EFFICIENT EXPLOITATION OF MULTIPLE INFORMATION SOURCES IN MULTI-OBJECTIVE CONSTRAINED BAYESIAN OPTIMIZATION

A Dissertation

by

DANIAL KHATAMSAZ

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Douglas L. Allaire
Committee Members,	Raymundo Arroyave
	Daniel A. McAdams
	Ankit Srivastava
Head of Department,	Guillermo Aguilar

December 2022

Major Subject: Mechanical Engineering

ABSTRACT

With the technological advancements made in recent years, more powerful computational machines are built. The ability to handle large computational problems, both in terms of storage requirements and processing time has brought the opportunity to simulate complex systems more accurately than ever. Engineers, in particular, are interested to take advantage of using advanced computational machines for their design and modeling purposes. Although powerful computers can model complex systems to predict outcomes, budget allocations remain an issue in design and optimization tasks. Running large computational problems, simulations or experiments can be cost-prohibitive in terms of time or computational resources, for example, when highly accurate and reliable results are expected. Therefore, it is desired to spend available computational resources wisely toward a design goal to avoid running futile calculations.

Recently, Bayesian optimization techniques have been developed and are widely employed to solve design problems in many engineering fields. The popularity of Bayesian optimization frameworks comes from the fact that they are able to work with minimal information and use a heuristic-based search strategy to probe the design space, looking for potentially informative experiments about optimal designs. Most of the Bayesian optimization frameworks are developed to use a single model to collect mapping information from design space to objective space and find the optimal design region, if not exactly a point. However, in cases of high-dimensional design spaces or complex objective functions, these frameworks still need to evaluate a large number of designs that can be costly and almost not practical considering computational resource limitations. Simplified models have been developed to lower the expenses related to evaluations of complex models but this comes at the price of accuracy loss. A simplified model has a lower fidelity but is able to give estimations of a quantity of interest at a much lower cost. The potential here is to employ several of these simplified models to obtain as much information as possible regarding an expensive complex model, known as the ground truth, by fusing the models and considering the correlation between the models and the ground truth. This approach opens up new ideas about

extending Bayesian optimization frameworks to employ multiple sources of information instead of relying on a single expensive model to achieve a less costly design process.

Although several approaches exist to fuse different sources in multifidelity Bayesian optimization frameworks, there are not many of them developed to deal with multi-objective functions. In many engineering design problems, there are multiple quantities of interest to consider when looking for the optimal design. The issue here is the solution to such design problems is not a single design but a set of designs to be discovered. This can be computationally demanding owing to the fact it is searching for a region instead of a single point.

Another issue with Bayesian optimization frameworks is that they tend to underperform when the dimensionality of the design space increases. While statistical techniques can be employed to define the most important design variables and discard the remaining unimportant ones to reduce the dimensionality of the design space, it comes at the price of losing accuracy. Also, it may be misleading since there is a possibility of having the optimal design laid somewhere close to the regions ignored to be searched.

The next problem is, most often, there are several design constraints in an engineering design problem. It is crucial to know the constraints and recognize feasible regions before spending resources to search the design space. Sometimes, the models representing the constraints are also expensive to query and it is computationally prohibitive to sample a large number of points to identify the feasibile region boundaries. Additionally, in some problems, a constraint can be in form of a binary check to see if a condition is satisfied or not. In such cases, there is not a continuous function to be modeled and check if its values lie within specific bounds. Additionally, the uncertainty in the feasibility prediction of a design is important in decision making processes once we know some constraints are known as hard constraints and they should be treated more conservatively.

In this study, we aim to address the issues mentioned above: first, we develop a multifidelity Bayesian optimization framework suitable to optimize multi-objective functions. The expected hypervolume improvement is employed as the criterion to look for optimal design regions while balancing the exploitation of the present system's information and the exploration of new regions

in the design space. Second, we propose a novel framework to use an adaptive active subspace method to efficiently recognize the important directions in the design space to form a lower dimensional space and reduce the dimensionality of the problem while it still allows searching all design variables but in different degrees. Finally, a multifidelity Bayesian classification framework is proposed to be employed within an optimization framework to solve constrained optimization problems more efficiently by actively learning the feasible region boundaries beforehand. Shannon Entropy definition is used to quantitatively determine the uncertainty, which is usually larger closer to the boundaries. We show the performance of our proposed frameworks on different test problems and then, its application to some engineering problems.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my academic advisor Dr. Douglas L. Allaire for his unconditional support, encouragement, and advising. He generously provided knowledge and expertise and made it possible to undertake this journey. I also would like to express my deepest appreciation to my committee members Dr. Raymundo Arroyave, Dr. Daniel A. McAdams, and Dr. Ankit Srivastava for their constructive comments and supports. Lastly, I would like to appreciate my family and friends whom supports and encouragements helped me through this stage of my life.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Allaire and Dr. McAdams of the Department of Mechanical Engineering and Dr. Arroyave and Dr. Srivastava of Department of Materials Science and Engineering. All work for the dissertation was completed by the student, under the advisement of Dr. Allaire of the Department of Mechanical Engineering.

Funding Sources

My graduate study was supported by the AFOSR MURI on multi-information sources of multiphysics systems under Award Number FA9550-15-1-0038, and by the National Science Foundation under grant no. CMMI-1663130 and DMR-2001333. I also acknowledge the support from the ARPA-E ULTIMATE Program through Project DE-AR0001427.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

The goal in a design problem is to find a set of design variables that minimize or maximize a property of interest while satisfying some constraints. Due to the complexity of the models created to represent different real world systems, that usually are in the form of black-box functions, numerical methods are suitable. Lately, Bayesian optimization techniques are getting more attention than ever and it is seen as a powerful and efficient tool to optimize expensive to evaluate objective functions. An important advantage of Bayesian optimization frameworks is that they are capable of working with minimal data. Usually, dealing with black-box functions requires sufficient number of function evaluations to build the proper mapping from a design space to an objective space and it needs large amount of computational resources. However, designers are interested to minimize the number of queries to make less but more informative queries from their models. Bayesian optimization frameworks model a target objective function using surrogate models, usually, Gaussian process regressions and use acquisition functions, for instance, Expected Improvement (EI) [2] or Knowledge Gradient (KG) [3, 4], to search the design space to suggest the best next experiment while balancing the exploration versus exploitation of current system's information about the optimum design variables.

Optimizing a black-box function can be challenging, particularly if the design space is high-dimensional. Although heuristic-based search helps in the sense that it searches for the optimal design more effectively, but still the computational resources to make sufficient queries to obtain a good estimate of the optimum design can be the bottleneck of any design application. Therefore, a potential solution to deal with such expensive optimization tasks is to create cheaper models estimating the same quantity of interest with some simplifying assumptions. The simplifications made in modeling a ground truth quantity of interest lead to simpler and cheaper models but the trade-off is they are lower in accuracy or fidelity of the information they provide. Different simplifications create different lower fidelity models, also know as information sources and the designer is able to extract useful information about the ground truth quantity of interest from each source

assuming every single information source have some useful information about the ground truth in some regions in the design space [5, 6, 7, 8]. In presence of multiple information sources, information fusion techniques are employed to collect information from different sources to build a fused model that is capable of estimating the ground truth quantity of interest in a less costly manner. This is possible owing to the fact that information collected from information sources costs less than directly querying the ground truth. Several different techniques can be employed to fuse multiple sources of information. Among these are approaches such as Bayesian modeling averaging [9, 10, 11, 12, 13, 14], the use of adjustment factors [15, 16, 17, 18], covariance intersection methods [19], and fusion under known correlation [20, 21, 22]. One of the goals in this study is to employ such information fusion methods to propose more efficient multifidelity optimization techniques to optimize expensive to evaluate functions.

As of now, multi-information source optimization frameworks are developed mainly to optimize single objective functions. However, when designing real-world systems, usually, there are multiple properties to consider. For example, in designing aerostructures, one goal can be to minimize the weight but also, minimizing the fuel burn at the same time can be the other objective. The challenge in multi objective optimization problems is that, mostoften, the objectives are in conflict with each other and there is not a single solution that optimizes all objectives simultaneously. The solution to such design applications is a set of designs that are not dominating each other, which means no design is superior to other designs in that set. This set of non-dominated designs forms the Pareto front in the objective space. Estimating the Pareto front is a typical approach to find non-dominated designs and the goal is to improve the approximation of the solution by comparing the respective Pareto fronts. To have a criteria to compare different solution sets, hypervolume-based techniques suggest to define a reference point in the objective space, then, the volume between Pareto front and the reference point is calculated referred to as hypervolume. Note that since in optimization of multi objective functions, the goal is to discover a region as the solution and not a single design, more of high fidelity information is needed which can be computationally demanding. There are several techniques to approximating Pareto fronts and followingly, the

hypervolume in multi-objective optimization problems such as weighted sum approach [23], the adaptive weighted sum approach [24], normal boundary intersection methods [25], hypervolume indicator methods [26, 27, 28, 29, 30, 31, 32], and others. The hypervolume indicator technique is well-suited to expected improvement based algorithms, which have been shown to work well in presence of multiple information sources (see, e.g., Refs. [33, 34, 6, 35]). Another goal of this study is to integrate hypervolume indicator improvement within a multi-information source Bayesian optimization framework to enable efficient optimization of expensive multi objective functions.

While developing multifidelity Bayesian optimization frameworks offer a more efficient approach to attack design problems, they may still require a large number of evaluations from information sources, in particular, if the design space is high-dimensional [36, 37]. Additionally, Bayesian optimization frameworks tends to underperform in high-dimensional spaces [38]. Therefore, to address the issues that stem from the curse of dimensionality, methods to reduce the dimensionality of the input space can be implemented within design frameworks. A simple idea to decrease the dimensionality of the input space is to keep the most effective design variables and discard less effective ones, for example, by fixing a value or defining the objective function on a lower dimensional design space. Among existing dimensionality reduction techniques are Global Sensitivity Analysis and Principal Component Analysis. Global sensitivity analysis determine the importance of each design variable from variation in the objective function [39, 40, 41, 42, 43]. Principal component analysis [44] defines a lower dimensional design space by finding the principal components that take the majority of the variance in the available samples in the high-dimensional design space to keep as much information as possible [45, 46, 47]. Note that PCA decreases the dimensionality of the design space by only considering the correlations between design variables and does not consider the variability of the objective value. Most of the times, a sufficient number of data is required to use such statistical tests and it may not be accessible for a particular system due to computational resource limitations.

