

Digital Discovery

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: C. Maguire, C. Hardcastle, T. Hastings, R. Arroyave and B. Vela, *Digital Discovery*, 2026, DOI: 10.1039/D5DD00517E.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Good Enough is Better: Feasibility vs. Pareto-Optimality in Alloy Design

Cayden Maguire^a, Christofer Hardcastle^a, Trevor Hastings^a, Raymundo Arróyave^{a,b,c}, Brent Vela^{a,*}

^aDepartment of Materials Science and Engineering, Texas A&M University, College Station, TX 77843, USA

^bJ. Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, TX 77843, USA

^cWm Michael Barnes '64 Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843, USA

Abstract

In alloy design, the search for candidate materials is often framed as an optimization problem, with the goal of identifying Pareto-optimal solutions across multiple objectives. However, Pareto-optimal solutions do not necessarily satisfy all minimum performance thresholds required for practical deployment. An alternative approach is to treat alloy design as a constraint satisfaction problem, in which the goal is to identify any solution that meets all bare minimum requirements across multiple quantities of interest. These approaches have yet to be benchmarked against each other in the context of realistic alloy design problems. In this work, we demonstrate that, in realistic alloy design campaigns involving multiple objectives and constraints, the constraint satisfaction framework yields a higher likelihood of finding viable alloys than optimization-based approaches. Furthermore, constraint-satisfaction approaches find the first viable alloy solutions earlier than optimization. Our results suggest that focusing on feasibility rather than optimality can lead to more actionable outcomes in materials discovery, particularly in highly constrained applications.

Keywords: Constraint Satisfaction-Based Alloy Design, Gaussian Process Classification, Bayesian Optimization, Active Learning

1. Introduction

For applications in aerospace, energy, and other safety-critical domains, new alloys must undergo rigorous qualification and certification before deployment [1]. Traditionally, alloy discovery efforts have been framed as optimization problems, seeking to identify Pareto-optimal solutions that balance multiple competing objectives [2]. However, in many real-world scenarios, the practical value of such “optimal” candidates is limited if they fail to satisfy strict minimum performance thresholds imposed by application requirements [3]. This raises a fundamental question: Does alloy design need optimality, or should the focus shift toward identifying any alloy that is simply good enough to meet all necessary constraints?

In the materials design literature, multi-objective optimization has been the dominant paradigm, enabled by advances in Bayesian optimization, evolutionary algorithms, and other surrogate-modeling techniques [1, 2, 4–6]. However, the emphasis on optimal trade-offs often overlooks the reality that many candidate materials, while Pareto-efficient, are unsuitable for deployment due to failing one or more hard constraints (e.g., corrosion resistance, manufacturability, phase stability, cost, etc.) [3]. The constraints may be binary [7] or thresholds [8] on continuous properties.

The alternative design strategy is constraint satisfaction frameworks [7, 9–13] which explicitly seek any solution that

satisfies all predefined requirements, regardless of whether it is optimal in the multi-objective sense. While this perspective aligns more directly with the decision-making processes of industrial alloy qualification, its systematic evaluation within the context of high-dimensional, uncertainty-aware Bayesian materials design has been largely absent, with few exceptions [8, 14, 15].

For example, in our prior work [1], we deployed a feasibility-first active-learning pipeline: we devoted the early budget to learning the constraint surfaces until the Shannon entropy of the feasibility map plateaued, indicating a stable decision boundary. We then pruned candidates with low probability of satisfying all constraints and performed multi-objective optimization within the resulting feasible subspace to trace the relevant trade-offs. This two-stage strategy avoids spending effort in infeasible regions and yields Pareto sets that are directly actionable for qualification. Despite the efficiency gain with respect to a brute-force querying of constraints, there were inefficiencies in this method. There is no need to accurately map the decision boundary in regions of the design space where objective values are poor i.e. there is no need to finely resolve the boundary in regions with poor objective values, but to rapidly identify alloys that satisfy all constraints.

This issue was recently addressed with the ANUBIS method [15], which uses a modified optimization acquisition function to account for the probability of feasibility (PoF), e.g. $EI \times PoF$ where EI is expected improvement. ANUBIS adopts a feasibility-aware BO paradigm, learning the unknown constraints on-the-fly with a variational Gaussian process classifier

*Corresponding author email: brentvela@tamu.edu



while simultaneously learning the objectives. This joint treatment focuses effort where high performance and plausible feasibility coincide, learning just enough of the boundary to discriminate promising candidates rather than mapping it everywhere.

However, this paradigm still seeks optimal (albeit feasible) points. Furthermore, all examples provided were single constraints; in practice, alloy design imposes multiple, often coupled constraints. In this paper, we raise the question: “*in realistic alloy design scenarios, are we truly interested in finding the feasible Pareto set?*” We argue that, in closed-loop alloy design, the practical goal is neither to train the most accurate regressor nor to chase a notional “utopia” composition, but to minimize the time it takes to find the *first* feasible alloy to identify a material that satisfies all non-negotiable requirements rapidly. We substantiate this perspective with a benchmark comparing our proposed Bayesian multi-constraint satisfaction against Bayesian multi-objective optimization and a modified Bayesian multi-objective optimization scheme that accounts for feasibility ($EI \times PoF$), with a random search as a baseline. Additionally, we show that the time to find the first feasible alloy can be reduced by equipping Gaussian Process Classifiers (GPCs) with informative prior mean functions.

Constraint satisfaction naturally aligns with stakeholder priorities: certify that a candidate quickly meets pre-registered, application-specific minimum thresholds. Faster feasibility discovery reduces program risk and preserves the runway for scale-up, qualification, and integration [16–18]. Moreover, in black-box, high-dimensional, discrete spaces, global Pareto optimality is not practically provable; at best, one can certify local non-dominance within the sampled set [19], or one can run statistical tests for “Bayesian advantage” by comparing principled optimization outcomes against a baseline random experiment selection policy [20]. By contrast, feasibility is auditable: a candidate either satisfies the thresholds (under uncertainty-aware acceptance tests) or does not. This reframes success around certifiable outcomes, aligns with decision-maker needs, and provides a natural, decision-relevant stopping rule for alloy discovery.

2. Methods

2.1. Gaussian Process

Gaussian processes (GPs) are common surrogates for Bayesian active learning and Bayesian constraint satisfaction. A surrogate model in this context is trained on a limited set of expensive ground-truth evaluations and replaces the expensive black-box experiments or simulations by providing a calibrated posterior over functions over the design space \mathcal{X} . For any design $\mathbf{x} \in \mathcal{X}$, the GP predictive distribution is Gaussian with mean $\mu(\mathbf{x})$ and standard deviation $\sigma(\mathbf{x})$, quantifying both our best estimate and its uncertainty. These uncertainties enable

principled exploration, risk-aware selection, and efficient allocation of experimental and computational budgets to optimize objectives and learn constraints. In other words, the output of a GP is a normal distribution at any point $\mathbf{x} \in \mathcal{X}$, and based on all the normal distributions within \mathcal{X} principled exploration and design of experiments can be done.

