2021;135:107216. DOI
201. Zhang Q, Xu H, Tan X, et al. The effects of phase constitution on magnetic and mechanical properties of FeCoNi(CuAl) (x = 0-1.2) high-entropy alloys. *J Alloys Compd* 2017;693:1061-7. DOI
202. Borkar T, Chaudhary V, Gwalani B, et al. A combinatorial approach for assessing the magnetic properties of high entropy alloys: role of Cr in $AlCo_xCr_{1-x}FeNi$. *Adv Eng Mater* 2017;19:1700048. DOI
203. Li P, Wang A, Liu C. Composition dependence of structure, physical and mechanical properties of $FeCoNi(MnAl)_x$ high entropy alloys. *Intermetallics* 2017;87:21-6. DOI
204. Taylor CD, Lu P, Saal J, Frankel GS, Scully JR. Integrated computational materials engineering of corrosion resistant alloys. *NPJ Mater Degrad* 2018:2. DOI
205. Taylor SR. The investigation of corrosion phenomena with high throughput methods: a review. *Corros Rev* 2011;29:135-51. DOI
206. Muster T, Trinchi A, Markley T, et al. A review of high throughput and combinatorial electrochemistry. *Electrochim Acta* 2011;56:9679-99. DOI
207. Whitfield MJ, Bono D, Wei L, Van Vliet KJ. High-throughput corrosion quantification in varied microenvironments. *Corros Sci* 2014;88:481-6. DOI
208. White P, Smith G, Harvey T, et al. A new high-throughput method for corrosion testing. *Corros Sci* 2012;58:327-31. DOI
209. Liu J, Liu N, Sun M, Li J, Sohn S, Schroers J. Fast screening of corrosion trends in metallic glasses. *ACS Comb Sci* 2019;21:666-74. DOI PubMed
210. Xiang C, Fu H, Zhang Z, et al. Effect of Cr content on microstructure and properties of $Mo_{0.5}VNbTiCr_x$ high-entropy alloys. *J Alloys Compd* 2020;818:153352. DOI
211. Renčiuková V, Macák J, Sajdl P, Novotný R, Krausová A. Corrosion of zirconium alloys demonstrated by using impedance spectroscopy. *J Nucl Mater* 2018;510:312-21. DOI
212. Qiu X. Corrosion behavior of $Al_2CrFeCo_xCuNiTi$ high-entropy alloy coating in alkaline solution and salt solution. *Results Phys* 2019;12:1737-41. DOI
213. Qiu X, Liu C. Microstructure and properties of $Al_2CrFeCoCuTiNi_x$ high-entropy alloys prepared by laser cladding. *J Alloys Compd* 2013;553:216-20. DOI
214. Hua N, Wang W, Wang Q, et al. Mechanical, corrosion, and wear properties of biomedical Ti-Zr-Nb-Ta-Mo high entropy alloys. *J Alloys Compd* 2021;861:157997. DOI

215. Elias CN, Lima JHC, Valiev R, Meyers MA. Biomedical applications of titanium and its alloys. *JOM* 2008;60:46-9. DOI

Park *et al. J Mater Inf* 2023;3:5
**DOI:** 10.20517/jmi.2022.37

**Journal of Materials Informatics**

**Research Article**

**Open Access**

Check for updates

# Thermodynamic modeling of the Fe-Sn system including an experimental re-assessment of the liquid miscibility gap

**Won-Bum Park[1], Michael Bernhard[1,2], Peter Presoly[2], Youn-Bae Kang[1,3,*]**

[1]Graduate Institute of Ferrous and Energy Materials Technology, Pohang University of Science and Technology, Pohang 37673, Korea.
[2]Chair of Ferrous Metallurgy, Montanuniversität Leoben, Leoben 8700, Austria.
[3]Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang 37673, Korea.

[*]**Correspondence to:** Prof. Youn-Bae Kang, Graduate Institute of Ferrous and Energy Materials Technology, Pohang University of Science and Technology, 77 Cheongamro, Namgu, Pohang 37673, Korea. E-mail: ybkang@postech.ac.kr

**How to cite this article:** Park WB, Bernhard M, Presoly P, Kang YB. Thermodynamic modeling of the Fe-Sn system including an experimental re-assessment of the liquid miscibility gap. *J Mater Inf* 2023;3:5. https://dx.doi.org/10.20517/jmi.2022.37

**Received:** 24 Nov 2022 **First Decision:** 10 Feb 2023 **Revised:** 2 Mar 2023 **Accepted:** 13 Mar 2023 **Published:** 23 Mar 2023

**Academic Editors:** Xingjun Liu, Yu Zhong **Copy Editor:** Ke-Cui Yang **Production Editor:** Ke-Cui Yang

## Abstract

The usage of low-grade ferrous scrap has increased over decades to decrease $CO_2$ emissions and to produce steel products at a low cost. A serious problem in melting post-consumer scrap material is the accumulation of tramp elements, e.g., Cu and Sn, in the liquid steel. These tramp elements are difficult to remove during conventional steelmaking processes. Sn is considered as one of the most harmful tramp elements because, together with Cu, it sometimes induces the liquid metal embrittlement in high-temperature ferrous processing, e.g., continuous casting and hot rolling. Furthermore, the chemical interaction between Fe and Sn plays an important role in the Sn smelting process. The raw material used in the Sn smelting process is $SnO_2$ (cassiterite), in which $Fe_3O_4$ is a gangue in the Sn ore. In the process, the reduction of $Fe_3O_4$ is unavoidable, which results in forming a Fe-Sn alloy (hardhead). The recirculation of the hardhead decreases the furnace capacity and increases the energy consumption in the smelting. The need to efficiently recover Sn from secondary resources is therefore inevitable. The CALculation of PHAse Diagrams (CALPHAD) approach helps to predict the equilibrium state of the multicomponent system. Previously reported studies of the Fe-Sn system show inconsistencies in the calculations and the experimental results. Mainly the miscibility gap in the liquid phase was under debate, as experimental data of the phase boundary are scattered. Experimental study and re-optimization of model parameters were carried out with emphasis on the correct shape of the miscibility gap. Three different experimental techniques were employed:

differential scanning calorimetry, electromagnetic levitation, and contact angle measurement. The present thermodynamic model has higher accuracy in predicting the solubility of Sn in the body-centered cubic (bcc), compared to previous assessments. This is achieved by re-evaluating the Gibbs energies of the FeSn and $FeSn_2$ compounds and the peritectic reaction related to $Fe_5Sn_3$. Also, the inconsistencies related to the miscibility gap around $X_{Sn}$ = 0.31-0.81 were resolved. The database developed in the present study can contribute to the development of a large CALPHAD database containing tramp elements.

**Keywords:** Fe-Sn, thermodynamics, CALPHAD, miscibility gap, contact angle measurement, DSC

## INTRODUCTION

Over the decades, there has been an increasing demand for steel due to building, construction, and transportation development. According to the forecast, worldwide steel consumption will reach 2500 million tons by 2050[1]. In addition to the rising costs of raw materials required to produce pig iron[2], the global society is going towards a carbon-neutral state. Consequently, iron and steelmaking industries are forced to reuse ferrous scrap materials to ensure competitiveness.

A severe issue in melting post-consumer scrap material is the accumulation of tramp elements, e.g., Cu and Sn, in the liquid steel. These tramp elements are difficult to remove during conventional steelmaking processes[3-6]. The main quality issue in high-temperature processes, e.g., continuous casting and hot rolling, related to Sn and Cu, is forming a liquid Cu-rich phase at the scale-steel interface. The phenomenon can be briefly summarized as follows: elements like Sn and Cu show less affinity to oxygen than Fe. As Fe gets selectively oxidized, the relative concentration of Cu and Sn increases, leading to enrichment at the surface or along the austenite (γ-Fe) grain boundaries. Though the solubility of Cu in austenite is high in the temperature range of 1,000-1,300 °C (mole fraction $X_{Cu}$ ~0.07-0.11)[7], Sn significantly reduces the maximum solubility. If the saturation limit is exceeded, a Cu-rich liquid phase will form under oxidizing conditions. The low-melting liquid may penetrate the steel along the austenite grain boundaries. This phenomenon is well-known as liquid metal embrittlement[8]. Further, Sn decreases the liquidus temperature of the Cu-rich melt and favors its stability down to even lower temperatures.

The chemical interaction between Fe and Sn also plays an important role in conventional Sn production. In the smelting process, Sn is extracted from $SnO_2$ (cassiterite). $SnO_2$ ore is often found with a gangue of $Fe_3O_4$ during mining. The reduction of $Fe_3O_4$ results in the formation of a Fe-Sn alloy, known as hardhead[9,10]. The recirculation of hardhead decreases the furnace capacity and increases the energy consumption in the smelting[11]. The need to recover Sn efficiently from secondary resources is now drawing attention[12].

Thermodynamic databases have been recently developed by the CALculation of PHase Diagrams (CALPHAD) method[13,14] for a variety of alloying systems. Within the framework of this approach, extensive experimental data, either from the literature or as the results of own measurements, are taken into account to model the stable phases of one, two or three-component systems. The great advantage of self-consistent CALPHAD databases is their accurate extrapolation to multicomponent alloys and the fast calculation runtime using available thermochemical software packages[15-17]. Therefore, a reliable description of the Fe-Sn system can be used in future multicomponent databases to describe the mutual influence of Cu and Sn on the high-temperature solubility limit in austenite or to predict the thermodynamic behavior of Fe and Sn in the smelting process.

In the past, researchers assessed the Fe-Sn system[18-22] using the Bragg-Williams (BW) model for the liquid phase. The Gibbs energies of the solid solutions were formulated by the Compound Energy Formalism (CEF)[23], and intermetallic phases were considered as stoichiometric compounds. Particularly for the miscibility gap in the liquid phase, information reported in the Fe-Sn system shows inconsistencies between the calculated results and the experimental data. Also, experimental results are scattered, most likely due to the experimental difficulty such as fast crystallization during the quenching of liquid alloys. Further, Kang and Pelton[24] demonstrated that the BW model is often inadequate for elucidating the miscibility gap as the calculated phase boundaries show higher and rounded shapes. The present study used three different experimental techniques to determine the binodal of the miscibility gap. Also, the Fe-Sn binary system was re-optimized using the Modified Quasichemical Model (MQM) in the pair approximation[25,26] for the liquid phase. The MQM was chosen due to its better performance in optimizing systems showing positive deviations from ideal mixing and forming a miscibility gap with fewer parameters[24]. The CEF was employed to describe the fcc and bcc solid solution phases[23]. Polynomial functions were used for the temperature dependence of the Gibbs energies of the stoichiometric compounds, FeSn, $FeSn_2$, $Fe_3Sn_2$, and $Fe_5Sn_3$.

The crystallographic data of all stable phases in the binary Fe-Sn system are listed in Table 1[27]. The phase diagram consists of the liquid solution, the face-centered cubic (fcc) and body-centered cubic (bcc) Fe solid solutions, pure Sn, as well as four intermetallic compounds (FeSn, $FeSn_2$, $Fe_3Sn_2$, and $Fe_5Sn_3$).

The binary Fe-Sn system optimized in the present study is shown in Figure 1, along with the literature data[28-46]. An essential feature of the phase diagram is a stable miscibility gap ($Liquid_1$ + $Liquid_2$). During cooling in the composition range of mole fraction Sn ($X_{Sn}$) between 0.31 and 0.80, the homogeneous melt separates into a Fe-rich melt ($Liquid_1$) and an Sn-rich melt ($Liquid_2$). The assessed consolute temperature is 1,365 °C at $X_{Sn}$ = 0.542; the monotectic temperature is 1,140 °C . In the Fe-rich part, bcc and fcc form a closed fcc single-phase region, also known as "γ-loop"[47]. The fcc phase shows a maximum solubility of $X_{Sn}$ = 0.0079 at T = 1,167 °C. The solid/liquid equilibria on the Fe side are characterized by a bcc/liquid two-phase region between $X_{Sn}$ = 0.099-0.313 above the monotectic temperature. Below this temperature, the bcc + liquid phase region exists over a wide composition range. Complex phase equilibria can be identified between the intermetallic compounds FeSn, $FeSn_2$, $Fe_3Sn_2$, and $Fe_5Sn_3$, and liquid Sn or the Sn-rich melt, respectively. Several transitions of the stoichiometric compounds are observed, discussed in detail in the Section "RESULTS OF THERMODYNAMIC OPTIMIZATION".

## MATERIALS AND METHODS

In the present study, the liquid phase miscibility gap was measured to provide key data for the modeling of the liquid phase. The sample preparation and methods were carefully selected. The present authors conducted well-established Differential Scanning Calorimetry (DSC)[48,49] in combination with the less common electromagnetic levitation technique[50] and contact angle measurement[51]. This section summarizes the sample preparation and experimental approaches used to evaluate the miscibility gap.

### Sample preparation

Master alloys were prepared using an electromagnetic levitation furnace, as shown in Figure 2. Thanks to the small chamber size of a fused silica tube (outer diameter 17.5 mm × inner diameter 16.9 mm × height 300 mm), it was easier to control the inner atmosphere against oxidation of the alloy. Conventional induction or resistance furnaces were rejected to prepare the master alloys because of the considerable chamber volume to maintain the reducing atmosphere since oxygen can prohibit the accurate determination of the phase equilibria. The miscibility gap reported in the previous studies[18-22] exists in XSn = 0.31-0.81. Detailed information on sample composition is listed in Table 2.

**Table 1. Crystallographic data of stable phase in the Fe-Sn system[27]**

| Phase | Pearson symbol | Space group | Strukturbericht | Prototype |
|---|---|---|---|---|
| Liquid | - | - | - | - |
| fcc | *cF*4 | $Fm\bar{3}m$ | A1 | Cu |
| bcc | *cI*2 | $Im\bar{3}m$ | A2 | W |
| bct-Sn | *tI*4 | $I4_1/amd$ | A5 | βSn |
| FeSn | *hP*6 | P6/mmm | B35 | CoSn |
| $FeSn_2$ | *tI*12 | I4/mcm | C16 | $Al_2Cu$ |
| $Fe_3Sn_2$ | *hR*10 | $R\bar{3}m$ | - | - |
| $Fe_5Sn_3$ | *hP*6 | $P6_3/mmc$ | $B8_2$ | $Ni_2In$ |

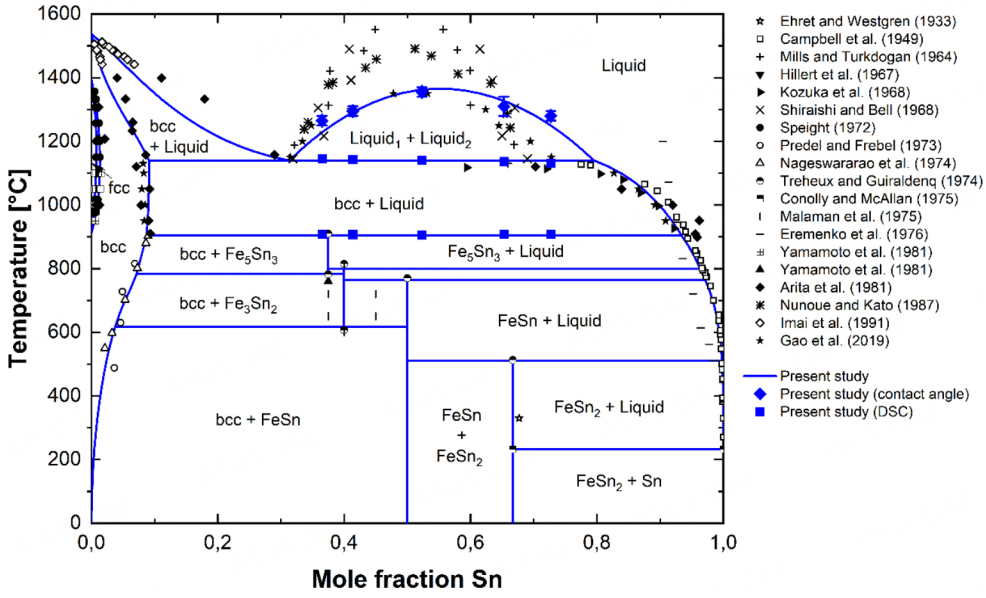


**Figure 1.** Calculated Fe-Sn phase diagram along with data from literature[28-46]. bcc: Body-centered cubic; DSC: Differential Dcanning Calorimetry; fcc: face-centered cubic.

Electrolytic Fe (99.9 mass pct., Blyth & Co., Ltd., Japan) and pure metal Sn (RND Korea Corp., Sn: 99.999 mass pct., Gwangmyeong-si, Korea) were charged in an alumina crucible (outer diameter: 8 mm × inner diameter 4 mm × height 10 mm). The crucible was placed in a fused silica cup in the chamber of the induction furnace equipped with an RF (Radio Frequency) generator (30 kW, 260 kHz, PSTEK, Korea). The power of the RF generator was turned on, and subsequently, the alloys were levitated and melted. During the levitation and subsequent melting, Ar-5 %$H_2$ (g) was supplied to the reaction chamber at a rate of 1.0 L $min^{-1}$ using a capillary flowmeter. The Ar-5 %$H_2$ (g) was purified by passing through the $CaSO_4$ column and heated Mg chips (500 °C). The oxygen partial pressure of the outlet gas was confirmed by a $ZrO_2$ oxygen sensor (Daiichi Nekken Co. Ltd, TB-IIF oxygen analyzer, Ashiya, Japan), showing ~$10^{-23}$ - ~$10^{-22}$ atm. The temperature of the levitated droplet was measured by a two-color pyrometer (Raytek, RAY2CBG, Santa Cruz, CA, USA) and was controlled by independently adjusting the RF generator power manually. The pyrometer was calibrated by measuring the melting temperature of pure Fe, pure Cu, and pure Ni with a B-type thermocouple. The uncertainty on the temperature was estimated to be ± 10 °C. In the experiments, the

**Table 2. Compositions used in the present experimental investigation**

| Sample | Composition ($X_{Sn}$) | Experimental technique | | |
|---|---|---|---|---|
| | | Differential scanning calorimetry | Contact angle measurement | Electromagnetic levitation technique |
| I | 0.365 | Yes | Yes | Yes |
| II | 0.414 | Yes | Yes | No |
| III | 0.523 | Yes | Yes | No |
| IV | 0.653 | Yes | Yes | No |
| V | 0.727 | Yes | Yes | No |

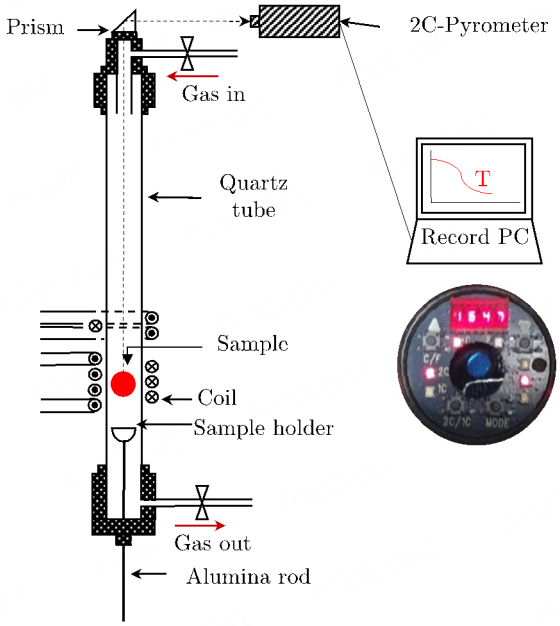


**Figure 2.** Experimental apparatus for the electromagnetic levitation test.

temperature was held at 1,600 °C for 2 min for mixing Fe and Sn. After a predetermined time passed, the RF generator was turned off and the droplet was cooled down to room temperature in the furnace to prevent oxidation.

### Differential scanning calorimetry

All differential scanning calorimetry (DSC) measurements in the present study were carried out in a NETZSCH DSC 404F1 Pegasus with an Rh furnace ($T_{max}$ = 1,650 °C) and a Pt sensor with thermocouples of type S. Alumina crucibles with lids (85 µl) were used for all experiments. The protective tube of the Rh furnace was purged permanently with Ar 5.0 (purity 99.999%). To minimize oxygen levels at high temperatures, a Zr getter was placed directly below the DSC sensor. By this, oxygen is absorbed by the getter material at temperatures above 350 °C. The experimental setup was calibrated by measuring the melting points of NETZCH's standards of pure metals In, Bi, Al, Ag, Au, Ni, and Co. More detailed information on the used equipment can be found in the references[48,49].

The DSC signal strongly depends on the (i) sample mass, (ii) the applied heating rate (HR) and (iii) the heat consumption/release (ΔH) during the phase transformation[48]. It was expected from thermodynamic calculations[19] that by crossing the liquidus phase boundary ($Liquid_1$ + $Liquid_2$ ↔ Liquid) during scanning,

only a minor amount of heat would be exchanged. Hence, pre-trials were carried out to identify proper parameters for achieving a sufficient resolution in the DSC signal. Alloys I and V were selected to optimize the DSC settings. The chemical analysis ($X_{Sn}$ = 0.365 and $X_{Sn}$ = 0.727) are close to the composition limit of the miscibility gap ($X_{Sn}$ = 0.31-0.81) with the smallest Liquid $_1$ + Liquid$_2$ two-phase region. Therefore, the maximum change of ΔH per time ($t$ = ΔT/HR) was expected to intensify the signal. The defined time-temperature profiles are graphically shown in Figure 3. Samples of alloys I and V with different masses (200 mg and 400 mg) were heated up at 30 °C min$^{-1}$ to 400 °C and held isothermally for 15 min to activate the Zr getter. Then, three different DSC settings were tested for each alloy to determine the actual phase transformations of interest: (i) sample mass of 200 mg with the most commonly applied HR of 10 °C min$^{-1}$[48]; (ii) sample mass of 200 mg with increased HR of 20 °C min$^{-1}$; and (iii) sample mass of 400 mg with HR of 20 °C min$^{-1}$. In the first time-temperature cycle, the samples were pre-melted at the respective HR to a final temperature of 1,500 °C to guarantee perfect contact between the sample droplet surface and the alumina crucible for the second cycle. Subsequently, the samples were cooled to 450 °C at -30 °C min$^{-1}$ and reheated to 1,500 °C with the defined HR to record the transition temperatures.

After the pre-trials, the actual DSC measurements for all samples I-V were based on the most proper experimental parameters, and analyzed in detail within the Section "RESULTS OF EXPERIMENTAL INVESTIGATIONS".

### Electromagnetic levitation technique

A previous study[50] reported that the metastable miscibility gap in Cu-Co-Fe liquid alloys could be measured using the electromagnetic levitation technique upon undercooling. The experiment was conducted to confirm the liquidus, solidus, and metastable-liquid separation temperature in Ar/He (g) atmosphere. Ar (g) was provided for the inert atmosphere, and He (g) was added to increase the thermal conductivity attributed to its low atomic weight. They observed a temperature change over time and confirmed the peritectic temperature, liquidus temperature, and metastable-liquid separation temperature. This study demonstrated that the slope change from the time-temperature profile could represent the phase transformation temperature of the alloys. In the present study, this technique was employed to measure the miscibility gap in the Fe-Sn binary liquid using electromagnetic levitation equipment.

0.6 g of Fe-Sn alloy sample were charged in an alumina crucible (outer diameter: 8 mm × inner diameter 4 mm × height 10 mm) and placed inside the fused silica sample holder. A reaction chamber made up of a fused silica tube was permanently purged by purified Ar(g). Ar(g) was purified by passing silica gel and drierite® for moisture and ascarite® for a trace of $CO_2$(g). The flow rate was controlled by Mass Flow Controller (KOFLOC, KOFLOC 3660, Kyoto, Japan). The MFC was preliminarily calibrated using a soap-bubble-column technique. The melting procedure was the same as that described in Section "Sample preparation" (also see Figure 2). The RF generator was turned on a current at 97 A under an Ar(g) flow rate of 1.2 L min$^{-1}$. As soon as the RF generator power was turned on, the sample was immediately levitated and melted. After the time-temperature profile showed a steady state, the injection of Ar(g) was replaced by a subsequent injection of He(g) at a rate of 2 L min$^{-1}$. At the same time, the current of the RF generator was controlled to cool down the sample to observe the slope change of the profile. When the sample was cooled down and showed a steady state, the RF generator was turned off. The time-temperature profile during each measurement was recorded by a PC connected to the pyrometer.

### Contact angle measurement

Min *et al.*[51] measured the contact angle change upon cooling of Bi-Cu-Sn liquid alloys using the constrained drop method and showed that this technique was capable of detecting the binodal temperature of the alloys that separated into two liquid phases. They confirmed that the contact angle was sensitive to

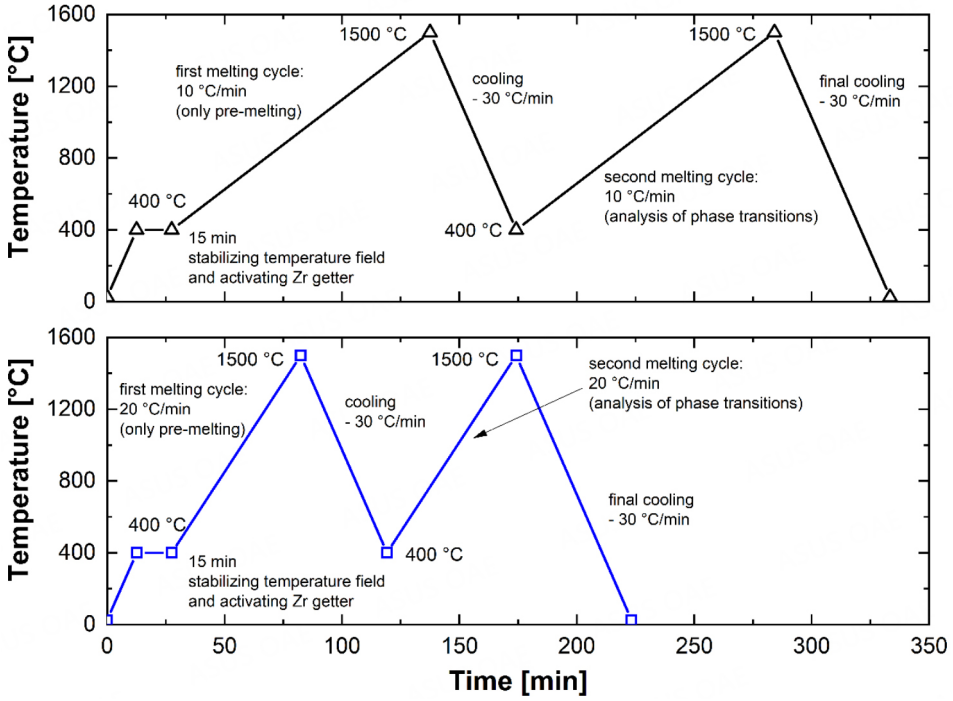


**Figure 3.** Time-temperature cycles defined in the DSC with two different heating rates of 10 °C $min^{-1}$ and 20 °C $min^{-1}$.

the surface tension of the droplet, which changed significantly upon the two-phase separation. In the present study, the contact angle of the liquid Fe-Sn binary alloy was measured in order to detect the binodal temperature of the alloy.

The apparatus to measure the contact angle is schematically shown in Figure 4. A sample mass of 2 g of the Fe-Sn alloy was placed on an alumina substrate, and charged into the hot zone of a horizontal resistance-heating furnace (Lenton, LTF-17, Parsons Lane, UK with $MoSi_2$ heating elements of Kanthal U-type) under an Ar-5% $H_2$(g) atmosphere. Ar-5% $H_2$(g) was purified by passing silica gel and $CaSO_4$ column for moisture and ascarite for a trace of $CO_2$(g). Since oxygen content could affect the contact angle, Ar-5% $H_2$(g) was also passed through the Mg and Cu turnings at 500 °C. The oxygen partial pressure was measured using a $ZrO_2$ oxygen sensor and its pressure range was ~$10^{-23}$-~$10^{-22}$ atm. Ar-5% $H_2$(g) was supplied into the furnace with a flow rate of 0.2 L $min^{-1}$ to minimize possible vaporization of the elements. The temperature of the horizontal furnace was raised to 1,400 °C at 2 °C $min^{-1}$ rate and held for 30 min. The melt temperature was checked using a B-type thermocouple calibrated with the standard thermocouple (CHINO Corp., Model C800-65, Type B, 0.5 mm in diameter, 1,500 mm in length, Tokyo, Japan). The temperature was lowered to 1,100 °C at 30 °C intervals. After holding temperature for 20 min at each temperature, a photograph was taken with Nikon Camera (Nikon, D3100, Tokyo, Japan) with a Canon lens (Canon, SIGMA APO 70-300 mm F4-5.6 DG, Tokyo, Japan). After obtaining the photographs, the image was converted into a greyscale image using Image J. The image was digitized and the digitized data were handled using Microsoft Excel to run an algorithm developed by Lee *et al.*[52] based on *k*-Means clustering[53,54]. This algorithm is an objective algorithm that can be adopted without being biased by human decisions, and it was shown to be effective in extracting the contact angle of a droplet. Details of this algorithm can be found elsewhere[52].

## RESULTS OF EXPERIMENTAL INVESTIGATIONS

### Phase equilibria obtained by DSC

Figure 5A and B present the DSC response during the heating of samples I and V (note that the results are corrected by an arbitrary shift of the voltage). Detailed information on the interpretation of DSC signals to determine phase transformations can be found in the reference[55]. The first deviation from the baseline is observed between 231.5-235.6 °C, corresponding to the melting of pure Sn (231.91 °C)[56]. According to the phase diagram in Figure 1, this phase transformation is predicted only for sample V from Table 2, indicating that due to non-equilibrium cooling during the preparation of the samples, the microstructure did not reach the equilibrium state at ambient temperature. Above 400 °C, only the second heating cycle of the remelted sample is plotted in Figure 5. Both alloys show identical phase transformations. Again, this transition sequence is calculated under equilibrium conditions only for a composition of $X_{Sn}$ = 0.727. It can be therefore concluded that even by slow cooling in the calorimeter (-30 °C $min^{-1}$), the strong unmixing tendency of Fe and Sn could not be avoided. The first onset at 517-520 °C may be assigned to FeSn + $FeSn_2$ ↔ FeSn + Liquid and/or $FeSn_2$ + Liquid ↔ FeSn + Liquid in the phase diagram [Figure 1]. However, this assumption is based only on the calculated phase equilibria. Between 778 °C and 804 °C, the $Fe_3Sn_2$ + Liquid and/or $Fe_3Sn_2$ + $Fe_5Sn_3$ two-phase regions exist. Above 907.5 °C, the liquid is in equilibrium with the bcc solid solution, representing the first phase boundary stable for all investigated compositions in the Fe-Sn phase diagram [Figure 1]. As expected, the onset temperatures in Figure 5 do not systematically depend on the used sample mass or the applied HR.

The monotectic temperature (bcc + Liquid ↔ $Liquid_1$ + $Liquid_2$) in Figure 5 is characterized by a peak in the DSC signal. It is well known that peak temperatures are generally shifted to higher temperatures with increased HR and larger sample mass[48,49]. This fact was also observed in the present results. The peak temperature depends on the liquid and bcc fractions below the monotectic temperature, which are transformed into two liquid phases when the monotectic line is reached. According to the lever rule in Figure 1, less bcc is stable in the bcc + liquid two-phase region with increasing $X_{Sn}$. Due to the lower amount of melting solid phase, the peak intensity was reduced and the peak temperature was detected at lower values of about 1,130 °C; see for comparison 1,150 °C at $X_{Sn}$ = 0.365. Therefore, the actual aim of the pre-measurements of identifying the liquidus temperature ($Liquid_1$ + $Liquid_2$ ↔ Liquid) could not be achieved. Even increasing the HR to 20 °C $min^{-1}$ and using twice the mass of the samples (400 mg) did not result in a detectable signal change by crossing the liquidus line above 1,200 °C.