In this study, we aim to use the Active Subspace Method to reduce the dimensionality of large

design spaces to solve optimization problems in a more efficient manner [48, 49, 50]. This method, in contrast to PCA, use the variability of a function to map an objective function to a lower dimensional subspace that captures the largest variability of the objective function. Then, the obtained subspace, also known as the active subspace, is searched using an acquisition function. Lower dimensionality of the active subspace makes the learning process in machine learning tasks easier by reducing the required computational cost and improving the performance of the Bayesian optimization frameworks. Since the active subspace is defined to represent the largest variability of an objective function, it is considered as a highly informative region in the full design space. As far as now, there has not been done any significant work to integrate the active subspace method within multi-information source Bayesian optimization frameworks. Another goal of this study is to build such a framework to further increase the efficiency and performance of optimization frameworks in high-dimensional design problems.

In this research, the overall goal is to develop more efficient optimization methods in comparison to the existing ones but dealing with design constraints is the other part of the story. When performing Bayesian optimization, a more efficient approach is to check the constraints beforehand to prevent searching infeasible regions and wasting computational resources on utility function computations and further, real evaluations from information sources or ground truth. In some cases, the functions representing the constraints may not be computationally cheap to evaluate and it is not possible to check the feasibility of every single design. A simple approach here is to build and train surrogate models to allow cheaper way of checking the feasibility. However, it need a large set of training data to cover the entire design space which is computationally prohibitive to generate. Another point here is, what if a constraint is not a continuous function and has discrete values, for example, it just tells if a design passes a constraint or not? To address these issues, we propose a Bayesian classification framework that uses Gaussian process classification to model the constraint boundaries and predict the feasibility[51]. While Bayesian optimization frameworks search for optimal designs, a Bayesian classification framework searches for boundaries to recognize feasible and infeasible regions in the design space. The uncertainty of assigning a label is

higher in regions closer to boundaries. A quantitative way of comparing the overall uncertainty of a classifier is to define it as entropy. Therefore, entropy measure can be employed to identify highly uncertain regions and how it would change if new training data is added. This way, it is possible to make effective queries from the constraint functions to efficiently learn the boundaries. A big advantage of using Gaussian process classification is that it provides not only the probability of feasibility, but also, the uncertainty of the prediction Which is a necessary element in a Bayesian approach. Due to similarity of Bayesian optimization and Bayesian classification concepts, in this work, there is the opportunity to develop a multifidelity approach for the classification part as well, if there are more than one model to represent the same constraint. Our multifidelity Bayesian classification framework can be implemented within a Bayesian optimization framework to create a novel framework that is capable of learning the constraints and objective functions efficiently, suitable for solving any constraint Bayesian optimization problems in single and multi-information source systems.

To the best of our knowledge, there have not been any significant works done toward developing hypervolume indicator-based multi-objective multi-fidelity Bayesian optimization frameworks. Additionally, multifidelity Bayesian classification frameworks or implementation of the active subspace method in multi-information source settings are not widely studied yet. Thus, our goal is to propose such methods and show their performance by applying them on synthetic functions in addition to some real-world engineering design problems. In Chapter 2, a multifidelity Bayesian optimization technique is introduced and its application on designing a dual-phase high-strength steel is shown. In chapter 3, a multifidelity multi-objective Bayesian optimization framework is presented and its performance is compared to some similar approaches in a wing design problem. The implementation of the active subspace method within our multifidelity framework is introduced in chapter 4. The framework is applied to the same material design problem introduced in chapter 2 and it is shown how it has resulted in a more efficient way of finding the optimal design region. In chapter 5, Bayesian classification framework and its extension to multifidelity cases is established. Then, this framework is combined with a multifidelity Bayesian optimization

framework to allow for optimization in presence of multiple constraints. Again, in an example, it is emphasized on the advantage of using multiple information sources, but this time, in a Bayesian classification scheme. Then, the combination of Bayesian classification and Bayesian optimization frameworks is applied on designing high-entropy alloys in presence of some constraints. In the final chapter, we wrap up the works done in this study and discuss possible future works.

2. BAYESIAN OPTIMIZATION IN MULTI-FIDELITY SETTINGS*

2.1 Overview

Materials design calls for the (inverse) exploitation of Process-Structure-Property (PSP) relationships to produce materials with targeted properties. Unfortunately, most materials design frameworks are not optimal, given resource constraints. Bayesian Optimization (BO)-based frameworks are increasingly used in materials design as they balance the exploration and exploitation of design spaces. Most BO-based frameworks assume that the design space can be queried by a single information source (e.g. experiment or simulation). Recently, we demonstrated microstructure-sensitive design of alloys with a BO framework capable of exploiting multiple information sources. While promising, the previous framework is limited as it assumes that the optimal microstructure is always feasible and considers microstructural features as the design space. Herein, we sidestep this unwarranted assumption and instead consider that chemistry and processing conditions constitute the design space amenable to optimization. We demonstrate the efficacy of our expanded framework by optimizing the mechanical performance of a ferritic/martensitic dual-phase material by adjusting composition/processing parameters. The framework uses thermodynamic results to predict microstructural attributes which are then used to predict the mechanical properties using a variety of micromechanical models and a microstructure-based finite element model. The final stage involves implementing model reification and information fusion, and a knowledge-gradient acquisition function to determine the next best design point and information sources to query. A detailed discussion of the various components and demonstration of how the framework can be implemented under three sets of cost-based constraints is presented.

*Reprinted with permission from "Efficiently exploiting process-structure-property relationships in material design by multi-information source fusion" by D. Khatamsaz, A. Molkeri, R. Couperthwaite, J. James, R. Arróyave, D. Allaire, and A. Srivastava., 2021. Acta Materialia, Copyright 2021 Acta Materialia Inc. Published by Elsevier Ltd.[37]

2.2 Introduction

Materials design through ICME approaches [52] calls for the integration of physics (or data [53, 54])-based models and experiments in order to establish quantitative Process-Structure-Property (PSP) relationships [55], which can then be exploited in order to elucidate the mechanisms by which the (multi-scale, hierarchical) microstructure of a material responds to external stimuli (i.e. its property), or the mechanisms by which processing/synthesis conditions alters the way materials are (hierarchically) organized [56]. By 'inverting' such relationships, it is then possible to identify the combinations of chemistry and processing *necessary* to produce (multi-scale) microstructures that meet target measures of performance. [57].

Unfortunately, fully integrated ICME frameworks with quantitative predictive accuracy still remain out of reach, primarily due to the complex, highly coupled, multi-scale nature of linkages along the PSP chain. This complexity makes it very difficult to computationally emulate such PSP chains. Even if one ignores such challenges and focuses instead on much simpler single-scale or effective models, the explicit integration of multiple tools within a single ICME framework constitutes a major challenge [58] that has remained mostly unresolved, although limited recent works [59, 60] have achieved some success in deploying fully integrated PSP model chains to carry out the design/optimization of materials. Major challenges arise when considering the considerable cost associated with querying PSP relationships. Recently, closed-loop Bayesian Optimization (BO) approaches [61], capable of efficiently balancing the exploration and exploitation of materials design spaces, have already been shown to be quite effective in materials optimization tasks under resource constraints [62].

A major limitation of most approaches to date is the fact that they tend to use a single model per linkage along PSP chains. This is an unnecessary restriction as often there are multiple computational models or information sources to choose from when trying to establish such quantitative linkages. A further limitation is the fact that most 'classical' ICME frameworks are not capable of directly integrating experimental information into the design/optimization loop, other than for calibration, validation or verification purposes. Potential valuable experimental information could

instead be directly incorporated into the ICME framework in a co-design scheme but this remains largely unexplored. In fact, even state-of-the-art BO-based materials design [61, 62] tends to be limited to a single probe (experimental or computational) to query specific linkages of the PSP chain, although other engineering fields have developed sophisticated approaches for the integration of multiple information sources within optimization schemes [63].

In Refs. [6, 64], a material design framework is developed that addresses most of the issues highlighted above. In order to avoid explicit I/O interfaces between different models connecting microstructure to mechanical behavior—for a demonstration problem involving microstructural design of a dual-phase material—each of the models, including the 'ground truth' (in that case a microstructure-based finite element model) were converted into a Gaussian Process (GP). Since all models used are attempting to describe the same underlying behavior (i.e. the connection between microstructural features and mechanical response), it is to be expected that they would exhibit some degree of correlation among themselves and also with the 'ground truth'. By exploiting such statistical correlations through the so-called 'reification' process [5, 8], we generated fused models that were capable of maximizing agreement with available information about the response of the 'ground truth' model, while minimizing responses clearly at odds with observations. The fused model was then used to answer two questions: (i) where to sample next in the design space, and (ii) which information source should be used to query the design space. To answer those questions, we constructed experimental utility functions based on the Knowledge Gradient (KG) [65]. In Ref. [6], it is demonstrated how this multi-information fusion BO framework was more efficient than state-of-the-art BO approaches that used a single (ground truth) information source. In Ref. [64], it is further demonstrated that explicitly accounting for the cost of individual information sources was essential if there are hard constraints imposed on the budget allocated to carry out optimization.

While promising, the previous framework was admittedly limited in that the complete PSP chains were not exploited and only a microstructure sensitive design problem was addressed, assuming that the design space consisted of a universe of microstructures that could in principle be accessed through suitable processing-chemistry combinations. This is a strong and unwarranted

assumption, as *optimal microstructures may not necessarily be feasible* (i.e. accessible through the available chemistry and/or processing controls). It is thus important to explore possible ways of extending our multi-information source fusion BO framework in such a way that the design space consists of degrees of freedom that are truly amenable to modification without concerns on whether they are feasible or not. Such a framework would then truly realize the ICME program which relies, as mentioned above, on the exploitation of PSP relationships in order to carry out the design of materials. Solving such a problem is not trivial as we would have to fuse information not only in parallel, as we have shown before (all models used connected the same microstructural input to the same target mechanical response), but also in series, as we would have to provide explicit connections between processing-chemistry prescriptions and properties/performance through intermediate microstructure information.