The posterior predictive distribution (i.e. the output of a trained GP) is defined by the equations below. An untrained GP is specified by a prior mean function $m(\mathbf{x})$ and a positive-definite covariance (kernel) $k(\mathbf{x}, \mathbf{x}')$. Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N \subset \mathcal{X}$ denote the training designs with noisy observations $\mathbf{y} \in \mathbb{R}^N$ (i.i.d. Gaussian noise with variance σ_n^2). For any test design $\mathbf{x}^* \in \mathcal{X}$, the posterior is Gaussian with mean and variance

$$\mu(\mathbf{x}^*) = m(\mathbf{x}^*) + \mathbf{k}_*^T [\mathbf{K} + \sigma_n^2 \mathbf{I}]^{-1} (\mathbf{y} - m(\mathbf{X})), \quad (1)$$

$$\sigma^2(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{k}_*^T [\mathbf{K} + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{k}_*, \quad (2)$$

where $m(\mathbf{X})$ is the prior mean function evaluated at the training inputs, $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel matrix over the training inputs, and $\mathbf{k}_* = [k(\mathbf{x}_1, \mathbf{x}^*), \dots, k(\mathbf{x}_N, \mathbf{x}^*)]^T$ is the vector of covariances between the training inputs and the test input. A nonzero prior mean can be handled directly via $m(\cdot)$ or equivalently by fitting a zero-mean GP to residuals $\mathbf{y} - m(\mathbf{X})$; both yield the above expressions. In this work, kernel hyperparameters are estimated by maximizing the log marginal likelihood.

2.2. Gaussian Process Kernels

We model covariance with the radial basis function (RBF, squared-exponential) kernel due to its smoothness and robustness for nonlinear response surfaces over composition. In its ARD (automatic relevance determination) form,

$$K_{\text{RBF}}(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}\right), \quad (3)$$

where σ_f^2 is the signal variance and ℓ_d is a per-feature length-scale.

We use ARD to assign an independent length-scale to each elemental fraction (feature). Compositions are expressed in at.%; accordingly, we initialize ℓ_d in composition units and constrain them to broad bounds (e.g., 30–100 at.%), letting marginal-likelihood optimization tune each ℓ_d to the data and grid resolution.

To capture observation noise, we add an independent white-noise term:

$$k_{\text{total}}(x, x') = k_{\text{RBF}}(x, x') + \sigma_n^2 \delta_{xx'}, \quad (4)$$

where σ_n^2 is the noise variance and $\delta_{xx'}$ is the Kronecker delta.



2.3. Gaussian Process Classifier (Categorical)

For categorical constraints, feasibility probabilities are obtained using conventional Gaussian Process Classification (GPC), referred to as categorical GPCs in this work. In this setting, we assume that for any design $\mathbf{x} \in \mathcal{X}$, the probability that the constraint is satisfied is

$$p(y = 1 | \mathbf{x}) = \sigma(f(\mathbf{x})), \quad (5)$$

where $y \in \{0, 1\}$ is a binary feasibility indicator, with $y = 1$ corresponding to the constraint being satisfied. The latent function $f(\mathbf{x})$ is modeled as a Gaussian Process,

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (6)$$

The link function $\sigma(\cdot)$ maps the GP output to probabilities i.e. the link function bounds the latent function between 0 and 1 so it can be interpreted as a probability. Here we use the logistic sigmoid,

$$\sigma(f) = \frac{1}{1 + e^{-f}}. \quad (7)$$

A one-dimensional example illustrating the latent GP and its categorical GP counterpart is shown in Figure 1.

In this work, categorical GPs are used to model binary feasibility constraints during both constraint satisfaction and constrained optimization. This approach is appropriate for modeling single phase stability, as demonstrated in previous work [7]. Furthermore, this implementation is amenable to the incorporation of physics-based prior knowledge, further accelerating the learning process. More details on Gaussian Process classification with informative priors can be found in Ref. [7].

2.4. Gaussian Process Classifier (Continuous)

For continuous constraints, feasibility probabilities are obtained using Gaussian Process (GP) regression on a real-valued quantity of interest. In this setting, for any design $\mathbf{x} \in \mathcal{X}$ we model the latent response $y(\mathbf{x}) \in \mathbb{R}$ with a GP and compute the probability that a thresholded constraint is satisfied. Let the constraint be $y(\mathbf{x}) \geq t$ (the case $y(\mathbf{x}) \leq t$ follows analogously).

We place a GP prior on the latent function,

$$y(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (8)$$

Given data, the GP yields a Gaussian posterior predictive distribution over \mathbf{x} with mean $\mu(\mathbf{x})$ and variance $\sigma^2(\mathbf{x})$. The feasibility probability is then the posterior probability that the threshold is met:

$$p(\text{feasible} | \mathbf{x}) = \mathbb{P}(y(\mathbf{x}) \geq t | \mathbf{x}) = 1 - \Phi\left(\frac{t - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right), \quad (9)$$

Where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. For an upper-bound constraint $y(\mathbf{x}) \leq t$,

$$p(\text{feasible} | \mathbf{x}) = \Phi\left(\frac{t - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right). \quad (10)$$

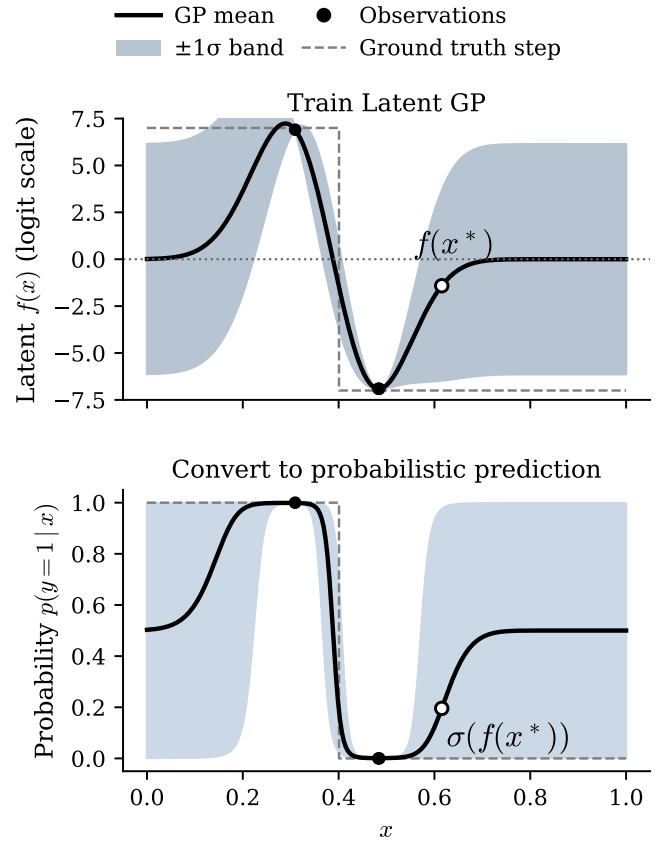


Figure 1: Categorical Gaussian Processes (GPCs) are used to model *binary feasibility constraints* during both constraint satisfaction and constrained optimization. Latent GP $f(x)$ is fit to binary labels. Feasibility probability $p(y=1 | x) = \sigma(f(x))$ with the same uncertainty propagated via $\sigma(\mu_f \pm \sigma_f)$. This approach is appropriate for modeling single-phase stability, as demonstrated previously [7], and is amenable to incorporating physics-based prior knowledge via informative prior mean functions to accelerate learning; see [7] for details.