Based on the pre-test, a reliable DSC analysis was possible for melting the intermetallic compound $Fe_5Sn_3$ (bcc + $Fe_5Sn_3$ ↔ bcc + Liquid and/or $Fe_5Sn_3$ + Liquid ↔ bcc + liquid). Both phase equilibria are stable at all chemical compositions investigated in the phase diagram. The formation of non-equilibrium microstructure components can therefore be neglected. The sample mass of 200 mg was selected, and a heating rate of 10 °C $min^{-1}$ was defined. The DSC analysis of samples I to V is summarized in Figure 6A. The bcc + $Fe_5Sn_3$ ↔ bcc + Liquid and/or $Fe_5Sn_3$ + Liquid ↔ bcc + Liquid phase transformation is obtained at 907 ± 1 °C. More significant deviations are observed in the peak of the monotectic temperature, which can be explained by the different bcc fractions melting during heating, influencing the intensity of the DSC signal. The present work proposes a temperature of 1,139 ± 5 °C for the bcc + Liquid ↔ $Liquid_1$ + $Liquid_2$ phase transformation. Figure 6B shows the enlarged section of the miscibility gap. Unfortunately, the binodal line could not be obtained for any sample composition.

### Results of the electromagnetic levitation test

Figure 7 shows the time-temperature profile change for sample I ($X_{Sn}$ = 0.365). The sample was held at 1,760 °C for some time and subsequently was cooled by blowing He(g). The temperature of the sample decreased to 1,385 °C. During the cooling, there was no visible change in the slope to the time-temperature profile, as was expected. Below 1,385 °C, the sample could not be levitated anymore due to the loss of the levitation force. According to some of the previous reports[30,33,44], the binodal temperature was higher than

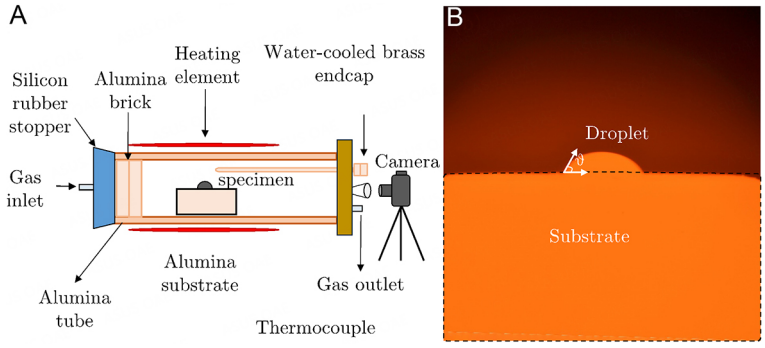


**Figure 4.** (A) A schematic diagram of the experimental apparatus and (B) a droplet image of sample I at 1,400 °C.

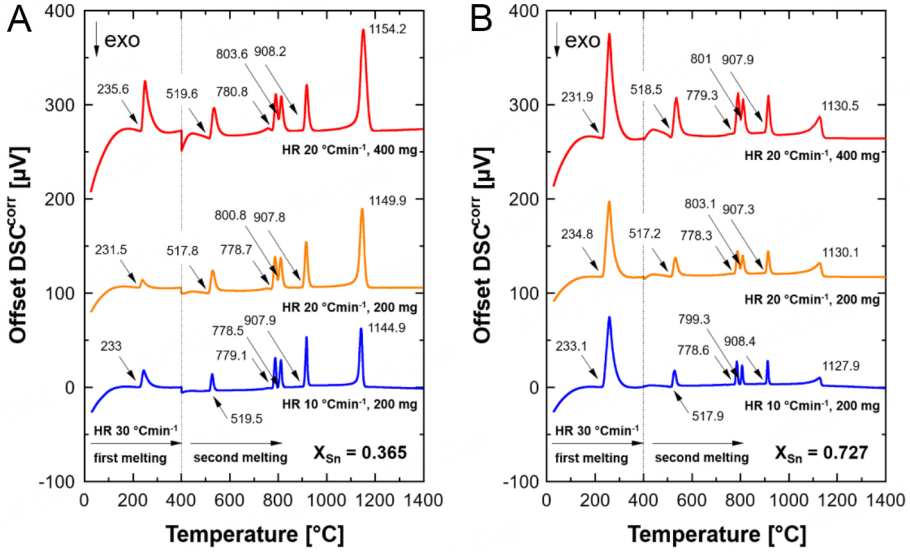


**Figure 5.** DSC response by measuring samples I (A) and V (B) with three different setting combinations (200 mg with 10 °C $min^{-1}$, 200 mg with 20 °C $min^{-1}$, and 400 mg with 20 °C $min^{-1}$). DSC: Differential Scanning Calorimetry; $DSC^{corr}$: Differential Scanning Calorimetry results corrected by arbitrary shift of voltage; exo: exothermic heat change; HR: Heating Rate.

1,385 °C. Therefore, the present electromagnetic levitation test did not support the previously reported higher binodal temperature. Additionally, it might be attributed to the fact that heat evolution during the two-phase separation might be too low to detect in the present study, contrary to the case of the Cu-Co-Fe liquid alloy[50]. It is consistent with the present DSC results. Therefore, the electromagnetic levitation test was not applicable in the present system.

### Characterization of the miscibility gap by contact angle measurement

The measured contact angles of the five samples are shown in Figure 8. The measured monotectic temperature by the DSC measurement was also marked. During the contact angle measurement, the temperature was varied from 1,400 °C to 1,100 °C.

The temperature interval was 30 °C, and therefore any abrupt change corresponds to the two-phase separation of the liquid phase with an uncertainty of ± 15 °C. Samples I and IV [Figure 8A and D] showed noticeable changes in the contact angle at 1,265 ± 15 °C and 1,310 ± 30 °C, respectively. The other samples II, III, and V [Figure 8B, C, and E] showed weak changes at 1,295 ± 15 °C, 1,355 ± 15 °C, and 1,295 ± 15 °C,

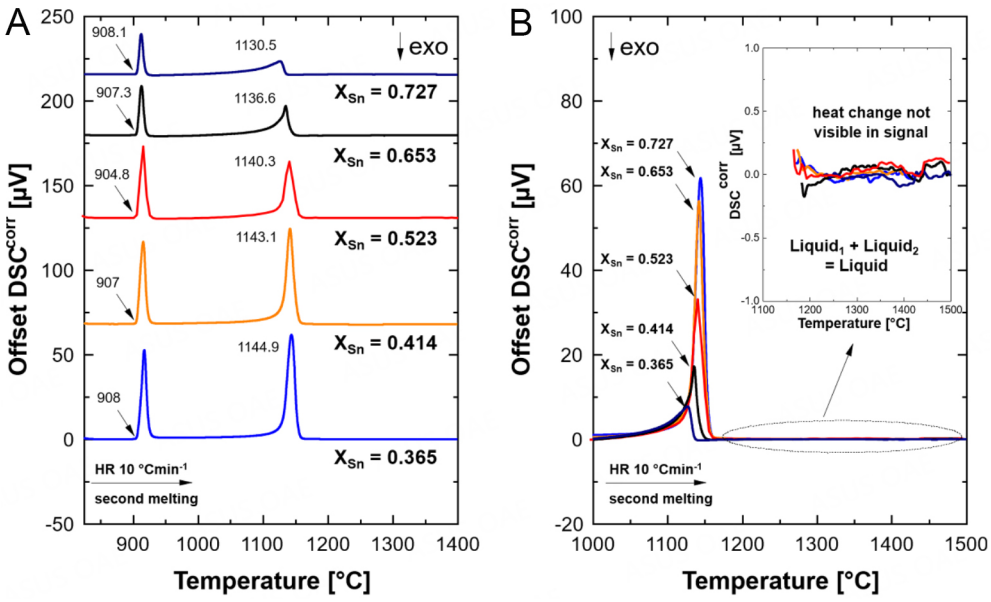


**Figure 6.** DSC signal obtained during heating of sample I to V with a heating rate of 10 °C $min^{-1}$: (A) above 400 °C only the second heating cycle is plotted; (B) enlarged section of the Liquid1 + Liquid2 ↔ Liquid region. DSC: Differential Scanning Calorimetry; $DSC^{corr}$: Differential Scanning Calorimetry results corrected by arbitrary shift of voltage; exo: exothermic heat change; HR: Heating Rate.

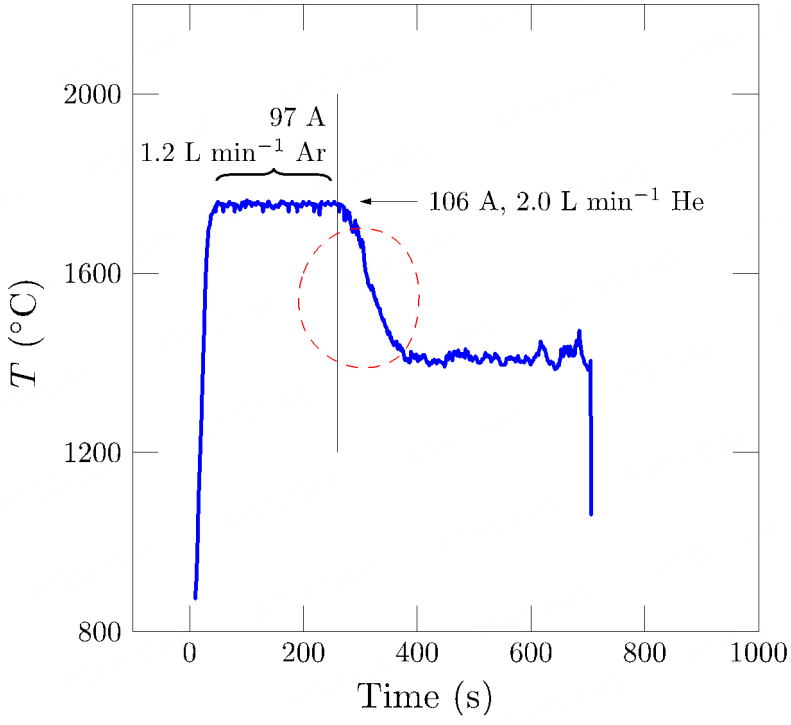


**Figure 7.** Time-temperature profile for the electromagnetic levitation test using sample I ($X_{Sn}$ = 0.365).

respectively. Nevertheless, the contact angle changes upon cooling were more noticeable than those observed using DSC upon heating and the electromagnetic levitation technique upon cooling. In the subsequent thermodynamic optimization, the detected temperatures were regarded as the binodal temperature consisting of the miscibility gap.

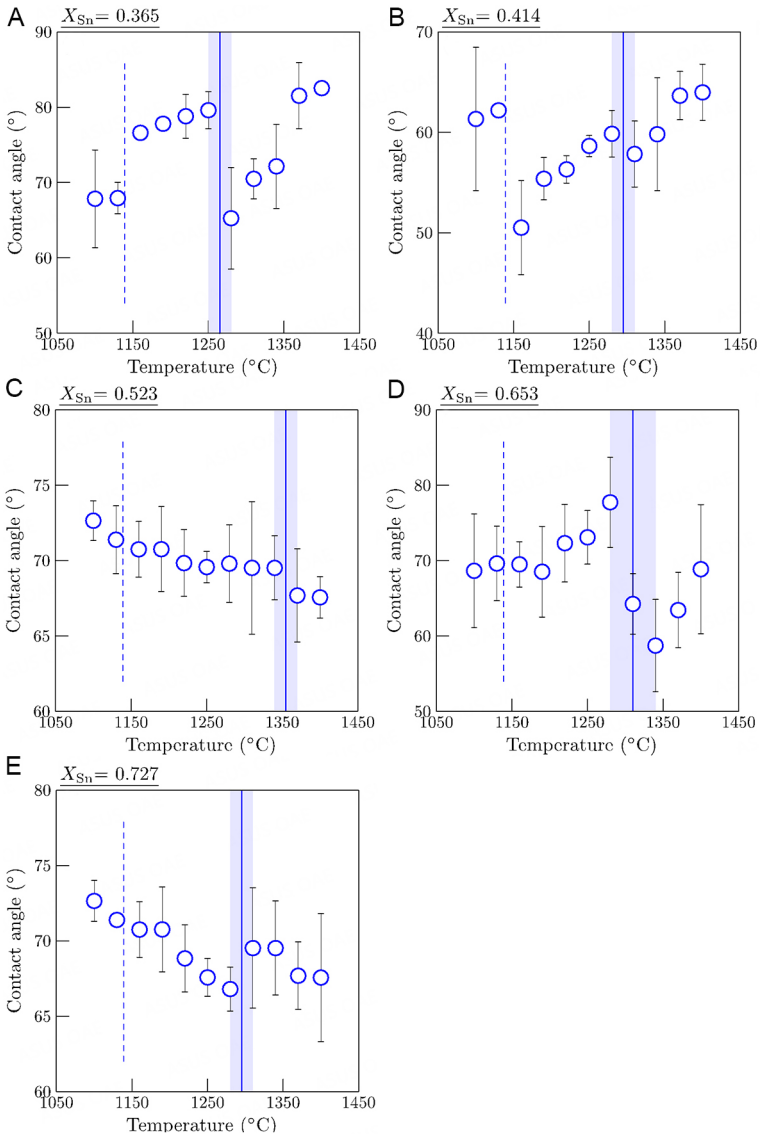


**Figure 8.** The contact angle of the five samples (I to V) measured at each temperature: (A) $X_{Sn}$ = 0.365; (B) $X_{Sn}$ = 0.414; (C) $X_{Sn}$ = 0.523; (D) $X_{Sn}$ = 0.653; and (E) $X_{Sn}$ = 0.727 Sn. Vertical dashed lines are the measured monotectic temperature by DSC (1139 ± 5 °C). Vertical thick lines with shade are the estimated binodal temperature with uncertainty.

The contact angle between pure liquid Fe and the alumina substrate varies in the range of 110 to 141[57-59] at 1,550-1,635 °C, and in the case of pure liquid Sn, the values are 131 to 165[58,60] at 338-1,000 °C. The contact angle of the Fe-Sn liquid alloys measured in the present study was lower than the reported values. Possible reasons are (1) impurities in the $Al_2O_3$ substrate or (2) the roughness in the $Al_2O_3$ substrate. Although oxygen contamination may happen at the surface of the alloys, the authors assumed that a significant change in the contact angle could be interpreted as the occurrence of a phase transformation.

## THERMODYNAMIC MODELS

### Liquid phase

In previous assessments of the Fe-Sn system[18-22], the BW random mixing model was used for describing the positive deviation of the enthalpy of mixing ($\Delta h > 0$) of the liquid phase. Kang and Pelton[24] showed that unless many empirical temperature-dependent parameters are used in the BW model, the shape of the miscibility gap is often too high and rounded. The limited reproducibility of flat-shaped miscibility gaps results from neglecting short-range ordering (SRO) as well as clustering. In this case, it was demonstrated[24] that the accuracy of the liquidus phase boundary could be significantly improved using the Modified Quasichemcal Model (MQM). Detailed information on SRO in liquid solutions can be found elsewhere[61]. In optimizing Fe-based phase diagrams, Shubhank and Kang[7] and Tafwidli and Kang[62] successfully applied the MQM modeling of the miscibility gap in the Fe-Cu and the Fe-C-S system, respectively. The miscibility gap in the present Fe-Sn system may not be seen as a significantly "flattened"-shape, compared to those discussed by Kang and Pelton[24]. However, Kang and Pelton also showed that the MQM predicts the miscibility gap (positive deviation) in ternary systems better than the BW model does[24,61]. Since the present study is a part of developing a thermodynamic database for a larger Fe-alloy system with tramp elements, the MQM was employed in the present study.

A detailed description of the MQM can be found elsewhere[25]. Hence, the model will be explained only briefly in this section. In the pair approximation for a binary solution consisting of Fe and Sn atoms, the following pair exchange reaction on the sites of a quasi-lattice is considered:

$$(\mathrm{Fe}-\mathrm{Fe})+(\mathrm{Sn}-\mathrm{Sn})=2(\mathrm{Fe}-\mathrm{Sn});\ \Delta g_{\mathrm{FeSn}} \tag{1}$$

where (i - j) represent the first nearest neighbor pair and $\Delta g_{FeSn}$ is the non-configurational Gibbs free energy change forming two moles of (Fe - Sn) pairs. If $n_{Fe}$ and $n_{Sn}$ are the numbers of moles Fe and Sn, $n_{ij}$ is the number of (i - j) pairs and $Z_{Fe}$ and $Z_{Sn}$ are the coordination numbers of Fe and Sn, then the following mass balances are considered:

$$Z_{\mathrm{Fe}}n_{\mathrm{Fe}}=2n_{\mathrm{FeFe}}+n_{\mathrm{FeSn}} \tag{2}$$
$$Z_{\mathrm{Sn}}n_{\mathrm{Sn}}=2n_{\mathrm{SnSn}}+n_{\mathrm{FeSn}} \tag{3}$$

The pair fractions, mole fractions and coordination-equivalent fractions are defined as:

$$X_{ij}=n_{ij}/(n_{\mathrm{FeFe}}+n_{\mathrm{SnSn}}+n_{\mathrm{FeSn}}) \tag{4}$$
$$X_{\mathrm{Fe}}=n_{\mathrm{Fe}}/(n_{\mathrm{Fe}}+n_{\mathrm{Sn}})=1-X_{\mathrm{Sn}} \tag{5}$$
$$Y_{\mathrm{Fe}}=Z_{\mathrm{Fe}}n_{\mathrm{Fe}}/(Z_{\mathrm{Fe}}n_{\mathrm{Fe}}+Z_{\mathrm{Sn}}n_{\mathrm{Sn}})=Z_{\mathrm{Fe}}X_{\mathrm{Fe}}/(Z_{\mathrm{Fe}}X_{\mathrm{Fe}}+Z_{\mathrm{Sn}}X_{\mathrm{Sn}})=1-Y_{\mathrm{Sn}} \tag{6}$$

The Gibbs energy of the solution is given by:

$$G=(n_{\mathrm{Fe}}g^{o}_{\mathrm{Fe}}+n_{\mathrm{Sn}}g^{o}_{\mathrm{Sn}})-T\Delta S^{\mathrm{config}}+(n_{\mathrm{FeSn}}/2)\Delta g_{\mathrm{FeSn}} \tag{7}$$

where $g^{o}_{Fe}$ and $g^{o}_{Sn}$ are the molar Gibbs energies of the pure components and $\Delta S^{config}$ is the configurational entropy of mixing given by randomly distributing the (Fe-Fe), (Sn-Sn) and (Fe-Sn) pairs in the one-dimensional Ising approximation[63]:

$$\Delta S^{config} = -R(n_{Fe}\ln X_{Fe} + n_{Sn}\ln X_{Sn}) \\ -R[n_{FeFe}\ln(X_{FeFe}/Y_{Fe}^2) + n_{SnSn}\ln(X_{SnSn}/Y_{Sn}^2) + n_{FeSn}\ln(X_{FeSn}/2Y_{Fe}Y_{Sn})] \quad (8)$$

The $\Delta g_{FeSn}$ may be expanded in terms of pair fractions:

$$\Delta g_{FeSn} = \Delta g^0_{FeSn} + \sum_{i\geq 1} g^{i0}_{FeSn} X^i_{FeFe} + \sum_{j\geq 1} g^{0j}_{FeSn} X^j_{SnSn} \quad (9)$$

where $\Delta g^0_{FeSn}$, $g^{i0}_{FeSn}$ and $g^{0j}_{FeSn}$ are the model parameters that can be functions of temperature.

The equilibrium pair distribution is determined by setting

$$(\partial G/\partial n_{FeSn})_{n_{Fe},n_{Sn}} = 0 \quad (10)$$

which leads to the equilibrium constant for the pair formation in Reaction (1):

$$X^2_{FeSn}/(X_{FeFe}X_{SnSn}) = 4\exp(-\Delta g_{FeSn}/RT) \quad (11)$$

The composition of maximum SRO is defined by the ratio of the coordination numbers $Z_{Sn}/Z_{Fe}$, as given in the following equations:

$$\frac{1}{Z_{Fe}} = \frac{1}{Z^{Fe}_{FeFe}}\left(\frac{2n_{FeFe}}{2n_{FeFe}+n_{FeSn}}\right) + \frac{1}{Z^{Fe}_{FeSn}}\left(\frac{n_{FeSn}}{2n_{FeFe}+n_{FeSn}}\right) \quad (12)$$

$$\frac{1}{Z_{Sn}} = \frac{1}{Z^{Sn}_{SnSn}}\left(\frac{2n_{SnSn}}{2n_{SnSn}+n_{FeSn}}\right) + \frac{1}{Z^{Sn}_{SnFe}}\left(\frac{n_{FeSn}}{2n_{SnSn}+n_{FeSn}}\right) \quad (13)$$

where $Z^{Fe}_{FeFe}$ and $Z^{Fe}_{FeSn}$ are the values of $Z_{Fe}$ when all neighbors of Fe are Fes and when all nearest neighbors of Fe are Sns, respectively, and where $Z^{Sn}_{SnSn}$ and $Z^{Sn}_{SnFe}$ are defined similarly. $Z^{Fe}_{FeSn}$ and $Z^{Fe}_{SnFe}$ represent the same quantity and are interchangeable. The coordination numbers $Z^{Fe}_{FeFe}$ and $Z^{Sn}_{SnSn}$ were set to 6[61], whereas the ratio of $Z_{Sn}/Z_{Fe}$ was set to 1 with $Z^{Fe}_{FeSn} = Z^{Sn}_{FeSn} = 6$ in the present study.

### Solid solutions and stoichiometric compounds

Bcc and fcc solid solutions were modeled using the Compound Energy Formalism (CEF)[23] with two sublattices. Fe and Sn are located on the substitutional sites, while vacancy (Va) occupies the interstitial sites. If all the sites in all but one of the sublattices are vacant, the CEF reduces to the BW random mixing model[23]. Note that the vacancy sublattice is only considered to incorporate the present descriptions into a multicomponent steel database. The Gibbs energy for fcc and bcc is therefore defined by

$$G = (X_{Fe}G^o_{Fe} + X_{Sn}G^o_{Sn}) + RT(X_{Fe}\ln X_{Fe} + X_{Sn}\ln X_{Sn}) + X_{Fe}X_{Sn}\sum_{v\geq 0} L^v_{Fe,Sn}(X_{Sn} - X_{Fe})^v + G^{mo}_m \quad (14)$$

where $X_i$ and $G^o_i$ are the mole fraction and the molar Gibbs energy of component i. $L^i_{Fe,Sn}$ represent the adjustable model parameters which can be a function of temperature. $G^{mo}_m$ is the contribution due to magnetic ordering. Its expression per mole of atoms was proposed by Hillert and Jarl[64]:

$$G_m^{mo} = RT\ln(\beta_0 + 1)g(\tau) \quad (15)$$

where $\tau = T/T_{Cr}$ and $T_{Cr}$ is the critical temperature (Néel or Curie) and $\beta_0$ is the average magnetic moment per atom.

The intermetallic phases - FeSn, $FeSn_2$, $Fe_3Sn_3$ and $Fe_5Sn_3$ - were treated as stoichiometric compounds.

## RESULTS OF THERMODYNAMIC OPTIMIZATION

Table 3 shows the assessed parameters in the present study. Data for pure elements were taken from the work of Dinsdale[65]. The calculated results are critically compared with the selected Fe-Sn assessments of Kumar *et al.*[19] and Miettinen[20], which have been widely accepted until recently.

### Thermodynamic properties of the liquid phase

The enthalpy of mixing in the liquid phase ($\Delta h$) was measured using calorimetry. Batalin *et al.*[67] determined the $\Delta h$ in the whole composition at 1,600 °C. Also, a full range of compositions of $\Delta h$ was investigated by Petrushevskiy *et al.*[68] at 1,677 °C. Lück and Predel[69] measured $\Delta h$ at 1,547 °C in a limited composition range ($X_{Sn} < 0.229$). They mentioned that it was difficult to obtain the data because of the high evaporation rate of Sn in the liquid alloy.

Figure 9A shows the calculated partial enthalpy of mixing ($\Delta h_{Fe,}$ $\Delta h_{Sn}$) and Figure 9B the integral enthalpy of mixing ($\Delta h$) of the liquid alloy along with experimental data and calculated results. Petrushevskiy *et al.*[68] measured slightly negative values for the $\Delta h_{Fe}$ on the Fe side. The data show inconsistencies with the values obtained by Batalin *et al.*[67]. The data of Batalin *et al.*[67] were considered more reliable. As also for $\Delta h$, a reasonable agreement can be identified with the measurements of Lück and Predel[69]. The large positive enthalpy of mixing indicates the stability of the miscibility gap. In the present study, both data sets by Batalin *et al.*[67] and Lück and Predel[69] were considered to model the liquid phase at $X_{Sn} < 0.25$, representing the main difference from the previous work of Kumar *et al.*[19] and Miettinen[20].

Activities of Fe and Sn in liquid alloys were investigated with Electromotive Force (EMF) measurement, vapor pressure measurement, mass spectrometric method, and transportation method. Kozuka *et al.*[32] conducted the EMF measurement with Fe, FeO, and $ZrO_2$-CaO solid electrolytes and used Sn-Fe alloys to measure the activities of Fe and Sn at 1,100 °C and 1,200 °C, respectively. Wagner and St. Pierre[70] conducted the mass spectrometric method to measure the ionic intensities of the Fe and Sn and obtained the activities of Fe and Sn at 1,287 °C and 1,537 °C. Also, Nunoue and Kato[44] and Yamamoto *et al.*[42] used a mass spectrometric method to determine the activity at 1,550 °C and 1,600 °C. By vapor-liquid equilibrium, Maruyama and Ban-ya[71] conducted a transportation method, which was used to measure the concentration ratios of the Fe and Sn in the vapor and metal phases to obtain the activity. Eremenko *et al.*[72] measured the vapor pressure with the effusion technique and Federenko and Brovkin[73] and Shiraishi and Bell[74] measured with the transportation method at 1,100-1,300 °C. Yazawa and Koike[75] determined the activity of Sn by distribution of Sn between Fe and Pb alloys at 1,350 to 1,500 °C.

The activities of each component in the liquid alloys at 1,100-1,600 °C are represented as $\Delta\mu_M = RT \ln a_M$ in Figure 10. As shown in Figure 10A, the present calculation and previous studies[19,20] show good accordance with data from Kozuka *et al.*[32] at 1,200 °C. Data from Wagner and St. Pierre[70] seem to differ up to $X_{Sn} = 0.15$ due to the fcc phase appearance. Data from Shiraishi and Bell[74] show a slight deviation up to $X_{Sn} = 0.3$, and this might have been attributed to the condensation of SnO during the experiment, as mentioned

**Table 3. Model parameters for the Fe-Sn system optimized in the present study (J mol$^{-1}$ or J mol$^{-1}$K$^{-1}$)**

| Phase | Model |
|---|---|
| **Liquid** | **MQM (Fe,Sn)** |
| $Z^{Fe}_{FeFe} = Z^{Sn}_{SnSn} = Z^{Fe}_{FeSn} = Z^{Sn}_{FeSn} = 6$ | |
| $\Delta g_{FeSn}$ | $14{,}434.8 - 2.21752T - 7322X_{FeFe} - 4184X_{SnSn} - 2928.8(X_{FeFe})^2$ |
| **bcc (α-Fe, δ-Fe)** | **CEF (Fe,Sn)$_1$(Va)$_3$** |
| $\mathbf{G_{Fe:Va}}$ | GHSERFE |
| $\mathbf{G_{Sn:Va}}$ | GBCCSN |
| $\mathbf{L_{Fe,Sn:Va}}$ | $23{,}012 + 8.9956T + 23{,}012(X_{Sn}-X_{Fe})$ |
| $\mathbf{\beta_{Fe:Va}}$ | $2.22X_{Fe}$ [64] |
| $\mathbf{T_{C(Fe:Va)}}$ | $1043X_{Fe}$ [65] |
| **fcc (γ-Fe)** | **CEF (Fe,Sn)$_1$(Va)$_1$** |
| $\mathbf{G_{Fe:Va}}$ | GFCCFE |
| $\mathbf{G_{Sn:Va}}$ | GFCCSN |
| $\mathbf{L_{Fe,Sn:Va}}$ | 23012 - 1.8828T |
| $\mathbf{\beta_{Fe:Va}}$ | $-2.1X_{Fe}$ [66] |
| $\mathbf{T_{C(Fe:Va)}}$ | $-201X_{Fe}$ [66] |
| **Stoichiometric compounds** | |
| FeSn | $C_p$ (1 K < $T$ < 41 K) = $1.0 \times 10^{-4}T + 5.1 \times 10^{-3}T^2 + 3.0 \times 10^{-5}T^3$<br>$C_p$ (41 K < $T$ < 135 K) = $-12.051834 + 6.36436 \times 10^{-1}T - 3.101986 \times 10^{2}T^{-2} - 1.92343 \times 10^{-3}T^2$<br>$C_p$ (135 K < $T$ < 250 K) = $25.827026 + 1.57530 \times 10^{-1}T - 6.659823 \times 10^{4}T^{-2} - 2.55233 \times 10^{-4}T^2$<br>$C_p$ (250 K < $T$ <1000 K) = $80-500T^{-0.5}$<br>$H°_{298}$ = -8600, $S°_{298}$ = 79.2675, $T_{Cr}$ = 365 K, $_0$ = 0.37 |
| $FeSn_2$ | $C_p$ (250 K < $T$ < 505.08 K) = $55.4363 + 8.427584 \times 10^{-2}T + 9.3122 \times 10^{4}T^{-2} - 3.710044 \times 10^{-5}T^2$<br>$C_p$ (505.08 K < $T$ < 800 K) = $40.032397 + 7.605276 \times 10^{-2}T + 4.170258 \times 10^{6}T^{-2} - 3.112401 \times 10^{-5}T^2 + 2.21526 \times 10^{27}T^{-10}$<br>$C_p$ (800 K < $T$ < 1811 K) = $80.4167 + 8.79504 \times 10^{-3}T - 1.54718 \times 10^{5}T^{-2} + 3.53562 \times 10^{-7}T^2 + 2.21526 \times 10^{27}T^{-10}$<br>$H°_{298}$ = -14750, $S°_{298}$ = 128.5 |
| $Fe_3Sn_2$ | $C_p$ (1 K < $T$ < 50 K) = $1.96 \times 10^{-2}T + 6.6 \times 10^{-3}T^2 + 1.0 \times 10^{-4}T^3$<br>$C_p$ (50 K < $T$ < 150 K) = $-23.162953 + 1.286666 \times 10^{0}T - 8.076442 \times 10^{3}T^{-2} - 3.253589 \times 10^{-3}T^2$<br>$C_p$ (150 K < $T$ < 250 K) = $51.084372 + 4.385531 \times 10^{-1}T - 1.400109 \times 10^{5}T^{-2} - 6.38662 \times 10^{-4}T^2$<br>$C_p$ (250 K < $T$ < 400 K) = $61.585963 + 2.462198 \times 10^{-1}T + 3.24163 \times 10^{5}T^{-2} - 1.55945 \times 10^{-4}T^2$<br>$C_p$ (400 K < $T$ < 505.08 K) = $103.4649 + 1.018659 \times 10^{-1}T^1 - 2.16314 \times 10^{5}T^{-2} - 3.63933 \times 10^{-5}T^2 - 2.28882 \times 10^{-5}T^{-0.5} + 1.83105 \times 10^{-4}T^1$<br>$C_p$ (505.08 K < $T$ < 800 K) = $346.009961 - 5.53886 \times 10^{-5}T - 1.308346 \times 10^{6}T^{-2} - 7.38 \times 10^{-6}T^2 - 8.954057 \times 10^{3}T^{-0.5} + 1.031169 \times 10^{5}T^{-1}$<br>$C_p$ (800 K < $T$ < 1811 K) = $130.4253 + 2.638512 \times 10^{-2}T - 4.64154 \times 10^{5}T^{-2} + 1.060686 \times 10^{-6}T^2 + 2.21526 \times 10^{27}T^{-10}$<br>$H°_{298}$ = -2100, $S°_{298}$ = 206 |
| $Fe_5Sn_3$ | $C_p$ (250 K < $T$ < 505.08 K) = $165.4545 + 1.571964 \times 10^{-1}T - 4.01830 \times 10^{5}T^{-2} - 5.441320 \times 10^{-5}T^2$<br>$C_p$ (505.08 K < $T$ < 800 K) = $142.348646 + 1.448618 \times 10^{-1}T + 5.713874 \times 10^{6}T^{-2} - 4.544854 \times 10^{-5}T^2 + 3.32289 \times 10^{27}T^{-10}$<br>$C_p$ (800 K < $T$ < 1811 K) = $202.9251 + 4.39752 \times 10^{-2}T - 7.73590 \times 10^{5}T^{-2} + 1.76781 \times 10^{-6}T^2 + 3.32289 \times 10^{27}T^{-10}$<br>$H°_{298}$ = 30,128.19, $S°_{298}$ = 356.0416 |

bcc: Body-centered cubic; CEF: Compound Energy Formalism; fcc: face-centered cubic; MQM: Modified Quasichemical Model.

in this study. In Figure 10B, the experimental data show inconsistencies with the assessments. In the experiments of Eremenko *et al.*[72], Fe and Sn show similar vapor pressure, and this might cause a deviation in the activity data. Yazawa and Koike[75] distributed Sn between Pb-Sn and Fe-Sn alloys, expecting slight mutual solubility between the two alloys. However, it is a practical problem during the experiment that Fe could be entrapped and suspended into the Pb melt. This phenomenon might be the reason that shows the deviation in the liquid activity since the activities diverged from fundamental activity behavior. In Figure 10C and D, the data of Nunoue and Kato[44] and Yamamoto *et al.*[42] at 1,600 °C are in good

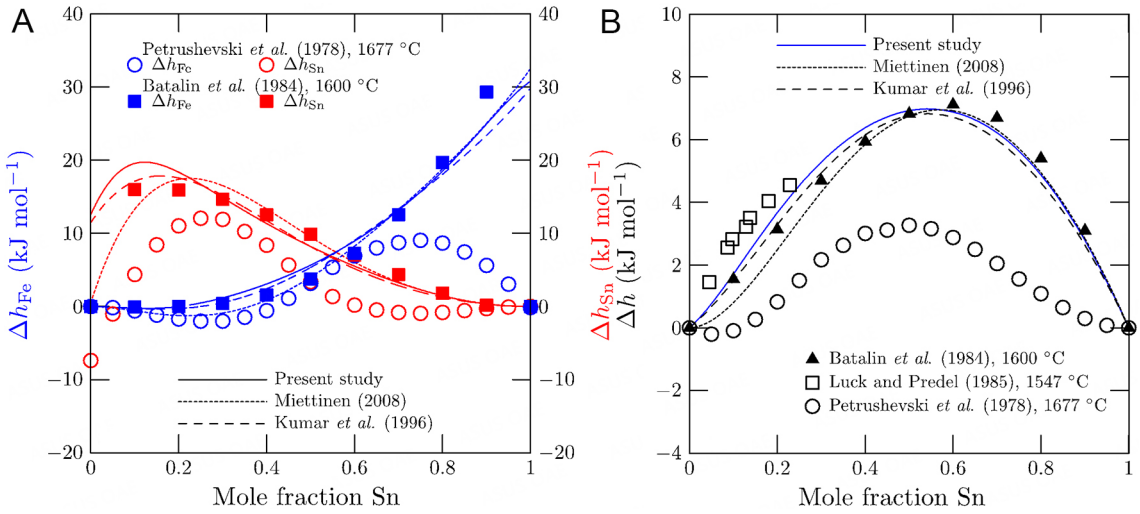


**Figure 9.** Enthalpy of mixing in the Fe-Sn alloys: (A) partial enthalpies of mixing and (B) integral enthalpy of mixing. Curves are the calculated enthalpies at 1,600 °C in the present study (solid), Kumar *et al.* (dashed)[19], and Miettinen (dotted)[20], along with experimental data[67-69]. Reference states are pure liquid Fe and Sn.

agreement with the calculated results. The data from Maruyama and Ban-Ya[71] and Wagner and St. Pierre[70] deviates from the other observations. In general, the calculations of $\Delta\mu_{Fe}$ reasonably agree with the overall activity data. Slightly lower values were calculated in the range of the equimolar composition, most likely related to the miscibility gap in the phase diagram.