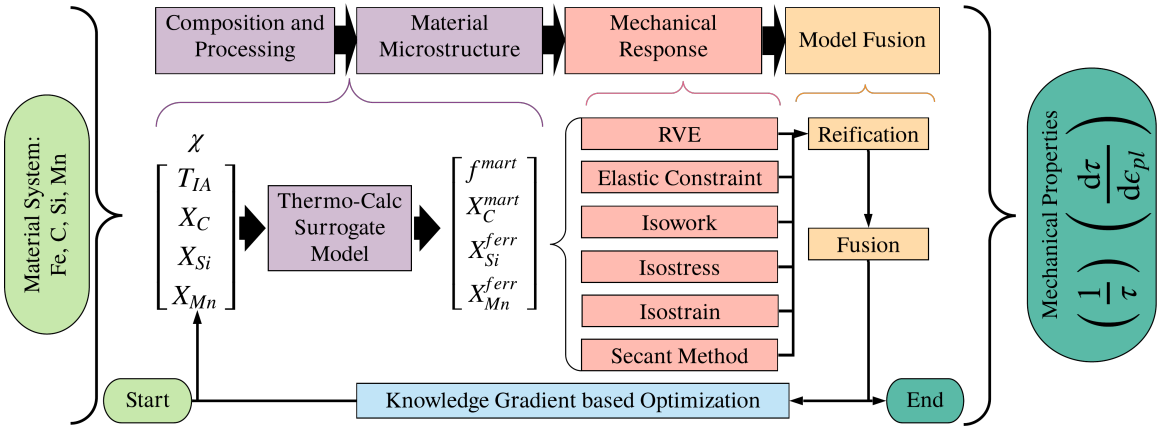

Figure 2.1: Schematic representation of the process-structure-property relationship used for the design of a dual-phase (ferrite-martensite) steel using our multi-information source fusion Bayesian optimization framework. Here, χ is the set of input variables with T_{IA} being the intercritical annealing temperature, X_C, X_{Si} and X_{Mn} being the carbon, silicon and manganese content, respectively, while the targeted output is the strength normalized strain hardening rate, $(1/\tau)\,(\mathrm{d}\tau/\mathrm{d}\epsilon_{pl})$.

In this work, we expand on our prior multi-information source fusion BO framework to account for cases in which it is possible to establish, at least in principle, quantitative PSP relationships. As in the previous works [6, 64], it demonstrates the framework by attempting to optimize the

performance of a dual-phase steel. In this case, however, the degrees of freedom, amenable for optimization, that we consider are chemistry and processing (in this simplified case, heat treatment temperature), rather than microstructure characteristics. In this way we avoid the limitations of our previous works (and that of others) as in this case the optimum microstructure is always feasible since, it is a consequence of chemistry and processing protocol. We represent the proposed framework in Fig. 2.1. This framework optimizes the normalized strain hardening rate $(1/\tau)\,(\mathrm{d}\tau/\mathrm{d}\epsilon_{pl})$ of a dual-phase (ferrite-martensite) steel by adjusting the composition and heat-treatment temperature of the material. The framework uses a surrogate model of the thermodynamic results to predict the phase volume fraction and composition in the material microstructure after single-stage heat-treatment (intercritical annealing followed by quenching). This data is then used to predict the mechanical properties of the dual-phase material using a variety of micromechanical models and a high through-put microstructure-based finite element model that utilizes a three-dimension representative volume element (RVE) of the material microstructure. After obtaining fused models through reification, the next point to query in the design space as well as the information source used to query it are determined using the KG acquisition function. What follows is a discussion of the various components of this framework and a demonstration of how the framework can be implemented under three separate sets of constraints, related to when queries to the 'ground truth' are made.

2.3 Methods

In this part, all ingredients and concepts required to build the framework are discussed in detail.

2.3.1 Connecting Chemistry and Processing to Microstructure

The current work considers a material system that contains C ($0.05 - 1$ wt%), Si ($0.1 - 2$ wt%), Mn ($0.15 - 3$ wt%) and Fe being the balance, and is subjected to a single-stage intercritical annealing heat-treatment at temperatures ranging from 650-850°C, followed by rapid quenching to produce a dual-phase (ferrite-martensite) microstructure. The prediction of the microstructure space is carried out through a surrogate model built from Thermo-Calc™ predictions in the region

of interest. This surrogate modeling approach was used to ensure that the calculations can be completed quickly, as well as to ensure that it would be possible to carry out optimization process on a computing resource without access to Thermo-Calc™ license.

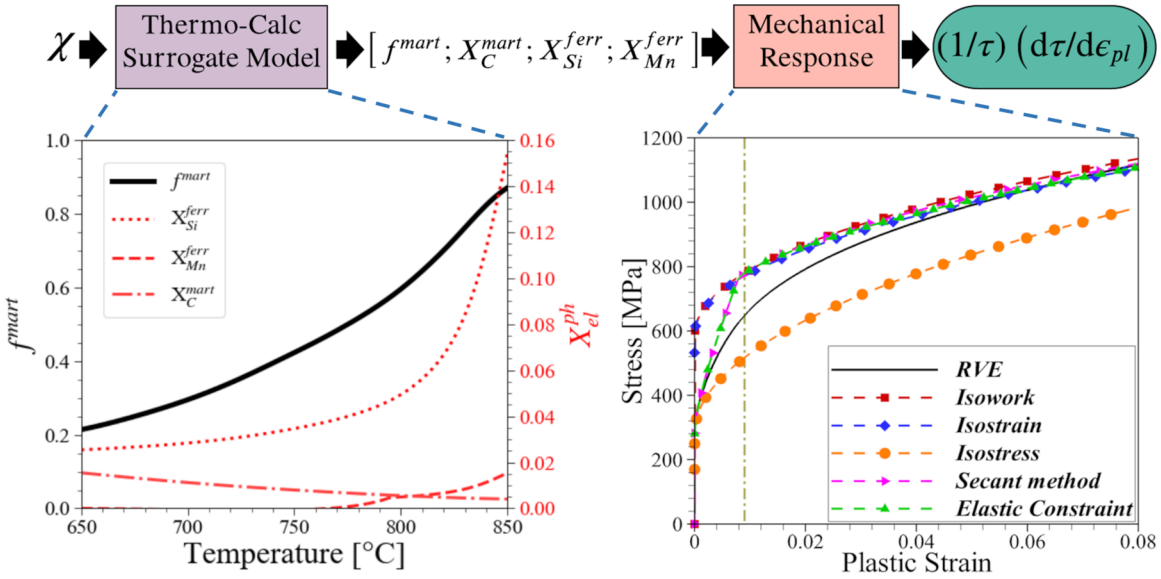

Figure 2.2: Schematic showing representative modeling results along the process-structure-property chain used herein for the design of a dual-phase (ferrite-martensite) steel. (Bottom left) Evolution of the martensite volume fraction, f^{mart}, and partitioning of alloying elements among the two constituent phases with temperature. (Bottom right) Evolution of stress with plastic strain in a dual-phase material as predicted by microstructure-based finite element model (RVE) and a variety of micromechanical models.

A detailed description on the construction of the surrogate model can be found in ref. [66]. Briefly, the Thermo-Calc™ model was uniformly sampled in the design space to obtain the austenite volume fraction and composition. The volume fraction of martensite was obtained using the Koistinen-Marburger equation [67]. This data was fit with a Gaussian Process (GP) model. The composition and volume fraction of the ferrite phase are defined by a mass balance. This thermodynamic surrogate model was linked to the mechanical models (described in the subsequent section) as shown in Fig. 2.2. As is often the case when using machine learning models, the accuracy of the GP surrogate is ultimately limited by the accuracy of the underlying ground truth as

well as the ability of the model form itself to reproduce the data used to train it. In general, we have found [66] that GPs are able to model the outcome of phase equilibria calculations with reasonable accuracy.

2.3.2 Connecting Microstructure to Mechanical Response

In this work, we assume that we have at our disposal a set of models (or sources of information) of varying complexity (or computational cost) and fidelity that establish quantitative relationships between microstructure and properties.. The models specifically take as input the quantitative features of the dual-phase microstructure in order to predict its mechanical response. For the purpose of demonstration, we consider a microstructure-based finite element model as 'ground truth' and a variety of reduced-order and micromechanical models as alternate cheap sources of information. All the mechanical models predict the mechanical response of the material by explicitly incorporating the effect of both the chemistry and processing conditions—resulting in specific phase constitution—by utilizing the predictions of the Thermo-Calc™ model, Fig. 2.2. Note that the stress-strain response of a composite dual-phase microstructure can in theory be accurately modeled using high-fidelity microstructure-based finite element calculations [68, 69, 70]. However, these calculations are computationally expensive and their direct use in an iterative material optimization scheme will be prohibitively time consuming. Reduced-order mechanical models [71, 72, 73] and more sophisticated micromechanical models [74] are computationally cheap, but this comes at a price, since their many simplifications result in loss of fidelity with regards to the 'ground truth'—all the mechanical models are briefly described below.

2.3.2.1 Microstructure-based Finite Element Model (Ground truth)

The microstructure-based finite element calculations utilize a three-dimensional RVE of the dual-phase microstructure to compute the uniaxial tensile stress-strain response of the material [70, 6, 64]. To this end, we first generate a three-dimensional ferrite grain structure in the RVE using a simulated annealing process [75] with brick meshing. Next, a fixed number of martensite particles are nucleated randomly on the ferrite grain boundaries and then grown at a fast rate along

the grain boundaries and a slow rate along other directions. The growth of the martensite particles is terminated when the overall volume fraction of the martensite reaches the prescribed target value. In the final RVE the crystallographic orientations of individual ferrite grains and that of martensite blocks are smeared out. This results in a composite dual-phase microstructure with two discretely modeled phases where each phase is modeled as an isotropic elastic-plastic material with Young's modulus, $E = 200$GPa, and Poisson's ration, $\nu = 0.3$, and constitutive relations (relating stress, τ, equivalent plastic strain, ϵ^{pl}, and composition, X_i^p, in weight fraction),

$$\tau = \tau_0^M + C_c(X_C^{mart})^{1/3} + K^M(\epsilon^{pl})^{n^M} \tag{2.1}$$

for the martensite phase with $\tau_0^M = 400$MPa, $C_c = 10^5$MPa, $K^M = 450$MPa and $n^M = 0.06$, and

$$\tau = \tau_0^F + C_{Si}(X_{Si}^{ferr})^{1/2} + C_{Mn}(X_{Mn}^{ferr})^{1/2} + K^F(\epsilon^{pl})^{n^F} \tag{2.2}$$

for the ferrite phase with $\tau_0^F = 200$MPa, $C_{Si} = 732$MPa, $C_{Mn} = 213$MPa, $K^F = 2200$MPa and $n^F = 0.5$. The constitutive relations in Eqs. (2.1) and (2.2), and the choice of the (representative) constitutive parameters are based on the assumptions that: (i) the strength of the martensite phase depends on its carbon content and the martensite phase doesn't exhibit significant strain-hardening, and (ii) the ferrite phase is softer than the martensite phase, exhibits significant strain-hardening and its strength depends on silicon and manganese content.

A typical finite element mesh of the RVE of the dual-phase microstructure employs 27,000 C3D8 brick elements from the ABAQUS/standard element library [76]. In the calculations, fully periodic boundary conditions are imposed on all six faces of the RVE. Next, to simulate uniaxial tensile deformation, the RVE is subjected to a monotonically increasing tensile strain along one of the axis, while maintaining zero resultant forces on the faces normal to the other two axes as in ref. [69]. Note that since the martensite particles are nucleated at random, *different realization of the same microstructure (i.e. the same overall phase constitution) in the RVE is different*. Thus, the

mechanical response predicted by different realizations of the same microstructure in the RVE is also slightly different as demonstrated in ref. [64].