Figure 2 illustrates how a Gaussian-process regressor (GPR) can be used for continuous classification via thresholding. The left panel shows observations of y , the posterior mean prediction $\mu(x)$, and a credibility band (e.g., $\pm 2\sigma(x)$). The horizontal dashed line marks the feasibility threshold $y = 5$. The marked point at $x = 11$ highlights the model's prediction and its uncertainty (vertical error bar). Under a Gaussian predictive marginal $y(x) \sim \mathcal{N}(\mu(x), \sigma^2(x))$, the probability of satisfying the constraint is

$$\mathbb{P}(y(11) \leq 5) = \Phi\left(\frac{5 - \mu(11)}{\sigma(11)}\right),$$

which converts the continuous surrogate into a probabilistic classifier relative to the constraint.

2.5. Probability of Satisfying All Constraints (PoF)

Let the design $\mathbf{x} \in \mathcal{X}$ be subject to C constraints. For each constraint $c \in \{1, \dots, C\}$, let

$$p_c(\mathbf{x}) = \mathbb{P}(\text{constraint } c \text{ is satisfied} | \mathbf{x})$$



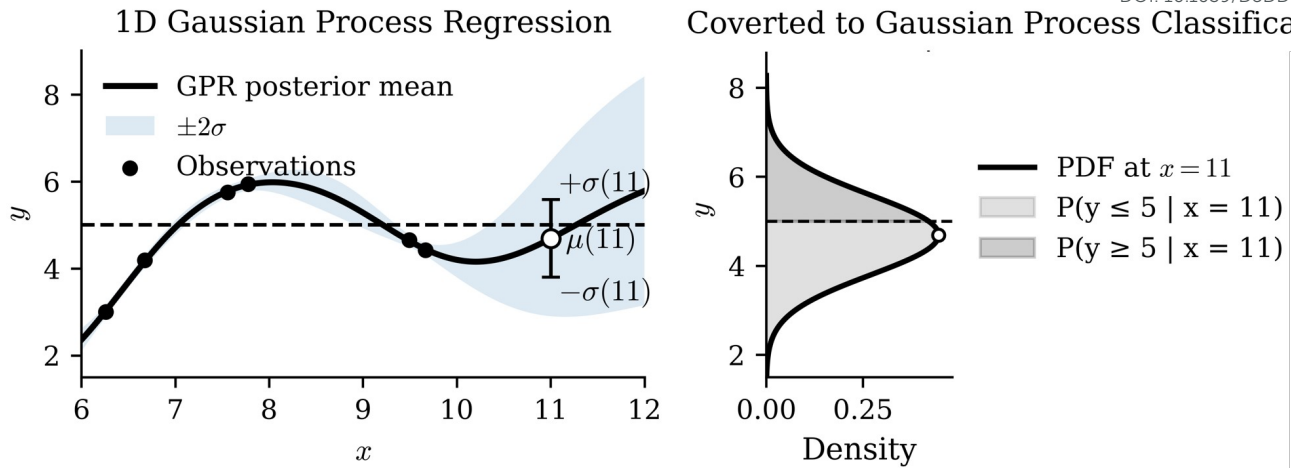


Figure 2: Classification of continuous properties using a GPR. The figure on the left shows a 1D regression problem, where the black dots represent the data used to train the model and the dashed line represents an objective constraint ($y \geq 5$). The mean prediction $\mu(x)$ and standard deviation $\pm\sigma(x)$ is highlighted for the arbitrary input $x = 11$, and its associated posterior predictive distribution is shown on the righthand figure. Since the posterior predictive distribution is a normal distribution, the CDF can be used to determine the probability of meeting the constraint.

denote the (calibrated) feasibility probability produced by its model: for categorical constraints via GPC, $p_c(\mathbf{x}) = \sigma(f_c(\mathbf{x}))$, and for continuous, thresholded constraints via GPR, $p_c(\mathbf{x}) = \Phi((t_c - \mu_c(\mathbf{x}))/\sigma_c(\mathbf{x}))$ or $1 - \Phi(\cdot)$ depending on the inequality direction.

Assuming conditional independence of the constraints given \mathbf{x} and the learned surrogates, the overall *probability of feasibility* (PoF) is the product of the individual probabilities:

$$\text{PoF}(\mathbf{x}) = \prod_{j=1}^C p_j(\mathbf{x}). \quad (11)$$

Figure 3 shows an example of PoF calculations across the design domain. The product in (11) is simple and intuitive; however, it assumes conditional independence of constraints given \mathbf{x} and the fitted surrogates. When constraints are strongly correlated (e.g., share underlying physics or data), (11) can over- or underestimate the joint feasibility and become overconfident. In future work, we will replace the pure product with dependence-aware aggregators, however, the current implementation is shown to be effective for alloy design.

2.6. Hypervolume as an Optimization Metric

In order to benchmark constraint satisfaction against constraint optimization, meaningful metrics must be defined to compare these two campaigns. The natural metric associated with optimization is the size of the current hypervolume, which quantifies the volume of objective space dominated by the current Pareto set relative to a reference point. Hypervolume growth reflects how efficiently an optimization strategy discovers designs that improve upon previously evaluated solutions.

Formally, let $\mathbf{y}_i := \mathbf{f}(\mathbf{x}_i)$ denote the objective vector for design \mathbf{x}_i , and write $\mathcal{P} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ for the current Pareto set. Let $\mathbf{r} \in \mathbb{R}^d$ be a chosen reference point that is dominated by all feasible designs. The hypervolume indicator is defined as the d -dimensional Lebesgue measure (i.e., length in 1D, area in 2D, volume in 3D, and hypervolume in d D) of the region dominated by \mathcal{P} and bounded by \mathbf{r} [22]:

$$\text{HV}(\mathcal{P}, \mathbf{r}) = \lambda_d \left(\bigcup_{\mathbf{y} \in \mathcal{P}} [\mathbf{y}, \mathbf{r}] \right), \quad (12)$$

where $[\mathbf{y}, \mathbf{r}] = \{\mathbf{z} \in \mathbb{R}^d : y_i \leq z_i \leq r_i \forall i\}$ denotes the axis-aligned hyperrectangle between \mathbf{y} and \mathbf{r} . In other words, take each Pareto point \mathbf{y} , draw the axis-aligned box to the reference point \mathbf{r} , take the union of all such boxes, and measure its size—area in 2D, volume in 3D, and hypervolume in d D.

A bi-objective example of this is shown in Figure 4. In Figure 4, the two axes correspond to the objectives f_1 and f_2 , both to be minimized. The black points $\mathbf{y}_1, \dots, \mathbf{y}_4$ represent nondominated solutions that together form the Pareto front. The dashed lines mark the reference point \mathbf{r} , which defines the bounding box for hypervolume calculation. The shaded staircase-shaped region highlights the portion of objective space dominated by the Pareto set and bounded by \mathbf{r} . The total area of this region corresponds to the hypervolume indicator $\text{HV}(\mathcal{P}, \mathbf{r})$, and it increases as additional nondominated points are discovered closer to the origin. This visual emphasizes how hypervolume quantifies simultaneous improvements in both objectives by capturing the volume of objective space excluded from domination by the reference point.