### Thermodynamic properties of solid solutions and stoichiometric compounds

Activity data for the bcc phase was only reported by Arita *et al.*[43] by analyzing the concentration of Sn in the interface between bcc and Ag-Sn liquid alloy. The activity coefficient of the Sn in Ag is essential to obtain the activity data itself. The measured composition by Arita *et al.*[43] was adopted and the calculated activity coefficients at this composition were taken from Wang *et al.*[76] The re-calculation was carried out as follows:

$$a_{\mathrm{Sn(in\ Fe)}} = a_{\mathrm{Sn(in\ Ag)}} = \gamma_{\mathrm{Sn(in\ Ag)}} \times X_{\mathrm{Sn(in\ Ag)}} \tag{16}$$

where $a_{\mathrm{Sn(in\ Fe)}}$ and $a_{\mathrm{Sn(in\ Ag)}}$ denote the activity of Sn in bcc-Fe and liquid Ag, $\gamma_{\mathrm{Sn(Ag)}}$ is the activity coefficient of Sn in liquid Ag from Wang *et al.*[76], and $X_{\mathrm{Sn(in\ Ag)}}$ denotes the mole fraction Sn in liquid Ag. In Figure 11, the data from Arita[43] was recalculated, and this experimental data was compared with the present and previous studies' calculated results. As shown, the experimental data and the calculated results from all assessments at 1,200 °C accorded well.

Lafaye *et al.*[22] employed Density Functional Theory (DFT) calculations for enthalpy of formation ($\Delta h$) of compounds and solid solutions (fcc and bcc). The calculated enthalpy of mixing of solid solutions was compared with the DFT calculations from Lafaye *et al.*[22] in Figure 12A for bcc and Figure 12B for fcc. The $\Delta h$ data of fcc and bcc in DFT calculations at 0 K show the highest value at $X_{Sn}$ = 0.5, but these were asymmetric toward Sn-rich side. The present calculations (curves) were not fitted to these DFT data, but similar asymmetric nature was predicted after the assessment of the model parameters. Considering that the DFT calculations also provide predicted results, the present results may be regarded as reliable estimations. The data of Jannin *et al.* (concentration dependence of $T_{Cr}$ and $\beta_0$) for bcc phases[77] were not considered in the present study due to insufficient information.

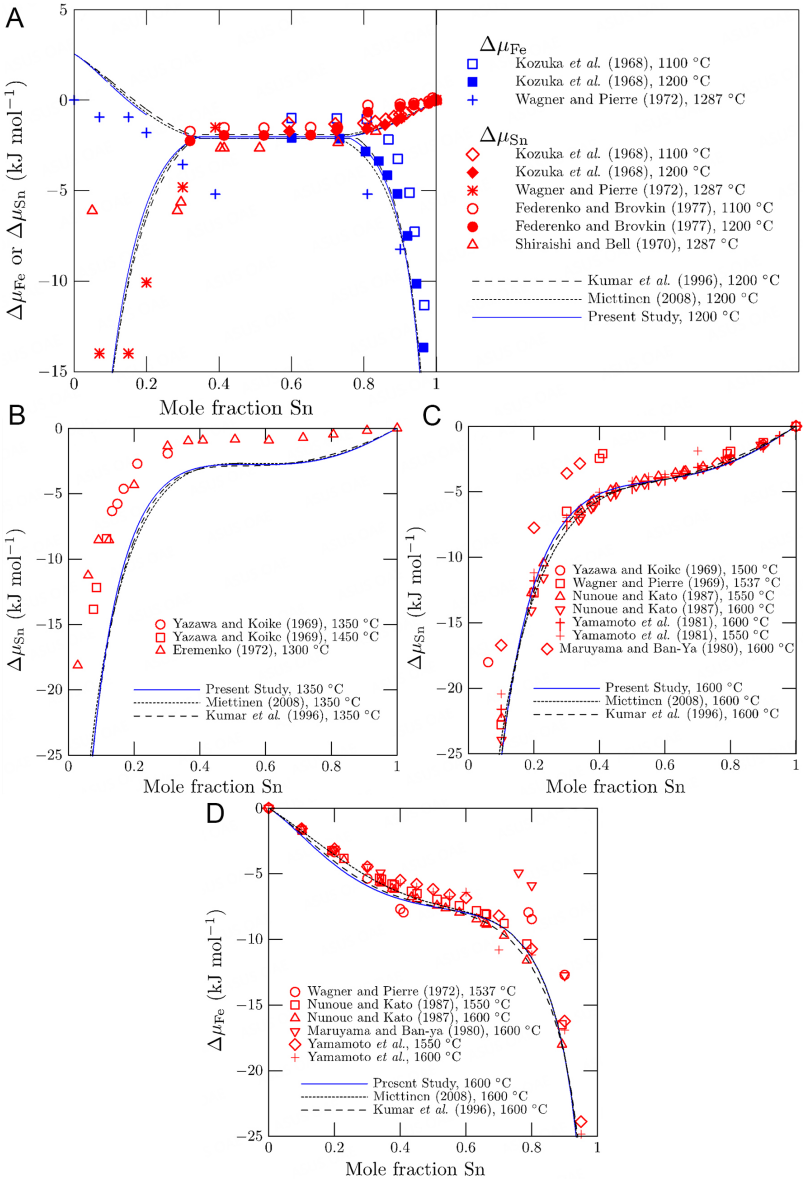


**Figure 10.** Relative chemical potential $\Delta\mu_M$ (= $RT$ ln $a_M$) of Fe and/or Sn at (A) 1,200 °C; (B) 1,350 °C; and (C); and (D) 1,600 °C along with experimental data[32,42,44,70-75] and thermodynamic assessments[19,20] from literature. Reference states are pure liquid Fe and Sn, except for (A) where pure fcc Fe was used.

Zabyr and Fitzner[78] measured the Gibbs energies of FeSn and $FeSn_2$ using the EMF method at 374 to 578 °C with $FeCl_2$ in $(KCl\text{-}LiCl)_{eut}$ as an electrolyte. Eremenko *et al.*[40] measured the Gibbs energy formation of FeSn and $FeSn_2$ compounds by equilibrium distribution of Fe-Pb-Sn and FeSn, and Fe-Pb-Sn and $FeSn_2$ at 350 to 600 °C. The heat capacities of FeSn and $Fe_3Sn_2$ were reported by Wu[79] and Sales[80], who measured the heat capacity from 0 to 380 K. The Néel temperature of FeSn was reported at 91.85 °C[80].

The heat capacity data of FeSn and $Fe_3Sn_2$ from Sales[80] and Wu[79] were adopted, as shown in Figure 13A and B. The heat capacity of $Fe_3Sn_2$ shows a similar trend in all calculations. The $C_p$ function of FeSn was

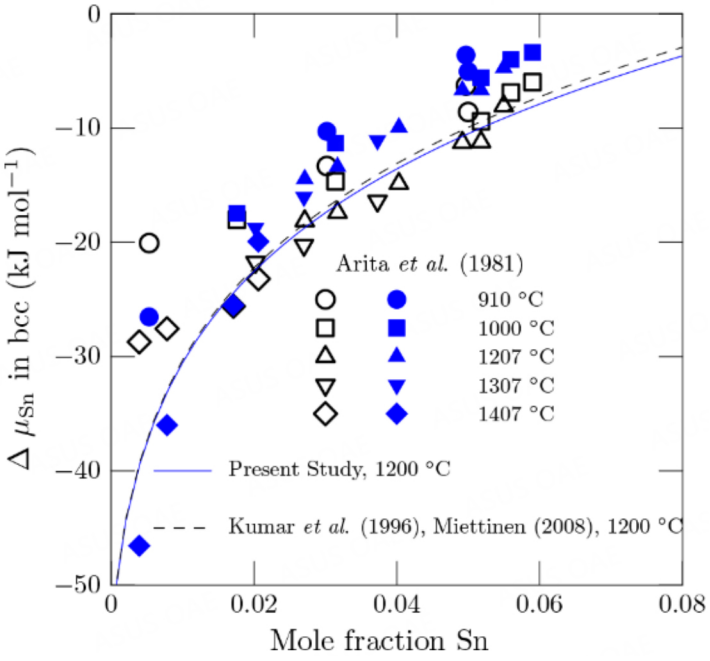


**Figure 11.** Relative chemical potential $\Delta\mu_{Sn}$ in the bcc phase, along with experimental data[43] and thermodynamic assessments from literature[19,20]. The reference state is pure liquid Sn. Open symbols were reported by Arita *et al.*[43], while closed symbols were recalculated using $\gamma_{Sn(in\ Ag)}$ of Wang *et al.*[76].

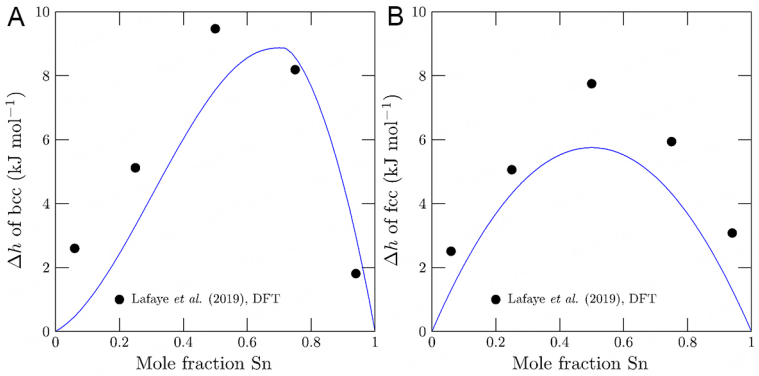


**Figure 12.** Enthalpy of mixing at 25 °C of solid solutions in the present study compared with DFT calculations[22]: (A) bcc and (B) fcc. bcc: Body-centered cubic; DFT: Density Functional Theory; fcc: face-centered cubic.

significantly improved, also demonstrated by the accurate prediction of the Néel temperature at 91.85 °C[80]. To evaluate the stability range in the phase diagram, the Gibbs energy information reported by Zabyr and Fitzner[78], see Figure 13C, was used in the present study. Previous assessments[19,20] evaluated the Gibbs energies of $FeSn_2$ and FeSn as somewhat negative, resulting in higher EMF values (dashed lines in Figure 13C). In the present study, these discrepancies were resolved. This impacts the solubility of Sn in bcc phase in equilibrium with these phases, which will be discussed later.

## Calculation of the phase diagram

A number of researchers[27,81-83] reported the general constitution of the Fe-Sn phase diagram. For the two-phase field of bcc and liquid on the Fe-rich side, Arita *et al.*[43] equilibrated Ag-Sn and Fe-Sn and analyzed the concentration of Sn with EPMA. Imai *et al.*[45] conducted the thermal analysis to determine the liquidus

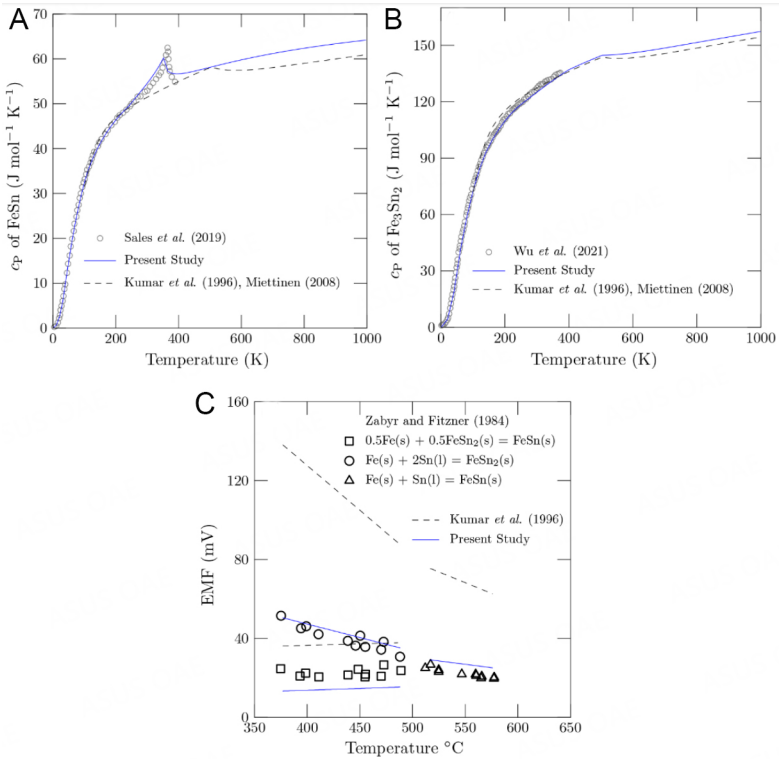


**Figure 13.** Thermodynamic properties of intermetallic compounds: (A) heat capacity of FeSn; (B) heat capacity of $Fe_3Sn_2$; and (C) EMF data relevant to FeSn and $FeSn_2$, plotted along with experimental data[78-80] and thermodynamic assessments from literature[19,20]. EMF: Electro Motive Force.

temperature from 1,474-1,512 °C and applied the melt equilibrium method at 1,450-1,500 °C to obtain the liquidus and solidus line on the Fe-rich side. At low temperatures (500-900 °C) of the Fe solvus line, Predel and Frebel[35], and Nageswararao *et al.*[36] measured the solubility limit for Sn in bcc from the lattice parameters measurement by X-ray diffraction. Yamamoto[41] investigated the solvus line of bcc and the γ-loop in the Fe-rich side using the diffusion couple method between 800-1,100 °C. Hillert[31] and Speight[34] measured the Sn solubility of bcc and fcc using EPMA. Kozuka *et al.*[32] estimated the liquidus line on the Sn-rich side by applying the EMF method from 900-1,100 °C. Eremenko *et al.*[40] determined the Sn-rich liquidus line by high-temperature melt equilibrium and quenching technique with the Sn addition in the liquid alloy in the temperature range of 270-1,130 °C. Also, Campbell[29] determined the Sn-rich liquids with the melt equilibrium technique.

The miscibility gap was measured by high-temperature melt equilibrium accompanied by quenching, and mass spectrometry method. Mills and Turkdogan[30], Shirashi and Bell[33], Gao *et al.*[46], and Campbell[29] measured the phase boundary of the miscibility gap at 1,125-1,550 °C by high-temperature melt equilibrium technique. Yamamoto *et al.*[42], Nunoue and Kato[44] and Wagner and St. Pierre[70] used mass spectrometric methods between 1,200 °C and 1,500 °C. Singh and Bhan[84] investigated the compounds (FeSn, $Fe_3Sn_2$ $Fe_{1.3}$ Sn) stability range with high-temperature X-ray diffraction (HT-XRD) at 498-923°C. Ehret and Westgren[28], Trehux and Guiraldenq[37], and Malaman[39] evaluated the stability range of compounds as phase diagram information. Connolly and McAllan[38] measured the eutectic reaction of $FeSn_2$ using resistance thermometry with a resolution of 0.1 mK.

**Table 4. Invariant reactions in the present Fe-Sn system**

| Invariant reaction | Reaction | Calculated temperature (°C) | Literature data (°C) |
|---|---|---|---|
| $Liquid_1 \rightarrow bcc + Liquid_2$ | Monotectic | 1,140 | 1,117[32]<br>1,134[19]<br>1,139 ± 5 [Present study] |
| $Liquid + bcc \rightarrow Fe_5Sn_3$ | Peritectic | 904 | 910[37]<br>895[19]<br>907 ± 1 [Present study] |
| $Liquid + Fe_5Sn_3 \rightarrow Fe_3Sn_2$ | Peritectic | 799 | 806[37]<br>804[84]<br>807[19] |
| $Fe_5Sn_3 \rightarrow bcc + Fe_3Sn_2$ | Eutectoid | 784 | 782[37]<br>750[41]<br>768[19] |
| $Fe_3Sn_2 \rightarrow FeSn + Liquid$ | Peritectic | 764 | 770[37]<br>761[19] |
| $Fe_3Sn_2 \rightarrow bcc + FeSn$ | Eutectoid | 617 | 607[37]<br>597[84]<br>607[19] |
| $FeSn + Liquid \rightarrow FeSn_2$ | Peritectic | 511 | 513[37]<br>513[19] |
| $Liquid \rightarrow FeSn_2 + Sn$ | Eutectic | 231 | 232.11[38]<br>232.06[19] |

Figure 1 shows the calculated phase diagram using the model in the present study. Currently obtained experimental results (DSC data for the monotectic reaction and the incongruent melting of $Fe_5Sn_3$) are plotted along with the literature data. Overall experimental data show good agreement with the calculation of the present study. The assessed phase diagrams in the previous studies[19,20] are compared in Figures 14 and 15 for the Fe-rich side, including the γ-loop, Sn-rich side, and miscibility gap in the liquid phase. In Figure 14A, the solubility of Sn in bcc in the range of 400-900 °C calculated in the present study showed better agreement compared with the results of Kumar *et al.*[19] and Miettinen[20], due to the less stable FeSn compound [Figure 13C]. DSC measurements carried out in the present study, bcc + $Fe_5Sn_3$ = bcc + Liquid and/or $Fe_5Sn_3$ + Liquid = bcc + Liquid phase transformation, showed more increased results and this was reproduced in the modeling results. In Figure 14B, the γ-loop in the Fe-rich region shows better agreement with the data of Speight[34]. In Figure 14C, the calculated Sn-rich liquidus shows favorable agreement with the available experimental data[29,32]. The description was also dependent on the description of the miscibility gap.

### Re-assessment of the miscibility gap

Figure 15 shows the enlarged section ($X_{Sn}$ = 0.30-0.90) in the Fe-Sn system, including the miscibility gap. The monotectic temperature from the DSC results was well reproduced in the present study. The slightly higher binodal temperature of two samples (I and V) may be attributed to the low phase fraction of Liquid or bcc beneath the monotectic temperature, yielding less change of the contact angle. The miscibility gap data reported in the literature showed inconsistent results, thereby less agreement with the previous assessments. In a mass spectrometry study[44], it might be difficult to detect noticeable ion intensity changes during cooling upon passing the binodal temperature. High-temperature melt equilibrium followed by quenching showed solid compounds precipitated in two-separated solidified liquids[30]. This was likely to induce inaccuracies in the composition analysis of each liquid phase. In the present study, the contact angle measurement could be a supplemental technique to identify the phase transition ($Liquid \rightarrow Liquid_1 + Liquid_2$). Therefore, the probable shape of the miscibility gap in the Fe-Sn binary liquid was elucidated. Subsequently, its description was assisted by the modeling using MQM, as shown in Figures 1 and 15. Invariant reactions calculated in the present study are listed in Table 4.

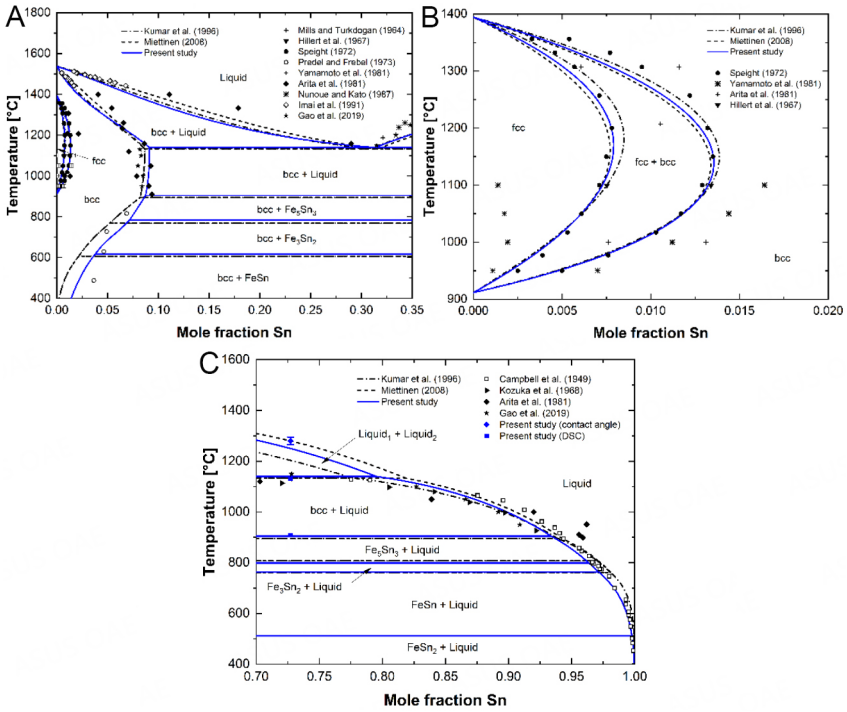


**Figure 14.** Enlarged sections of the Fe-Sn system: (A) Fe-rich region; (B) γ-loop; (C) Sn-rich region along with experimental data[29-32,34,35,41,43-46] and thermodynamic assessments[19,20] from literature. bcc: Body-centered cubic; DSC: Differential Scanning Calorimetry; fcc: face-centered cubic.

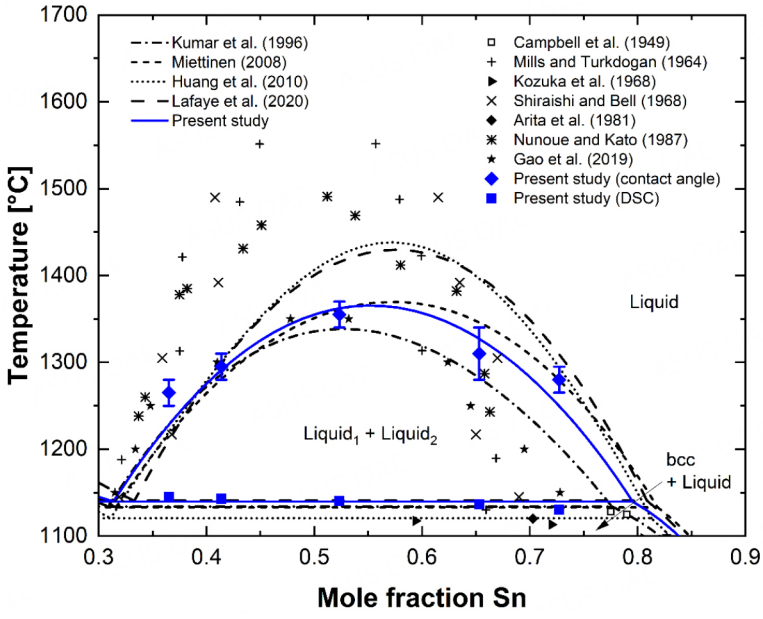


**Figure 15.** Enlarged section of the miscibility gap in the Fe-Sn system, along with experimental data from literature[29,30,32,33,43,44,46] and thermodynamic assessments[19-22]. bcc: Body-centered cubic; DSC: Differential Scanning Calorimetry.

In Figure 15, all the miscibility gap data reported in the literature[29,30,32,33,43,44,46] as well as those measured in the present study are shown, along with the available model calculations[19-22]. The present study was marked by a solid curve. It is seen that the present study and that of Miettinen[20] resulted in the best explanation for

the miscibility gap data reported in the present study. It is concluded that the previously reported "sharp-peaked" miscibility gap was in error. Kang and Pelton discussed that many immiscible alloys exhibit a "flattened" miscibility gap, which should be explained by considering non-random mixing (clustering) between atoms[24]. The MQM is a suitable model for treating non-random mixing. The BW random mixing model might be used, but with more adjustable model parameters. Furthermore, Kang and Pelton[24] demonstrated that the MQM performs well when extrapolating the binary miscibility gap to higher-order systems, which is essential in developing a multicomponent thermodynamic database for numerous tramp elements in steel.

## CONCLUSIONS

Thermodynamic modeling of the Fe-Sn binary system was carried out with new experimental investigations employing DSC, electromagnetic levitation technique, and contact angle measurement. The experiments provided key experimental data which helped the modeling of liquid alloy's Gibbs energy. The monotectic temperature ($Liquid_1 \rightarrow bcc + Liquid_2$) and binodal for the liquid miscibility gap were measured, which could resolve discrepancies reported in the literature. The liquid phase was modeled using the MQM in the pair approximation, which was known to be superior in modeling solutions exhibiting demixing tendencies. The following points were improved during the present thermodynamic modeling:

(1) Description of the previously controversial miscibility gap

(2) Gibbs energies of FeSn and $FeSn_2$

(3) Low temperature solubility of Sn in bcc

The developed database can be used as a part of a larger database for steel systems containing tramp elements such as Sn.

## DECLARATIONS

### Authors' contributions

Performed experimental works: Park WB, Bernhard M, Presoly P
Performed literature survey: Park WB, Bernhard M
Performed thermodynamic modeling Park WB, Bernhard M
Writing manuscript: Bernhard M, Park WB, Kang YB
Fund acquisition: Kang YB

### Availability of data and materials

Not applicable.

### Financial support and sponsorship

This research was supported by Basic Research program (NRF-2021R1F1A1049973), Brain Pool program (NRF-2022H1D3A2A01081708), both funded by the Ministry of Science and ICT through the National Research Foundation of Korea.

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.