2.3.2.2 Reduced-order and micromechanical models (Information sources)

As in refs. [6, 33], low-fidelity models are used as information sources to link material microstructure to overall mechanical response. The reduced-order mechanical models used here are based on three assumptions on how the work, stress, and strain partition among the constituent phases: (i) an *isostrain model* is built on the assumption that both the constituent phases undergo the same amount of strain under deformation [71]; (ii) an *isostress model* is constructed following the assumption of equal partitioning of stresses among both the phases [72]; and (iii) an *isowork model*, that assumes that both the phases undergo the same amount of work of deformation upon the deformation of the dual-phase microstructure [73]. In addition, we employ two slightly more sophisticated micromechanical models linking material microstructure to overall mechanical response. These models are constructed based on the homogenization schemes proposed in ref. [74]. We refer to the first micromechanical model as the 'secant method', where the prediction of overall mechanical response is based on Hill's weakening constraint power in a plastically-deforming matrix. The second micromechanical model is referred to as the 'elastic constraint' model, which is based on Kröner's treatment of the matrix-inclusion system under elastic constraints [74]. All these reduced-order and micromechanical models use the same constitutive relations given in Eqs. (2.1) and (2.2) for the two constituent phases.

2.3.3 Material Design Framework

An optimization problem can be written as

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}), \tag{2.3}$$

where f is the objective function and $\mathbf{x}^*$ is the optimal design vector in the feasible input space $\mathcal{X}$. Oftentimes, the analytical form of the objective function f is unknown or is very expensive to evaluate through either physical experiments or high fidelity simulations. Indeed, resource

constraints—i.e. the total budget available to the design campaign—place strong limits on the number of times the objective function can be queried. In many cases, however, it is likely that there exist cheaper information sources that can potentially be used to *estimate* the response of the much more expensive information source, at varying degrees of fidelity and cost. Such information sources could be models constructed with simplifying assumptions or evaluated at much lower resolution than the 'ground truth'. In cases in which the 'ground truth' is an experiment, the cheap sources could be (physics-based or machine learning) models, or could be other experiments that are much faster/cheaper to carry out but that are correlated to the more expensive 'ground truth'.

Here, we propose a multi-fidelity approach to seek the solution to Eq. (2.3). Assuming that every source contains some useful information about the quantity of interest—different sources may approximate the 'ground truth' better than others in some regions of the design space—, we aim to efficiently fuse all these sources in order to estimate the objective quantity of interest as accurately as possible. Following the previous work [33, 6], we employ a multi-information source optimization framework. In this approach, the first step is to construct intermediate surrogates for each of the sources according to the prior knowledge about the connection between the design space and their response. Here we use GPs as surrogates given their mathematical properties, including their ability to predict not only the mean value, but also the variance of the quantity of interest in the design space, and the straightforward manner in which the causal correlation between points in the design space can be modeled, among others [51]. Using a reification-based information fusion approach developed in ref. [5, 8], a fused model is built from all the information sources by exploiting the degree to which they are correlated with each other and with the 'ground truth'. The fused model is then used to determine the next design vector and information source to query. Here, we use the Knowledge Gradient (KG) acquisition function as the utility function to rank the next best design point to explore, and the best information source to query it, taking into account the cost of evaluating said information source as well as the uncertainty associated to the model predictions. This policy balances the exploration of the input space and exploitation of the current state of knowledge about the objective function, while also accounting for the impact of

such exploration/exploitation on the available resources.

By fitting GPs to data from previous evaluations of the individual information sources, the prior GP distribution for each information source is represented as

$$f_{\text{GP},i}(\mathbf{x}) \sim \mathcal{GP}\left(m_i(\mathbf{x}), k_i(\mathbf{x}, \mathbf{x}')\right), \tag{2.4}$$

where $k_i(\mathbf{x}, \mathbf{x}')$ is a real-valued kernel function over the input space and $m_i(\mathbf{x})$ defines the mean function. Here, we employ the squared exponential covariance function—note that other kernels may be used depending on our knowledge of the properties of the design space—as the kernel function specified as

$$k_i(\mathbf{x}, \mathbf{x}') = \sigma_s^2 \exp\left(-\sum_{h=1}^{d} \frac{(x_h - x_h')^2}{2l_h^2}\right), \tag{2.5}$$

where variables d, σ_s^2, and l_h are the dimension of the input space, signal variance, and characteristic length-scale. The latter controls the degree to which two points in the input space, separated by a specific distance, are correlated. The behavior of the GP surrogate is controlled by these hyperparameters and it is necessary to tune them based on the available data. Here, we optimize the hyperparameters by maximizing the log marginal likelihood.

Assuming N_i evaluations of information source i are available and indicated by $\{\mathbf{X}_{N_i}, \mathbf{y}_{N_i}\}$, where $\mathbf{X}_{N_i} = (\mathbf{x}_{1,i}, \ldots, \mathbf{x}_{N_i,i})$ is the N_i input samples to information source i and $\mathbf{y}_{N_i} = (f_i(\mathbf{x}_{1,i}), \ldots, f_i(\mathbf{x}_{N_i,i}))$ represents the outputs from information source i relatively, the posterior GP distribution of information source i at a design point $\mathbf{x}$ is given as

$$f_{\text{GP},i}(\mathbf{x}) \mid \mathbf{X}_{N_i}, \mathbf{y}_{N_i} \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma_{GP,i}^2(\mathbf{x})\right), \tag{2.6}$$

where

$$\mu_i(\mathbf{x}) = K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma_{n,i}^2 I]^{-1} \mathbf{y}_{N_i}, \tag{2.7}$$

$$\sigma_{\text{GP},i}^2(\mathbf{x}) = k_i(\mathbf{x}, \mathbf{x}) - K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma_{n,i}^2 I]^{-1} K_i(\mathbf{X}_{N_i}, \mathbf{x}), \tag{2.8}$$

and for information source i, $K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i})$ is a $N_i \times N_i$ matrix with (m, n) entry as $k_i(\mathbf{x}_{m,i}, \mathbf{x}_{n,i})$, and $K_i(\mathbf{X}_{N_i}, \mathbf{x})$ is the $N_i \times 1$ vector whose m^{th} entry is $k_i(\mathbf{x}_{m,i}, \mathbf{x})$. Here, the term $\sigma^2_{n,i}$ is used to model observation error for information source i found by experimental data.

We then quantify the discrepancy corresponding to each information source surrogate with respect to the 'ground truth' (i.e. the RVE model) as

$$\sigma_i^2(\mathbf{x}) = \sigma^2_{GP,i}(\mathbf{x}) + \sigma^2_{f,i}(\mathbf{x}), \tag{2.9}$$

where $\sigma^2_{f,i}(\mathbf{x})$ is the variance related to the fidelity of information source i, which is added to the uncertainty associated with the corresponding GP to find the total uncertainty of the information source surrogate i.

Since all information sources are estimating the same quantity of interest, they are expected to be correlated to the 'ground truth' to varying degrees. Every information source is assumed to have some useful information about the true objective function. In the proposed approach of refs. [33, 6], and contrary to conventional multi-fidelity approaches, it is not required to determine the hierarchy of fidelity according to agreement with the 'ground truth', rather *all information sources are brought together* into a single fused model. Here, we follow the same approach. Specifically, for fusion of normally distributed data—which is the case here since the information sources are represented by GPs—the method presented by ref. [77] is followed. This leads to fused mean and fused variance estimates at a given design point $\mathbf{x}$ as:

$$E[\hat{f}(\mathbf{x})] = \frac{\mathbf{e}^{\mathrm{T}}\tilde{\Sigma}(\mathbf{x})^{-1}\boldsymbol{\mu}(\mathbf{x})}{\mathbf{e}^{\mathrm{T}}\tilde{\Sigma}(\mathbf{x})^{-1}\mathbf{e}}, \tag{2.10}$$

$$\mathrm{Var}\left(\hat{f}(\mathbf{x})\right) = \frac{1}{\mathbf{e}^{\mathrm{T}}\tilde{\Sigma}(\mathbf{x})^{-1}\mathbf{e}}, \tag{2.11}$$

where $\mathbf{e} = [1, \ldots, 1]^{\mathrm{T}}$, $\boldsymbol{\mu}(\mathbf{x}) = [\mu_1(\mathbf{x}), \ldots, \mu_S(\mathbf{x})]^{\mathrm{T}}$ are the mean values of S sources, and $\tilde{\Sigma}(\mathbf{x})^{-1}$ is the inverse of the covariance matrix between the information sources. For example, in the

presence of two information sources, Eq. (2.10) is defined as

$$E[\hat{f}(\mathbf{x})] = \frac{(\sigma_2^2 - \rho\sigma_1\sigma_2)\mu_1 + (\sigma_1^2 - \rho\sigma_1\sigma_2)\mu_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}, \tag{2.12}$$

where σ_1^2 and σ_2^2 are the total variances of sources 1 and 2 respectively and the coefficient ρ shows the correlation between the information sources at a specific point $\mathbf{x}$. To estimate the correlations between the errors of two information sources i and j, we use reification [8, 5]. First, information source i is reified and the correlation coefficient is computed as

$$\rho_{ij}(\mathbf{x}) = \frac{\sigma_i^2(\mathbf{x})}{\sigma_i(\mathbf{x})\sigma_j(\mathbf{x})} = \frac{\sigma_i(\mathbf{x})}{\sqrt{\left(\mu_i(\mathbf{x}) - \mu_j(\mathbf{x})\right)^2 + \sigma_i^2(\mathbf{x})}}. \tag{2.13}$$

Then, information source j is reified to estimate $\rho_{ji}(\mathbf{x})$. Finally, the variance-weighted average of the two estimated correlation coefficients is computed to estimate the correlation between the models as

$$\bar{\rho}_{ij}(\mathbf{x}) = \frac{\sigma_j^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})}\rho_{ij}(\mathbf{x}) + \frac{\sigma_i^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})}\rho_{ji}(\mathbf{x}). \tag{2.14}$$

The average correlations are used in Eqs. (2.10) and (2.11) to estimate the fused mean and variance.