We will track HV as a function of iteration, wanting to maximize it as fast as possible. To ensure a fair comparison between objectives (so that those with large magnitudes do not arbitrarily dominate HV), we compute HV in a scaled objective space



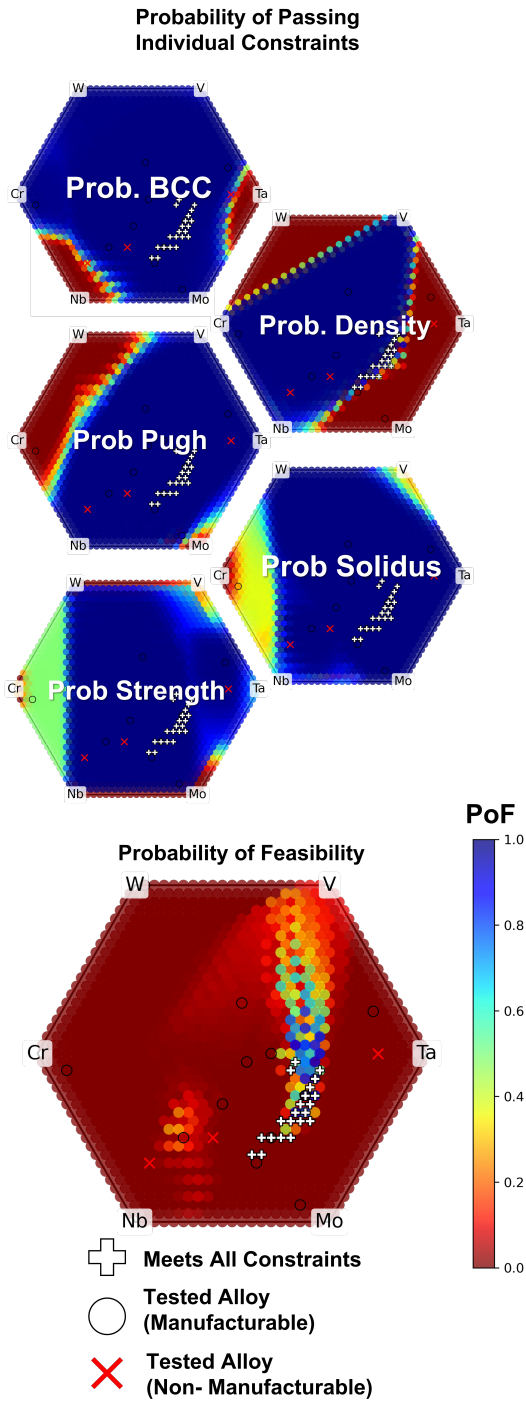


Figure 3: PoF calculations across the design space. The Nb–Mo–Ta–V–W–Cr system is represented using a barycentric projection. The markers show the data used to train the GP. After training, the probability of meeting each individual constraint is computed across the design space (upper five projections). Assuming conditional independence, the PoF is the product of the individual probabilities (bottom projection). Note, overlap occurs on barycentric projections when datapoints are mapped closely together. This limitation is discussed in our previous work [21].

using fixed objective ranges. This is shown in Eq. 13,

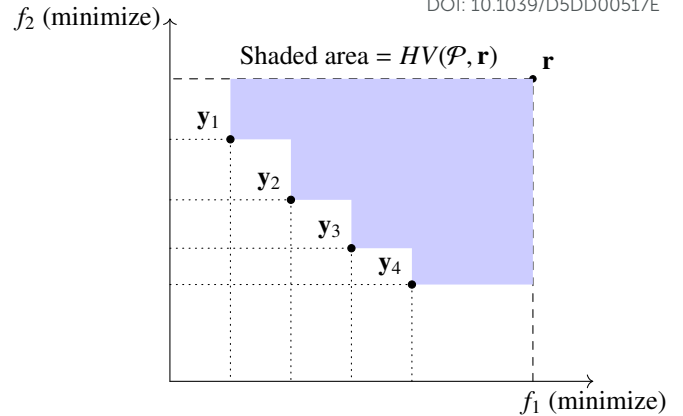


Figure 4: Bi-objective hypervolume (minimization). The shaded staircase is the region dominated by the Pareto set \mathcal{P} and bounded by the reference point \mathbf{r} .

$$HV_{\text{scaled}} = HV\left(\left\{\frac{\mathbf{y}_i - \mathbf{r}}{\Delta}\right\}_{i \in \mathcal{P}}, \mathbf{0}\right) \quad (13)$$

where \mathcal{P} is the current measured Pareto set, $\mathbf{y}_i = [ST_i, -\text{Density}_i, YS_i, \text{Pugh}_i]$ is the vector of measured objective values for input i , $\mathbf{r} = [0, -30, 0, 0]$ is the vector of reference points for each objective, Δ is the vector of objective ranges (max - min), and $HV(\cdot, \cdot)$ is the dominated hypervolume.

2.7. Time to First Feasible Alloy as a Constraint Satisfaction Metric

For constraint-satisfaction active learning, the natural progress metric is the time to discover the first feasible design. Let C constraints be written as $g_c(\mathbf{x}) \leq 0$ for $c = 1, \dots, C$. Define the ground-truth feasibility as

$$F(\mathbf{x}) = \prod_{c=1}^C \mathbb{I}[g_c(\mathbf{x}) \leq 0] \in \{0, 1\}, \quad (14)$$

where $\mathbb{I}[\cdot]$ is the indicator (Iverson) function. This is the deterministic analogue of the *probability of feasibility* $\text{PoF}(\mathbf{x}) = \prod_{c=1}^C p_c(\mathbf{x})$ introduced in Section 2.5: here, the product aggregates *truth values* rather than probabilities of passing. If any of the constraint c is not met, the term $\mathbb{I}[g_c(\mathbf{x}) \leq 0] = 0$ and thus the entire product $F(\mathbf{x}) = 0$ for alloy \mathbf{x} .

With this definition, the time required to identify the first feasible alloy (TTFF) can be expressed either in terms of iterations in the sequential case, or in the batch case, in terms of total alloys investigated. We take the perspective that the iteration-based metric is more significant, as time is the true limiting factor in practical discovery workflows. When operating in batches, the process already leverages economies of scale, reducing the marginal cost per experiment and emphasizing the importance of minimizing the number of sequential iterations rather than the total number of evaluations. However,



for simplicity in this work, we benchmark sequential optimization against sequential constraint satisfaction. For the sequential case, we define TTFF as the iteration where the average cumulative number of feasible queries is greater than or equal to 1.

2.8. Acquisition Function for Constraint Satisfaction Campaigns

Acquisition functions determine *where* the campaign samples next, often trading off exploitation (sampling where success is most likely) and exploration (sampling to reduce uncertainty).

In our constraint-satisfaction setting, we target the feasible set of alloys with an acquisition function,

$$a(\mathbf{x}) = \text{PoF}(\mathbf{x}), \quad (15)$$

where $\text{PoF}(\mathbf{x}) \in [0, 1]$ by definition.