## REFERENCES

1. The Swedish Steel Producers' Association. Environmental evaluation of steel and steel structures. Available from: https://www.jernkontoret.se/globalassets/publicerat/handbocker/stalkretsloppet_slutrapport_miljohandbok_engelsk_web.pdf [Last accessed on 16 Mar 2023].
2. Jégourel Y. The global iron ore market: from cyclical developments to potential structural changes. *Extr Ind Soc* 2020;7:1128-34. DOI
3. Zhang X, Ma G, Liu M, Li Z. Removal of residual element tin in the ferrous metallurgy process: a review. *Metals* 2019;9:834. DOI
4. Nachtrab WT, Chou YT. Grain boundary segregation of copper, tin and antimony in C-Mn steels at 900 °C. *J Mater Sci* 1984;19:2136-44. DOI
5. Kim SW, Lee HG. Effect of oxide scale formation on the behaviour of Cu in steel during high temperature oxidation in $O_2$-$N_2$ and $H_2O$-$N_2$ atmospheres. *Steel Res Int* 2009;80:121-9. DOI
6. Yin L, Sridhar S. Effects of residual elements arsenic, antimony, and tin on surface hot shortness. *Metall and Materi Trans B* 2011;42:1031-43. DOI
7. Shubhank K, Kang Y. Critical evaluation and thermodynamic optimization of Fe-Cu, Cu-C, Fe-C binary systems and Fe-Cu-C ternary system. *Calphad* 2014;45:127-37. DOI
8. Melford DA. The influence of residual and trace elements on hot shortness and high temperature embrittlement. *Phil Trans R Soc Lond A* 1980;295:89-103. DOI
9. Yu Y, Li L, Wang J. Sn recovery from a tin-bearing middling with a high iron content and the transformation behaviours of the associated As, Pb, and Zn. *Sci Total Environ* 2020;744:140863. DOI PubMed
10. Su Z, Zhang Y, Liu B, Lu M, Li G, Jiang T. Extraction and separation of tin from tin-bearing secondary resources: a review. *JOM* 2017;69:2364-72. DOI
11. Bunnakkha C, Jarupisitthorn C. Extraction of tin from hardhead by oxidation and fusion with sodium hydroxide. *J Met Mater Miner* 2012;22:1-6.
12. Lee S, Lee M, Kim HY. Recovery of high purity Sn by multi-step reduction of Sn-containing industrial wastes. *J Korean Inst Resour Recyc* 2015;24:11-5. DOI
13. Spencer P. A brief history of CALPHAD. *Calphad* 2008;32:1-8. DOI
14. Sundman B, Lukas HL, Fries SG. Computational thermodynamics: the Calphad method. 1st ed. Cambridge University Press; 2007. pp. 1-296. DOI
15. Bale C, Bélisle E, Chartrand P, et al. FactSage thermochemical software and databases, 2010-2016. *Calphad* 2016;54:35-53. DOI
16. Andersson J, Helander T, Höglund L, Shi P, Sundman B. Thermo-Calc & DICTRA, computational tools for materials science. *Calphad* 2002;26:273-312. DOI
17. Cao W, Chen S, Zhang F, et al. PANDAT software with PanEngine, PanOptimizer and PanPrecipitation for multi-component phase diagram calculation and materials property simulation. *Calphad* 2009;33:328-42. DOI
18. Nüssler H, von Goldbeck O, Spencer P. A thermodynamic assessment of the iron-tin system. *Calphad* 1979;3:19-26. DOI
19. Kumar K, Wollants P, Delaey L. Thermodynamic evaluation of Fe-Sn phase diagram. *Calphad* 1996;20:139-49. DOI
20. Miettinen J. Thermodynamic description of the Cu-Fe-Sn system at the Cu-Fe side. *Calphad* 2008;32:500-5. DOI
21. Huang YC, Gierlotka W, Chen SW. Sn-Bi-Fe Thermodynamic modeling and Sn-Bi/Fe interfacial reactions. *Intermetallics* 2010;18:984-91. DOI
22. Lafaye P, Toffolon-masclet C, Crivello J, Joubert J. Thermodynamic modelling of the Fe-Sn-Zr system based on new experiments and first-principles calculations. *J Alloys Compd* 2020;821:153200. DOI
23. Hillert M. The compound energy formalism. *J Alloys Compd* 2001;320:161-76. DOI
24. Kang YB, Pelton AD. The shape of liquid miscibility gaps and short-range-order. *J Chem Thermodyn* 2013;60:19-24. DOI
25. Pelton AD, Degterov SA, Eriksson G, Robelin C, Dessureault Y. The modified quasichemical model I - binary solutions. *Metall Mater Trans B* 2000;31:651-9. DOI
26. Pelton AD, Chartrand P. The modified quasi-chemical model: Part II. Multicomponent solutions. *Metall Mater Trans A* 2001;32:1355-60. DOI
27. Okamoto H. Phase diagrams of binary iron alloys. Metals park, Ohio: American Society for Metals; 1993:385-92.
28. Ehret WF, Westgren AF. X-ray analysis of iron-tin alloys. *J Am Chem Soc* 1933;55:1339-51. DOI
29. Campbell AN, Wood JH, Skinner GB. The system iron-tin: liquidus only. *J Am Chem Soc* 1949;71:1729-33. DOI

30. Mills KC, Turkdogan ET. Liquid miscibility gap in iron-tin system. *Trans Metall Soc AIME* 1964;230:1202-3.
31. Hillert M, Wada T, Wada H. The alpha-gamma equilibrium in Fe-Mn, Fe-Mo, Fe-Ni, Fe-Sb, Fe-Sn and Fe-W systems. *J Iron Steel Inst* 1967;205:539-46.
32. Kozuka Z, Shidahara Y, Sugimoto E, et al. Thermodynamic study of hardhead (tin-iron alloy). *Nippon Kogyo Kaishi* 1968;84:1657-62. DOI
33. Shiraishi SY, Bell HB. Miscibility gap in liquid iron-tin alloys. *Trans Inst Min Metall Sect C* 1968;77:104-5.
34. Speight EA. The gamma loop in the iron-Tin system. *Met Sci J* 1972;6:57-60. DOI
35. Predel B, Frebel M. Precipitation behavior of α-solid solutions of the Fe-Sn system. *Metall Trans* 1973;4:243-9. DOI
36. Nageswararao M, Mcmahon CJ, Herman H. The solubility and solution behavior of antimony and tin in α-lron and the effects of nickel and chromium additions. *Metall Trans B* 1974;5:1061-8. DOI
37. Treheux D, Guiraldenq P. Etude des diagrammes d'equilibre binaires par la methode des couples de diffusion application au systeme fer-etain. *Scr Metall* 1974;8:363-6. DOI
38. Connolly J, Mcallan J. The tin-iron eutecticL'eutectique etain-ferDas Zinn-eisen-eutektikum. *Acta Metallurgica* 1975;23:1209-14. DOI
39. Malaman B, Roques B, Courtois A, Protas J. Structure cristalline du stannure de fer $Fe_3Sn_2$. *Acta Crystallogr B Struct Sci* 1976;32:1348-51. DOI
40. Eremenko VN, Churakov MM, Pechentkovskaya LE. Conditions of stannide formation during the interaction of Fe with a Sn-Pb melt and their thermodynamic properties. *Russ Metall* 1976;4:58-62.
41. Yamamoto T, Takashima T, Nishida K. Interdiffusion in the α-solid solutions of the Fe-Sn system. *J Jpn Inst Met* 1981;45:985-90. DOI
42. Yamamoto M, Mori S, Kato E. Mass spectrometric study of the thermodynamic properties of liquid Fe-Sn, Fe-Sn-Cu alloys. *Tetsu-to-Hagane* 1981;67:1952-61. DOI
43. Arita M, Ohyama M, Goto KS, Someno M. Measurements of activity, solubility, and diffusivity in α and γ Fe-Sn alloys between 1183 and 1680 K. *Int J Mater Res* 1981;72:244-50. DOI
44. Nunoue S, Kato E. Mass spectrometric determination of the miscibility gap in the liquid Fe-Sn system and the activities of this system at 1550 °C and 1600 °C. *Tetsu-to-Hagane* 1987;73:868-75. DOI
45. Imai N, Tanaka T, Yuki T, Iida T, Morita Z. Equilibrium distribution of Sn between solid and liquid phases in Fe-Sn and Fe-C-Sn alloys. *Tetsu-to-Hagane* 1991;77:224-30. DOI
46. Gao J, Li C, Guo C, Du Z. Investigation of the stable and the metastable liquidus miscibility gaps in Fe-Sn and Fe-Cu binary systems. *Int J Miner Metall Mater* 2019;26:1427-35. DOI
47. Bernhard M, Fuchs N, Presoly P, Angerer P, Friessnegger B, Bernhard C. Characterization of the γ-loop in the Fe-P system by coupling DSC and HT-LSCM with complementary in-situ experimental techniques. *Mater Charact* 2021;174:111030. DOI
48. Bernhard M, Presoly P, Fuchs N, Bernhard C, Kang Y. Experimental study of high temperature phase equilibria in the iron-rich part of the Fe-P and Fe-C-P systems. *Metall Mater Trans A* 2020;51:5351-64. DOI
49. Bernhard M, Presoly P, Bernhard C et al. An assessment of analytical liquidus equations for Fe-C-Si-Mn-Al-P-alloyed steels using DSC/DTA techniques. *Metall Mater Trans B* 2021;52:2821-30. DOI
50. Kim DI, Abbaschian R. The metastable liquid miscibility gap in Cu-Co-Fe alloys. *J Phase Equilibria Diffus* 2000;21:25-31. DOI
51. Min S, Park J, Lee J. Surface tension of the 60% Bi-24% Cu-16%Sn alloy and the critical temperature of the immiscible liquid phase separation. *Maters Lett* 2008;62:4464-6. DOI
52. Lee D, Cho Y, Kim JH, Hwang I, Chung Y, Kang Y. Application of k-means clustering to material research: measurement of layer thickness and contact angle. *Met Mater Int* ;2023:1-12. DOI
53. Lee S. Comparison of initial seeds methods for K-means clustering. *J Internet Comput Serv* 2012;13:1-8. DOI
54. Morissette L, Chartier S. The k-means clustering technique: general considerations and implementation in Mathematica. *Tutor Quant Methods Psychol* 2013;9:15-24. DOI
55. Boettinger WJ, Kattner UR, Moon K, Perepezko JH. DTA and heat-flux DSC measurements of alloy melting and freezing. In: Zhao ZC, editor. Methods for phase diagram determination. Amsterdam: Elsevier Science; 2006. pp. 151-205. DOI
56. Barin I. Thermochemical data of pure substances. Part I and Part II. NewYork: Verlag Chemie; 1989, pp. 1392. DOI
57. Humenik M, Kingery WD. Metal-ceramic interactions: III, surface tension and wettability of metal-ceramic systems. *J Am Ceramic Soc* 1954;37:18-23. DOI
58. Chidambaram PR, Edwards GR, Olson DL. A thermodynamic criterion to predict wettability at metal- alumina interfaces. *Metall Mater Trans B* 1992;23:215-22. DOI
59. Kapilashrami E, Jakobsson A, Seetharaman S, Lahiri AK. Studies of the wetting characteristics of liquid iron on dense alumina by the X-ray sessile drop technique. *Metall and Materi Trans B* 2003;34:193-9. DOI
60. Nikolopoulos P. Surface, grain-boundary and interfacial energies in Al2O3 and Al2O3-Sn, Al2O3-Co systems. *J Mater Sci* 1985;20:3993-4000. DOI
61. Pelton AD, Kang Y. Modeling short-range ordering in solutions. *Int J Mater Res* 2007;98:907-17. DOI
62. Tafwidli F, Kang Y. Thermodynamic modeling of Fe-C-S ternary system. *ISIJ Int* 2017;57:782-90. DOI
63. Pelton AD, Blander M. Thermodynamic analysis of ordered liquid solutions by a modified quasichemical approach - application to silicate slags. *Metall Mater Trans B* 1986;17:805-15. DOI
64. Hillert M, Jarl M. A model for alloying in ferromagnetic metals. *Calphad* 1978;2:227-38. DOI

65. Dinsdale A. SGTE data for pure elements. *Calphad* 1991;15:317-425. DOI
66. Gustafson P. A thermodynamic evaluation of Fe-C system. *Scand J Metall* 1985;14:259-67.
67. Batalin GI, Sudavtsova VS, Kurach VP. Thermodynamic properties of liquid Fe-Sn alloys. *Izv Akad Nauk SSSR* 1984;4:50-1.
68. Petrushevskiy MS, Esin YuO, Bayev VM, et al. Influence of short-range ordering on the concentration dependence of the enthalpies of formation of liquid of iron-tin alloys. *Russ Metall* 1978;1:61-3.
69. Lück R, Predel B. The enthalpy of mixing of liquid iron-tin alloys determined by means of a new high-temperature calorimeter. *Z Metallkd* 1985;76:684-6.
70. Wagner S, St. pierre GR. Thermodynamics of the liquid binary iron-tin by mass spectrometry. *Metall Trans B* 1972;3:2873-8. DOI
71. Maruyama N, Ban-Ya S. Measurement of activities in liquid Fe-Cu, Fe-Cr and Fe-Sn alloys by a transportation method. *J Japan Inst Met Mater* 1980;44:1422-31. DOI
72. Eremenko VN, Lukashenko GM, Pritula VL. Thermodyanmic properties of Fe-Sn melts. *Russ Metall* 1972;1:72-5.
73. Fedorenko AN, Brovkin VG. Vapor pressure of tin and thermodynamic properties of the tin and iron system. *Sb Nauchn Tr-Gos Proektn Nauchno-Issled Inst* 1977;3:83-9.
74. Shiraishi SY, and Bell HB. Thermodynamic study of tin smelting. PT. 1. Iron-tin and iron-tin-oxygen alloys. *Inst Mining Met Trans Sect C* 1970;79:C120-7.
75. Yazawa A, Koike K. Tin smelting. II. Activity measurements in molten tin-iron alloy. *Nippon Kogyo Kaishi* 1969;85:39-42. DOI
76. Wang J, Hudon P, Kevorkov D, Chartrand P, Jung I, Medraj M. Thermodynamic and experimental study of the Mg-Sn-Ag-in quaternary system. *J Phase Equilib Diffus* 2014;35:284-313. DOI
77. Jannin C, Michel A, Lecocq P. Magnetism and properties of different phases in the Fe-Sn system. *Comptes Redus Hebomadaires Seances Acad Sci* 1963;257:1906-7.
78. Zabyr L, Fitzner K. Gibbs free energy of formation of iron antimonide ($FeSb_2$), iron-tin (FeSn), and iron-tin ($FeSn_2$) intermetallic phases. *Arch Hutn* 1984;29:227-33.
79. Wu P, Song J, Yu X, et al. Evidence of spin reorientation and anharmonicity in kagome ferromagnet $Fe_3Sn_2$. *Appl Phys Lett* 2021;119:082401. DOI
80. Sales BC, Yan J, Meier WR, Christianson AD, Okamoto S, Mcguire MA. Electronic, magnetic, and thermodynamic properties of the kagome layer compound FeSn. *Phys Rev Mater* 2019;3:1-8. DOI
81. Kubaschewski O. Iron binary phase diagrams. 1st ed. Berlin Heidelberg: Springer Science & Business Media; 1982, pp.139-42. DOI
82. Hultgren R, Desai PD, Hawkins DT, et al. Selected values of the thermodynamic properties of binary alloys. Metals Park, Ohio: American Society for Metals; 1973, pp. 884-7. DOI
83. Hansen M, Anderko K. Constitution of binary alloys. New York: McGraw-Hill; 1958, pp.718-20.
84. Singh M, Bhan S. Contribution to the Fe - Sn system. *J Mater Sci Lett* 1986;5:733-5. DOI

Gao *et al. J Mater Inf* 2023;3:6
**DOI:** 10.20517/jmi.2023.03

**Journal of Materials Informatics**

**Research Article**

**Open Access**

# Development of an accurate "composition-process-properties" dataset for SLMed Al-Si-(Mg) alloys and its application in alloy design

**Tianchuang Gao[1], Jianbao Gao[1,*], Jinliang Zhang[2], Bo Song[2], Lijun Zhang[1,*]**

[1]State Key Laboratory of Powder Metallurgy, Central South University, Changsha 410083, Hunan, China.
[2]State Key Laboratory of Materials Processing and Die & Mould Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China.

[*]**Correspondence to:** Dr. Jianbao Gao, State Key Laboratory of Powder Metallurgy, Central South University, Lushannan Road No. 932, Changsha 410083, Hunan, China. E-mail: jianbao.gao@csu.edu.cn; Prof. Lijun Zhang, State Key Laboratory of Powder Metallurgy, Central South University, Lushannan Road No. 932, Changsha 410083, Hunan, China. E-mail: lijun.zhang@csu.edu.cn

**How to cite this article:** Gao T, Gao J, Zhang J, Song B, Zhang L. Development of an accurate "composition-process-properties" dataset for SLMed Al-Si-(Mg) alloys and its application in alloy design. *J Mater Inf* 2023;3:6. https://dx.doi.org/10.20517/jmi.2023.03

**Received:** 18 Jan 2023 **First Decision:** 13 Feb 2023 **Revised:** 19 Feb 2023 **Accepted:** 17 Mar 2023 **Published:** 28 Mar 2023

**Academic Editors:** Xiao-Dong Xiang, Sergei V Kalinin, Xingjun Liu **Copy Editor:** Ke-Cui Yang **Production Editor:** Ke-Cui Yang

## Abstract

Al-Si-Mg series alloys are the most common alloys available for additive manufacturing forming with low cracking tendency. However, there is no systematic study on the computational design of SLMed Al-Si-(Mg) alloys due to the huge parameter space of composition and processes. In this paper, a high-quality dataset of SLMed Al-Si-(Mg) alloys containing 176 pieces of data from 50 publications was first established, which recorded the information, including alloy compositions, process parameters, test conditions, and mechanical properties. A threshold value of 35 J/mm$^3$ for energy density (Ed) was then proposed as a criterion to clean the data points with lower ultimate tensile strength (UTS) and elongation (EL). The cleaned dataset consists of a first training/testing dataset with 142 data for model construction and a second testing dataset with 9 data for model verification. After that, four machine learning models were applied to establish the quantitative relation of "composition-processes-properties" in SLMed Al-Si-(Mg) alloys. The MLPReg model was chosen as the optimal one considering its best performance and subsequently utilized to design novel compositions and process parameters for SLMed Al-Si-(Mg) alloys. The UTS and EL of the designed alloy with a maximum comprehensive mechanical property are 549 MPa and 16%, both of which are higher than all the available experimental data. It is anticipated that the present design strategy based on the machine learning method should generally be applicable to other SLMed alloy systems.

**Keywords:** Selective laser melting, Al-Si-Mg alloy, machine learning, alloy design, mechanical properties

## INTRODUCTION

Due to their low density, high specific strength, high specific stiffness, and good plasticity, aluminum alloys are the preferred lightweight structure materials in aerospace, automotive, ships, and other fields[1]. The rapid development of marine, aerospace, and automotive transportation has put forward strict requirements for the performance of alloys. Selective laser melting (SLM) forming technology, as a promising additive manufacturing (AM) technology[2], has the advantages of a fast cooling rate and grain refinement. It can improve the mechanical properties of alloys and prepare parts with complex shapes, which greatly broadens the application range of aluminum alloys. Among different types of aluminum alloys, the Al-Si-(Mg) (i.e., Al-Si and Al-Si-Mg) series alloys are one of the few aluminum alloy systems suitable for the additive manufacturing process, including Al4Si[3,4], Al12Si[5-7], AlSi10Mg[8-10], and AlSi7Mg[11,12], etc. In particular, alloys with near-eutectic compositions over a very narrow solidification temperature range of 40K~50K can greatly reduce the risk of cracking during the laser additive manufacturing process and enable the preparation of nearly fully dense products, thus receiving extensive research attention.

The mechanical properties of SLMed Al-Si-(Mg) alloys are closely related to their composition and complex processing parameters. For instance, 0.2~0.4 wt.% Mg content may enhance the strength of alloys through precipitation hardening due to the precipitation of the fine $Mg_2Si$ phase during the aging heat treatment or the AM process[9,11,13]. Furthermore, the specific process parameters of the SLM process significantly affect the alloy properties, including laser power[14-16], scanning speed[16,17], powder layer thickness[15], scanning spacing[18,19], build direction[20-22], scanning strategy[23,24], and post heat-treatment processes[13,11,20]. Though a number of experimental investigations have been devoted to different SLMed Al-Si-(Mg) alloys in the literature, the factors that affect the properties and performance of alloys are too many, and thus there is still no systematic study on the relation among the "composition-process-properties" of the SLMed Al-Si-(Mg) alloys. Thus, there is an urgent need to solve this problem because it may block the efficient design of high-performance SLMed Al-Si-(Mg) alloys over high-dimensional parameter space.

To solve this problem, a variety of computational methods at different scales, including CALculation of PHAse Diagram (CALPHAD)[25-30], phase-field modeling[31-33], finite element simulation[34], and machine learning (ML)[28,29,35,36] can be utilized. Among them, ML is one of the most efficient computational methods. It can be utilized to establish the quantitative relation of "composition-processes-properties", and even accelerate the design of high-performance alloys over a high dimensional parameter space. Mondal *et al.*[37] employed a physical information-based machine learning method to systematically investigate the cracking mechanism of SLMed 6061Al, 2024Al, and AlSi10Mg alloys. The decision trees, support vector machines, and logistic regression techniques were used to predict crack formation conditions, and the cracking susceptibility maps were established for optimizing the process parameters. Yu *et al.*[38] used AdaBoost, gradient tree boosting, K-nearest neighbors, decision tree, and Extra Trees regressors to successfully predict the hardness and relative mass density of LPBFed CNTs/AlSi10Mg nanocomposites with cellular structure features as input. The relative errors of the predicted hardness and relative mass density due to the optimal model are as low as 3.61% and 1.42%, respectively. He *et al.*[39] applied a gaussian process regression-based machine learning approach to establish a processing window for a high-density additive manufactured 2-vol% TiCN reinforced AlSi10Mg composite. Though the ML approach has been widely used to design process parameters for SLMed Al alloys, there is still a lack of reports on the establishment of a high-quality "composition-process-properties" dataset of the basic SLMed Al-Si-(Mg) alloys using the ML approach. Such a situation urgently needs to be improved.

Consequently, based on the ML approach together with an exhausting collection of literature data, an accurate "composition-process-properties" dataset for the basic but technically important SLMed Al-Si-(Mg) alloys is to be developed, and then applied to design the SLMed Al-Si-(Mg) alloys with better mechanical properties than the literature reports. In the next section, a brief introduction to the ML approach is given, followed by the data collection and pre-processing of the dataset, including the data cleaning and feature analysis. After that, four ML models are employed to establish the quantitative relation of "composition-processes-properties" in the SLMed Al-Si-(Mg) alloys, and the optimal one for best reproducing the training and testing sets is to be selected. The selected ML model is subsequently applied to discover the novel compositions and processing parameters of as-built Al-Si-(Mg) alloys with better mechanical properties than the literature reports. Finally, a conclusion of this paper is drawn.

## MACHINE LEARNING APPROACH

The machine learning approach can establish an intrinsic relation between the input and output layers by systematically analyzing each characteristic variable in the data, thus realizing the prediction of new data. The basic ML workflow in alloy design mainly includes data collection, data cleaning, feature analysis, model selection, model training, model verification, and prediction analysis. In this paper, four machine learning models were built and comprehensively compared with each other to select the one with the highest accuracy, including linear regression (LinearReg), multi-layer perceptron regression (MLPReg), random forest regression (RFReg), and k-nearest neighbors (K-NN) regression.

LinearReg is widely utilized to construct a linear relationship between the target values and the characteristic variables by adjusting the regression coefficients. MLPReg is an artificial neural network that has a single input layer, one or more hidden layers, and a single output layer of perceptron, which is constituted by numerous neurons. The quantified relation between target values and variables is obtained by different weights and deviations in each neuron. The RFReg model consists of multiple mutually unrelated decision trees, each of which yields a prediction result from randomly selected samples and features, and the prediction result is obtained by combining the results of all trees and taking the average. The model has the function of calculating the importance of features. When the K-NN regressor is used for prediction, the mean value of the nearest data point is chosen as the prediction value.

In order to avoid the order of magnitude difference between each dimension of the dataset, resulting in inaccurate prediction results, all the variables need to be standardized before constructing the model. The data then obey a normal distribution with a mean of 0 and a variance of 1 (i.e., standard normal distribution). The equation of standardization treatment is given as

$$Z = (X - \mu)/\sigma \tag{1}$$

where $Z$ is the return value, $\mu$ is the mean of the training samples, and $\sigma$ is the standard deviation of the training samples.

Two metrics methods were utilized to evaluate the quality of machine learning models, i.e., the mean absolute error (MAE) and the coefficient of determination ($R^2$). The MAE measures the relative magnitude of deviation, while the $R^2$ can be used to characterize the fitness level of the model. They are respectively defined as

$$MAE\ (y, y') = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2 \quad (2)$$

$$R^2(y, y') = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

In Equations (2) and (3), $y_i'$ is the prediction value of the $i^{th}$ data, $y_i$ is the actual value corresponding to it, while $n$ is the size of the dataset.

## DATA COLLECTION AND PRE-PROCESSING

### Data collection

The mechanical properties [including ultimate tensile strength (UTS), yield strength (YS), and elongation (EL)] of the SLMed alloys do not only depend on the compositions[40-47] and processes[48-54] but also the testing condition[55-59]. Therefore, in order to study the influencing factors affecting the mechanical properties of SLMed Al-Si-(Mg) alloy, the compositions[60-62] (i.e., Si and Mg contents), the manufacturing processes (including laser power[63-65], scanning speed[66,67], scanning spacing[68-70], powder layer thickness[71,72], hatching space[73], and rotation angle[74]), and the testing direction[75,76] should be considered at the same time. However, it is quite challenging to analyze data in high-dimensional characteristic variable space, and the dimension of the characteristic variable thus needs to be reduced. The energy density (Ed)[77-79] is commonly used to describe the combined effect of laser power, scanning speed, scanning spacing, powder layer thickness, and hatching space on alloy properties. Ed can be calculated by using the following equation,

$$Ed = P/(v \times h \times t) \quad (4)$$

where $P$ is laser powder, $v$ is scan speed, $h$ is hatch space, and $t$ is layer thickness. Besides, the testing direction and the rotation angle should also be considered during data analysis. The testing direction describes the test situation. The tensile direction is regarded as 0° when it is parallel to the substrate plane, while 90° when it is vertical to the substrate plane. Moreover, the rotation angle represents the successive layer rotation during the additive manufacturing process.

In order to make the dataset more reliable and accurate, the following guidelines were obeyed during the data collection:

(*i*) Only the data of Al-Si-(Mg) alloys manufactured by SLM techniques were collected;

(*ii*) The data reported in the literature must be complete, and all the information on the composition, process parameters, and mechanical properties should be included;

(*iii*) If no actual alloy composition was given in the literature, the nominal composition of the alloy was used instead.

The data sets of as-built SLMed Al-Si-(Mg) alloys, which include 176 pieces of data from 50 publications, were established in the present work and displayed in the Supplementary Materials. Figure 1A and B shows the distribution of UTS, YS, and EL for different SLMed Al-Si-(Mg) alloys, mainly including Al4Si[3,4], Al7SiMg[12,13,24,49-60], AlSi10Mg[9-11,61-79], and Al12Si[4,5,14,40-48]. It is clearly observed in the figure that the strength

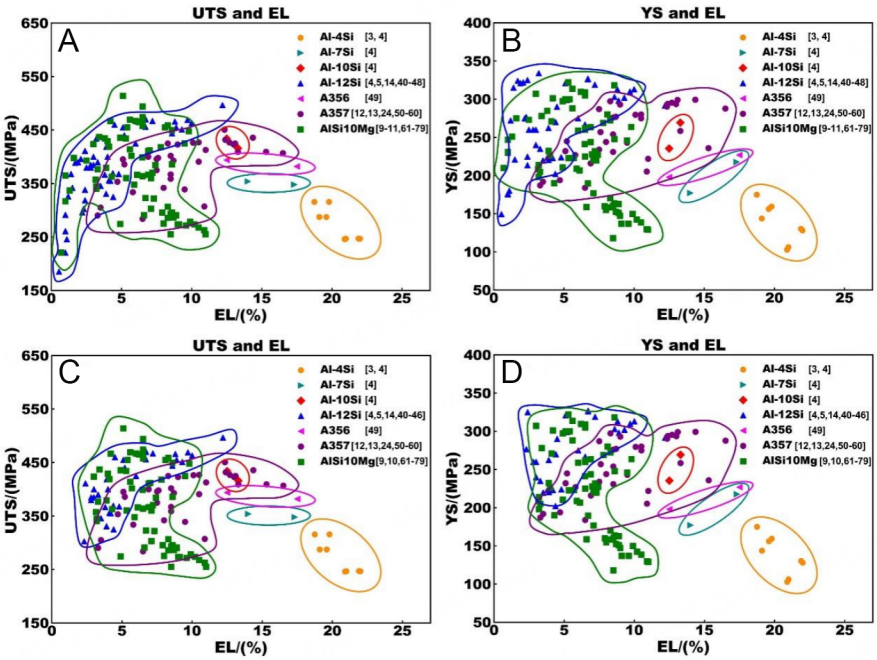


**Figure 1.** Distribution of UTS-EL and YS-EL data in different SLMed Al-Si-(Mg) alloys including Al4Si, Al7Si, Al10Si, Al12Si, A356, A357, and AlSi10Mg: (A) and (B) raw data; (C) and (D) data with cleaning. EL: elongation; UTS: ultimate tensile strength; YS: yield strength.

of the alloys tends to increase as the Si content increases while the elongation of the alloys decreases. The maximum UTS value is 514MPa from the AlSi10Mg (wt.%) alloy[77], while the maximum EL value is 22% from the Al4Si (wt.%) alloy[4]. However, there is a region with both poor strength and elongation, which is the lower left corner in Figure 1A. The detailed statistical results of all data are shown in Supplementary Table 1. It shows the statistical results of the composition of each element, Ed, rotation angle, tensile direction, UTS, YS, and EL among the 176 data. The range of Ed is from 13.33 J/mm$^3$ to 176.19 J/mm$^3$. The rotation angle mainly includes 45°, 67°, 73°, and 90°.

### Data cleaning

In order to precisely design alloys with excellent performance, the data with relatively high mechanical properties need to be retained as much as possible while avoiding the impact of low-performance data. Before constructing the model, a data cleaning by abandoning the data with lower mechanical properties is thus needed. The cleaned datasets are detailly shown in Figure 1C and D and Supplementary Table 2 and are to be used for constructing and validating the ML model. As shown in Figure 1C and D, the data for Al12Si and AlSi10Mg alloys with both poor strength and ductility were cleaned according to their lower Ed (≤ 35 J/mm$^3$). Finally, the cleaned dataset includes a first training/testing dataset with 142 data for model construction and a second testing dataset with 9 data for model verification.

To investigate the key factors affecting the mechanical properties of the SLMed Al-Si-(Mg) alloys, three main characteristic variables (i.e., Ed, layer rotation, and test direction) and their effect on the properties in three representative Al-Si-(Mg) alloys (i.e., Al12Si, AlSi10Mg, and Al7SiMg) were systematically analyzed. As shown in Figure 2A and B, the Ed is one of the most critical parameters in the fabrication process and also plays a key role in mechanical properties. The results indicate that the alloy with lower Si content requires higher Ed, which may be caused by the lower energy absorption of Al. Interestingly, there are also some very limited cases for the AlSi10Mg alloy with higher UTS and EL fabricated by high Ed (≥ 130 J/mm$^3$) , which needs further consideration and verification. For the Al12Si and AlSi10Mg alloys, most of them tend

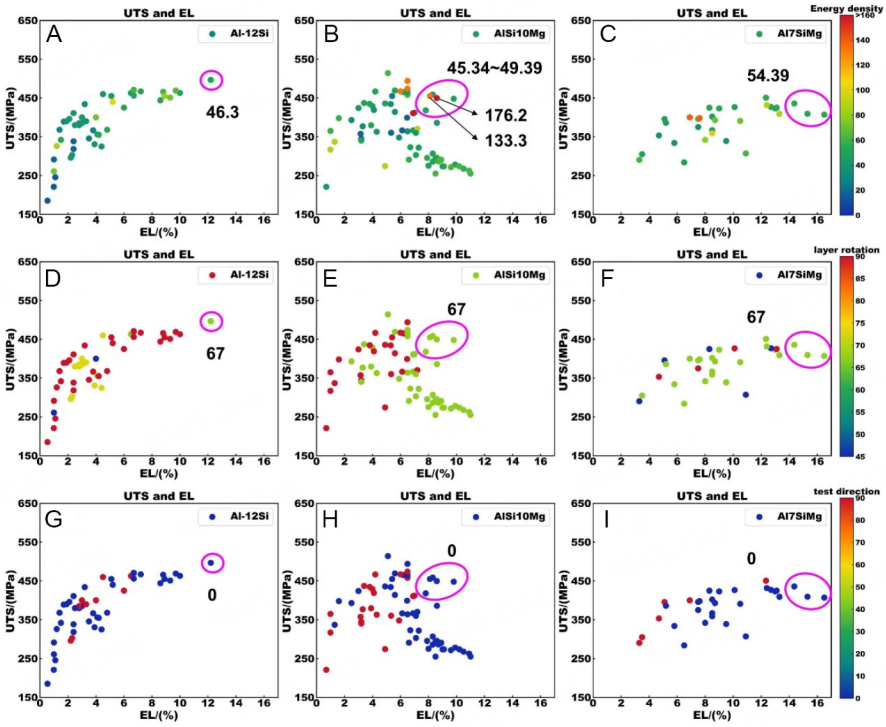


**Figure 2.** Effect of different characteristic variables (i.e., Ed, layer rotation, and test direction) on the mechanical properties of different SLMed Al-Si-(Mg) alloys: (A), (D), and (G) Al12Si alloys; (B), (E), and (H) AlSi10Mg alloys; (C), (F), (I) Al7SiMg alloys. Ed: energy density; EL: elongation; UTS: ultimate tensile strength.

to behave with low tensile strength and weak ductility when the Ed is low (≤ 30~40 J/mm$^3$). Therefore, in order to design high-performance alloys, some data points with Ed lower than 35 J/mm$^3$, especially for Al12Si alloy in the lower left region, should be deleted during the data cleaning.

Figure 2D-F and Figure 2G-I show the effect of rotation angle and testing direction on the properties. The results demonstrate that the UTS and EL of the alloys are much lower when both the rotation angle and the testing direction are 90°. Moreover, all the high-performance data points occur with a combination of layer rotation of 67° and test direction of 0°. This fact may be associated with the grain size and combination of neighboring powder layers during the additive manufacturing process[23]. With an optimal layer rotation, the fully densified parts with refined grain size can be fabricated. Meanwhile, directional solidification in SLM leads to preferential grain growth along the <100> direction, and the resulting intense texture is the main reason for the anisotropy of the alloy[80]. Thijs *et al.*[80] found a strong <100> texture along the scanning direction (i.e., parallel to the substrate plane) in SLMed AlSi10Mg alloy. It led to higher UTS and EL in alloys along the 0° test direction. However, an interesting observation was also found in some literature reports[12,49,53,54,56,66,68-70,72,74,75], as shown in Supplementary Figure 1. Under the same preparation conditions, the alloys with a test direction of 90° can also exhibit higher UTS. Kimura *et al.*[49] found the presence of coarsened microstructure encompassing the borders of laser scan tracks, which led to a lower UTS for horizontal alloys. Meanwhile, Kumar *et al.*[81] discovered that the distribution of molten particles along the vertical direction has a strong inter-particle bond, which plays a role in enhancing the strength of the alloy, but also increases the brittle behavior. As a result, they show a higher UTS but with a lower EL compared with the other alloys. Hence, in the Al-Si-(Mg) alloys, the influence of different features on the alloy properties is complex, and it is difficult to obtain the ideal target alloy by simple data analysis or limited experiments.