We use the fused means and variances to construct the fused model. Assuming N_f samples from the design space $\mathcal{X}$ are available, the fused mean vector and the diagonal matrix of fused variances are $\mu_{\text{Wink}}(\mathbf{x}_{1:N_f})$ and $\sum_{(\mathbf{x}_{1:N_f})}$ = diag $(\sigma^2_{\text{Wink}}(\mathbf{x}_1), ..., \sigma^2_{\text{Wink}}(\mathbf{x}_n))$ respectively. The posterior predictive distribution of the fused model is specified as

$$\hat{f}^{\,\text{fused}}(\mathbf{x}) \sim \mathcal{N}(\mu^{\text{fused}}(\mathbf{x}), \Sigma^{\text{fused}}(\mathbf{x})). \tag{2.15}$$

Next, the fused GP, which is considered as our best predictive model containing all current knowledge about the 'ground truth', is used to determine the next design point and information source to query. This is done by employing a utility function that considers the cost and the value of such a query relative to the optimization of the objective function in hand.

The value-gradient utility, which is a cost-aware KG utility, is used to evaluate different information sources to compare the amount of knowledge gained about the objective function's maximum and the cost to determine which model and design point adds the most value regarding Eq. (2.3). The utility function searches for immediate improvement to the knowledge state of the system in one step and at the same time looks for the expected improvement in two steps where it has the highest gradient in knowledge. Using the fused GP and its maximum mean function value, the immediate expected improvement can be defined.

At a design point $\mathbf{x}$, defining information sources by $(i_{1:N}, \mathbf{x}_{1:N}, y_{1:N})$ with design points and corresponding objective values for the first N queries and taking $\hat{f}$ as the posterior distribution of the fused model, the expected improvement is given as

$$\begin{aligned} \textit{EI}(\mathbf{x}) &= E\left[\max_{\mathbf{x}' \in \mathcal{X}} E[\hat{f}(\mathbf{x}')|i_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_{N+1} = \mathbf{x}, y_{1:N}] - \max_{\mathbf{x}' \in \mathcal{X}} E[\hat{f}(\mathbf{x}')|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]\right] \\ &= E\left[\max_{\mathbf{x}' \in \mathcal{X}} E[\hat{f}(\mathbf{x}')|i_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_{N+1} = \mathbf{x}, y_{1:N}]\right] - \max_{\mathbf{x}' \in \mathcal{X}} E[\hat{f}(\mathbf{x}')|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}], \end{aligned} \tag{2.16}$$

where the last expression can be removed from the expectation operator since it is known when conditioned on the first N queries. Next, we use KG policy [3, 4, 65] to maximize this expectation. Defining the value of being at the knowledge state H^N as $V^N(H^N) = \max_{\mathbf{x} \in \mathcal{X}} H^N$, where the knowledge state itself is presented by $H^N = E[\hat{f}(\mathbf{x}')|i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]$, the KG as a measure of expected improvement is

$$\nu^{\text{KG}}(\mathbf{x}) = E[V^{N+1}(H^{N+1}(\mathbf{x})) - V^N(H^N)|H^N], \tag{2.17}$$

where the design point chosen is the point that maximizes $\nu^{KG}(\mathbf{x})$. To consider the immediate and expected improvement, we use the value-gradient utility given as

$$U = \mu^*_{\text{fused}} + \max_{\mathbf{x} \in \mathcal{X}} \nu^{\text{KG}}(\mathbf{x}). \tag{2.18}$$

Considering Eq. (2.18), the value-gradient policy chooses the next point to query such that

value-gradient utility is maximized.

Searching for the next point to query begins by generating Latin hypercube samples (LHS) from the input space as alternative points. When an alternative point is queried from the GP_i, based on Eq. (2.4), the output is presented as a normal distribution with a mean and variance. Then, N_q independent samples are drawn out of this normal distribution

$$f_i^q(\mathbf{x}) \sim \mathcal{N}(\mu_i(\mathbf{x}), \sigma^2_{\text{GP},i}(\mathbf{x})) \quad \text{for } q = 1, \ldots, N_q, \tag{2.19}$$

with $f_i^q(\mathbf{x})$ as the sample q for alternate point $\mathbf{x}$ and information source i. To calculate the value-gradient utility for each alternate point, we augment GP_i with a sample drawn from GP_i to build the $GP_{temp,i}$. This approach assumes that the alternate point and the predicted sample value are part of the system's knowledge and a new fused model is created. Using Eq. (2.18), the value-gradient utility is measured. The process is repeated for all samples drawn from the GP_i for the same alternative point. For every alternative point and information source, the expected value-gradient utility is calculated by

$$EU_{\mathbf{x},i} = \frac{1}{N_q} \sum_{q=1}^{N_q} U^q_{\mathbf{x},i}. \tag{2.20}$$

To consider the cost of querying information sources, the expected utility function can be found per unit cost. Finally, the alternative point with the highest expected value gradient utility per unit cost and the information source to which that alternative point was added is chosen for the next query. This process is repeated and the fused GP is updated until the budget is exhausted or convergence requirements are fulfilled. Then, by searching the fused GP's mean function, the optimum solution of Eq. (2.3) is estimated.

2.3.4 Case studies

There are many ways to approach an optimization problem based on how the search policy is employed and how the problem is constrained. In our work, three approaches are considered.

The first two approaches are cost-controlled, but with different triggers to call the RVE model (or ground truth) and which components of the calculation are considered in the cost. The third approach has no cost consideration in executing different information sources. For every particular case-study, five realizations of the entire process are obtained and the results presented are the averaged values over these five realizations. The three different approaches are described in the following subsections.

2.3.4.1 Cost Consideration - Iteration Controlled 'Ground Truth' Query (CC-IC)

When querying information sources while limited resources are available for the design process, it is important to compute the value that each query adds to the system, while also considering the resources needed to execute the query. For our cost consideration - iteration controlled (CC-IC) approach, at each step, we select an alternative point and an information source to query that adds the highest value per unit cost to the system's state of knowledge. Considering the cost of queries influence the decision making process with exploitation of the design space being favored over its exploration. The decision-maker in this case acts conservatively in order to avoid the risk of wasting the (constrained) resources on highly uncertain queries unless it can expect a very good 'return on the investment' of querying that source. To update the model discrepancies, the 'ground truth'—i.e. the RVE—is queried once after every ten evaluations of the information sources. In Table 2.1, the costs assigned to each information source are shown based on their computation time [64].

Table 2.1: Computational cost of various information sources.

Information source	Cost (Seconds)	Normalized cost
Isostrain	2.3×10^{-4}	1
Isostress	1.0×10^{-3}	4.4
Isowork	4.7×10^{-1}	2.0×10^{3}
Secant method	8.4×10^{1}	3.7×10^{5}
Elastic constraint	3.6×10^{1}	1.6×10^{5}
RVE	7.2×10^{3}	3.1×10^{7}

2.3.4.2 *Cost Consideration - Cost Controlled 'Ground Truth' Query (CC-CC)*

We also study a case in which the cost of the entire process is considered. This is referred to here as the cost consideration - cost controlled (CC-CC) approach. While querying information sources is costly, so too are constructing GP surrogates and calculating the fused model and utility function, especially as the number of data points increases over time. We call this the modeling cost and measure it by calculating the time required for (i) generating alternative points and samples, (ii) building temporary GPs, (iii) performing the reification process, information fusion, and building the fused GP, (iv) evaluating the utility function, and (v) deciding on the next point to query. The measurement of the modeling cost, including the computational cost of the individual sources and the cost of generating points to sample and selecting the sources to query, is completed on each iteration and subtracted from the budget.

Additionally, we define a cost-controlled criterion to query the 'ground truth' model. Basically, instead of querying the 'ground truth' after a specific number of queries of the (cheaper) information sources, a query is made when a specific amount of the budget is spent. This strategy allows the decision-maker to use the cheaper information sources before querying the expensive 'ground truth' model, as long as it is finding value in querying the cheap information sources. This policy is in a sense the closest analogue to a realistic materials design campaign, where the budget available greatly determines the querying (i.e. experimentation or simulation) protocol.

Considering the modeling cost *prevents the framework from querying a very large number of points from the cheaper information sources before querying the 'ground truth'*. For instance, thousands of queries are needed to pass any defined amount of budget if the isostrain information source is chosen each time. This has the potential of significantly increasing the overall cost. By considering the modeling cost, fewer calls to the cheaper models are made, which helps to moderate the increase in the modeling cost.

2.3.4.3 No Cost Consideration (NCC)

When the cost of evaluating information sources is not considered as part of the decision making process, the system always chooses the alternative point and the information source to query with the highest expected value gradient utility, no matter how expensive the query is. In other words, there is no constraint on the decision making process concerning costs. In this approach, which we refer to as no cost consideration (NCC), the 'ground truth' is queried after every ten information source evaluations in the step to update the model discrepancy terms. This strategy leads to increased exploration of the design space, thus adding the maximum knowledge from the evaluation to the system after each update. Although this approach is not considered practical in real-world optimization problems, it is used as a comparison with the cost-controlled approaches in the current work.

2.4 Results

We now present the results of applying our material design framework to optimize the normalized strain hardening rate, $(1/\tau)(d\tau/d\epsilon^{pl})$ at a plastic strain level of 0.9%, of a dual-phase (ferrite-martensite) steel by adjusting the composition and heat-treatment temperature of the material. The design problem follows the three decision-making policies described in Section 2.3.4. To compare these decision-making policies, we consider how quickly each case reaches what is considered as the maximum mechanical property region of the design space. In our previous work ref. [64], for a fixed material chemistry, the maximum objective value was found to be 30.5, with a certain level of uncertainty due to stochastic nature of the process—related to the different test points arising from Latin hypercube sampling. In the current work, we seek to find an objective value as good as the best in ref. [64] or better. Accordingly, a normalized strain hardening rate ($[1/\tau][\mathrm{d}\sigma/\mathrm{d}\epsilon]$) greater than 30.5 is considered as the target objective value to reach and this is shown by the shaded area in Figs. 2.3 and 2.4.

To quantify the uncertainty in the current design process, five realizations of the design process were generated. Each realization utilized the same initial data and was allowed to run to comple-

tion. Through this approach we were able to quantify the combined effect of several sources of uncertainty in the optimization process. The first is that the RVE calculations rely on a stochastic approach to generate the dual-phase microstructure. As a result, the prediction from RVE calculation can be different even for the exact same inputs that include volume fraction and composition of the constituent phases. The next source of uncertainty that needs to be accounted for is the determination of the points at which the *KG* is evaluated. This is done by Latin Hypercube sampling and as a result will vary for each realization of the framework. The final source of uncertainty arises while calculating the *KG*, since a set of alternate points are sampled randomly from the normal distribution defined by the GP at the chosen test point. The *KG* is calculated for all these alternate points and then averaged. This average *KG* value is then used when calculating the test point with the maximum *KG*. Given the considerable computational cost of the scheme it is not feasible, at this time, to deconvolute these uncertainty sources, and we are thus only estimating the total uncertainty and depicting it as shaded uncertainty bounds in Figs. 2.3 and 2.4.