Since only a small percentage of alloys (0.05% of the design space) with similar properties are feasible, an exploration term is omitted from the acquisition function.

2.9. Acquisition Function for Optimization

Regarding multi-objective constrained optimization, if the goal is to find Pareto optimal points and expand/improve the Pareto front, then the AQF should expect hypervolume improvement. For a candidate designs \mathbf{x} with predictive distribution

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma}(\mathbf{x})),$$

the hypervolume improvement relative to the current Pareto set \mathcal{P} and reference point \mathbf{r} is

$$\Delta\text{HV}(\mathbf{y}) = \max\{0, \text{HV}(\mathcal{P} \cup \{\mathbf{y}\}; \mathbf{r}) - \text{HV}(\mathcal{P}; \mathbf{r})\}.$$

The EHVI is the expectation of this improvement under the multivariate normal predictive distribution:

$$\text{EHVI}(\mathbf{x}) = \int_{\mathbb{R}^m} \Delta\text{HV}(\mathbf{y}) \mathcal{N}(\mathbf{y}; \boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma}(\mathbf{x})) d\mathbf{y}.$$

Several forms and approximations exist to calculate EHVI; however for speed (due to the computation expense), we opt to use pEHVI, achieving near-exact accuracy with markedly reduced runtime [23].

The *pointwise* EHVI (pEHVI) computes, for each anchor point $\mathbf{y}^{(i)} \in \mathcal{P}$, the EHVI as if $\mathbf{y}^{(i)}$ were the only Pareto point, then takes the minimum across anchors:

$$\text{pEHVI}(\mathbf{x}) = \min_{\mathbf{y}^{(i)} \in \mathcal{P}} \text{EHVI}_{1\text{pt}}(\mathbf{x}; \mathbf{y}^{(i)}, \mathbf{r}), \quad (16)$$

where $\text{EHVI}_{1\text{pt}}(\cdot)$ denotes the EHVI computed with respect to the single anchored box $[\mathbf{y}^{(i)}, \mathbf{r}]$ [23]. Because this reduces to

independent truncations along each objective, $\text{EHVI}_{1\text{pt}}$ admits a closed form in terms of univariate Gaussian CDF/PDF terms, yielding a cost that is *linear in d* for each anchor.

$$\text{pEHVI}(\mathbf{x}; \mathcal{P}, \mathbf{r}) = \min_{1 \leq j \leq n} \text{EHVI}_{1\text{pt}}(\mathbf{x}; \mathbf{y}^{(j)}, \mathbf{r}), \quad (17)$$

$$\text{EHVI}_{1\text{pt}}(\mathbf{x}; \mathbf{y}, \mathbf{r}) = \prod_{i=1}^m \text{EI}_i(r_i) - \prod_{i=1}^m [\text{EI}_i(r_i) - \text{EI}_i(y_i)], \quad (18)$$

$$\text{EI}_i(t) = (t - \mu_i(\mathbf{x})) \Phi\left(\frac{t - \mu_i(\mathbf{x})}{\sigma_i(\mathbf{x})}\right) + \sigma_i(\mathbf{x}) \phi\left(\frac{t - \mu_i(\mathbf{x})}{\sigma_i(\mathbf{x})}\right), \quad (19)$$

2.10. Models for Ground-Truths and Priors

Constraint-satisfaction and multi-objective constrained-optimization campaigns used the same ground-truth models and prior models, described below.

The solidus temperature and density were calculated in Thermo-Calc using the TCHEA7 thermodynamic database using the *Property* module. The prior models for these properties are rule of mixtures melting temperature and rule of mixtures density, respectively.

Phase stability was evaluated at 600 °C with the *Equilibrium* module, selected as the target temperature for in-house hot rolling. Alloys predicted to be single-phase BCC at 600 °C were deemed manufacturable (a simplifying assumption used here to illustrate a simple manufacturability constraint i.e. if this constraint is not met, other quantities of interest cannot be queried). We define “single-phase BCC” as a BCC mole fraction $\geq 99\%$. The prior for BCC phase stability in this work is

$$m_{\text{BCC}}(\mathbf{x}) = \begin{cases} +5, & \text{VEC}(\mathbf{x}) < 6.87, \\ -5, & \text{otherwise,} \end{cases}$$

which is subsequently bound between 0 and 1 with the logistic sigmoid. $\text{VEC} < 6.87$ is an experimental indicator for single phase BCC [24].

We used the Maresca–Curtin model for yield strength, which posits edge-dislocation glide as the rate-controlling mechanism in BCC HEAs at elevated temperatures [25]. This model provides a conservative (lower-bound) estimate of yield strength for BCC RHEAs [26] and is therefore an appropriate ground-truth model for our in silico alloy-design exercise. The yield strength was queried at 700 °C. The associated prior model was the Maresca-Curtin yield strength queried at 25 °C

As a first-order proxy for ductility, we used the Pugh ratio (B/G), computed from rule-of-mixtures estimates of the bulk and shear moduli. Prior work supports B/G as a reasonable



Table 1: Feasibility criteria used in this work.

Property	Criterion
Solidus temperature	$T_{\text{solidus}} > 2473 \text{ K (2200}^\circ\text{C)}$
Density	$\rho < 9.0 \text{ g cm}^{-3}$
Yield strength	$\sigma_y > 700 \text{ MPa}$
Pugh ratio	$B/G > 2.5$
Phase stability	$x_{\text{BCC}} \geq 99\% \text{ at } 600^\circ\text{C}$

first-approximation indicator of ductility, making it suitable for our in silico design campaigns. We did not have a prior model for ductility in this work and instead relied on a zero-mean prior for the GP.

2.11. Alloy Design Domain

The design domain comprises the 6-element, 5-dimensional, Nb–Mo–Ta–V–W–Cr system. We sampled the composition simplex on a 5 at.% grid, restricting to binaries through quaternaries which yields 40,553 unique candidate alloys.

Within these 40,553 alloys, 22 (0.054%) meet the constraints listed in Table 1, and 12,591 (31.05%) are Pareto-optimal regarding the objectives in Table 2. There are 18 (0.044%) points that are both feasible and Pareto-optimal. These 22 feasible points constitute less than .1% of the initial design space. This is consistent with prior works [12, 27] that application-relevant constraints drastically narrow the apparent breadth of the high-entropy-alloy design space

3. Results and Discussion

3.1. Constraint Satisfaction Campaign

In constraint satisfaction alloy design campaigns, the goal is to rapidly find alloys that meet all bare minimum operational constraints, i.e., minimize the TTFF. In this case study, the goal is to find alloys that meet all of the constraints in Table 1. This set of constraints was chosen to loosely reflect the requirements of real-world alloy design efforts [28].

To initiate the active learning scheme, first an alloy is sampled from the design space and evaluated for BCC phase stability. In our in-silico setup, we treat “processable” as single-phase BCC. If the alloy is single-phase BCC, we then measure the remaining properties in Table 1 and update all classifiers with the new results. If it is not single-phase BCC, we stop there and update only the BCC classifier; no other data are collected for that alloy. This mirrors real world scenarios where before investing in property tests, the alloy must be shown to be processable.