Some data points with good performance in three alloy systems were selected and analyzed, as marked by the purple circles in Figure 2. In general, from the current dataset, the promising Ed for Al-Si-(Mg) alloy is roughly distributed between 45 J/mm$^3$ and 55 J/mm$^3$. As the Si content increases, the suitable Ed value decreases. It seems that the combination of 67° layer rotation and 0° test direction provide great opportunities to obtain high-performance alloy. Based on the existing experimental dataset, the effect of more manufacturing parameters (laser power, scan speed, hatch space, and layer thickness) on the mechanical properties of SLMed Al-Si-(Mg) alloys were further investigated and analyzed, as shown in Supplementary Figure 2. The suitable layer thickness is 30 μm~40 μm, while the suitable hatch space is 100 μm~130 μm. With the Si content decreasing, the suitable laser power varies from 250 W to 370 W, while the suitable scan speed can vary from 1,800 mm/s to 1,300 mm/s.

### Feature analysis

After the data cleaning, the correlations between the characteristic variables and the mechanical properties were analyzed. As shown in Figure 3, Pearson's correlation coefficient for the characteristic and alloy properties was calculated. The results show that the correlation coefficient between any two variables is totally lower than 0.8, except for UTS and YS. The correlation coefficients between UTS and YS are as high as 0.92, with a strong linear relationship. Thus, the strength of the alloy can be described by using only UTS. Moreover, the correlation coefficients between the input variables (i.e., W(Si) and W(Mg), Ed, layer rotation, and test direction) and mechanical properties (i.e., UTS, YS, and EL) are very low. It means that there is no strong linear relationship between any two of them, and the alloy properties are not obtained by a simple linear summation of each variable's contribution. Therefore, the machine learning model (either MLPReg model or RFReg model), which can describe the complex unknown relationship for multi-variables, should be used to construct the relation of "composition-process-properties" in SLMed Al-Si-(Mg) alloys.

## RESULTS AND DISCUSSION

### Model construction and selection

Based on the constructed data set, the next key is to establish a quantitative relationship among alloy compositions, manufacturing parameters, and mechanical properties by machine learning technique. Five feature variables (W(Si), W(Mg), Ed, layer rotation, and test direction) were used as the input layer to predict UTS, YS, and EL simultaneously.

In order to select the most suitable model, four machine learning models were used to establish the quantitative relationship, including LinearReg, MLPReg, RFReg, and K-NN regressor. The randomized search strategy combining 5-fold cross-validation was utilized to find the optimal hyperparameters, based on which the optimal combinations of model parameters were selected. All the data were randomly divided into training and test sets with a 9:1 ratio.

Figure 4 presents the MAE value and $R^2$ score of the above four models based on the cleaned dataset. The lower MAE and the larger the $R^2$ score, the better the predictive performance. As shown in Figure 5, the LinearReg presents the poorest performance with $R^2$ less than 0.5 and MAE higher than 30. For the K-NN model with the lowest MAE value, it may be overfitting because the $R^2$ deviation between the testing set and the training set is large while the $R^2$ score of the training set is approaching 1. Both MLPReg and RFReg have good performance in training and testing sets; the $R^2$ score of the training set for these two models are 0.8962, 0.8216, and 0.8985, 0.8452 for the testing set, respectively. To further investigate the accuracy and stability of the model, the size of the testing set changed from 10% to 40%, and the evolution of the $R^2$ score was recorded in Supplementary Figure 3. The results also show that the testing set $R^2$ score of K-NN

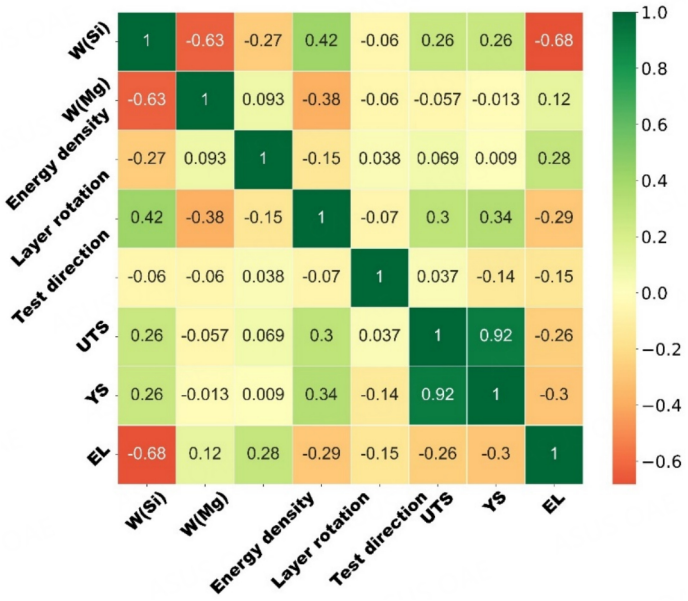


**Figure 3.** Pearson's correlation coefficients between characteristic variables and alloy properties. EL: elongation; UTS: ultimate tensile strength; YS: yield strength.

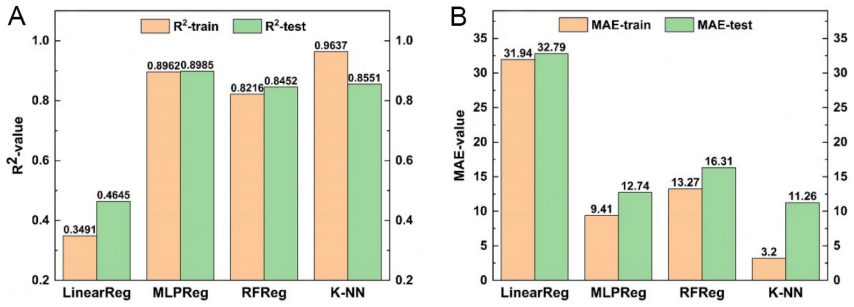


**Figure 4.** (A) $R^2$ score and (B) MAE value of different models on 90% training set and 10% testing set. MAE: mean absolute error.

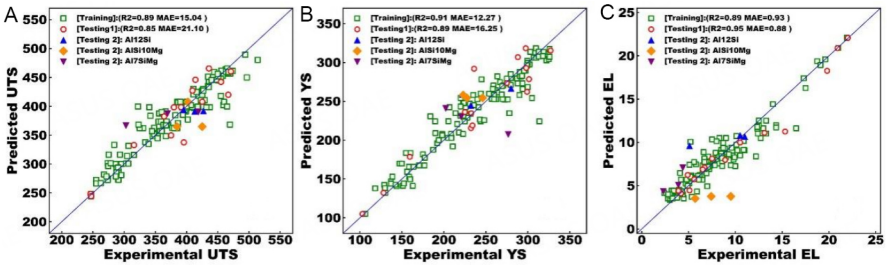


**Figure 5.** Performance of the MLPReg model predicted (A)UTS; (B)YS; and (C)EL on the training set and the testing set. EL: elongation; UTS: ultimate tensile strength; YS: yield strength.

dropped dramatically with the increased size due to the overfitting. Moreover, compared with the RFReg model, the MLPReg has a higher score overall. Therefore, the MLPReg model was selected as the optimal model to construct the quantitative relation of "composition-process-properties" in SLMed Al-Si-(Mg) alloy.

### Validation and prediction for the selected ML model

Based on the selected MLPReg model, the model hyperparameters were optimized by a random search strategy with a 10% testing dataset. The structure of the optimized MLPReg model was 5 × 100 × 200 × 3 (1 input layer with 5 features, 2 hidden layers with 100 and 200 neurons, and 1 output layer with 3 output features). The detailed optimization results are displayed in Supplementary Table 3. The comparison between the predicted and experimental properties of the trained MLPReg model is shown in Figure 5. When the predicted result is very close to the actual value, the data point should be in the vicinity of the line $y = x$. The results show that the predicted UTS, YS, and EL are in good agreement with the experimental results on both the training and testing set. The average $R^2$ score of training and testing sets can reach 0.8962 and 0.8985.

Besides, different operators and experimental environments can affect the mechanical properties of the target alloy. In order to test the generality of the model and further verify the model accuracy, 3 representative alloys were selected from each of the three systems[5,48,58,59,78,79] (i.e., Al7SiMg, AlSi10Mg, Al12Si) from the cleaned dataset as the second test set. The composition and manufacturing parameters of those 9 alloys are shown in Supplementary Table 4. The comparison of predicted and experimental results of the second test set is also shown in Figure 5. The results showed that the selected ML model can predict the UTS, YS, and EL of all the alloys simultaneously, especially reproducing the mechanical properties of the 9 alloys in the second test dataset well. Therefore, it indicates that the presently selected ML model can simultaneously predict the UTS, YS, and EL of all the SLMed Al-Si-(Mg) alloy with high accuracy and will be utilized to design the alloy with high performance further.

Based on the trained and validated MLPReg model, the mechanical properties of alloys over a wide range of composition and process parameters can be predicted. A method for randomly generating composition and processes over the parameter space was used to create the big input dataset in this work. The range of Si is between 4.1wt.% and 12.1wt.% with $\Delta w(Si)$ = 1wt.%, and that of Mg varies from 0wt.% to 0.6wt.% with $\Delta w(Mg)$ = 0.1wt.%. The value of Ed varies from 26 J/mm$^3$ to 171 J/mm$^3$ with a step of 5 J/mm$^3$. Supplementary Table 2 shows that over 95% of the high-performance alloys were prepared using 67°, 73°, or 90° as the layer rotation angle. To achieve the best performance and increase the computational efficiency, three values (i.e., 67°, 73°, and 90°) were selected as candidates to make predictions. Figure 6 shows the predicted results based on the trained MLPReg model for random 11,300 data (combinations of compositions and processes) in SLMed Al-Si-(Mg) alloy. The plot has been colored with Quality Index (QI). The strength-ductility trade-off relationship indicates that monolithic materials have a limitation in achieving simultaneously high strength and elongation. In Figure 6, the alloys with high performance (i.e., high QI) usually presented a high strength with medium elongation.

To further validate the accuracy of the model, the predicted data with 7 wt.%, 10 wt.%, and 12 wt.% Si were extracted respectively and compared with the experimental results, as shown in Figure 7. The results indicated that all the experimental data locates in the region of the predicted results, inferring that the model can not only describe the properties of the existing experimental data very well but also predict the mechanical properties of the alloys over the unknown region(s).

### Alloy design with desired mechanical properties

Based on the accurate prediction results over wide composition and process parameter space, one can design the SLMed Al-Si-(Mg) alloy with desired machinal properties. However, how to design an alloy with simultaneously high strength and ductility needs to be considered. The multi-objective to single-objective optimization strategy, which converts a weighted linear combination of multi-objective performance into a single objective for optimizing the design for a single objective quantity (Ashby's method[82]), is one of the

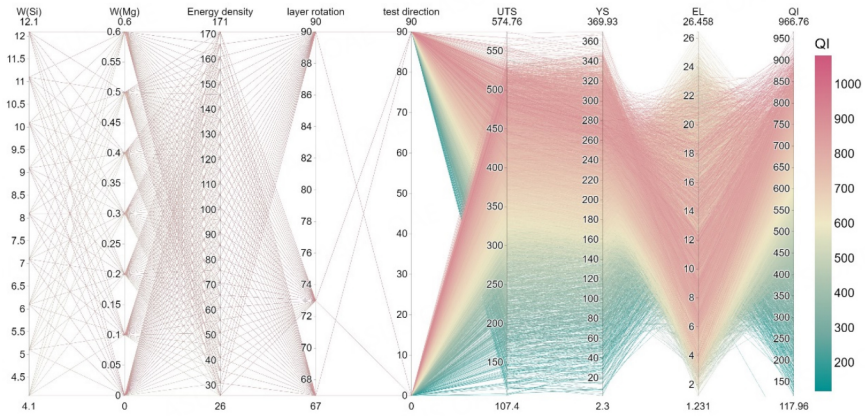


**Figure 6.** Results from MLPReg model prediction displayed as a parallel coordinate plot for 11300 combinations of compositions and process of Al-Si-(Mg) alloy. The plot is colored with Quality Index (QI). EL: elongation; UTS: ultimate tensile strength; YS: yield strength.

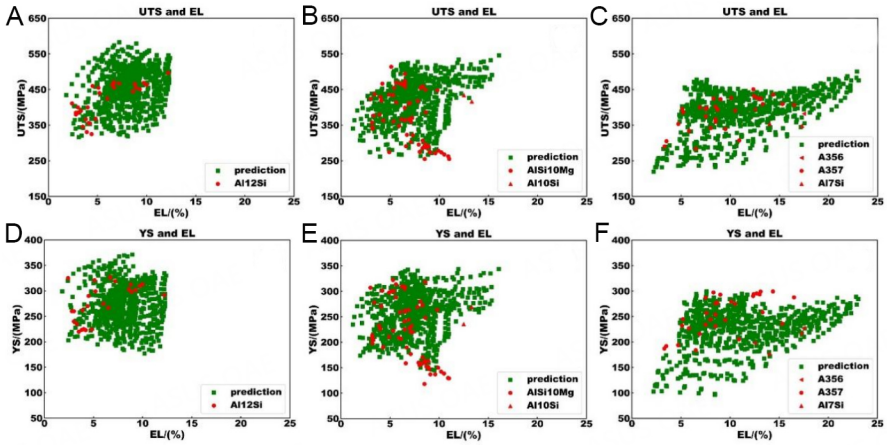


**Figure 7.** Distribution of predicted and experimental data (i.e., UTS, YS, and EL) in alloys with different Si contents: (A) and (D)12 wt.%; (B) and (E) 10 wt.%; (C) and (F) 7 wt.%. EL: elongation; UTS: ultimate tensile strength; YS: yield strength.

commonly used multi-objective optimization strategies and thus utilized in this work. The QI, which represents the evaluation criterion for alloy properties considering both strength and ductility, was employed in this work to design the alloy with optimal comprehensive mechanical properties. QI can be calculated by the following formula,

$$QI = UTS + YS \times log_{10}(EL) \quad (5)$$

While the machine learning model, which can accurately predict the mechanical properties of the SLMed Al-Si-(Mg) alloys, has been established in this work, the QI can be used to find novel SLMed Al-Si-(Mg) alloys with higher comprehensive mechanical properties than all the experimental data reported in the literature based on the trained MLPReg model.

In order to make the recommended alloys as reasonable as possible, the contents of alloying elements and manufacturing parameter values of all the recommended alloys should not exceed the range of the training data set. According to the above discussion, the Ed of 35 J/mm$^3$ is the threshold value for good mechanical

**Table 1. Presently designed alloys with respective maximum UTS, maximum YS, maximum EL, and maximum QI under as-build conditions**

| Group | W(Si) (wt%) | W(Mg) (wt%) | Ed (J/mm$^3$) | Rotation (°) | Direction (°) | UTS (MPa) | YS (MPa) | EL (%) | QI (MPa) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 12.1 | 0.6 | 36 | 67 | 90 | 559.0 | 344.8 | 6.2 | 831.5 |
| 2 | 12.1 | 0.6 | 36 | 73 | 0 | 548.9 | 362.8 | 7.4 | 863.7 |
| 3 | 4.1 | 0 | 96 | 90 | 0 | 307.4 | 162.9 | 24.5 | 533.8 |
| 4 | 10.1 | 0.1 | 66 | 90 | 0 | 549.4 | 346.1 | 16.1 | 966.8 |

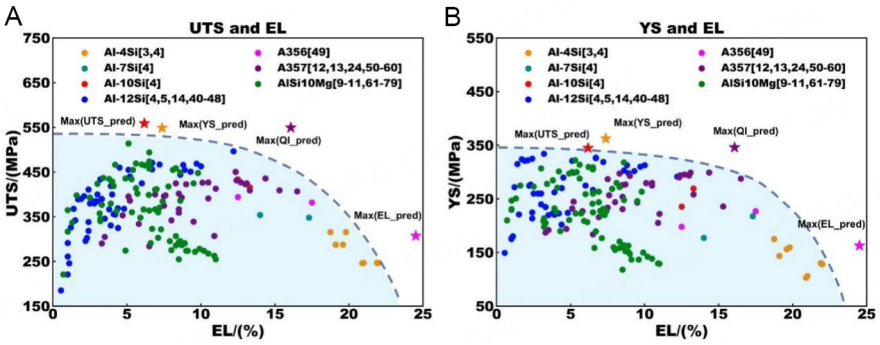


**Figure 8.** Presently designed alloys with respective maximum UTS, YS, EL, and QI, compared with as-built experiment data in different plots: (A) UTS-EL; (B)YS-EL. EL: elongation; QI: Quality Index; UTS: ultimate tensile strength; YS: yield strength.

properties. The analysis of the dataset in Supplementary Table 2 shows that 75% of the data points have an Ed below 62 J/mm$^3$, with only a few special points larger than 100 J/mm$^3$. Therefore, the alloys with Ed lower than 35 J/mm$^3$ (threshold value) and higher than 100 J/mm$^3$ (extreme value) were excluded here.

After predictions with random combinations of compositions and processes, the selected alloys with maximum UTS, maximum YS, maximum EL, or maximum QI are listed in Table 1. It can be clearly seen from the predicted results that the alloy with higher Si content owns higher strength. On the contrary, the highest elongation is achieved when the Si content is only 4%. The high-performance alloy with maximum QI can be obtained when the Si content is 10%. Such phenomena are consistent with the previously known experimental results[11,57,83], which further reinforces the reliability of the dataset and the accuracy of the model. This fact also indicates that the model not only deepens our understanding of the SLMed Al-Si-(Mg) system but also helps us to design new alloys.

The comparison between the design results and the original experimental dataset is shown in Figure 8. In the cleaned experimental datasets, the maximum value of UTS is 514 MPa. While in the predicted dataset based on the trained MLPReg model, the UTS of alloys can reach 560 MPa with the EL higher than 6%. The test direction for this alloy is 90°, which is consistent with some literature[12,68,70] and the data analysis above. This also indicates that the predicted results are reasonable. Moreover, for the designed alloy with a maximum QI of 966 MPa, its UTS is 549 MPa, and EL is 16%, both of which are higher than all the available experimental data. Although the designed results generally conform to the current experimental rules, their accuracy still needs further experimental validation as the major target for the next publication.

## CONCLUSIONS

• Based on an exhausting collection from 50 publications, a high-quality dataset for SLMed Al-Si-(Mg) alloys (including Al12Si, Al10Si, AlSi10Mg, Al7Si, Al7SiMg, and Al4Si) with 176 pieces of data was

established. The dataset records the information on alloy composition, process parameters, test conditions, and mechanical properties in detail. The effect of three main variables (i.e., Ed, layer rotation, and test direction) on the mechanical properties was systematically analyzed in the three representative alloys (Al12Si, AlSi10Mg, and Al7SiMg). A threshold value of 35 J/mm$^3$ for Ed was used as a criterion to clean the data points with lower UTS and EL. The cleaned dataset consists of a first training/testing dataset with 142 data for model construction and a second testing dataset with 9 data for model verification.

• Four ML models were employed to establish the quantitative relation of "composition-processes-properties" in SLMed Al-Si-(Mg) alloys. After a comprehensive comparison, the MLPReg model was selected due to its best performance on both training and testing sets. The selected MLPReg model was then utilized to design novel compositions and process parameters for SLMed Al-Si-(Mg) alloys with maximum UTS, YS, EL, and comprehensive mechanical property index QI. For the alloy with maximum QI, its UTS can reach 549.4MPa, which is ~40MPa higher than the best one over the available experimental data, while its elongation can still retain 16%.

• The successful demonstration in this paper indicates that the present design strategy driven by the ML technique should generally be applicable to other SLMed alloy systems.

## DECLARATIONS

### Authors' contributions

Establishment, selection, and verification of machine learning model: Gao T
Made contributions to the conception and design of the study: Gao T, Gao J, Zhang L
Literatures data collation, interpretation, analysis, cleaning, and construction of data set, drafted the manuscript: Gao T, Gao J, Zhang J
Supervised, revised, and finalized the manuscript: Gao T, Gao J, Zhang J, Song B, Zhang L

### Availability of data and materials

The using data set and the trained machine learning model can be seen in Supplementary Materials.

### Financial support and sponsorship

The financial support from the Science and Technology Program of Guangxi province, China (Grant No. AB21220028), the Natural Science Foundation of Hunan Province for Distinguished Young Scholars (Grant No. 2021JJ10062), and the Lvyangjinfeng Talent program of Yangzhou is acknowledged.

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

## REFERENCES


1. Zhang J, Song B, Wei Q, Bourell D, Shi Y. A review of selective laser melting of aluminum alloys: Processing, microstructure, property and developing trends. *J Mater Sci Technol* 2019;35:270-84. DOI
2. Ghio E, Cerri E. Additive manufacturing of AlSi10Mg and $Ti_6Al_4V$ lightweight alloys via laser powder bed fusion: a review of heat treatments effects. *Materials* 2022;15:2047. DOI PubMed PMC
3. Kimura T, Nakamoto T, Ozaki T, Sugita K, Mizuno M, Araki H. Microstructural formation and characterization mechanisms of selective laser melted Al-Si-Mg alloys with increasing magnesium content. *Mater Sci Eng A* 2019;754:786-98. DOI
4. Kimura T, Nakamoto T, Mizuno M, Araki H. Effect of silicon content on densification, mechanical and thermal properties of Al-xSi binary alloys fabricated using selective laser melting. *Mater Sci Eng A* 2017;682:593-602. DOI
5. Suryawanshi J, Prashanth K, Scudino S, Eckert J, Prakash O, Ramamurty U. Simultaneous enhancements of strength and toughness in an Al-12Si alloy synthesized using selective laser melting. *Acta Mater* 2016;115:285-94. DOI
6. Wang P, Lao C, Chen Z, et al. Microstructure and mechanical properties of Al-12Si and Al-3.5Cu-1.5Mg-1Si bimetal fabricated by selective laser melting. *J Mater Sci Technol* 2020;36:18-26. DOI
7. Awd M, Siddique S, Walther F. Microstructural damage and fracture mechanisms of selective laser melted Al-Si alloys under fatigue loading. *Theor Appl Fract Mech* 2020;106:102483. DOI
8. Cerri E, Ghio E, Bolelli G. Effect of the distance from build platform and post-heat treatment of AlSi10Mg alloy manufactured by single- and multi-laser selective laser melting. *J Mater Eng Perform* 2021;30:4981-92. DOI
9. Chen B, Moon S, Yao X, et al. Strength and strain hardening of a selective laser melted AlSi10Mg alloy. *Scr Mater* 2017;141:45-9. DOI
10. Larrosa N, Wang W, Read N, et al. Linking microstructure and processing defects to mechanical properties of selectively laser melted AlSi10Mg alloy. *Theor Appl Fract Mech* 2018;98:123-33. DOI
11. Hwang WJ, Bang GB, Choa S. Effect of a stress relief heat treatment of AlSi7Mg and AlSi10Mg alloys on mechanical and electrical properties according to silicon precipitation. *Met Mater Int* 2022. DOI
12. Denti L. Additive manufactured A357.0 samples using the laser powder bed fusion technique: shear and tensile performance. *Metals* 2018;8:670. DOI
13. Yang KV, Rometsch P, Davies C, Huang A, Wu X. Effect of heat treatment on the microstructure and anisotropy in mechanical properties of A357 alloy produced by selective laser melting. *Mater Des* 2018;154:275-90. DOI
14. Suzuki A, Miyasaka T, Takata N, Kobashi M, Kato M. Control of microstructural characteristics and mechanical properties of AlSi12 alloy by processing conditions of laser powder bed fusion. *Addit Manuf* 2021;48:102383. DOI
15. Gheysen J, Marteleur M, van der Rest C, Simar A. Efficient optimization methodology for laser powder bed fusion parameters to manufacture dense and mechanically sound parts validated on AlSi12 alloy. *Mater Des* 2021;199:109433. DOI
16. Mei J, Han Y, Zu G, et al. Achieving superior strength and ductility of AlSi10Mg alloy fabricated by selective laser melting with large laser power and high scanning speed. *Acta Metall Sin* 2022;35:1665-72. DOI
17. Dai D, Gu D, Zhang H, et al. Influence of scan strategy and molten pool configuration on microstructures and tensile properties of selective laser melting additive manufactured aluminum based parts. *Opt Laser Technol* 2018;99:91-100. DOI
18. Giovagnoli M, Silvi G, Merlin M, Di Giovanni MT. Optimisation of process parameters for an additively manufactured AlSi10Mg alloy: limitations of the energy density-based approach on porosity and mechanical properties estimation. *Mater Sci Eng A* 2021;802:140613. DOI
19. Salandari-rabori A, Wang P, Dong Q, Fallah V. Enhancing as-built microstructural integrity and tensile properties in laser powder bed fusion of AlSi10Mg alloy using a comprehensive parameter optimization procedure. *Mater Sci Eng A* 2021;805:140620. DOI
20. Wang C, Zhu J, Wang G, et al. Effect of building orientation and heat treatment on the anisotropic tensile properties of AlSi10Mg fabricated by selective laser melting. *J Alloys Compd* 2022;895:162665. DOI
21. Li X, Yi D, Wu X, et al. Effect of construction angles on microstructure and mechanical properties of AlSi10Mg alloy fabricated by selective laser melting. *J Alloys Compd* 2021;881:160459. DOI
22. Maconachie T, Leary M, Zhang J, et al. Effect of build orientation on the quasi-static and dynamic response of SLM AlSi10Mg. *Mater Sci Eng A* 2020;788:139445. DOI
23. Gupta MK, Singla AK, Ji H, et al. Impact of layer rotation on micro-structure, grain size, surface integrity and mechanical behaviour of SLM Al-Si-10Mg alloy. *J Mater Res Technol* 2020;9:9506-22. DOI
24. Yadav P, Rigo O, Arvieu C, Lacoste E. Microstructural and mechanical aspects of AlSi7Mg0.6 alloy related to scanning strategies in L-PBF. *Int J Adv Manuf Technol* 2022;120:6205-23. DOI
25. Lu Z, Zhang L. Thermodynamic description of the quaternary Al-Si-Mg-Sc system and its application to the design of novel Sc-additional A356 alloys. *Mater Des* 2017;116:427-37. DOI
26. Lu Z, Zhang L, Wang J, Yao Q, Rao G, Zhou H. Understanding of strengthening and toughening mechanisms for Sc-modified Al-Si-(Mg) series casting alloys designed by computational thermodynamics. *J Alloys Compd* 2019;805:415-25. DOI
27. Liu G, Gao J, Che C, Lu Z, Yi W, Zhang L. Optimization of casting means and heat treatment routines for improving mechanical and corrosion resistance properties of A356-0.54Sc casting alloy. *Mater Today Commun* 2020;24:101227. DOI
28. Gao J, Li Z. Current situation and prospect of computationally assisted design in high-performance additive manufactured aluminum alloys: a review. *Acta Met Sin* 2022;59:87-105. DOI
29. Yi W, Liu G, Lu Z, Gao J, Zhang L. Efficient alloy design of Sr-modified A356 alloys driven by computational thermodynamics and

machine learning. *J Mater Sci Technol* 2022;112:277-90. DOI
30. Gao J, Zhong J, Liu G, et al. A machine learning accelerated distributed task management system (Malac-Distmas) and its application in high-throughput CALPHAD computation aiming at efficient alloy design. *Adv Powder Mater* 2022;1:100005. DOI
31. Wei M, Tang Y, Zhang L, Sun W, Du Y. Phase-field simulation of microstructure evolution in industrial A2214 alloy during solidification. *Metall and Mat Trans A* 2015;46:3182-91. DOI
32. Gao J, Malchère A, Yang S, et al. Dewetting of Ni silicide thin film on Si substrate: in-situ experimental study and phase-field modeling. *Acta Mater* 2022;223:117491. DOI
33. Yang S, Zhong J, Wang J, Gao J, Li Q, Zhang L. A novel computational model for isotropic interfacial energies in multicomponent alloys and its coupling with phase-field model with finite interface dissipation. *J Mater Sci Technol* 2023;133:111-22. DOI
34. Zhang J, Yuan W, Song B, et al. Towards understanding metallurgical defect formation of selective laser melted wrought aluminum alloys. *Adv Powder Mater* 2022;1:100035. DOI
35. Yi W, Liu G, Gao J, Zhang L. Boosting for concept design of casting aluminum alloys driven by combining computational thermodynamics and machine learning techniques. *J Mater Inf* 2021;1:11. DOI
36. Zhang S, Yi W, Zhong J, Gao J, Lu Z, Zhang L. Computer alloy design of Ti modified Al-Si-Mg-Sr casting alloys for achieving simultaneous enhancement in strength and ductility. *Materials* 2022;16:306. DOI PubMed PMC
37. Mondal B, Mukherjee T, Debroy T. Crack free metal printing using physics informed machine learning. *Acta Mater* 2022;226:117612. DOI
38. Yu T, Mo X, Chen M, Yao C. Machine-learning-assisted microstructure-property linkages of carbon nanotube-reinforced aluminum matrix nanocomposites produced by laser powder bed fusion. *Nanotechnol Rev* 2021;10:1410-24. DOI
39. He P, Liu Q, Kruzic JJ, Li X. Machine-learning assisted additive manufacturing of a TiCN reinforced AlSi10Mg composite with tailorable mechanical properties. *Mater Lett* 2022;307:131018. DOI
40. Prashanth K, Scudino S, Klauss H, et al. Microstructure and mechanical properties of Al-12Si produced by selective laser melting: Effect of heat treatment. *Mater Sci Eng A* 2014;590:153-60. DOI
41. Prashanth K, Scudino S, Eckert J. Tensile properties of Al-12Si fabricated via selective laser melting (SLM) at different temperatures. *Technologies* 2016;4:38. DOI
42. Li X, Wang X, Saunders M, et al. A selective laser melting and solution heat treatment refined Al-12Si alloy with a controllable ultrafine eutectic microstructure and 25% tensile ductility. *Acta Mater* 2015;95:74-82. DOI
43. Prashanth K, Scudino S, Eckert J. Defining the tensile properties of Al-12Si parts produced by selective laser melting. *Acta Mater* 2017;126:25-35. DOI
44. Liu M, Wada T, Suzuki A, Takata N, Kobashi M, Kato M. Effect of annealing on anisotropic tensile properties of Al-12%Si alloy fabricated by laser powder bed fusion. *Crystals* 2020;10:1007. DOI
45. Wang X, Zhang L, Fang M, Sercombe T. The effect of atmosphere on the structure and properties of a selective laser melted Al-12Si alloy. *Mater Sci Eng A* 2014;597:370-5. DOI
46. Rashid R, Masood S, Ruan D, et al. Effect of energy per layer on the anisotropy of selective laser melted AlSi12 aluminium alloy. *Addit Manuf* 2018;22:426-39. DOI
47. Zhang S, Ma P, Jia Y, et al. Microstructure and mechanical properties of Al-(12-20)Si Bi-material fabricated by selective laser melting. *Materials* 2019;12:2126. DOI PubMed PMC
48. Siddique S, Imran M, Wycisk E, Emmelmann C, Walther F. Influence of process-induced microstructure and imperfections on mechanical properties of AlSi12 processed by selective laser melting. *J Mater Process Technol* 2015;221:205-13. DOI
49. Kimura T, Nakamoto T. Microstructures and mechanical properties of A356 (AlSi7Mg0.3) aluminum alloy fabricated by selective laser melting. *Mater Des* 2016;89:1294-301. DOI
50. Rao H, Giet S, Yang K, Wu X, Davies CH. The influence of processing parameters on aluminium alloy A357 manufactured by selective laser melting. *Mater Des* 2016;109:334-46. DOI
51. Rao JH, Zhang Y, Fang X, Chen Y, Wu X, Davies CH. The origins for tensile properties of selective laser melted aluminium alloy A357. *Addit Manuf* 2017;17:113-22. DOI
52. Casati R, Vedani M. Aging response of an A357 Al alloy processed by selective laser melting. *Adv Eng Mater* 2019;21:1800406. DOI
53. de Menezes JT, Castrodeza EM, Casati R. Effect of build orientation on fracture and tensile behavior of A357 Al alloy processed by selective laser melting. *Mater Sci Eng A* 2019;766:138392. DOI
54. Zou T, Ou Y, Zhu H, Li L. Effects of heat treatment on microstructure and tensile properties of AlSi7Mg alloy fabricated by selective laser melting. *Hot Work Technol* 2019;48:154-7. DOI
55. Tang G, Feng T, Duan G, et al. Process and properties of AlSi7Mg alloy fabricated by laser selected melting. *Foundry Technol* 2020;41:219-22. DOI
56. Zou T, Ou Y, Zhu H, Qin J. Microstructure and mechanical properties of selective laser melted AlSi7Mg alloy. Available from: http://www.mater-rep.com/EN/abstract/abstract2647.shtml [Last accessed on 28 Mar 2023].
57. Cerri E, Ghio E. Aging profiles of AlSi7Mg0.6 and AlSi10Mg0.3 Alloys manufactured via laser-powder bed fusion: direct aging versus T6. *Materials* 2022;15:6126. DOI PubMed PMC
58. Cacace S, Gökhan Demir A, Sala G, Mattia Grande A. Influence of production batch related parameters on static and fatigue resistance of LPBF produced AlSi7Mg0.6. *Int J Fatigue* 2022;165:107227. DOI
59. Rao JH, Zhang Y, Zhang K, Wu X, Huang A. Selective laser melted Al-7Si-0.6Mg alloy with in-situ precipitation via platform heating