As a preliminary step to set up the design problem, it is important to study how the prior belief about the 'ground truth' result provided to the system can affect the final results. This prior information is defined as the number of initial RVE evaluations (or 'ground truth' information) used to train the intermediate GP for the RVE information source. The goal is to use an initial training set size that sufficiently adds information while avoiding any extra expense on generating unnecessarily large initial training data sets.

The results of the parametric study conducted to address this question are presented in Table 2.2. Specifically, the effect of different initial training set sizes of the RVE on iterations and cost required to reach the initial 'guess' maximum objective value is investigated. The parametric study was carried out using the NCC decision-making policy, where during the optimization, the RVE is queried every ten iterations or queries to the cheaper information sources. The cost shown is divided into cost of generating the initial training set and the cost incurred during the optimization process. The final objective value reached for all the cases is the maximum normalized strain hardening rate found after reaching 300 iterations. We observe that from an initial training set size

Table 2.2: Comparing the number of iterations and cost incurred to find the initial 'guess' maximum objective value of 30.5 for various initial training set sizes i.e. RVE results. The parametric study was carried out using the NCC decision-making policy. The final objective values obtained after reaching 300 iterations are also presented.

RVEs used to train		Iterations to exceed	Cost (ks)		Final objective
Initial	Optimization	known max	Initial	Optimization	value reached
10	14	141	72	110.811	31.6071
20	17	171	144	137.106	31.3176
40	8	81	288	60.921	32.6166
70	8	81	504	60.921	32.7012
100	5	51	720	37.326	32.0356

of 20 to 40, the iterations needed to reach the 'guess' maximum objective value decreases by half. However, as we increase the size of the initial training set from 40 to 70, we observe no significant change in the number of iterations required to reach the objective value and *it is thus a waste of resources to generate training sets greater than* 40. Similarly, we do not see any major improvement in the final objective value reached at the end of the optimization by increasing the initial training set from 40 to 70. A similar effect is observed when a training set size of 100 is considered. Therefore, to balance the cost of training and optimization, for all calculations presented hereafter, we use an initial training set size of 40 RVE points.

Figure 2.3 compares the performance of three decision-making policies described in Section 2.3.4. To this end, the maximum objective value found by each optimization process at end of an iteration is plotted. As seen in Fig. 2.3, after about 50 iterations, all approaches come up with better design points resulting in higher objective values. Basically, after a certain number of evaluations from lower fidelity models, the system is learning about the location of the maximum of the 'ground truth' model and evolves toward more and more optimal design points.

We note from this figure that the performance of both the NCC and CC-IC policies is similar. Out of all three policies, the NCC policy explores more since it is the least conservative. The CC-IC policy, on the other hand, prefers to exploit the current system knowledge rather than to explore highly uncertain regions. The CC-CC policy, which considers the calculation of the full

modeling cost in addition to the information source cost, affects the decision making process in that it shows slower convergence to the target properties. This is likely due to the variable number of iterations between RVE calls, which influences how extensively the framework is exploring the design space. Additionally, by defining the cost-controlled calls to the RVE model, we let the system exhaust the cheap information sources before each call, allowing for more exploration of the design space. However, we balance the number of queries from the cheap information sources by considering the modeling cost as well in this policy. Note that for a fixed number of iterations, the CC-CC policy is not calling the RVE as much as the other policies. Thus, due to the fewer calls to the RVE, it makes sense that the CC-CC policy does not reach high objective values as fast as the NCC and CC-IC policies.

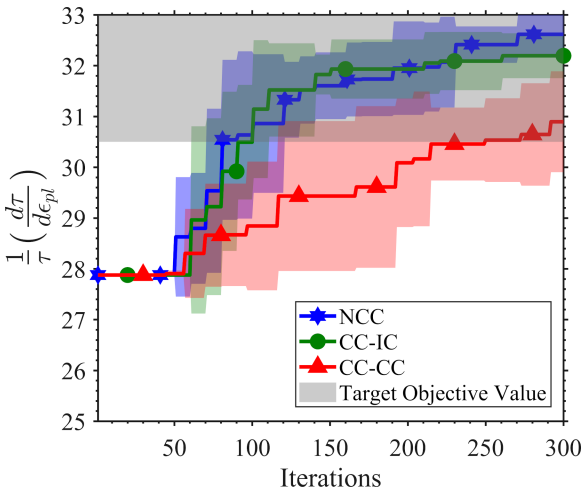

Figure 2.3: Comparing the maximum objective value found through the optimization process as a function of number of iterations for different decision-making policies: No Cost Consideration (NCC), Cost Consideration - Iteration Controlled (CC-IC) and Cost Consideration - Cost Controlled (CC-CC) 'ground truth' query.

Since the number of calls to the RVE and the number of information sources queried in each policy are different, every policy obtains an optimal solution (with respect to our target) at different overall cost. A comparison is done between all policies, Fig. 2.4, in order to show the objective value obtained as a function of the cost. The cost is calculated based on the time required to

complete the defined number of iterations in kiloseconds (ks). As shown in Fig. 2.4, the CC-CC policy is suggesting *significantly better design points than* NCC and CC-IC policies *at the same cost*. As stated before, the advantage of the CC-CC policy is that it favors the extraction of the most information out of a cheap source before querying the RVE. The CC-CC policy is thus the most cost-effective one, which is important in cases in which there are hard constraints the budget available to carry out the design. These cases are the norm rather than the exception in the context of materials design.

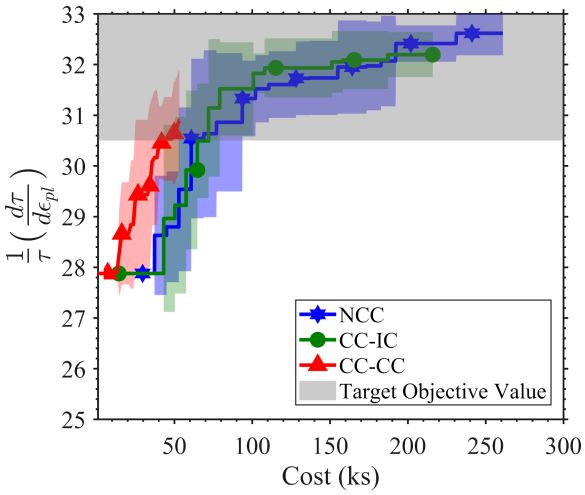

Figure 2.4: Comparing the maximum objective value found through the optimization process as a function of total cost for different decision-making policies: No Cost Consideration (NCC), Cost Consideration - Iteration Controlled (CC-IC) and Cost Consideration - Cost Controlled (CC-CC) 'ground truth' query.

To study the results in more detail, we investigate how each policy proceeds to an estimation of the optimal design point. Knowing that our framework employs multiple information sources to estimate the quantity of interest, the resource constraints on the decision-making process can have a large impact on how information sources are queried. By tracking the calls to different information sources during the optimization procedure, we can understand when and which information source had the highest value for Eq. (2.3) subject to the resource constraints of the given approach (CC-IC, CC-CC, and NCC). These results are shown in Fig. 2.5.

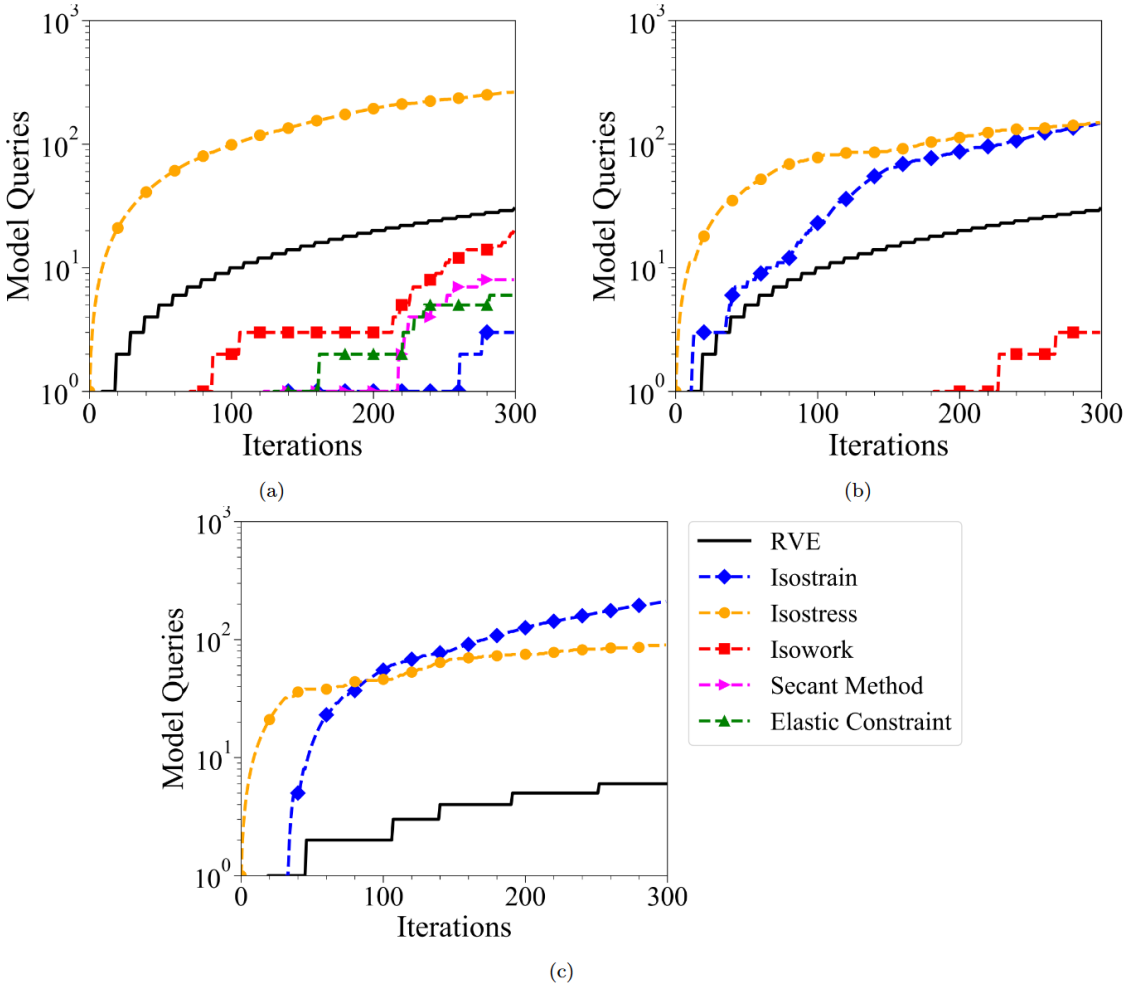

Figure 2.5: The progression of calls to different information sources as a function of the number of iterations for (a) No Cost Constraint (NCC), (b) Cost Constrained - Iteration Controlled (CC-IC) and (c) Cost Constrained - Cost Controlled (CC-CC) optimization procedures.