Using the trained GPCs, we estimate the probability of passing each constraint for every candidate alloy in the design

Table 2: Optimization objectives with the remaining manufacturability constraint.

Quantity	Objective
Solidus temperature	maximize T_{solidus} [K]
Density	minimize ρ [g cm^{-3}]
Yield strength	maximize σ_y [MPa]
Pugh ratio	maximize B/G

Constraint: $x_{\text{BCC}} \geq 99\% \text{ at } 600^\circ\text{C}$.

space. We then multiply these probabilities to obtain the overall probability of feasibility (the chance that an alloy satisfies all constraints at once). This product assumes the constraints are conditionally independent under the model and that the classifier outputs are well-calibrated; if constraints are correlated, the result can be biased, but it remains a simple and effective screening heuristic. See the supplemental section for an investigation of a PoF model that accounts for objective dependence.

We select the next alloy using a constraint-satisfaction acquisition function (Eqn. 15) that favors candidates with the highest probability of feasibility. Once this point is selected for querying, the process repeats. This closed-loop is summarized in Figure 5.

3.2. Optimization Campaign

In multi-objective optimization, the goal is to identify and improve the Pareto front. Progress is often measured by the increase in dominated hypervolume (see Section 2.6). In this case study, we aim to push the Pareto front outward with respect to the objectives in Table 2. Similarly, these objectives were chosen to loosely mirror the needs of real-world alloy design efforts [28].

To start the optimization loop, we again sample an alloy and apply the same BCC manufacturability gate. If it fails, we update only the phase-stability (BCC) classifier. If it passes, we observe the objective values.

With the current GPRs for objectives and the BCC GPC for the constraint, we evaluate predictions across the design space. For each alloy, we compute pointwise EHVI (pEHVI; see Section 2.9) and the probability of being single-phase BCC, p_{BCC} . Our acquisition is the product $\text{AQF}(\mathbf{x}) = p_{\text{BCC}}(\mathbf{x}) \times \text{pEHVI}(\mathbf{x})$, which prioritizes candidates that both expand the Pareto front and are likely manufacturable. We pick the alloy that maximizes this quantity and iterate in a closed loop. This process is summarized in Figure 6.

3.3. Constraint Satisfaction Minimizes TTFF

In order to provide a measure of statistical robustness and to make sure any effect of initialization, hyperparameter tuning,



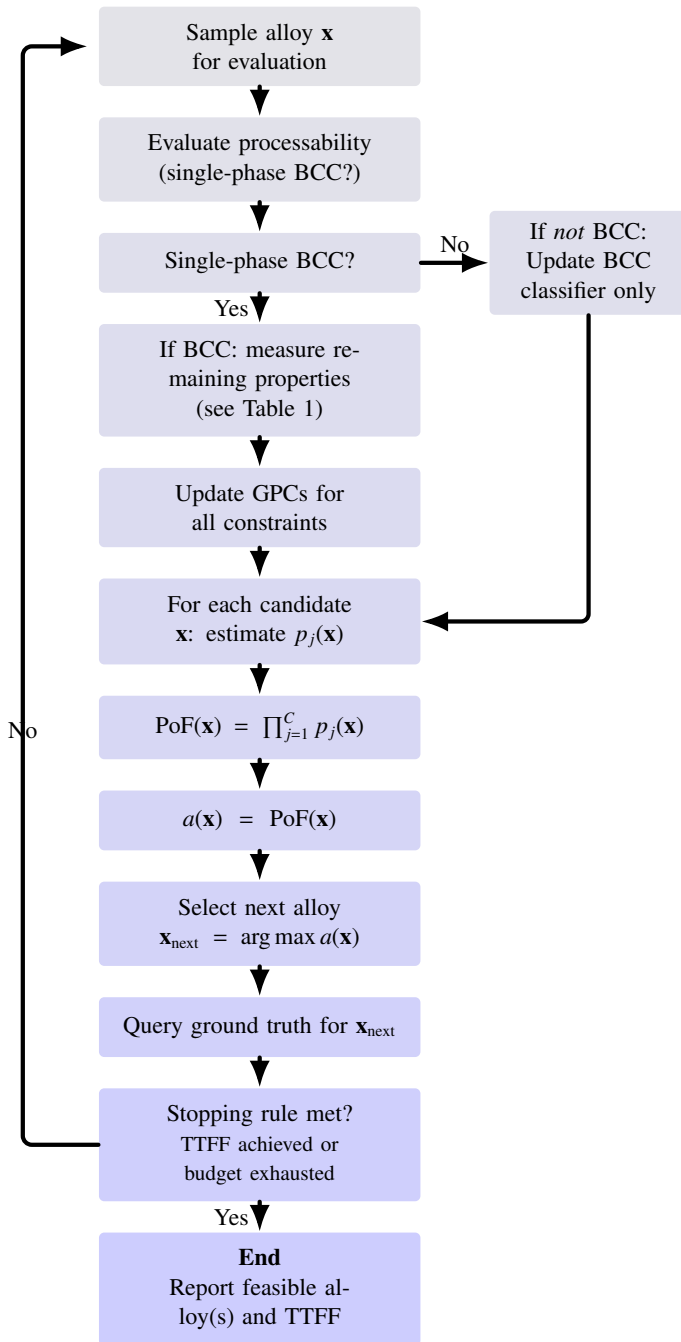


Figure 5: Closed-loop constraint-satisfaction workflow to minimize time-to-first-feasible (TTFF).

etc., is accounted for, we run 200 of these closed loops for optimization and constraint satisfaction and track the metrics listed in Sections 2.6 and 2.7 as a function of iteration. The mean and standard deviation of these metrics vs iteration is plotted in 7 and 8.

Figure 7 shows dominated hypervolume versus iteration, and Figure 8 shows the cumulative number of feasible alloys found. As expected in 7, the optimization strategy increases hypervolume more rapidly, whereas in Figure 8 constraint satisfaction discovers feasible points sooner and in greater numbers. By