for residual strain removal. *Mater Des* 2019;182:108005. DOI

60. Lorusso M, Trevisan F, Calignano F, Lombardi M, Manfredi D. A357 alloy by LPBF for industry applications. *Materials* 2020;13:1488. DOI PubMed PMC
61. Read N, Wang W, Essa K, Attallah MM. Selective laser melting of AlSi10Mg alloy: Process optimisation and mechanical properties development. *Mater Des* 2015;65:417-24. DOI
62. Bagherifard S, Beretta N, Monti S, Riccio M, Bandini M, Guagliano M. On the fatigue strength enhancement of additive manufactured AlSi10Mg parts by mechanical and thermal post-processing. *Mater Des* 2018;145:28-41. DOI
63. Li W, Li S, Liu J, et al. Effect of heat treatment on AlSi10Mg alloy fabricated by selective laser melting: microstructure evolution, mechanical properties and fracture mechanism. *Mater Sci Eng A* 2016;663:116-25. DOI
64. Tradowsky U, White J, Ward R, Read N, Reimers W, Attallah M. Selective laser melting of AlSi10Mg: influence of post-processing on the microstructural and tensile properties development. *Mater Des* 2016;105:212-22. DOI
65. Hitzler L, Janousch C, Schanz J, et al. Direction and location dependency of selective laser melted AlSi10Mg specimens. *J Mater Process Technol* 2017;243:48-61. DOI
66. Casati R, Hamidi Nasab M, Coduri M, Tirelli V, Vedani M. Effects of platform pre-heating and thermal-treatment strategies on properties of AlSi10Mg alloy processed by selective laser melting. *Metals* 2018;8:954. DOI
67. Kan WH, Nadot Y, Foley M, Ridosz L, Proust G, Cairney JM. Factors that affect the properties of additively-manufactured AlSi10Mg: porosity versus microstructure. *Addit Manuf* 2019;29:100805. DOI
68. Xiong Z, Liu S, Li S, Shi Y, Yang Y, Misra R. Role of melt pool boundary condition in determining the mechanical properties of selective laser melting AlSi10Mg alloy. *Mater Sci Eng A* 2019;740-741:148-56. DOI
69. Padovano E, Badini C, Pantarelli A, Gili F, D'aiuto F. A comparative study of the effects of thermal treatments on AlSi10Mg produced by laser powder bed fusion. *J Alloys Compd* 2020;831:154822. DOI
70. Fiocchi J, Biffi CA, Colombo C, Vergani LM, Tuissi A. Ad hoc heat treatments for selective laser melted Alsi10mg alloy aimed at stress-relieving and enhancing mechanical performances. *JOM* 2020;72:1118-27. DOI
71. Li Z, Li Z, Tan Z, Xiong D, Guo Q. Stress relaxation and the cellular structure-dependence of plastic deformation in additively manufactured AlSi10Mg alloys. *Int J Plast* 2020;127:102640. DOI
72. Sert E, Hitzler L, Hafenstein S, Merkel M, Werner E, Öchsner A. Tensile and compressive behaviour of additively manufactured AlSi10Mg samples. *Prog Addit Manuf* 2020;5:305-13. DOI
73. Park T, Baek M, Hyer H, Sohn Y, Lee K. Effect of direct aging on the microstructure and tensile properties of AlSi10Mg alloy manufactured by selective laser melting process. *Mater Charact* 2021;176:111113. DOI
74. Ou Y, Zhang Q, Wei Y, et al. Evolution of heterogeneous microstructure and its effects on tensile properties of selective laser melted AlSi10Mg alloy. *J Mater Eng Perform* 2021;30:4341-55. DOI
75. Paul MJ, Liu Q, Best JP, et al. Fracture resistance of AlSi10Mg fabricated by laser powder bed fusion. *Acta Mater* 2021;211:116869. DOI
76. Riener K, Oswald S, Winkler M, Leichtfried GJ. Influence of storage conditions and reconditioning of AlSi10Mg powder on the quality of parts produced by laser powder bed fusion (LPBF). *Addit Manuf* 2021;39:101896. DOI
77. Chen S, Tan Q, Gao W, et al. Effect of heat treatment on the anisotropy in mechanical properties of selective laser melted AlSi10Mg. *Mater Sci Eng A* 2022;858:144130. DOI
78. Bisht MS, Gaur V, Singh I. On mechanical properties of SLM Al-Si alloy: Role of heat treatment-induced evolution of silicon morphology. *Mater Sci Eng A* 2022;858:144157. DOI
79. Van Cauwenbergh P, Samaee V, Thijs L, et al. Unravelling the multi-scale structure-property relationship of laser powder bed fusion processed and heat-treated AlSi10Mg. *Sci Rep* 2021;11:6423. DOI PubMed PMC
80. Thijs L, Kempen K, Kruth J, Van Humbeeck J. Fine-structured aluminium products with controllable texture by selective laser melting of pre-alloyed AlSi10Mg powder. *Acta Mater* 2013;61:1809-19. DOI
81. Kumar MS, Javidrad HR, Shanmugam R, Ramoni M, Adediran AA, Pruncu CI. Impact of print orientation on morphological and mechanical properties of L-PBF based AlSi7Mg parts for aerospace applications. *Silicon* 2022;14:7083-97. DOI
82. Ashby M. Multi-objective optimization in material design and selection. *Acta Mater* 2000;48:359-69. DOI
83. Zhao L, Song L, Santos Macías JG, et al. Review on the correlation between microstructure and mechanical performance for laser powder bed fusion AlSi10Mg. *Addit Manuf* 2022;56:102914. DOI

Menon *et al. J Mater Inf* 2023;3:7
**DOI:** 10.20517/jmi.2022.38

**Journal of Materials Informatics**

**Research Article**

**Open Access**

Check for updates

# Linking processing parameters with melt pool properties of multiple nickel-based superalloys via high-dimensional Gaussian process regression

**Nandana Menon[1], Sudeepta Mondal[1,2], Amrita Basak[1,*]**

[1]Department of Mechanical Engineering, The Pennsylvania State University, University Park, PA 16802, USA.
[2]Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA.

[*]**Correspondence to:** Dr. Amrita Basak, Department of Mechanical Engineering, The Pennsylvania State University, 233 Reber Building, Burrowes Rd., University Park, PA 16802, USA. E-mail: aub1526@psu.edu

**How to cite this article:** Menon N, Mondal S, Basak A. Linking processing parameters with melt pool properties of multiple nickel-based superalloys via high-dimensional Gaussian process regression. *J Mater Inf* 2023;3:7. https://dx.doi.org/10.20517/jmi.2022.38

**Received:** 1 Dec 2022 **First Decision:** 8 Feb 2023 **Revised:** 27 Feb 2023 **Accepted:** 15 Mar 2023 **Published:** 31 Mar 2023

**Academic Editors:** Xingjun Liu, Ma Qian **Copy Editor:** Ke-Cui Yang **Production Editor:** Ke-Cui Yang

## Abstract

A physics-based model is used to predict the melt pool properties in the laser-directed energy deposition of several nickel-based superalloys for different process parameters. The input space is high-dimensional, consisting of a common 19-dimensional composition space for each alloy and the process parameters (laser power and scan velocity). Gaussian Process-based regression frameworks are developed by training surrogates on data generated by a validated analytical model. These surrogates are thereafter used to predict and define relationships between the composition, resultant thermophysical properties, process parameters, and the subsequent melt pool property. The probabilistic predictions are augmented by uncertainty quantification and sensitivity analysis to substantiate the findings further.

**Keywords:** Laser directed energy deposition, Gaussian process, nickel-based superallloys, melt pool properties, uncertainty quantification, sensitivity analysis

## INTRODUCTION

Laser-directed energy deposition (L-DED) is a subset of metal additive manufacturing (AM), where parts

are created via layer-by-layer addition of metal powder or wire feedstock into a melt pool generated by a high-intensity laser source. The melt pool undergoes a complex thermal history, which thereby controls the microstructures and mechanical properties of the as-fabricated part. Melt pool geometry can, thus, serve as a justified proxy to assess overall build quality, and has been studied extensively in the literature[1-4]. While L-DED is driven by numerous process parameters, laser power and scan velocity have been observed to profoundly affect melt pool geometry while being relatively easy to control. L-DED machines are sold as flexible platforms; thus, users must identify the optimum values based on application and material[5]. A good process model that connects these process parameters to the melt pool geometry can, therefore, serve as an efficient basis for process planning and quality control of an L-DED part over a wide processing window.

The depth of the melt pool is especially of interest as it directly affects the quality and integrity of the final part. Shallow melt pools result in a weak bond between the newly deposited layer and the previous layer, which can cause the build to fail, or create porosity or other defects in the final part. On the other hand, a deep melt pool can potentially lead to distortion or cracking of the final part. In summary, the depth of the melt pool must be carefully controlled to ensure that the material is properly melted and solidified, leading to a high-quality final part. Typically, process parameter development for L-DED relies on a design-of-experiments (DoE) in the process parameter space. Melt pool depths are iteratively evaluated by performing experiments or modeling melt pool geometry. While finite element approaches allow extensive modeling of the physical phenomena, they are rather complex and computationally expensive. On the contrary, analytical models make certain assumptions for simplification while obeying similar conservation laws. One such analytical model was developed by Eagar-Tsai[6]. With appropriate calibration, this model can be an inexpensive source of physically realistic low-fidelity solutions, thereby allowing dense sampling of the process parameter space.

More recently, researchers have leveraged machine learning algorithms to facilitate the development of robust process development models for L-DED. Feenstra *et al.*[7] used an Artificial Neural Network to model functional relationships between melt pool geometry and process parameters for three different alloys: SS316L, Inconel 625, and Hastelloy X. A Neural Network was also implemented to capture the influence of thermophysical properties. However, since the training data comprised only three alloys, these observations were not conclusive. Juhasz[8] used a selection of machine learning methodologies that included Neural Networks, Gaussian Process (GP), Support-Vector Machines, and Gradient Boosted Decision Trees to perform regression and classification studies relating to melt pool geometries. An ensemble regression accuracy of 70.5% and an ensemble classification accuracy of 72.3% were achieved across the dataset. Similar investigations were carried out by Akbari *et al.*[9], who also studied the influence of material properties in addition to process parameters. Ye *et al.*[10] designed a semi-dynamic machine learning model for melt pool depth prediction where artificial neural networks were trained on melt pool images. However, the combined influence of process parameters was studied, as a consequence of which individual relationships between parameters and outputs could not be established. Additionally, the use of deterministic models meant that uncertainty quantification was not possible. Probabilistic modeling using Gaussian Processes for melt pool prediction and optimization was carried out for CMSX-4® using an analytical model by Mondal *et al.*[11]. Menon *et al.*[12] extended the work by developing a GP-based multi-fidelity surrogate model to generate process maps of CMSX-4® with uncertainty quantification. The multi-fidelity model improved prediction accuracy with a 55% reduction in root mean squared error compared to a single-fidelity model, however, for a single alloy system.

While these papers have successfully used machine learning to enhance the modeling process for parameter development, the materials used are limited. More often than not, the models do not accommodate multiple

material systems, implying that for each material (i.e., for any new composition), a new model must be developed. A major challenge associated with L-DED, especially powder L-DED, is that the overall material selection is still relatively limited, which poses additional design difficulty as the available materials may not suit the application requirements[13]. Naturally, a significant thrust of L-DED research is developing new materials or modifying conventional alloys so that they are explicitly designed for L-DED to exploit its advantages to the maximum. Nickel-based superalloys are one such class of alloys regularly manufactured using L-DED. These alloys possess a unique combination of high-temperature strength and creep resistance[14,15]. However, there still exists much hesitancy in completely adopting these alloys due to concerns surrounding microstructural heterogeneities and defects that limit repeatability in parts produced[16]. Designing new materials would require efficiently exploring the vast search space for targeted properties. Many nickel-based and other superalloy developments are guided by high-throughput experiments[17-19]. Recently, researchers have explored machine learning strategies, such as active learning strategies, to iteratively conduct experiments facilitating the exploration of the search space[20-22]. To maximize the full potential of AM, alloy development for AM processing requires additional attention so that resultant AM parts meet desired specifications over a broad processing landscape. Currently, such an understanding is restricted due to limitations surrounding process-structure-property relationships that depend on faster, reliable, and standardized methods for testing potential AM materials.

Nickel-based superalloys contain as many as 19 elements added to achieve particular properties. While describing the effect of individual elements can involve time-consuming experiments and multiple iterations, describing interactions as a whole via such experiments is nearly impossible. Researchers have attempted to describe functional relationships between elements, resultant thermophysical properties, process parameters, and deposit characteristics using dimensionless numbers, however, at the expense of losing individual contributions[23]. Designing alloy compositions for targeted deposit requirements, as specified in Integrated Computational Materials Engineering, requires integrating databases with tools for modeling and simulation[24]. To date, most of the multiscale modeling for L-DED to capture the chemical and material research and the fundamental physics has been a sequential process, disregarding potential correlations between these scales. To address this outstanding knowledge gap, this paper proposes a *co-design* approach that allows for the integration of materials, processes, and systems by exploring a high-dimensional nonlinear design space. Linkages are developed for steady-state melt pool properties over an expansive process parameter space for different nickel-based superalloys by incorporating physics-based analytical modeling and data-driven analysis. Several nickel-based superalloys are selected, and their thermophysical properties are evaluated using JMatPro®[25]. All alloys are brought to a common, nineteen-dimensional composition space. An analytical model by Eagar and Tsai[6] is interrogated to obtain the steady-state melt pool depth for different alloy compositions and process parameter combinations. Finally, surrogate-assisted statistical learning frameworks based on Gaussian process (GP) multidimensional regression are developed to model the process-composition-melt pool property linkages. Two kinds of process development strategies (as depicted in Figure 1) are presented in this paper:

- One-Step Surrogate Modeling: Linking composition and process parameters to melt pool depths

- Two-Step Surrogate Modeling: First, the composition space of the alloys is linked to their respective thermophysical properties. These predicted thermophysical properties, along with process parameters, are then linked to melt pool depths.

While one-step surrogates allow for high-dimensional modeling incorporating the effect of individual elements, two-step surrogate allows for similar investigations with a lower degree of complexity by relating alloys to their thermophysical properties.

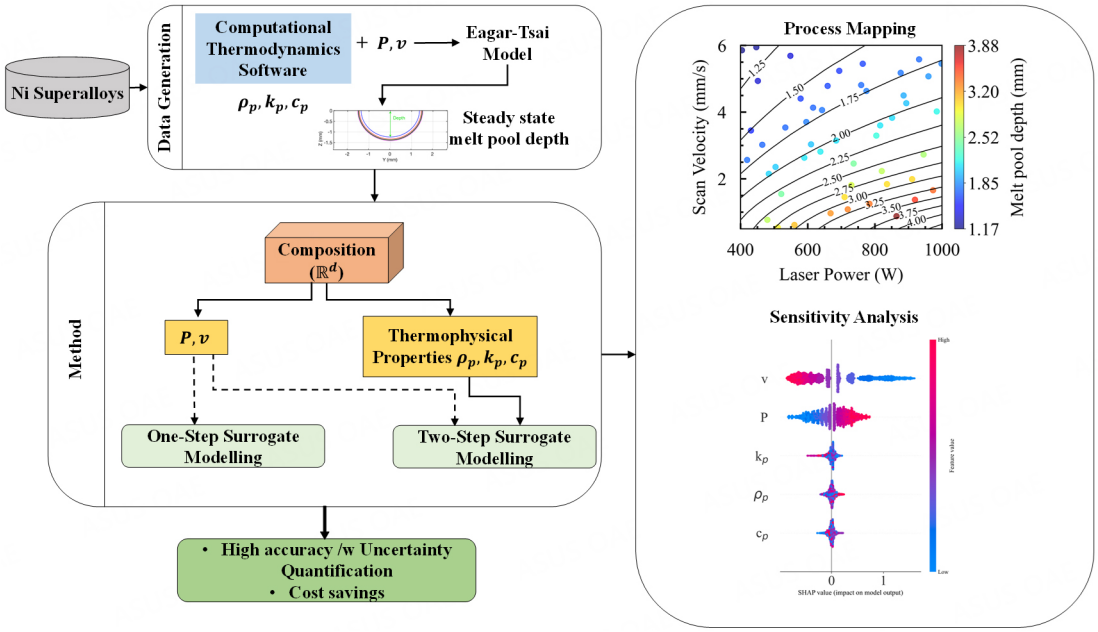


**Figure 1.** Overall framework for developing the linkages between composition, process parameter, and melt pool property. Here, $P$ is the laser power, $v$ is the scan velocity, and $\rho_p$, $k_p$, and $c_p$ are the density, thermal conductivity, and specific heat of the alloys, respectively.

## MATERIALS AND METHODS

This section briefly presents the procedure for estimating the thermophysical properties of nickel-based superalloys, followed by a brief discussion on the thermal model and GP surrogate development. The overall framework is shown in Figure 1.

### Material and thermophysical properties

Nickel-based superalloys contain up to 40 wt.% of a combination of five to ten elements in addition to nickel (Ni) to optimize their performances for high-temperature applications[26]. Supplementary Table 1 in the supplementary document lists the composition of the fifty-eight superalloys obtained from the Cannon Muskegon database[27], where every superalloy is represented as constituent fractions of 19 elements. The thermophysical properties - thermal conductivity ($k_p$), specific heat ($c_p$), and density ($\rho_p$) - are calculated at 300 K using JMatPro®. To better visualize the distribution of the 58 alloys in the 19-dimensional composition space, dimensionality reduction via Principal Component Analysis is carried out. The space is decomposed into two principal components. Applying a non-hierarchical cluster analysis, i.e., the k-mean method, helped identify clusters of alloys on a distance matrix in the multidimensional space. Figure 2A shows the results from the k-means algorithm for k = 4. Three distinct clusters are observed, while the two orange points show irregular behavior as indicated by the distance from their cluster center. Figure 2B shows a similar distribution of the 58 alloys over the thermophysical property space and their projections on pairwise property planes. Similar clusters are also observed in this 3D space, indicating that multiple combinations of elements can result in similar thermophysical properties. Such consistent behavior of thermophysical properties with composition is vital when determining the choice of the kernel for the formulation of the GP surrogate.

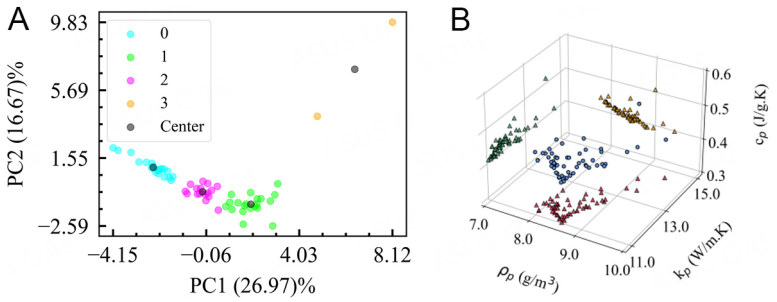


**Figure 2.** (A) Result of the k-means clustering algorithm where each color denotes a cluster in the reduced composition space. Here, PC1 and PC2 are the two principal components obtained after the Principal Component Analysis decomposition of the 19-dimensional space; (B) a 3D plot of the distribution of the alloys in the thermophysical property space of $\rho_p$, $k_p$, $c_p$ at 300 K, indicated by the blue circles. The projection of these points on the pairwise planes of $\rho_p$ - $c_p$, $c_p$ - $k_p$, $k_p$ - $\rho_p$ are indicated by the yellow, green, and red triangles.

### Thermal model for predicting melt pool properties

To obtain the melt pool property data, the present work implements a widely used analytical model developed by Eagar and Tsai[6]. By modifying Rosenthal's solution[28] for a steady-state point heat source, Eagar-Tsai's model solves for the three-dimensional temperature field produced by a traveling Gaussian heat source on a semi-infinite plate. The model significantly improves the prediction of the thermal field near the heat source over the parent model. The simplifying assumptions made by Eagar-Tsai's model are: (i) absence of convective and radiative heat flow; (ii) constant average thermal properties; (iii) quasi-steady state semi-infinite medium; and (iv) no phase change in the material. Figure 3A explains the coordinate system used in the Eagar-Tsai model where $(x_c, y_c, z_c)$ denotes a particular location. The heat source is traveling with a uniform speed of $v$ in the $X$-direction and is assumed to be a 2D surface Gaussian. According to Eagar-Tsai's model, the temperature $T(x_c, y_c, z_c, t)$ at a location, $(x_c, y_c, z_c)$, and time, $t$ is given by:

$$T(x_c, y_c, z_c, t) - T_0 = \frac{\alpha_L P}{\pi \rho_p c_p \left(4\pi a_p\right)^{1/2}} \int_0^t \frac{dt'(t-t')^{-1/2}}{2a_p(t-t') + {\sigma_L}^2} e^{-\frac{(x_c - vt')^2 + y_c^2}{4a_p(t-t') + 2\sigma_L^2} - \frac{z_c^2}{4a_p(t-t')}} \quad (1)$$

Here, $T_0$ is the initial temperature of the substrate, $P$ is the laser power, $v$ is the scan velocity, $\alpha_L$ is the absorptivity of the laser beam, $a_p \triangleq \frac{k_p}{\rho_p c_p}$ is the thermal diffusivity and $t'$ is a dummy integration variable. The distribution parameter ($\sigma_L$), which is the standard deviation of the Gaussian function, is kept constant at 0.85 in this study. The model is used to replicate single-bead deposits on a substrate of the same material as that being deposited. Once the temperature field is solved, the liquidus temperature is used to demarcate the 3D iso-surface where the temperature equals the liquidus temperature of the material of interest. The iso-surface represents the melt pool boundary, from which the melt pool dimensions are calculated. The theoretical and experimental data for steel, titanium, and aluminum were compared in the open literature and demonstrated by Eager and Tsai with good agreement[6]. Figure 3B shows the results using this model for the steady-state melt pool properties of a popular nickel-based superalloy, IN718, processed via L-DED[29].

### GP-based surrogate modeling

In this paper, surrogate modeling is done to link compositions and process parameters to the melt pool properties. GPs are a class of stochastic processes that assume any finite collection of random variables to follow a multivariate jointly Gaussian distribution. Mathematically, given a scalar-valued function and for a finite collection of inputs, $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T\}$, the corresponding function outputs, $\{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T\}$ are assumed to

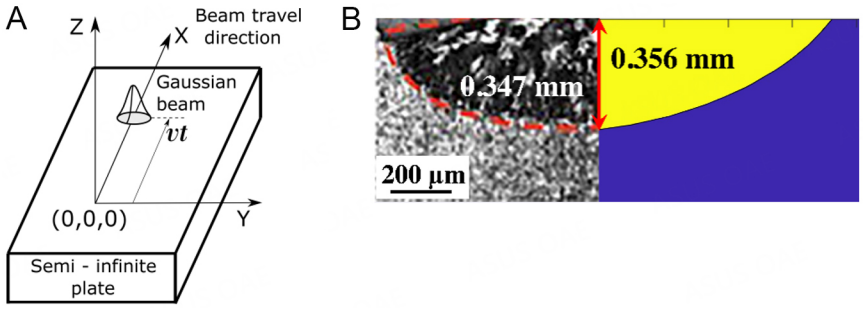


**Figure 3.** (A) Schematic illustrating the coordinate system of the analytical model[11]. This Figure is under the terms of the Creative Commons Attribution 4.0 License from http://creativecommons.org/licenses/by/4.0/; (B) a comparison of the experimental and calculated melt pool dimensions for IN718. This Figure is reproduced with permission from[29].

have a multivariate jointly Gaussian distribution defined by a mean function, $m(\mathbf{x}) \triangleq E[\mathbf{y}]$, where $E$ denotes the expectation, and a covariance function, $k(\mathbf{x}, \mathbf{x}') \triangleq E[(\mathbf{y} - m(\mathbf{x}))(\mathbf{y}' - m(\mathbf{x}'))]$[30]. Here, $\mathbf{x}'$ and $\mathbf{y}'$ denote an input other than $\mathbf{x}$ and the corresponding functional output of it, respectively. For a given training data set, $\mathcal{D}^{trn} = \{(\mathbf{x}_i^{trn}, \mathbf{y}_i^{trn})\}$, $i = 1, ..., N$ and a test dataset, $\mathcal{D}^{tst} = \{(\mathbf{x}^{tst}, \mathbf{y}^{tst})\}$, the conditional distribution of the outputs at the test locations is given by:

$$\mathbf{y}^{tst}|\mathbf{y}^{trn}, \mathbf{x}^{trn}, \mathbf{x}^{tst} \sim \mathcal{N}(\mu^{tst}, \Sigma^{tst}) \tag{2}$$

Here,

$$\mu^{tst} = \mathbf{K}(\mathbf{x}^{tst}, \mathbf{x}^{trn})[\mathbf{K}(\mathbf{x}^{trn}, \mathbf{x}^{trn}) + \sigma_\epsilon{}^2\mathbf{I}]^{-1}\mathbf{y}^{trn} \tag{3}$$

$$\Sigma^{tst} = \mathbf{K}(\mathbf{x}^{tst}, \mathbf{x}^{tst}) - \mathbf{K}(\mathbf{x}^{tst}, \mathbf{x}^{trn})[\mathbf{K}(\mathbf{x}^{trn}, \mathbf{x}^{trn}) + \sigma_\epsilon{}^2\mathbf{I}]^{-1}\mathbf{K}(\mathbf{x}^{trn}, \mathbf{x}^{tst}) \tag{4}$$

Here, $\mathbf{I}$ is the identity matrix, and $\mathbf{K}$ is the covariance matrix. Thus, the predicted posterior distribution of the outputs at every test data point is also a Gaussian distribution, defined by the mean, $\mu^{tst}$ and covariance, $\sum^{tst}$. The noise observed in outputs is modeled with independent and identically distributed (i.i.d.) Gaussian distributions with zero mean and variance, $\sigma_{\in}{}^2$. The accuracy of the prediction depends on the values of the mean and covariance function parameters, which are called hyper-parameters. The hyper-parameters ($\theta$) are learned by minimizing the negative log marginal likelihood (NLML)[30] via a gradient descent method called the limited-memory Broyden Fletcher Goldfarb Shanno (L-BFGS) algorithm[31]:

$$NLML = -log\mathcal{p}(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{2}log|\mathbf{K}| + \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} + \frac{N}{2}log2\pi \tag{5}$$

The Matérn covariance function with a shape parameter of 5/2 is used for each GP in this paper for two reasons: (i) the associated length scales are less susceptible to non-smooth regions in the data set[32]; and (ii) the Matérn kernel does not have the concentration of measure problem for high-dimensional inputs to the extent of the Radial Basis Function kernel[33]. In this paper, two GP surrogates are developed as described below:

• *One-Step Surrogate Modeling*: A single GP surrogate ($GP_{OneStep}$) that directly links composition and process parameters to melt pool depths. This GP explains individual elements and process parameters that

contribute to melt pool depth. The dimension of the input to this GP is 21 (19-dimensional composition + 2 process parameters).

• *Two-Step Surrogate Modeling*:

1. A set of GPs that links the composition space to the thermophysical properties- $\rho_p$, $k_p$, $c_p$. These GPs are indicated as $\mathrm{GP}_{\mathrm{TwoStep}_{[\mathrm{Prop}]}}$ where [Prop] represents $\rho_p$, $k_p$, or $c_p$. These GPs explain how individual elements affect thermophysical properties. The dimension of the input to this GP is 19 corresponding to the individual elements that make up the composition of the alloy.

2. A second GP, $\mathrm{GP}_{\mathrm{TwoStep}_{[\delta]}}$ that links predicted thermophysical properties and process parameters to the melt pool depths to explain how thermophysical properties and process parameters alter melt pool depths. The dimension of the input is 5 (3 thermophysical properties as predicted by $\mathrm{GP}_{\mathrm{TwoStep}_{[\mathrm{Prop}]}}$ +2 process parameters).

Python libraries, GPy[34], and scikit-learn[35] are used to implement the GP regression surrogate. The surrogates are evaluated on different sizes of the training data. This ensures the surrogates' generalizability and stability and helps determine the minimum training data sufficient to develop a robust surrogate. To quantify the performance, the following metrics are selected:

• Relative $L_2$ error norm expressed as $\mathrm{L}_2 \triangleq \frac{\|\hat{\lambda}-\lambda\|_2}{\|\lambda\|_2}$. Here, $\hat{\lambda}$ denotes the mean of the posterior predictive Gaussian distribution, and $\lambda$ corresponds to the true value of the variable. $\|\cdot\|_2$ corresponds to the Euclidean norm.

• Coefficient of Determination or $R^2$ indicates the variation of a dependent variable explained by the independent variable(s) and is indicative of a statistical measure of fit of a regression surrogate.

• 2σ-band coverage percentage is the proportion of test points for which the true value is within ±2σ of the mean of the predicted posterior Gaussian distribution, which approximately corresponds to the 95% confidence interval (CI). Here, σ corresponds to the standard deviation of the posterior predictive Gaussian distribution.

The prediction performance evaluation is further supplemented by examining each surrogate's residual plots for the prediction on the test points. Residual plots indicate the distribution of residuals or the difference between the predicted and true responses. Typically, a good surrogate would produce residual plots with the following characteristics: (i) residuals symmetrically distributed about "0" with a higher density of points clustered closer to "0"; (ii) absence of any distinct nonlinear pattern; (iii) absence of outliers or significantly large residuals; and (iv) non-constant variation of residuals or heteroscedasticity[36]. Since GP regression surrogates are developed assuming that residuals are Gaussian and have the same variance for all observations, such heteroscedastic residuals can affect the predictive performance of the GP regression surrogate[37].