From Fig. 2.5(a), it seems that the NCC policy favors the querying of the design space through the isostress reduced order model at the beginning of the optimization process. We note that this is not the cheapest source. Over time, the amount of knowledge gained from querying this source reaches a stage of diminishing returns. At this stage, the framework begins querying the isowork model—still not the cheapest source. As the optimization progresses, the NCC policy ends up encouraging the querying of the design space through all the available information sources. This is an expected result since there are no cost constraints and there is thus no penalty based on the information source queried. This behavior also reinforced the idea that this policy results in a more

exploratory behavior, as the framework can utilize information from all sources more consistently. We note, however, that since the more expensive information sources are queried more compared to CC-IC and CC-CC policies, this approach will be considerably more expensive than the cost-constrained policies. A final observation to make is that, despite the fact that the micro-mechanical models (i.e. 'secant method' and 'elastic constraint') are more sophisticated than the simpler isostress approximation, the framework favors, by far, querying from this cheaper source, perhaps because the system has 'discovered' that it can make very useful inferences about the 'ground truth' by querying this cheap, and perhaps less accurate, source.

In a similar fashion to the NCC policy (Fig. 2.5(a)), the CC-IC (Fig. 2.5(b)) and CC-CC (Fig. 2.5(c)) policies favor querying the isostress information source first, followed by the isostrain approximation—these are the cheapest information sources available. Although the isostress approximation is more expensive than the isostrain model, the system initially finds more value in querying the former. After some iterations, however, the value—i,e, the ratio of information gained per cost—of the isostrain information source increases relative to the isostress approximation. We note that when the framework follows the CC-CC policy (Fig. 2.5(c)), the transition in favor of the isostrain approximation (cheaper than the isostress model by a factor of $\sim$4) happens faster, as this is the most effective strategy to follow under hard resource constraints.

Considering the cost difference between the cheapest and most expensive information sources, the value of the information obtained from the more expensive information sources must be considerably higher to have a higher cost adjusted value in comparison to the cheaper information sources. As such, the framework is expected to exhaust all the information from the cheaper sources before moving on to the more expensive sources. In fact, it can be seen that none of the micro-mechanical models is queried at all under CC-IC and CC-CC and the isowork information source is only queried under CC-IC at the very last stages of the process.

A major challenge of any optimization framework is to detect when convergence has been reached. In this very specific case, the problem is even more challenging, as the evaluation of the 'ground truth' has a degree of noise arising from the stochastic nature of the process. The

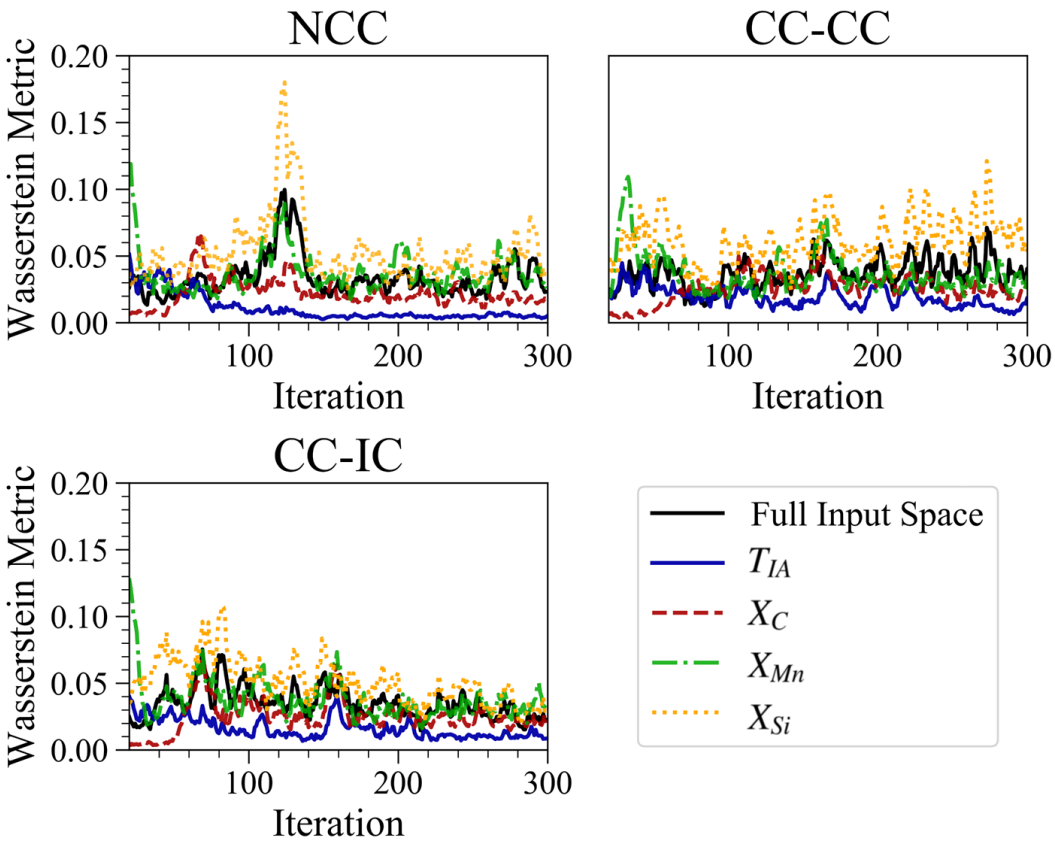

Figure 2.6: Evolution of the Wasserstein Metric for the full four-dimensional input space and each of the individual dimensions, temperature (T_{IA}), carbon (C), manganese (Mn) and silicon (Si) as a function of the number of iterations for No Cost Constraint (NCC), Cost Constrained - Cost Controlled (CC-CC) and Cost Constrained - Iteration Controlled (CC-IC) optimization procedures.

framework is thus *expected to converge into a region, rather than a single point*, in the design space. One possible approach to observe how the framework is able to approach the optimal point (or region) in the design space is to utilize the Wasserstein metric, also referred to as the Earth mover's distance. The Wasserstein metric can be understood as a distance function between two probability distribution functions in a given metric space. A more complete description of the theory and properties of the Wasserstein metric can be found in ref. [78]. This metric has been used in many applications, such as comparing the color distribution in images [79, 80], measuring the mixing and convergence of Markov Chains [81, 82] and as the loss function for the training of Generative Adversarial Neural Networks [83].

Here, we compute the Wasserstein metric between multiple one-dimensional distributions to understand how the framework approaches the optimal point or region in the multi-dimensional design space. To this end, we first compute the Euclidean distance between the current best point (that corresponds to the maximum of the fused GP) and a reference point. The reference point is taken to be $\chi_0 = [0.3, 0.3, 0.3, 0.3]$ in the unit hypercube design space. As the optimization progresses, the evolving distance between the current best point and the reference point form a distribution. We use the distribution of the distance in the four-dimensional space as well as the distributions of its projection along the individual dimensions to compute the Wasserstein metric. Specifically, starting from iteration 20, we take the 20 distance measurements prior to the iteration, split them into two sets of 10 each, and then compare the two sets using the Wasserstein metric. As the process converges, the Wasserstein metric is expected to approach zero, provided the sample size is large enough. In our case, having 10 samples per distribution means that the metric will be noisy even close to convergence. Here, we consider that a measure of convergence corresponds to a condition in which the Wasserstein metric maintains an almost constant value over multiple iterations.

As can be seen in Fig. 2.6, the Wasserstein metric computed from the distribution of the distance in the full four-dimensional input space show that the three different optimization policies have similar convergence (or lack thereof) patterns. However, the evolution of the Wasserstein metric computed from the distributions of the projections of the distance along the individual dimensions does provide a better understanding of the optimization process. These show quite clearly that all three methods converge rapidly in the temperature 'dimension'. In other words, the framework finds the optimal temperature very quickly. The next design input to reach convergence is the amount of carbon in the steel, followed by the manganese concentration. We note, however, that the silicon signal is very noisy and there is no indication of convergence. The slower convergence of manganese and non-convergence in silicon is somewhat expected since both elements do not play a significant role in controlling the normalized strain hardening rate of the material.

Finally, after all the evaluations are done and the system has built the final fused models, the

fused models mean functions are searched to find the highest objective value they suggest. However, the predicted objective value is not important in itself, but rather, the corresponding design point is more valuable. The optimal design point suggested by all three approaches is compared with the predictions of a single PSP chain involving thermodynamic calculations coupled with microstructure-based finite element calculations (RVE) in Fig. 2.7. In the figure, the contour plots show the variation of the maximum objective value where the maximum is taken over two (hidden) design variables, χ_3 and χ_4, for a given set of values of the remaining two (displayed) design variables, χ_1 and χ_2, i.e. $\max\left[f_{\chi_1,\chi_2}\left(\chi_3,\chi_4\right)\right]$, with χ_1 and χ_2, as obtained from the predictions of the aforementioned single PSP chain. Similarly, two out of four optimal design variables suggested by all three approaches are plotted at a time in Fig. 2.7. As expected, the CC-CC policy still makes predictions slightly different than the NCC and CC-IC policies. This may be because with a fewer number of queries from the RVE, the system has made less corrections to the fused model prediction compared to NCC and CC-IC policies. On the other hand, both the NCC and CC-IC policies are predicting the optimal design point very closely to the 'true' optimal design point. Furthermore, from the chemistry/processing (design) space exploration plots in Fig. 2.7 it can be noted that the true optimal solution is not unique. For example, in Fig. 2.7(a), we see that an objective value of $\left(1/\tau\right)\left(\mathrm{d}\tau/\mathrm{d}\epsilon_{pl}\right) \approx 32$ has multiple solution in temperature and carbon space, i.e. the optimal solution is a region in the temperature and carbon space. Similarly, in Figs. 2.7(b) and 2.7(c) we see that the solution in temperature and manganese space, and temperature and silicon space, respectively, is also not unique.