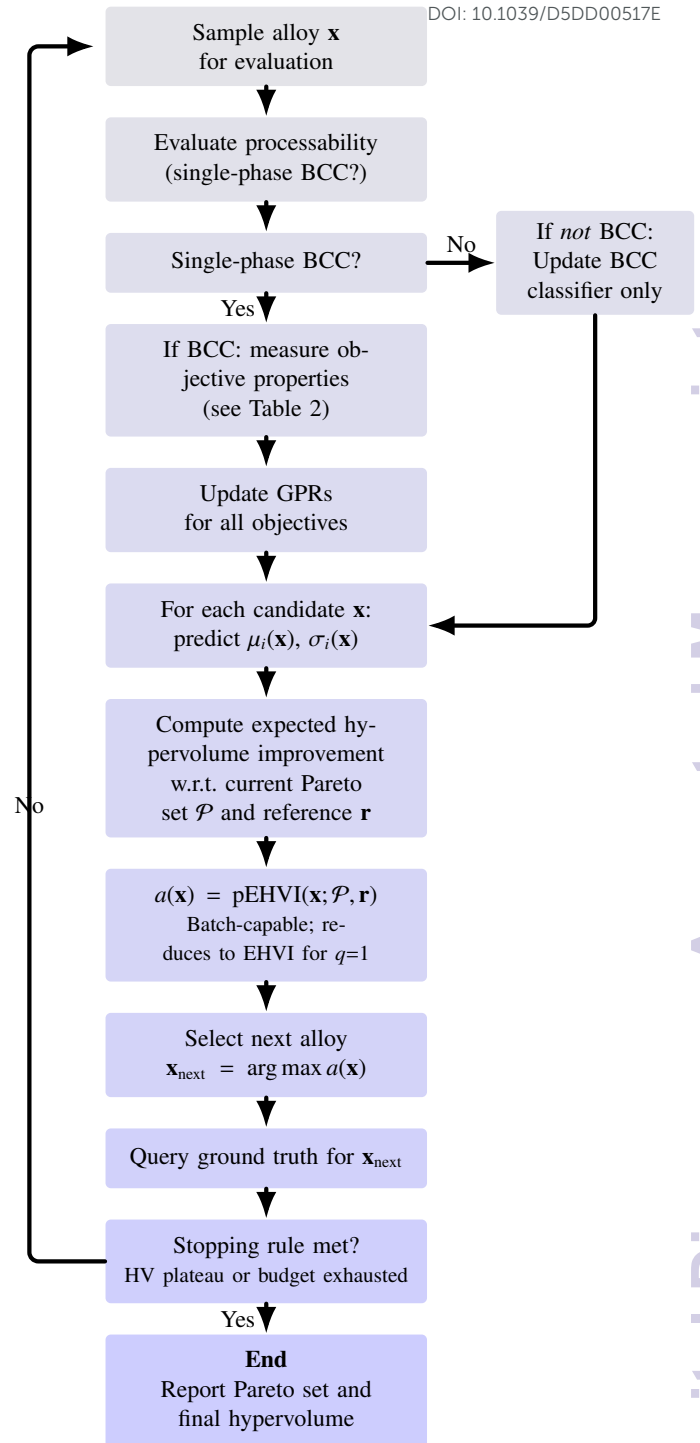


Figure 6: Closed-loop multi-objective optimization workflow using pEHVI under a single-phase BCC constraint.

the end of the experiment, on average, the constraint satisfaction loop identifies all but one of the feasible alloys and its average time-to-first-feasible (TTFF) is 9 iterations. Recall that only 22 out of 41,496 candidates are feasible (0.05% of the design space), making this a needle-in-a-haystack problem solved within a short campaign (half of the feasible alloys found by iteration 23; 21 out of 22 feasible alloys found by iteration 48).



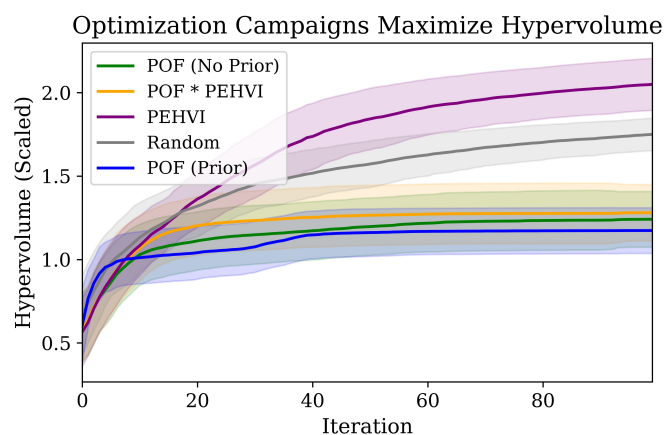


Figure 7: Average dominated hypervolume per iteration. The solid lines show the average metrics for each iteration, with the optimization shown in purple, the constraint satisfaction with a prior shown in blue, the constraint satisfaction without a prior shown in green, the combined model (PoF \times PEHVI) shown in orange, and the random search shown in grey. The shaded region shows one standard deviation above and below the mean.

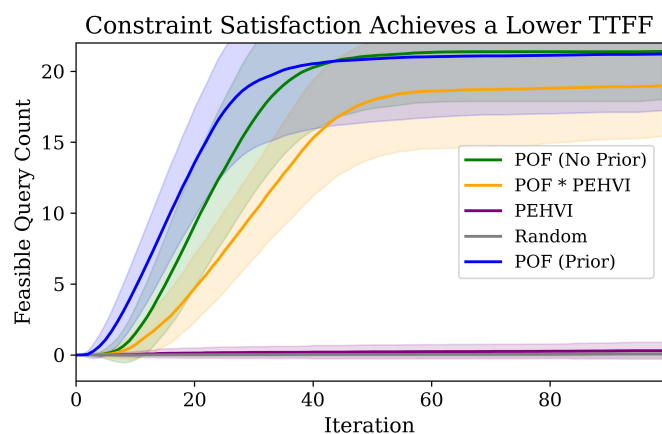


Figure 8: Average cumulative feasible alloys found per iteration. The solid line shows the average cumulative queries, with the optimization shown in purple, the constraint satisfaction with a prior shown in blue, the constraint satisfaction without a prior shown in green, the combined model (POF \times PEHVI) shown in orange, and the random search shown in grey. The shaded region shows one standard deviation above and below the mean. There are 22 feasible alloys out of 41,496 possible candidates.

In contrast, the optimization strategy struggles to find feasible candidates, performing as poorly as a random search, while the combined model performs between the two strategies (with a TTFF of 11 iterations and an average of 19 feasible alloys found by the end of the campaign).

3.4. Informative Prior Mean Functions Minimize TTFF

Having shown that constraint satisfaction achieves a lower time-to-first-feasible (TTFF) than optimization, we next tested whether adding informative priors to the GPCs (see Section 2.10) further helps. We ran an additional constraint-satisfaction campaign in which the GPCs used informative prior mean functions, repeating the closed loop 200 times and averaging metrics

by iteration. As shown in Figure 8, the prior-equipped variant yields a smaller TTFF than the baseline (5 iterations instead of 9) and also shortens the time required to discover all feasible points.

4. Discussion

The empirical results presented here strongly suggest that, in realistic alloy design scenarios, the definition of success depends heavily on the operational context. Design campaigns that focus on maximizing hypervolume are not guaranteed to find feasible candidates. This is evident from the fact that the combined model performs worse in terms of TTFF and total cumulative feasible queries, and that the random search results in a greater hypervolume increase despite finding only a fraction of the feasible queries that the constraint satisfaction models find. When the discovery objective is to rapidly identify any alloy—or any other material, as these results are materials-agnostic and broadly generalizable—meeting all minimum performance thresholds, a feasibility-driven strategy yields more decision-relevant outcomes than classical multi-objective optimization. The *time-to-first-feasible* (TTFF) metric introduced here captures this notion quantitatively, providing a direct measure of discovery efficiency aligned with realistic, stakeholder-driven priorities.