## RESULTS AND DISCUSSION

### Prediction of the melt pool depth using Eagar-Tsai model

The steady-state melt pool depths for fifty-eight different nickel-based superalloys are obtained using the Eagar-Tsai model for several combinations of laser power and scan velocity. The laser power ($P$) is varied from 400 W to 1,000 W with increments of 50 W, and the scan velocity ($v$) is varied from 0.5 mm/s to 6

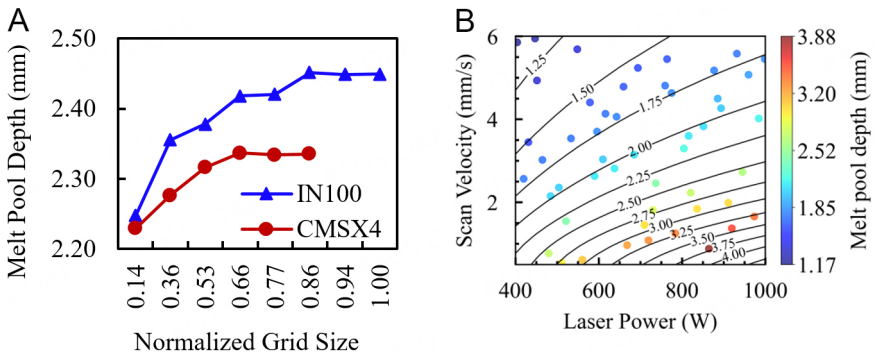


**Figure 4.** (A) Grid convergence results for CMSX-4® and IN100 for $P$ = 300 W and $v$ = 0.5 mm/s, (B) melt pool depth of IN100 superimposed on melt pool depth for CMSX-4®, over the same process parameter space. The black curves correspond to the melt pool depth contours for CMSX-4®, while the markers correspond to the melt pool depth of IN100 whose depth is matched to the color scale associated with the plot.

mm/s with increments of 0.5 mm/s. The power and velocity ranges are selected based on typical L-DED processing parameters. In total, the full-factorial DoE generates 156 distinct laser power and scan velocity combinations. Hence, for each alloy, a total of 156 simulations are performed, resulting in a total of 58 × 156 = 9,048 simulations for all fifty-eight alloys. Other parameters pertaining to the model are selected after suitable calibration with data available in the literature[11].

Grid convergence is a critical requirement of any computational analysis. It is observed that the domain discretization required to obtain convergent results for the melt pool depth differs with alloys. Figure 4A shows the grid convergence results for two different alloys demonstrating the necessity of performing an alloy-specific grid convergence study due to the differences in thermophysical properties among the alloys. From Figure 4A, it is evident that the melt pool depths for CMSX-4® and IN100 at the same $P$ and $v$ values (here, $P$ = 300 W and $v$ = 0.5 mm/s) are different, which is further demonstrated in Figure 4B by overlaying the output melt pool depths of both alloys over the same process parameter window. Hence, for the same processing parameters, the alloys yield different melt pool dimensions necessitating varied grid requirements within the melt pool for every $P$ and $v$ combination. The total computational time for one alloy ranges from 22 to 26 h. The simulations are performed on an Intel Xeon Gold 6230 CPU @ 2.1GHz computer with 128 GB of RAM. The computational simulations for all 58 alloys for 156 process parameter combinations are found to consume approximately 1,430 CPU h.

### One-step surrogate modeling

A GP surrogate, $GP_{OneStep}$, that links the composition of the 58 alloys and process parameters directly to the melt pool depth is trained, where the relationship between the process parameters, composition and melt pool depth is a function of the form, $f$: $\mathbb{R}^{21} \rightarrow \mathbb{R}^{1}$. The input is a 21-dimensional space consisting of a 19-dimensional composition space and 2-dimensional printing parameters (i.e., $P$ and $v$). The output is a 1-dimensional space consisting of the melt pool depth, $\hat{\delta}$. Figures 5 and 6 show the regression performance of $GP_{OneStep}$ for different train-test ratios- 80%:20%, 60%:40%, 40%:60%, and 20%:80%. Essentially, it shows the parity plots with the true melt pool depth, $\delta$ plotted against the predicted melt pool depth, $\hat{\delta}$. Typically, an ideal surrogate would have a predicted depth equal to the true depth, and, therefore, the predictions would lie on the $\hat{\delta} = \delta$ line. The vertical distance of the predictions from this line signifies the error. Therefore, a good fit is indicated by a higher density of predictions clustered around this line and scattered randomly about the diagonal line, with minimum bias. For every case, the parity plots in Figure 5 show low $L_2$ and high $R^2$ which are evidence of good statistical fits. There is not much variation in the $R^2$ values, reported as

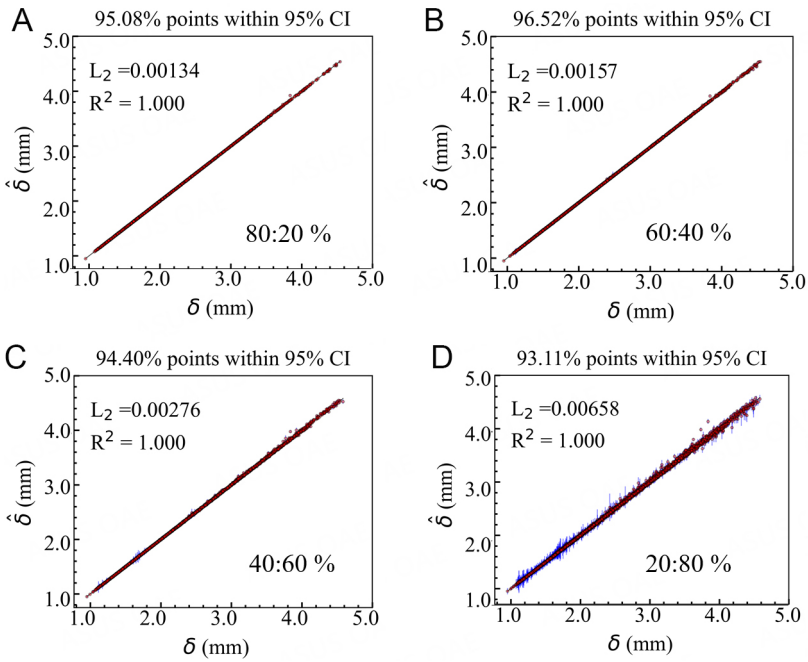


**Figure 5.** Parity plot indicating the prediction performance of $GP_{OneStep}$ on the test set, for different train- test ratios: (A) 80%:20%; (B) 60%:40%; (C) 40%:60%; and (D) 20%:80%. The predicted melt pool depth ($\hat{\delta}$) is reported on the Y-axis and the true melt pool depth ($\delta$) is on the X-axis. The red circles indicate the GP mean and the blue bars are the $\pm 2\sigma$ bands.

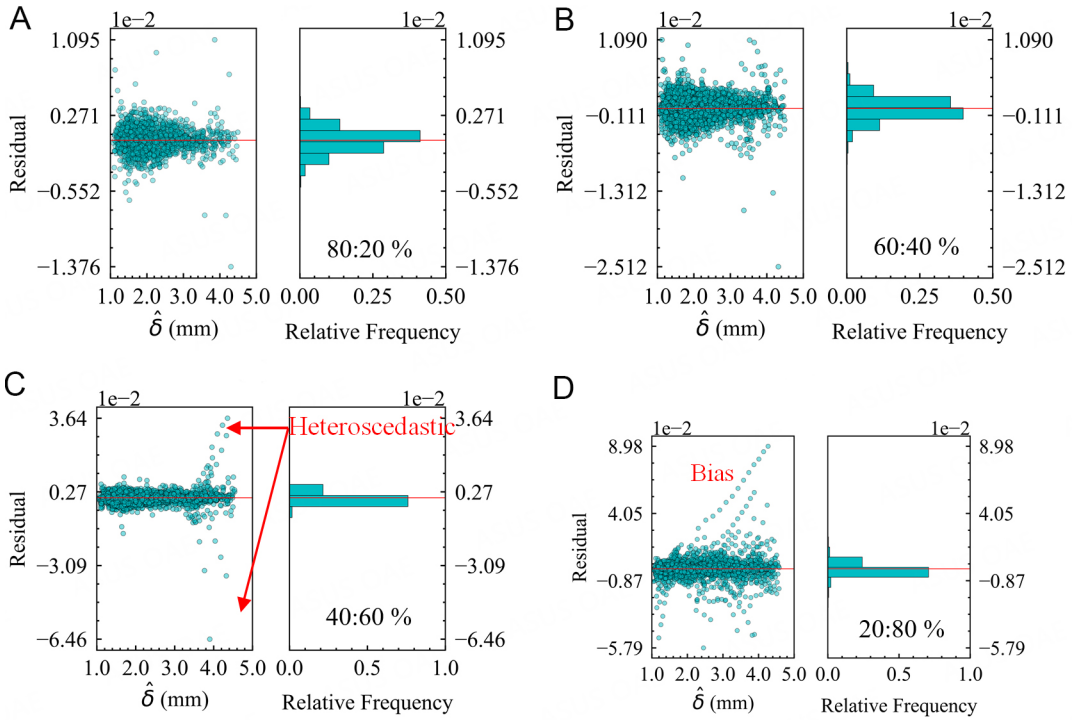


**Figure 6.** Residuals plotted for $GP_{OneStep}$ predictions on the test data for the train-test ratio: (A) 80%:20%; (B) 60%:40%; (C) 40%:60%; and (D) 20%:80%. In each sub-figure, the left-hand plot shows the residuals plotted against the predicted melt pool depth ($\hat{\delta}$) and the right-hand plot shows the frequency distribution of the residuals for each case of the train-test ratio.

~0.999 for all training data, implying the surrogate predicts well even with fewer data. The proportion of predictions located within 95% CI or $\pm 2\sigma$ band coverage decreased from 95% to 93%, with a reduction in training data from 80% to 20% of total data. The variance represented by the length of $\pm 2\sigma$ bar also increases with training data reduction. A higher number of points with bias are observed as training points are reduced for higher values of melt pool depth. Furthermore, bias and errors are investigated using the residual plots in Figure 6. Most of the residual points are clustered around the *Residual* = 0 line. Specifically, these plots are checked for heteroscedasticity, which none of the current predictions appears to suffer from, although outliers are present in almost all cases Figure 6. However, both underpredicted and overpredicted values are observed for higher melt pool depth in Figure 6C. For the train-test ratio of 20:80%, in Figure 6D, the bias is observed to increase as the predicted melt pool depth is higher than the true value. Such a behavior can be attributed to larger ranges of δ in the test data set of larger volume. The histogram of residuals observes a somewhat normal distribution, although with a non-zero mean.

Overall, it can be concluded that the trained GP regression-based approach successfully learns the relationship between input and output data to predict the steady-state melt pool depth for unseen input test data with good accuracy and confidence. The GP surrogate performs well even for the lowest train-test ratio of 20%:80%, highlighting its ability to provide accurate predictions with uncertainty quantification even with small datasets. However, a 60%:40% ratio would be recommended to obtain a more robust, reliable surrogate for further predictions and analysis.

### Two-step surrogate modeling

*Linking composition to the thermophysical properties*

For a given thermophysical property, the data for the multivariate GP surrogate consists of $N_{alloy}$ inputs where each input is the composition of an alloy in the space $\mathbb{R}^{19}$. These inputs are paired with the corresponding thermophysical property as the output. Since there are only $N_{alloy} = 58$ alloys, 20% of the data is held out as a test set on which the prediction performance of the trained GP surrogate is carried out. The performance is quantified using the relative $L_2$ error norm between the mean of the posterior predictive Gaussian distribution and the true thermophysical properties as obtained from JMatPro®. The parity plots for each $GP_{TwoStep_{[Prop]}}$ are shown in Figure 7. The relatively large amount of model uncertainty may be due to - (i) limited data- 80% of 58 alloys is not sufficient to learn in a space that is as large as the composition; (ii) the regression of clustered data is complicated by possible correlations between composition from the same cluster, as seen in Figure 2. The highest $L_2$ error is observed for surrogate predictions of thermal conductivity due to the larger variability in the thermal conductivity data compared to the other properties.

*Linking thermophysical properties to the melt pool depth*

This section elucidates how the GP surrogate learned to link the predicted thermophysical properties combined with process parameters $P$ and $v$ to steady-state melt pool depths. Figures 8 and 9 illustrate the performance of the $GP_{TwoStep_{[\delta]}}$. This surrogate is trained to link thermophysical properties and process parameters to the melt pool depth as a function $f: \mathbb{R}^5 \rightarrow \mathbb{R}^1$. The $R^2$ value saturates at 0.999 while the $L_2$ value increases from $4 \times 10^{-4}$ to $6 \times 10^{-3}$ upon reducing the train-test ratio from 80%:20% to 20%:80%. In comparison to the observations made for $GP_{OneStep}$ [Figure 5], a significant reduction in $L_2$ values is observed for 80% and 60% of the training data, by an order of magnitude. This can be associated with the reduction in the dimensionality of the input space. Additionally, the comparable level of accuracy for the train-test ratios of 40%:60% and 20%:80% indicates that the thermophysical properties are an appropriate feature representation of the 19-dimensional composition space for designing predictive surrogates. Figure 8 also demonstrates that 86%-96% of the predictions lie within the $\pm 2\sigma$ indicating reasonable confidence in the prediction.

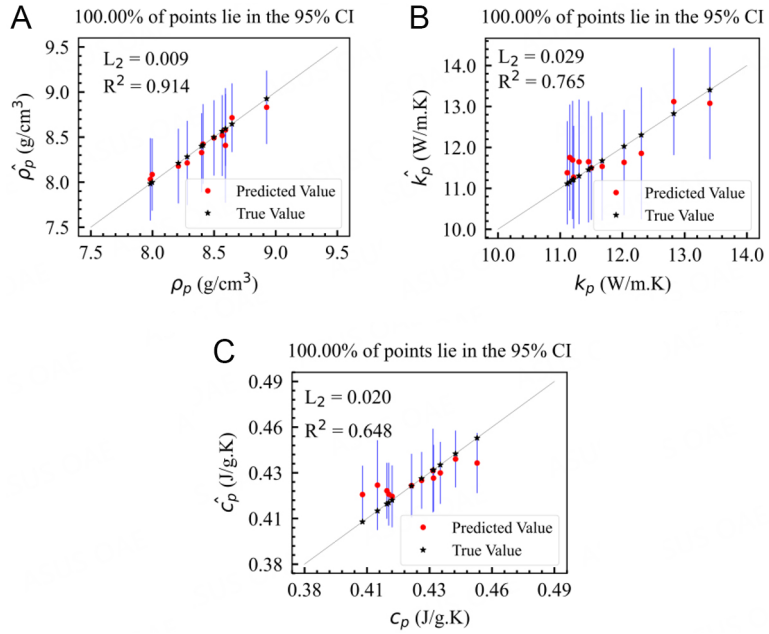


**Figure 7.** Parity plots indicating the prediction performance exhibited on the test set, containing 20% of total data, for the thermophysical properties: (A) density by $^{GP_{TwoStep_{[\rho_p]}}}$; (B) thermal conductivity by $^{GP_{TwoStep_{[k_p]}}}$; and (C) specific heat by $^{GP_{TwoStep_{[c_p]}}}$. The predicted and true properties are reported on the Y- and X- axes respectively. The red circles indicate the GP mean and the blue bars correspond to the ±2σ bands. The relative $L_2$ error norm and number of points within the 95% CI are also reported on each plot.

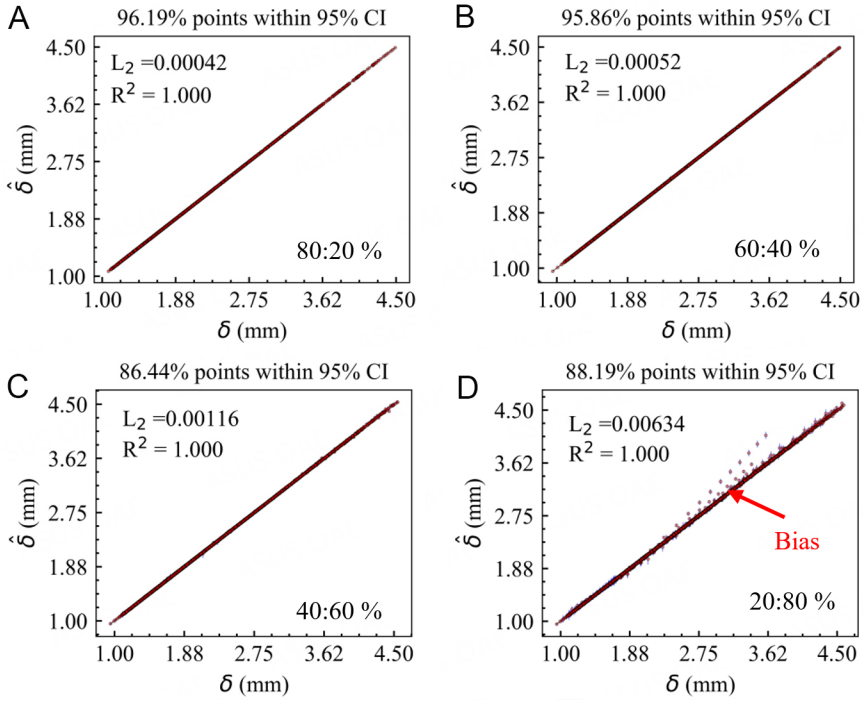


**Figure 8.** Parity plot indicating the prediction performance of $^{GP_{TwoStep_{[\delta]}}}$ on the test set for different train-test ratios: (A) 80%:20%; (B) 60%:40%; (C) 40%:60%; and (D) 20%:80%. The predicted melt pool depth ($\hat{\delta}$) is reported on the Y-axis and the true melt pool depth ($\delta$) is on the X-axis. The red circles indicate the GP mean and the blue bars are the ±2$\sigma$ bands, for each case of the train-test ratio.

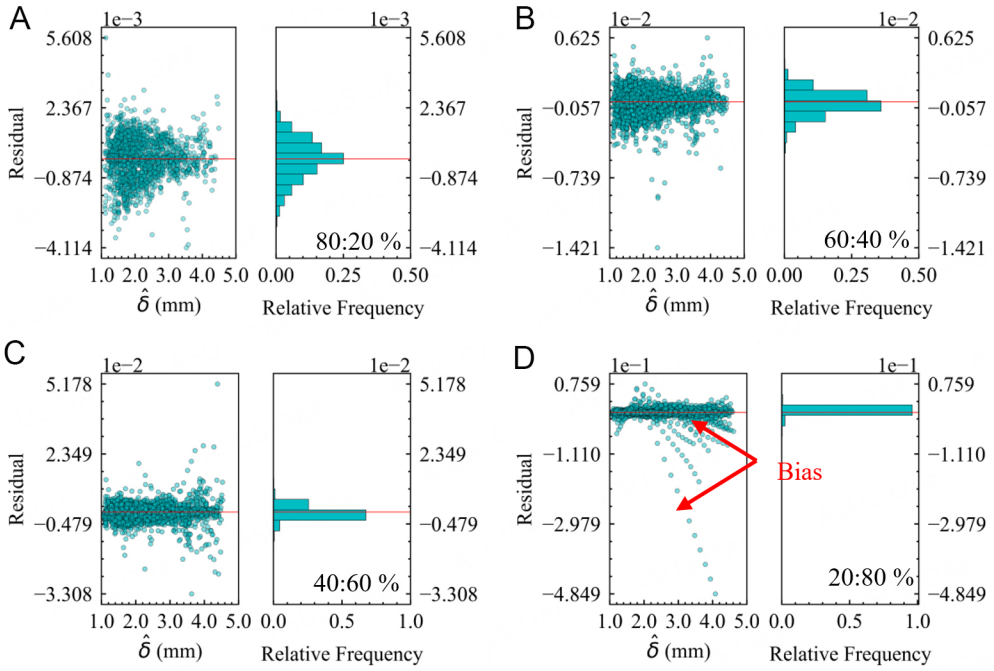


**Figure 9.** Residual plots for $GP_{TwoStep[\delta]}$ predictions on the test data for the train-test ratio: (A) 80%:20%; (B) 60%:40%; (C) 40%:60%; and (D) 20%:80%. In each sub-figure, the left-hand plot shows the residuals plotted against the predicted melt pool depth ($\delta$) and the right-hand plot shows the frequency distribution of the residuals for each case of the train-test ratio.

Unlike Figure 6, the residuals for $GP_{TwoStep[\delta]}$, as seen in Figure 9, show fewer outliers and lower variances, with most points distributed around the "0" line. The reduction in $L_2$ values, observed for 80% and 60% of training data, are also noticeable in the range of residuals observed for Figure 9A and B. For 40% and 20% of the training data, the $GP_{TwoStep[\delta]}$ reports metrics similar to those of $GP_{OneStep}$, with the exception of heteroscedasticity being absent in the former. The growing bias observed in Figure 8D for 20% of training data is confirmed by its corresponding residual plot in Figure 9D. The residuals roughly show a normal distribution as depicted in Figure 9 with Figure 9A showing a zero-mean residual distribution. Considering both the parity and residual plot behavior, a train-test ratio of 60%:40% can be deemed appropriate for $GP_{TwoStep[\delta]}$.

## Sensitivity analysis

The tested GP surrogates are used to perform sensitivity analysis to elucidate and rank uncertainty parameters/inputs to the model. First, a variance-based sensitivity analysis, referred to as the Sobol method[38], is adopted. The Sobol method measures global sensitivity across the entire input space. The influence of each uncertainty parameter on the output is evaluated from the ratio of its variance to the total variance. A game theory approach called the Shapley Additive exPlanations (SHAP) is also implemented. SHAP quantifies the importance of each uncertainty parameter through feature analysis by calculating Shapely values. Input variables with larger absolute Shapley values are considered more important.

The Sobol analysis is carried out using the SALib package[39]. Samples are generated using the Saltelli sampler[40]. The number of samples to be generated is determined via a convergence test such that Sobol calculations converge without any computing errors, as discussed in Supplementary Figure 1. The first-order index represents the contribution to the output variance by a single input variable, and the total-order index accounts for all contributions - by first and higher-order indices arising due to interactions between

the input variables[25]. Such an analysis would help identify which variables to focus on to reduce the overall uncertainty of the model.

Based on the initial analysis, the surrogates trained on 60% of the data are selected for sensitivity analysis. Figure 10A shows the results of Sobol calculations for $GP_{OneStep}$. The first-order ($S1$) and the total-order ($ST$) indices are plotted as a bar graph with corresponding 95% confidence levels indicated by the vertical black bars. Scan velocity and laser power show more impact than the elemental compositions. $S1$ sensitivities alone sum up to ~0.96, showing the effect of higher-order interactions is very low. The pairwise interactions, indicated by $S2$ values, are negligible in comparison but have been provided in Supplementary Table 2 of the supplementary document. Figure 10B presents the SHAP bar plot summarizing the uncertainty parameters' importance in descending order. Here, global significance is denoted by the average of absolute Shapley values per input variable across the data. There are some discrepancies between the order observed in the Sobol analysis [Figure 10A] and the plot in Figure 10B, which can be associated with the sparse, high-dimensional input space. A more informative description is provided in the summary plot, Figure 10C. The vertical dispersion of overlapping points indicates a higher concentration of observations with comparable impact. As expected, the scan velocity inversely affects the melt pool depth, while an increase in laser power increases the melt pool depth. Individual effects of the elements can also be deduced from Figure 10C. Elements such as B, Co, and W have a detrimental effect on the melt pool depth, while elements such as C, Hf, and Ta have a positive effect on the melt pool depth. These findings are supported by the literature, e.g., carbon content is known to decrease the thermal conductivity of a material. Thermal conductivity negatively affects the melt pool depth - higher thermal conductivity allows heat from the melt pool to be transferred along the sides reducing its depth. Thus, an increase in Carbon results in a decrease in $k_p$, thereby giving a deeper melt pool[41]. Similarly, the addition of Boron has been reported to widen melt pools reducing their depths and conforming to the observations made using the SHAP analysis[42].

The results of Sobol and SHAP analyses performed on $GP_{TwoStep[δ]}$ are shown in Figure 11. From the $S1$ values, scan velocity is again observed to have a strong first-order sensitivity, followed by the laser power, in the process parameter space being considered. The thermophysical properties such as $\rho_p$ and $k_p$ have relatively lower influence, while $c_p$ has negligible $ST$ sensitivity. The total-order indices, $ST$, are much larger than the first-order indices for $\rho_p$ and $k_p$, indicating the presence of strong higher-order interactions for these input parameters. Figure 11B shows each input's second-order $S2$ values plotted pairwise. Significant effects of pairwise interactions are observed for $k_p$ - $v$ and $\rho_p$ - $k_p$. The negative values observed for the $S2$ index of $P$ - $k_p$ and $\rho_p$ - $c_p$ can be ignored as these correspond to convergence errors inherent within the SALib package. Summing up, all first- and second-order Sobol indices result in 0.99. This implies that the variance in the output is almost entirely explained by the input parameters and their second-order interaction effects.

Figure 11C and D show that the average Shapely values of $v$ and $P$ dominate over other inputs. Other variables such as $\rho_p$, $k_p$, and $c_p$ show relatively small Shapely values. The importance rankings of $\rho_p$ and $k_p$ contradict those based on Sobol indices; however, they are almost similar in uncertainty parameter importance. The summary plot in Figure 11D shows the magnitude and direction of the impact. Lower values of $v$ have a larger overall influence on melt pool depth than higher values of $v$, indicated by the higher spread of points in the x-direction. Much of the thermophysical property points are clustered around "0" Shapely values except for a handful of high $k_p$ and $\rho_p$ instances resulting in lower melt pool depth and a few points of high specific heat increasing melt pool depth. This corroborates the low ranking attributed to the thermophysical properties in Figure 11C. The influence of $k_p$ is explained in the sensitivity analysis of $GP_{OneStep}$ earlier. A lower density could result in higher volumetric heat flux, increasing the melt pool depth. Meanwhile, the specific heat controls the response rate to heat flux, implying a lower specific heat would increase the melt pool size[43].

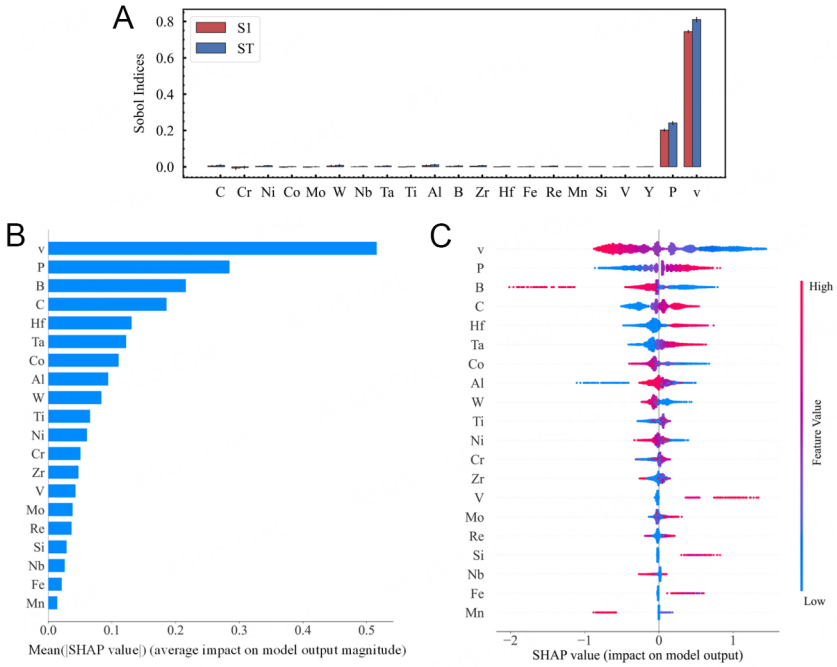


**Figure 10.** Sensitivity Analysis for $GP_{OneStep}$. (A) The first-order and total-order Sobol indices presented as bar graphs for each uncertainty parameter (input variable); (B) absolute mean SHAP plot where aggregate SHAP values for each uncertainty parameter are plotted in descending order; (C) summary of SHAP analysis presented as a beeswarm plot. Each point on the summary plot corresponds to a Shapley value for an input variable and an instance, where the input variables are provided on the y-axis and the Shapely values along the x-axis. For each group, the color of the points is determined by the value of the same uncertainty parameters. The uncertainty parameters are ordered by the mean SHAP values.

### Comparing one-step and two-step surrogate modeling

All the surrogates from each method are assessed for their accuracy, uncertainty, and computational cost. Accuracy is measured as $L_2$. Uncertainty measurements of the predictions are compared using global variance defined as $\sum_{i=1}^{N} \sigma^2(i)$, where $\sigma^2(i)$ is the variance of the posterior predictive Gaussian distribution at a test input indexed with $i$. Since a significant interest, in the development of surrogates, is directed at computational savings, the budget savings are calculated as $(t_{E-T_{\text{test}}} - t_{GP_X})/(\text{Total}\sim t_{E-T}) \times 100\%$. Here, for a given train-test ratio, $t_{E-T_{\text{test}}}$ refers to the time taken by Eagar-Tsai to simulate test points, $t_{GP_X}$ refers to the time taken by the surrogate for the same train-test ratio, and $X$ = OneStep or $\text{TwoStep}_{[\delta]}$. $\text{Total}\sim t_{E-T}$ is the time taken by Eagar-Tsai to simulate all the points in the DoE. Note the time taken by $GP_{\text{TwoStep}_{[\text{Prop}]}} << GP_{\text{TwoStep}_{[\delta]}}$, and, hence, the $GP_{\text{TwoStep}_{[\text{Prop}]}}$ is not included in the calculation. The results are plotted in Figure 12. $GP_{\text{TwoStep}_{[\delta]}}$ is observed to have lower $L_2$ values, with a significant improvement over $GP_{\text{OneStep}}$ as training points are increased. This is expected as the residuals observed for the $GP_{\text{OneStep}}$ were higher by an order of magnitude. This edge in accuracy for $GP_{\text{TwoStep}_{[\delta]}}$ over its counterpart can be attributed to the surrogate operating on a lower-dimensional input space. However, representing the 19-dimensional composition via a 3-dimensional thermophysical property space results in a loss of information, increasing variance, or uncertainty surrounding the predictions. Additionally, the high uncertainty associated with the predicted thermophysical properties also adds to the increase in $\sum_{i=1}^{N} \sigma^{2(i)}$ for $GP_{\text{TwoStep}_{[\delta]}}$. However, adding training points alleviates these issues, as seen from the reduction as the proportion of training points increases from 20% to 40%.

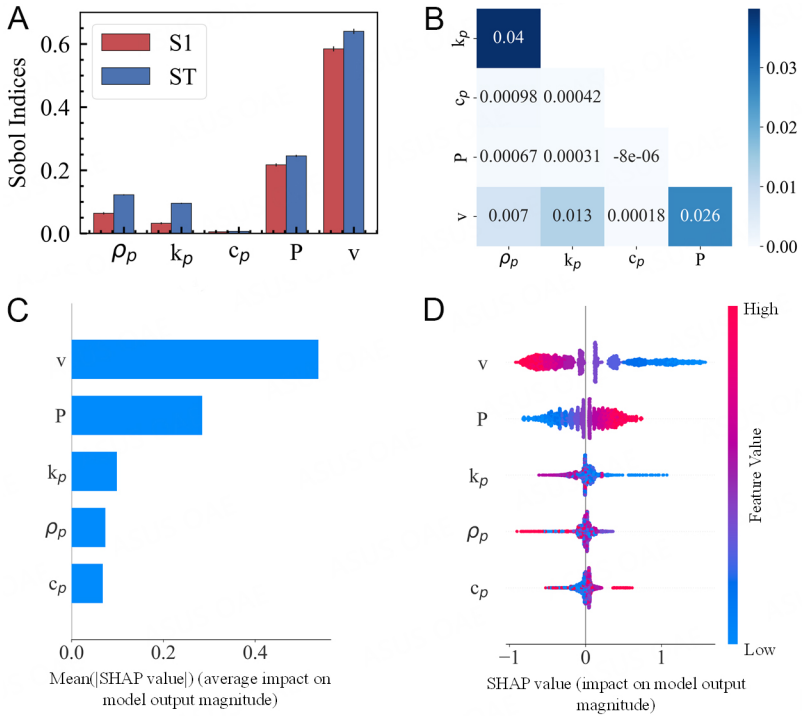


**Figure 11.** Sensitivity Analysis for $GP_{TwoStep_{[\delta]}}$. (A) The first-order and total-order Sobol indices presented as bar graphs for each uncertainty parameter (input variable); (B) second-order Sobol indices as a heat map; (C) absolute mean SHAP plot where aggregate SHAP values for each uncertainty parameter are plotted in descending order; (D) summary of SHAP analysis presented as a beeswarm plot. Each point on the summary plot corresponds to a Shapley value for an input variable and an instance, where the input variables are provided on the y-axis and the Shapely values along the x-axis. For each group, the color of the points is determined by the value of the same uncertainty parameters. The uncertainty parameters are ordered by the mean SHAP values.