Constraint satisfaction reframes alloy design around *certifiable feasibility* rather than *unverifiable optimality*. In practical alloy development, the ground-truth material response surfaces are never fully known; consequently, global Pareto optimality cannot be proven—it can only be approximated within the sampled design space. At best, one can certify local non-dominance among observed candidates, a fact that limits the decision value of optimization-based strategies in finite-budget campaigns. By contrast, feasibility is auditable: a candidate either satisfies the registered performance thresholds (within uncertainty bounds) or it does not. *The difference between feasibility (which we can test and certify) and optimality (which we can only approximate) marks the fundamental limit of what can be known in materials discovery—its epistemic boundary—when our information is incomplete or uncertain.*

The observed impact of informative prior mean functions further supports the value of embedding physical intuition into probabilistic classifiers. By incorporating thermodynamic or mechanical priors, we showed that Gaussian Process Classifiers (GPCs) can learn feasibility boundaries more efficiently and focus querying on physically meaningful regions of the design space. This approach naturally extends to multi-fidelity and cost-aware frameworks, in which prior knowledge derived from simulations or relevant prior experiments accelerates feasibility discovery.

Finally, given the unique strengths of constraint satisfaction and optimization, and the fact that the combined model



does not fully take advantage of either strategy, suggests that these paradigms should be used in a sequential, complementary framework. A hybrid strategy, beginning with feasibility discovery and transitioning to Pareto refinement, mirrors the logic of stage-gate design processes in engineering practice. In this structure, constraint satisfaction would provide the *go/no-go* certainties required for decision-making, while optimization explores trade-offs within the feasible subspace. Consistent with entropy-based active learning of constraints, continued constraint focused querying becomes progressively less effective once the feasibility boundary is sufficiently learned. The expected entropy reduction plateaus (“an elbow”), indicating diminishing returns. In this regime, reallocating evaluations to objective optimization is more efficient; accordingly, entropy-based frameworks explicitly trigger the transition to Bayesian optimization once the average entropy reduction over constraint models flattens and drops below a threshold (e.g., 3%) [29]. The combination of these complementary approaches constitutes a unified, fully-Bayesian workflow that effectively balances exploration and exploitation over the long range within a materials discovery campaign.

5. Conclusion

In this work, we benchmarked two Bayesian closed-loop discovery strategies on the same alloy design space: (i) *constraint satisfaction*, which targets rapid discovery of any alloy meeting all feasibility criteria, and (ii) *multi-objective optimization*, which seeks to expand the Pareto front. Across 200 independent runs per setting, we tracked dominated hypervolume and feasibility metrics as a function of iteration. **Key findings:**

- *Time-to-first-feasible (TTFF)*. Constraint satisfaction consistently achieved a lower TTFF than optimization, identifying feasible alloys earlier and in greater numbers. By the end of the campaign, the prior equipped model had, on average, recovered 21 out of 22 feasible candidates (0.05% of the 41,496 total), effectively resolving a needle-in-a-haystack problem within a limited number of iterations.
- *Hypervolume growth*. As expected, optimization expanded dominated hypervolume faster, reflecting its objective.
- *Hybrid strategy performance*. The combined model sits between optimization and constraint satisfaction in terms of performance, not fully taking advantage of either strategy.
- *Informative priors*. Equipping the GPCs with informative prior means further reduced TTFF and shortened the time required to find all feasible points relative to the baseline constraint-satisfaction loop.

When the priority is to obtain *any* viable alloy quickly (e.g., early-stage, resource-constrained campaigns), constraint satisfaction is preferable. When the goal shifts to improving trade-offs among properties, multi-objective optimization is the right tool. In practice, a hybrid schedule—using constraint satisfaction to secure the first feasible alloy, then handing off to optimization—balances speed with frontier improvement. When prior knowledge exists, informative priors are a low-cost way to accelerate feasibility discovery.

Limitations include the fact that our feasibility model assumes approximate independence among constraints and well-calibrated classifier outputs; deviations from either can lead to systematic over- or under-estimation of the probability of feasibility (*PoF*). The BCC constraint serves as a practical proxy for manufacturability *in silico*, but it may not fully capture the spectrum of processing risks encountered experimentally. Another limitation is that the scope of the work is limited to a sequential evaluation strategy; batch selection methods were not considered. These caveats motivate future work on joint (correlated) constraint models, improved probabilistic calibration, cost-aware and multi-fidelity acquisition strategies, and adaptive hybrid frameworks that transition from feasibility-first discovery to Pareto frontier expansion as evidence accumulates. More broadly, these findings delineate the epistemic limits of optimization in materials discovery: while feasibility can be demonstrated, optimality can only be inferred.

Code Availability

The code associated with this work is available at the following repository: <https://github.com/hardccw643/constraint-satisfaction.git>.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used GPT-5 in order to ideate/brainstorm alternative sentence structures and stylistic choices in limited sections of the paper. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

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



- v91y2025i1d10.1007_s10898-024-01436-7.html
- [24] S. Guo, C. Ng, J. Lu, C. T. Liu, Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys, *Journal of Applied Physics* 109 (10) (2011) 103505. arXiv:https://pubs.aip.org/aip/jap/article-pdf/doi/10.1063/1.3587228/15077138/103505_1_online.pdf, doi:10.1063/1.3587228.
URL <https://doi.org/10.1063/1.3587228>
- [25] F. Maresca, W. A. Curtin, Mechanistic origin of high strength in refractory bcc high entropy alloys up to 1900k, *Acta Materialia* 182 (2020) 235–249. doi:<https://doi.org/10.1016/j.actamat.2019.10.015>.
URL <https://www.sciencedirect.com/science/article/pii/S1359645419306755>
- [26] B. Vela, D. Khatamsaz, C. Acemi, I. Karaman, R. Arróyave, Data-augmented modeling for yield strength of refractory high entropy alloys: A bayesian approach, *Acta Materialia* 261 (2023) 119351. doi:<https://doi.org/10.1016/j.actamat.2023.119351>.
URL <https://www.sciencedirect.com/science/article/pii/S135964542300681X>
- [27] C. Acemi, B. Vela, E. Norris, W. Trehern, K. C. Atli, C. Cleek, R. Arróyave, I. Karaman, Multi-objective, multi-constraint high-throughput design, synthesis, and characterization of tungsten-containing refractory multi-principal element alloys, *Acta Materialia* 281 (2024) 120379.
- [28] C. Acemi, B. Vela, E. Norris, W. Trehern, K. C. Atli, C. Cleek, R. Arróyave, I. Karaman, Multi-objective, multi-constraint high-throughput design, synthesis, and characterization of tungsten-containing refractory multi-principal element alloys, *Acta Materialia* 281 (2024) 120379. doi:<https://doi.org/10.1016/j.actamat.2024.120379>.
URL <https://www.sciencedirect.com/science/article/pii/S1359645424007298>
- [29] D. Khatamsaz, B. Vela, P. Singh, et al., Bayesian optimization with active learning of design constraints using an entropy-based approach, *npj Computational Materials* 9 (2023) 49. doi:10.1038/s41524-023-01006-7.
URL <https://doi.org/10.1038/s41524-023-01006-7>



Openly Available Data

- The code associated with this work is available at the following DOI:
<https://doi.org/10.5281/zenodo.17780080>

Restricted Access Data

- Due to Thermo-Calc's terms of service, we cannot share the exhaustive Thermo-Calc data. We have obfuscated the data and provided it in the repository.