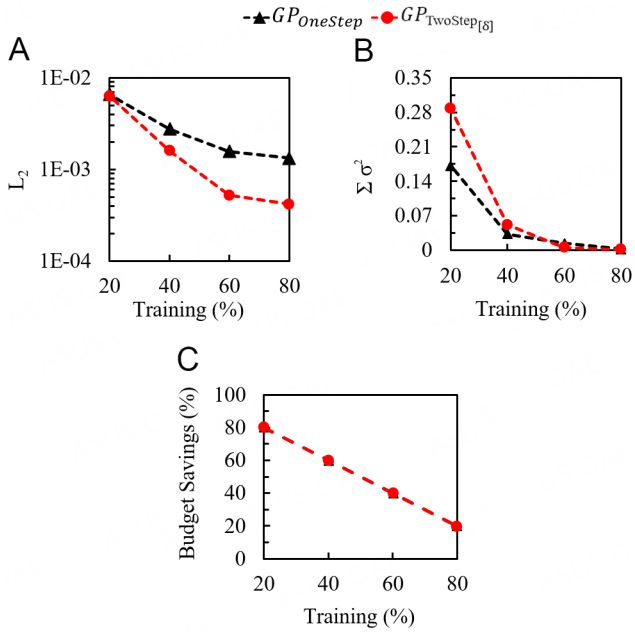


**Figure 12.** Comparison of $GP_{OneStep}$ and $GP_{TwoStep_{[\delta]}}$ based on the metrics (A) $L_2$; (B) global variance, $\sum\sigma^2$; and (C) Budget Savings (%).

Figure 12C shows that the use of surrogate results in 20% to 80% reduction in the cost of simulations, with an increase in the proportion of test points from 20% to 80%. Comparing the budget savings, the cost savings incurred in $GP_{OneStep}$ and $GP_{TwoStep[\delta]}$ are almost similar. While the $GP_{OneStep}$ took nearly 2.5 times the amount taken by $GP_{TwoStep[\delta]}$, the cost of running the surrogates is less by several orders of magnitude compared to running the Eagar-Tsai model (minutes *vs.* several h) over the entire process parameter space. This, therefore, results in similar budget savings for both surrogates.

Some limitations of the present study include the absence of experimental data in the development of the surrogate which could enhance the physical relevance of the GP models. Additionally, the work presented here employs a full-factorial DoE for the process parameters to ensure the reliable performance of the surrogate while handling high-dimensional data. This enables the surrogate to capture the influence of each input feature with lower uncertainty and higher accuracy. However, more efficient strategies of sampling such as Latin Hypercube Sampling need to be explored in the future.

## CONCLUSIONS

A co-design framework using Gaussian process-based regression surrogates is used to generate process parameter property linkages by predicting the steady-state melt pool depth over a process parameter space for 58 nickel-based superalloys. Two methods of surrogate modeling are carried out. Both methodologies are trained and tested on different train-test ratios to ensure the stability of the GPs. While both methodologies are successful in predicting the melt pool depth with excellent accuracy and confidence even with scarce training data, a 60%:40% train-test ratio is deemed appropriate for designing robust surrogates. Sensitivity analysis is important to the development of computationally efficient surrogates by determining the feature importance of input variables. Sobol and SHAP-based sensitivity analyses are carried out using the selected surrogates. Both reveal the scan velocity as the dominant input in the process parameter space being considered. The specific heat is observed to have a relatively negligible impact on the melt pool depth.

The GP framework can, therefore, be used to characterize the process parameter window for a wide range of nickel-based alloys. The framework can also serve as an effective tool to aid active learning techniques for alloy design and development, unearthing previously undiscovered combinations of compositions that yield novel nickel-based superalloys. They can be used to determine the composition of elements to obtain the target thermophysical properties that provide the desirable deposit characteristics. Such a high-dimensional co-design approach, therefore, presents a ubiquitous framework for material discovery, process characterization, and uncertainty quantification. In short, the results establish a path toward creating process development strategies for several different metallic alloys having similar compositions.

## DECLARATIONS

### Authors' contributions

Conceptualization, methodology: Menon N, Mondal S, Basak A
Writing - review and editing: Menon N, Basak A
Software, validation, formal analysis, investigation, data curation, writing - original draft preparation and visualization: Menon N, Mondal S
Resources, supervision, project administration, funding acquisition: Basak A
All authors have read and agreed to the published version of the manuscript.

### Availability of data and materials

The data that support the findings of this study are available from the corresponding author Amrita Basak, upon reasonable request.

### Financial support and sponsorship

The work reported in this paper is funded in part by the Pennsylvania State University, PA 16802, USA, and, in part, by the U.S. Army Engineer Research and Development Center through Contract Number (W912HZ21C0001). Any opinions, findings, and conclusions in this paper are those of the authors and do not necessarily reflect the views of the supporting institutions.

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.



## REFERENCES

1. Jelvani S, Shoja Razavi R, Barekat M, Dehnavi M. Empirical-statistical modeling and prediction of geometric characteristics for laser-aided direct metal deposition of inconel 718 superalloy. *Met Mater Int* 2020;26:668-81. DOI
2. Dinda G, Dasgupta A, Mazumder J. Laser aided direct metal deposition of Inconel 625 superalloy: microstructural evolution and thermal stability. *Mater Sci Eng A* 2009;509:98-104. DOI
3. Zhong C, Pirch N, Gasser A, Poprawe R, Schleifenbaum J. The influence of the powder stream on high-deposition-rate laser metal deposition with inconel 718. *Metals* 2017;7:443. DOI
4. Sreekanth S, Ghassemali E, Hurtig K, Joshi S, Andersson J. Effect of direct energy deposition process parameters on single-track deposits of alloy 718. *Metals* 2020;10:96. DOI
5. Gibson I, Rosen DW, Stucker B. Additive manufacturing technologies. Cham, Switzerland: Springer, 2021.
6. Eagar TW, Tsai NS. Temperature fields produced by traveling distributed heat sources. *Weld J* 1983;62:346-55. DOI
7. Feenstra D, Molotnikov A, Birbilis N. Utilisation of artificial neural networks to rationalise processing windows in directed energy deposition applications. *Mater Design* 2021;198:109342. DOI
8. Juhasz M. Machine learning predictions of single clad geometry in directed energy deposition. Available from: https://osf.io/vxdmz/ [Last accessed on 24 Mar 2023].
9. Akbari P, Ogoke F, Kao N, et al. MeltpoolNet: melt pool characteristic prediction in Metal Additive Manufacturing using machine learning. *Addit Manuf* 2022;55:102817. DOI
10. Ye J, Bab-hadiashar A, Hoseinnezhad R, et al. Predictions of in-situ melt pool geometric signatures via machine learning techniques for laser metal deposition. *Int J Comput Integr Manuf* 2022;Online ahead of print:1-17. DOI
11. Mondal S, Gwynn D, Ray A, Basak A. Investigation of melt pool geometry control in additive manufacturing using hybrid modeling. *Metals* 2020;10:683. DOI
12. Menon N, Mondal S, Basak A. Multi-fidelity surrogate-based process mapping with uncertainty quantification in laser directed energy deposition. *Materials* 2022;15:2902. DOI PubMed PMC
13. Lawrence JR, Waugh D. Laser surface engineering: processes and applications. Elsevier; 2014.
14. Basak A, Das S. Additive manufacturing of nickel-base superalloy IN100 through scanning laser epitaxy. *JOM* 2018;70:53-9. DOI
15. Angel NM, Basak A. On the fabrication of metallic single crystal turbine blades with a commentary on repair via additive manufacturing. *JMMP* 2020;4:101. DOI
16. Graybill B, Li M, Malawey D, Ma C, Alvarado-Orozco JM, Martinez-Franco E. Additive manufacturing of nickel-based superalloys. *Am Soc Mech Eng* 2018;51357:V001T01A015. DOI
17. Wang Z, Zhang L, Li W, et al. A high-throughput approach to explore the multi-component alloy space: a case study of nickel-based superalloys. *J Alloys Compd* 2021;858:158100. DOI
18. Qin Z, Wang Z, Wang Y, et al. Phase prediction of Ni-base superalloys via high-throughput experiments and machine learning. *Mater*

*Res Lett* 2021;9:32-40. DOI
19. Xiong W. Additive manufacturing as a tool for high-throughput experimentation. *J Mater Inf* 2022;2:12. DOI
20. Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *NPJ Comput Mater* 2019;5:21. DOI
21. Zhang Y, Xu X. Lattice misfit predictions via the gaussian process regression for Ni-based single crystal superalloys. *Met Mater Int* 2021;27:235-53. DOI
22. Yu J, Xi S, Pan S, et al. Machine learning-guided design and development of metallic structural materials. *J Mater Inf* 2021;1:9. DOI
23. Mukherjee T, Manvatkar V, De A, Debroy T. Dimensionless numbers in additive manufacturing. *J Appl Phys* 2017;121:064904. DOI
24. Wang W, Yin J, Chai Z, et al. Big data-assisted digital twins for the smart design and manufacturing of advanced materials: from atoms to products. *J Mater Inf* 2022;2:1. DOI
25. Sente Software - JMatPro®. Available from: https://www.sentesoftware.co.uk/jmatpro [Last accessed on 24 Mar 2023].
26. Pollock TM, Tin S. Nickel-based superalloys for advanced turbine engines: chemistry, microstructure and properties. *J Propuls Power* 2006;22:361-74. DOI
27. Cannon Muskegon. Available from: https://cannonmuskegon.com/ [Last accessed on 24 Mar 2023].
28. Rosenthal D. The theory of moving sources of heat and its application to metal treatments. *Trans ASME* 1946;68:849-65. DOI
29. Lee YS, Nordin M, Babu SS, Farson D. Influence of fluid convection on weld pool formation in laser cladding. Available from: https://www.researchgate.net/profile/Yousub-Lee-2/publication/268278864_Influence_of_Fluid_Convection_on_Weld_Pool_Formation_in_Laser_Cladding/links/549af80b0cf2b80371371766/Influence-of-Fluid-Convection-on-Weld-Pool-Formation-in-Laser-Cladding.pdf [Last accessed on 31 Mar 2023].
30. Rasmussen CE. Gaussian processes in machine learning. In: Bousquet O, von Luxburg U, Rätsch G, editors. Advanced Lectures on Machine Learning. Berlin: Springer Berlin Heidelberg; 2004. pp. 63-71.
31. Fletcher R. Practical methods of optimization. John Wiley & Sons; 2013.
32. Duvenaud D. Automatic model construction with Gaussian processes. University of Cambridge; 2014.
33. Le Q, Sarlós T, Smola A. Fastfood-approximating kernel expansions in loglinear time. Available from: http://proceedings.mlr.press/v28/le13-supp.pdf [Last accessed on 24 Mar 2023].
34. GPy by SheffieldML. Available from: https://sheffieldml.github.io/GPy/ [Last accessed on 24 Mar 2023].
35. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. Available from: http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https:/ [Last accessed on 31 Mar 2023].
36. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 1979;47:1287. DOI
37. Wang C, Neal RM. Gaussian process regression with heteroscedastic or non-Gaussian residuals. *arXiv preprint arXiv:1212.6246* :2012. DOI
38. Sobol′ I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comput Simul* 2001;55:271-80. DOI
39. Herman J, Usher W. SALib: an open-source python library for sensitivity analysis. *J Open Source Softw* 2017;2:97. DOI
40. Saltelli A, Ratto M, Andres T, et al. Global sensitivity analysis: the primer. John Wiley & Sons; 2008.
41. Zieliñska M, Yavorska M, Porêba M, Sieniawski J. Thermal properties of cast nickel based superalloys. Available from: https://www.researchgate.net/profile/Jan-Sieniawski/publication/50302351_Thermal_properties_of_cast_nickel_based_superalloys/links/543fade30cf21227a11a8218/Thermal-properties-of-cast-nickel-based-superalloys.pdf [Last accessed on 31 Mar 2023].
42. Chen X, Xie J, Fox P. Direct laser remelting of iron with addition of boron. *Mater Sci Technol* 2004;20:715-9. DOI
43. Ahmed SH, Mian A. Influence of material property variation on computationally calculated melt pool temperature during laser melting process. *Metals* 2019;9:456. DOI

# AUTHOR INSTRUCTIONS

## 1. Submission Overview

Before you decide to publish with *Journal of Materials Informatics* (*JMI*), please read the following items carefully and make sure that you are well aware of Editorial Policies and the following requirements.

### 1.1 Topic Suitability

The topic of the manuscript must fit the scope of the journal. Please refer to Aims and Scope for more information.

### 1.2 Open Access and Copyright

The journal adopts Gold Open Access publishing model and distributes content under the Creative Commons Attribution 4.0 International License. Copyright is retained by authors. Please make sure that you are well aware of these policies.

### 1.3 Publication Fees

The publication fee for each submission is $299. There are no additional charges based on color, length, figures, or other elements. OAE provides expense deduction for authors as appropriate. For more details, please refer to OAE Publication Fees.

### 1.4 Language Editing

All submissions are required to be presented clearly and cohesively in good English. Authors whose first language is not English are advised to have their manuscripts checked or edited by a native English speaker before submission to ensure the high quality of expression. A well-organized manuscript in good English would make the peer review even the whole editorial handling more smoothly and efficiently.

If needed, authors are recommended to consider the language editing services provided by Charlesworth to ensure that the manuscript is written in correct scientific English before submission. Authors who publish with OAE journals enjoy a special discount for the services of Charlesworth via the following two ways.

Submit your manuscripts directly at http://www.charlesworthauthorservices.com/~OAE;

Open the link http://www.charlesworthauthorservices.com/, and enter Promotion Code "OAE" when you submit.

### 1.5 Work Funded by the National Institutes of Health

If an accepted manuscript was funded by National Institutes of Health (NIH), the authors may inform Editors of the NIH funding number. The Editors are able to deposit the paper to the NIH Manuscript Submission System on behalf of the authors.

## 2. Submission Preparation

### 2.1 Cover Letter

A cover letter is required to be submitted accompanying each manuscript. It should be concise and explain why the study is significant, why it fits the scope of the journal, and why it would be attractive to readers, *etc.*

Here is a guideline of a cover letter for authors' consideration:

In the first paragraph: include the title and type (e.g., Original Article, Review, Case Report, *etc.*) of the manuscript, a brief on the background of the study, the question the author sought out to answer and why;

In the second paragraph: concisely explain what was done, the main findings and why they are significant;

In the third paragraph: indicate why the manuscript fits the Aims and Scope of the journal, and why it would be attractive to readers;

In the fourth paragraph: confirm that the manuscript has not been published elsewhere and not under consideration of any other journal. All authors have approved the manuscript and agreed on its submission to the journal. Journal's specific requirements have been met if any.

If the manuscript is contributed to a special issue, please also mention it in the cover letter.

If the manuscript was presented partly or entirely in a conference, the author should clearly state the background information of the event, including the conference name, time and place in the cover letter.

### 2.2 Types of Manuscripts

There is no restriction on the length of manuscripts, number of figures, tables and references, provided that the manuscript is concise and comprehensive. The journal publishes Original Article, Review, Meta-Analysis, Case Report, Commentary, *etc.* For more details about paper type, please refer to the following table.

| Manuscript Type | Definition | Abstract | Keywords | Main Text Structure |
|---|---|---|---|---|
| Original Article | An Original Article describes detailed results from novel research. All findings are extensively discussed. | Structured abstract including Aim, Methods, Results and Conclusion. No more than 250 words. | 3-8 keywords | The main content should include four sections: Introduction, Methods, Results and Discussion. |
| Review | A Review paper summarizes the literature on previous studies. It usually does not present any new information on a subject. | Unstructured abstract. No more than 250 words. | 3-8 keywords | The main text may consist of several sections with unfixed section titles. We suggest that the author includes an "Introduction" section at the beginning, several sections with unfixed titles in the middle part, and a "Conclusion" section in the end. |
| Case Report | A Case Report details symptoms, signs, diagnosis, treatment, and follows up an individual patient. The goal of a Case Report is to make other researchers aware of the possibility that a specific phenomenon might occur. | Unstructured abstract. No more than 150 words. | 3-8 keywords | The main text consists of three sections with fixed section titles: Introduction, Case Report, and Discussion. |
| Meta-Analysis | A Meta-Analysis is a statistical analysis combining the results of multiple scientific studies. It is often an overview of clinical trials. | Structured abstract including Aim, Methods, Results and Conclusion. No more than 250 words. | 3-8 keywords | The main content should include four sections: Introduction, Methods, Results and Discussion. |
| Systematic Review | A Systematic Review collects and critically analyzes multiple research studies, using methods selected before one or more research questions are formulated, and then finding and analyzing related studies and answering those questions in a structured methodology. | Structured abstract including Aim, Methods, Results and Conclusion. No more than 250 words. | 3-8 keywords | The main content should include four sections: Introduction, Methods, Results and Discussion. |
| Technical Note | A Technical Note is a short article giving a brief description of a specific development, technique or procedure, or it may describe a modification of an existing technique, procedure or device applied in research. | Unstructured abstract. No more than 250 words. | 3-8 keywords | / |
| Commentary | A Commentary is to provide comments on a newly published article or an alternative viewpoint on a certain topic. | Unstructured abstract. No more than 250 words. | 3-8 keywords | / |
| Editorial | An Editorial is a short article describing news about the journal or opinions of senior editors or the publisher. | None required | None required | / |
| Letter to Editor | A Letter to Editor is usually an open post-publication review of a paper from its readers, often critical of some aspect of a published paper. Controversial papers often attract numerous Letters to Editor | Unstructured abstract (optional). No more than 250 words. | 3-8 keywords (optional) | / |
| Opinion | An Opinion usually presents personal thoughts, beliefs, or feelings on a topic. | Unstructured abstract (optional). No more than 250 words. | 3-8 keywords | / |
| Perspective | A Perspective provides personal points of view on the state-of-the-art of a specific area of knowledge and its future prospects. Links to areas of intense current research focus can also be made. The emphasis should be on a personal assessment rather than a comprehensive, critical review. However, comments should be put into the context of existing literature. Perspectives are usually invited by the Editors. | Unstructured abstract. No more than 150 words. | 3-8 keywords | / |

## 2.3 Manuscript Structure

### 2.3.1 Front Matter

#### 2.3.1.1 Title

The title of the manuscript should be concise, specific and relevant, with no more than 16 words if possible. When gene or protein names are included, the abbreviated name rather than full name should be used.

#### 2.3.1.2 Authors and Affiliations

Authors' full names should be listed. The initials of middle names can be provided. Institutional addresses and email addresses for all authors should be listed. At least one author should be designated as corresponding author. In addition, corresponding authors are suggested to provide their Open Researcher and Contributor ID upon submission. Please note that any change to authorship is not allowed after manuscript acceptance.

#### 2.3.1.3 Abstract

The abstract should be a single paragraph with word limitation and specific structure requirements (for more details please refer to Types of Manuscripts). It usually describes the main objective(s) of the study, explains how the study was done, including any model organisms used, without methodological detail, and summarizes the most important results and their significance. The abstract must be an objective representation of the study: it is not allowed to contain results which are not presented and substantiated in the manuscript or exaggerate the main conclusions. Citations should not be included in the abstract.

#### 2.3.1.4 Keywords

Three to eight keywords should be provided, which are specific to the article, yet reasonably common within the subject discipline.

### 2.3.2 Main Text

Manuscripts of different types are structured with different sections of content. Please refer to Types of Manuscripts to make sure which sections should be included in the manuscripts.

#### 2.3.2.1 Introduction

The introduction should contain background that puts the manuscript into context, allow readers to understand why the study is important, include a brief review of key literature, and conclude with a brief statement of the overall aim of the work and a comment about whether the aim was achieved. Relevant controversies or disagreements in the field should be introduced as well.

#### 2.3.2.2 Methods

Methods should contain sufficient details to allow others to fully replicate the study. New methods and protocols should be described in detail while well-established methods can be briefly described or appropriately cited. Experimental participants selected, the drugs and chemicals used, the statistical methods taken, and the computer software used should be identified precisely. Statistical terms, abbreviations, and all symbols used should be defined clearly. Protocol documents for clinical trials, observational studies, and other non-laboratory investigations may be uploaded as supplementary materials.

#### 2.3.2.3 Results

This section contains the findings of the study. Results of statistical analysis should also be included either as text or as tables or figures if appropriate. Authors should emphasize and summarize only the most important observations. Data on all primary and secondary outcomes identified in the section Methods should also be provided. Extra or supplementary materials and technical details can be placed in supplementary documents.

#### 2.3.2.4 Discussion

This section should discuss the implications of the findings in context of existing research and highlight limitations of the study. Future research directions may also be mentioned.

#### 2.3.2.5 Conclusion

It should state clearly the main conclusions and include the explanation of their relevance or importance to the field.

### 2.3.3 Back Matter

#### 2.3.3.1 Acknowledgments

Anyone who contributed towards the article but does not meet the criteria for authorship, including those who provided professional writing services or materials, should be acknowledged. Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgments section. This section is not added if the author does not have anyone to acknowledge.

#### 2.3.3.2 Authors' Contributions

Each author is expected to have made substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data, or the creation of new software used in the work, or have drafted the work or substantively revised it.

Please use Surname and Initial of Forename to refer to an author's contribution. For example: made substantial contributions to conception and design of the study and performed data analysis and interpretation: Salas H, Castaneda WV; performed data acquisition, as well as provided administrative, technical, and material support: Castillo N, Young V.

If an article is single-authored, please include "The author contributed solely to the article." in this section.

#### 2.3.3.3 Availability of Data and Materials

In order to maintain the integrity, transparency and reproducibility of research records, authors should include this section in their manuscripts, detailing where the data supporting their findings can be found. Data can be deposited into data repositories or published as supplementary information in the journal. Authors who cannot share their data should state that the data will not be shared and explain it. If a manuscript does not involve such issue, please state "Not applicable." in this section.

#### 2.3.3.4 Financial Support and Sponsorship

All sources of funding for the study reported should be declared. The role of the funding body in the experiment design, collection, analysis and interpretation of data, and writing of the manuscript should be declared. Any relevant grant numbers and the link of funder's website should be provided if any. If the study is not involved with this issue, state "None." in this section.

#### 2.3.3.5 Conflicts of Interest

Authors must declare any potential conflicts of interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. If there are no conflicts of interest, please state "All authors declared that there are no conflicts of interest." in this section. Some authors may be bound by confidentiality agreements. In such cases, in place of itemized disclosures, we will require authors to state "All authors declare that they are bound by confidentiality agreements that prevent them from disclosing their conflicts of interest in this work.". If authors are unsure whether conflicts of interest exist, please refer to the "Conflicts of Interest" of *JMI* Editorial Policies for a full explanation.

#### 2.3.3.6 Ethical Approval and Consent to Participate

Research involving human subjects, human material or human data must be performed in accordance with the Declaration of Helsinki and approved by an appropriate ethics committee. An informed consent to participate in the study should also be obtained from participants, or their parents or legal guardians for children under 16. A statement detailing the name of the ethics committee (including the reference number where appropriate) and the informed consent obtained must appear in the manuscripts reporting such research.

Studies involving animals and cell lines must include a statement on ethical approval. More information is available at Editorial Policies.

If the manuscript does not involve such issue, please state "Not applicable." in this section.

#### 2.3.3.7 Consent for Publication

Manuscripts containing individual details, images or videos, must obtain consent for publication from that person, or in the case of children, their parents or legal guardians. If the person has died, consent for publication must be obtained from the next of kin of the participant. Manuscripts must include a statement that a written informed consent for publication was obtained. Authors do not have to submit such content accompanying the manuscript. However, these documents must be available if requested. If the manuscript does not involve this issue, state "Not applicable." in this section.

#### 2.3.3.8 Copyright

Authors retain copyright of their works through a Creative Commons Attribution 4.0 International License that clearly states how readers can copy, distribute, and use their attributed research, free of charge. A declaration "© The Author(s) 2023." will be added to each article. Authors are required to sign License to Publish before formal publication.

#### 2.3.3.9 References

Preferably original research articles that directly support the statements should be cited. Review articles could be cited when they specifically address the statement made in the manuscript. An abstract should not be used as a reference. Non-specific citations should be avoided.

References should be numbered in order of appearance at the end of manuscripts. In the text, reference numbers should be placed in square brackets and the corresponding references are cited thereafter. If the number of authors is less than or equal to six, we require to list all authors' names. If the number of authors is more than six, only the first three authors' names are required to be listed in the references, other authors' names should be omitted and replaced with "et al.". Abbreviations of the journals should be provided on the basis of Index Medicus. Information from manuscripts accepted but not published should be cited in the text as "Unpublished material" with written permission from the source.

| Types | Examples |
|---|---|
| Journal articles by individual authors | Weaver DL, Ashikaga T, Krag DN, et al. Effect of occult metastases on survival in node-negative breast cancer. *N Engl J Med* 2011;364:412-21. [PMID: 21247310 DOI: 10.1056/NEJMoa1008108] |
| Organization as author | Diabetes Prevention Program Research Group. Hypertension, insulin, and proinsulin in participants with impaired glucose tolerance. *Hypertension* 2002;40:679-86. [PMID: 12411462] |
| Both personal authors and organization as author | Vallancien G, Emberton M, Harving N, van Moorselaar RJ; Alf-One Study Group. Sexual dysfunction in 1,274 European men suffering from lower urinary tract symptoms. *J Urol* 2003;169:2257-61. [PMID: 12771764 DOI: 10.1097/01.ju.0000067940.76090.73] |
| Journal articles not in English | Zhang X, Xiong H, Ji TY, Zhang YH, Wang Y. Case report of anti-N-methyl-D-aspartate receptor encephalitis in child. *J Appl Clin Pediatr* 2012;27:1903-7. (in Chinese) |
| Journal articles ahead of print | Odibo AO. Falling stillbirth and neonatal mortality rates in twin gestation: not a reason for complacency. BJOG 2018; Epub ahead of print [PMID: 30461178 DOI: 10.1111/1471-0528.15541] |
| Books | Sherlock S, Dooley J. Diseases of the liver and billiary system. 9th ed. Oxford: Blackwell Sci Pub; 1993. pp. 258-96. |
| Book chapters | Meltzer PS, Kallioniemi A, Trent JM. Chromosome alterations in human solid tumors. In: Vogelstein B, Kinzler KW, editors. The genetic basis of human cancer. New York: McGraw-Hill; 2002. pp. 93-113. |
| Online resource | FDA News Release. FDA approval brings first gene therapy to the United States. Available from: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm574058.htm. [Last accessed on 30 Oct 2017] |
| Conference proceedings | Harnden P, Joffe JK, Jones WG, editors. Germ cell tumours V. Proceedings of the 5th Germ Cell Tumour Conference; 2001 Sep 13-15; Leeds, UK. New York: Springer; 2002. |
| Conference paper | Christensen S, Oppacher F. An analysis of Koza's computational effort statistic for genetic programming. In: Foster JA, Lutton E, Miller J, Ryan C, Tettamanzi AG, editors. Genetic programming. EuroGP 2002: Proceedings of the 5th European Conference on Genetic Programming; 2002 Apr 3-5; Kinsdale, Ireland. Berlin: Springer; 2002. pp. 182-91. |
| Unpublished material | Tian D, Araki H, Stahl E, Bergelson J, Kreitman M. Signature of balancing selection in Arabidopsis. *Proc Natl Acad Sci U S A*. Forthcoming 2002. |

For other types of references, please refer to U.S. National Library of Medicine.
The journal also recommends that authors prepare references with a bibliography software package, such as EndNote to avoid typing mistakes and duplicated references.

#### 2.3.3.10 Supplementary Materials

Additional data and information can be uploaded as Supplementary Materials to accompany the manuscripts. The supplementary materials will also be available to the referees as part of the peer-review process. Any file format is acceptable, such as data sheet (word, excel, csv, cdx, fasta, pdf or zip files), presentation (powerpoint, pdf or zip files), image (cdx, eps, jpeg, pdf, png or tiff), table (word, excel, csv or pdf), audio (mp3, wav or wma) or video (avi, divx, flv, mov, mp4, mpeg, mpg or wmv). All information should be clearly presented. Supplementary materials should be cited in the main text in numeric order (e.g., Supplementary Figure 1, Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2, *etc.*). The style of supplementary figures or tables complies with the same requirements on figures or tables in main text. Videos and audios should be prepared in English and limited to a size of 500 MB.

## 2.4 Manuscript Format

### 2.4.1 File Format

Manuscript files can be in DOC and DOCX formats and should not be locked or protected.

### 2.4.2 Length

There are no restrictions on paper length, number of figures, or amount of supporting documents. Authors are encouraged to present and discuss their findings concisely.

### 2.4.3 Language

Manuscripts must be written in English.

### 2.4.4 Multimedia Files

The journal supports manuscripts with multimedia files. The requirements are listed as follows:
Videos or audio files are only acceptable in English. The presentation and introduction should be easy to understand. The frames should be clear, and the speech speed should be moderate.
A brief overview of the video or audio files should be given in the manuscript text.
The video or audio files should be limited to a size of up to 500 MB.
Please use professional software to produce high-quality video files, to facilitate acceptance and publication along with the submitted article. Upload the videos in mp4, wmv, or rm format (preferably mp4) and audio files in mp3 or wav format.

### 2.4.5 Figures

Figures should be cited in numeric order (e.g., Figure 1, Figure 2) and placed after the paragraph where it is first cited;
Figures can be submitted in format of tiff, psd, AI or jpeg, with resolution of 300-600 dpi;
Figure caption is placed under the Figure;
Diagrams with describing words (including, flow chart, coordinate diagram, bar chart, line chart, and scatter diagram, *etc.*) should be editable in word, excel or powerpoint format. Non-English information should be avoided;
Labels, numbers, letters, arrows, and symbols in figure should be clear, of uniform size, and contrast with the background;
Symbols, arrows, numbers, or letters used to identify parts of the illustrations must be identified and explained in the legend;
Internal scale (magnification) should be explained and the staining method in photomicrographs should be identified;
All non-standard abbreviations should be explained in the legend;
Permission for use of copyrighted materials from other sources, including re-published, adapted, modified, or partial figures and images from the internet, must be obtained. It is authors' responsibility to acquire the licenses, to follow any citation instruction requested by third-party rights holders, and cover any supplementary charges.

### 2.4.6 Tables

Tables should be cited in numeric order and placed after the paragraph where it is first cited;
The table caption should be placed above the table and labeled sequentially (e.g., Table 1, Table 2);
Tables should be provided in editable form like DOC or DOCX format (picture is not allowed);
Abbreviations and symbols used in table should be explained in footnote;
Explanatory matter should also be placed in footnotes;
Permission for use of copyrighted materials from other sources, including re-published, adapted, modified, or partial tables from the internet, must be obtained. It is authors' responsibility to acquire the licenses, to follow any citation instruction requested by third-party rights holders, and cover any supplementary charges.

### 2.4.7 Abbreviations

Abbreviations should be defined upon first appearance in the abstract, main text, and in figure or table captions and used consistently thereafter. Non-standard abbreviations are not allowed unless they appear at least three times in the text. Commonly-used abbreviations, such as DNA, RNA, ATP, *etc*., can be used directly without definition. Abbreviations in titles and keywords should be avoided, except for the ones which are widely used.

### 2.4.8 Italics

General italic words like *vs*., *et al., etc*., *in vivo*, *in vitro*; *t* test, *F* test, *U* test; related coefficient as *r*, sample number as *n*, and probability as *P*; names of genes; names of bacteria and biology species in Latin.

### 2.4.9 Units

SI Units should be used. Imperial, US customary and other units should be converted to SI units whenever possible. There is a space between the number and the unit (i.e., 23 mL). Hour, minute, second should be written as h, min, s.

### 2.4.10 Numbers

Numbers appearing at the beginning of sentences should be expressed in English. When there are two or more numbers in a paragraph, they should be expressed as Arabic numerals; when there is only one number in a paragraph, number < 10 should be expressed in English and number > 10 should be expressed as Arabic numerals. 12345678 should be written as 12,345,678.

### 2.4.11 Equations

Equations should be editable and not appear in a picture format. Authors are advised to use either the Microsoft Equation Editor or the MathType for display and inline equations.

### 2.5 Submission Link

Submit an article via https://oaemesas.com/login?JournalId=jmi.

*Journal of Materials Informatics*
(*JMI*)
Los Angeles Office
245 E Main Street ste122, Alhambra,
CA 91801, USA
Tel: +1 323 9987086
E-mail: editorialoffice@jmijournal.net
Website: www.jmijournal.net