2.5 Discussion

In this chapter, we expanded on our prior work [6, 64] on multi-information source fusion BO framework to account for the material design space that are truly amenable to modifications without concerns on whether they are feasible or not. Note that while promising, a microstructure discovered by only considering material microstructure as design space can be optimal in principle but may not be accessible in practice. We have demonstrated our framework by optimizing the performance of a material with material chemistry and processing conditions as the design space, thus

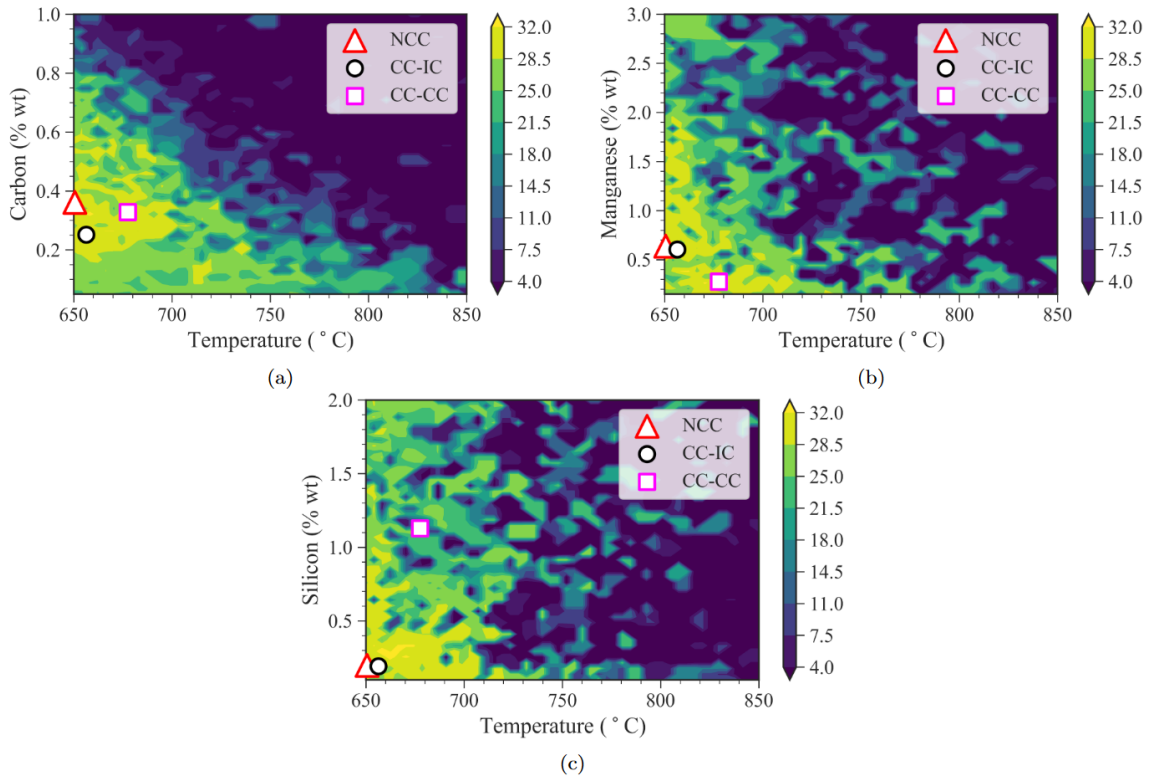

Figure 2.7: The optimal design point corresponding to the maximum objective value found by the three optimization processes, No Cost Constraint (NCC), Cost Constrained - Cost Controlled (CC-CC) and Cost Constrained - Iteration Controlled (CC-IC), are overlaid on the contour plots of the normalized strain hardening rate (objective) in the design space (temperature, carbon, manganese and silicon) generated by extensively exploring the 'ground truth' process-structure-property modeling chain involving thermodynamic calculations coupled with microstructure-based finite element calculations (RVE).

truly realizing the ICME paradigm. Specifically, we have utilized our material design framework to optimize the strength normalized strain hardening rate (a single metric that provides an indication of the ductility and formability) of a dual-phase (ferrite-martensite) material by adjusting the composition and heat-treatment temperature of the material. The framework utilized a surrogate model of the thermodynamic results to predict the phase volume fraction and composition in the material microstructure post single-stage heat-treatment (intercritical annealing followed by quenching), which was then used to predict the mechanical properties of the dual-phase material using a variety of micromechanical models (as cheap information sources) and a high throughput

microstructure-based finite element model (as expensive 'ground truth').

The material design framework developed herein was tested under three different conditions. These conditions were shown to affect how quickly the framework attained a sufficiently high value of the objective function and the information sources that were used to achieve that solution. For example, it was seen that when comparing the NCC (no cost considetation) to the CC-IC (cost consideration-iteration controlled) or CC-CC (cost consideration-cost controlled) policies, the NCC policy was able to utilize all the information sources while the other policies only favored queries from the cheaper models. The results indicate that querying extensively from all models leads to an increased exploration of the design space. This, however, may not be desired for real life design problems where it is often necessary to obtain an optimum material composition within a tight budget constraint. In this regard, it is worth mentioning that when studying different policies by comparing the amount of cost they have incurred and the objective values they have obtained, the CC-CC policy showed significantly better performance by using less resources compared to the other two policies. This highlights the value in extensively extracting information out of cheaper information sources.

One big challenge in the optimization of an unknown design space is how to ascertain that the solution obtained from the optimization framework has converged. To address this challenge, we utilized the distance of the suggested optimal design point to a fixed reference point at every iteration in order to track the convergence of the framework. This convergence is strongly tied to the influence that a given dimension(s) in the input space has on the final objective function. For example, in problem at hand, temperature and carbon have the biggest influence on the objective function and it is to be expected that the solution subspace corresponding to these to input variables converges relatively fast. On the other hand, silicon and manganese only have a weak influence on the objective, thus it is expected that a rather broad range of their values can lead to similar values of the objective function—the design space is almost degenerate with respect to silicon and manganese, once an optimal region within the temperature-carbon 'slice' has been identified. This makes the optimization of those two parameters extremely challenging, and highly 'oscillatory'.

The fact that not all the dimensions in the problem are equally influential (recall Fig. 2.6) points to a possible strategy to make this framework even faster and more efficient: *find and follow directions (i.e. subspaces) in the design space along which the objective function changes the most*. This approach, known as the active subspace method [50], decomposes the design space in such a way that at any time only the most 'influential' or 'active' subspaces are considered when trying to find the optimum in a multi-dimensional space. Effectively, this approach would reduce the dimension of the problem, accelerating the rate at which the solution is approached. As hinted above, in this problem we would expect that such an active subspace search would hone in on the temperature-carbon subspace prior to refining the optimization by exploiting the manganese and/or silicon coordinates. A multi-information source fusion BO framework with active subspace search will be presented in chapter 4.

3. MULTI-INFORMATION SOURCE BAYESIAN OPTIMIZATION OF MULTI-OBJECTIVE FUNCTIONS*

3.1 Overview

Multi-objective optimization is often a difficult task owing to the need to balance competing objectives. A typical approach to handling this is to estimate a Pareto frontier in objective space by identifying non-dominated points. This task is typically computationally demanding owing to the need to incorporate information of high enough fidelity to be trusted in design and decision-making processes. In this chapter, we present a multi-information source framework for enabling efficient multi-objective optimization. The framework allows for the exploitation of all available information and considers both potential improvement and cost. The framework includes ingredients of model fusion, expected hypervolume improvement, and intermediate Gaussian process surrogates. The approach is demonstrated on a test problem and an aerostructural wing design problem.

3.2 Introduction

When estimating a ground truth quantity of interest, for example, fuel burn for an aircraft or a particular material property, we can often consider different mathematical formulations of the analysis or prediction. This, in addition to experimental data and expert opinion, can give rise to the ability to use multiple different sources of information for the estimation task at hand. The different assumptions made lead to differing levels of fidelity among the sources, as well as different costs, both in terms of time and monetarily. In the presence of multiple sources of information, we seek analysis and design processes that exploit the extra information that would not be present if only a single source were available. The opportunity is the efficient selection of which information source to query and where to query it on the basis of cost and potential for improvement in the estimation of a quantity of interest. To do so, we employ a Bayesian optimization framework well-

*Reprinted with permission from "Bayesian Optimization of Multi Objective Functions Using Multiple Information Sources" by D. Khatamsaz, L. Peddareddygari, S. Friedman, and D. Allaire., 2021. AIAA Journal, Copyright 2021 by the American Institute of Aeronautics and Astronautics, Inc.[84]

suited to the optimization of black-box functions. These approaches generally use an acquisition function to search the design space effectively and efficiently through the tradeoff of exploration and exploitation. The challenge is ensuring proper fusion of information as it becomes available and a need for a rapid capability for moving from prior predictive information to posterior predictive information without necessarily executing a true information source. The standard practice of using Gaussian processes within Bayesian optimization frameworks as updatable surrogates provides an avenue for efficiently incorporating information source fusion within the search process. We exploit this here. We also note here that our goal is to estimate and optimize properties or other performance metrics of real systems. There is therefore, a notion of ground truth, which is the true quantity being estimated, which is likely only observable with noise. Often, this ground truth is represented as the information that can be acquired from the highest fidelity information source. This assumption may be reasonable in some circumstances, particularly if the information source is an experiment with the realized system. Here, we keep the term ground truth to ensure that the overall goal is clear, and we use our highest fidelity information source as a proxy for that ground truth. In the previous chapter, we handled the case where ground truth is measured with noise. Here, we do not include the noise in the ground truth for clarity, but the framework we present can incorporate this if known.

While a multi-information source capability can be applicable to a wide variety of contexts, our focus here is on multi-objective optimization. Previous works in this area, particularly with emphasis on multifidelity methods, includes for example, an efficient global optimization (EGO) approach based on the use of a hypervolume indicator or surrogate models creating for every objective [85]. In Ref. [86], a point-by-point approach is employed that considers the ends of the estimated the Pareto front in an effort to obtain better solutions via single objective optimizations. In Ref. [87] encourages the use of standard multi-objective evolutionary algorithm introduced in Refs. [88, 89] to apply on the lower fidelity information source to build a surrogate model to search and obtain a Pareto front. Then, the high fidelity information source is evaluated at those non-dominated designs to correct the surrogate model associated to the lower fidelity information

source. A similar approach is suggested in Ref. [90] using a surrogate model built with samples from a low fidelity information source to search the design space for potential non-dominated designs. A high fidelity information source is then evaluated at those design points and the approximation of the Pareto front is obtained by optimizing a cokriging model constructed with these new evaluations.

In this chapter, we present a novel framework for exploiting available information sources to identify non-dominated points in objective space to estimate the true Pareto front of a given problem. To handle fusion of information sources, we incorporate model reification introduced in [5], which builds off [22, 77]. Model reification is a fusion technique that learns correlations among information sources and guards against overconfidence that can occur when nearly identical sources are used. This fusion process is also non-hierarchical and allows fidelity to vary over a design space. To enable rapid assessment of posterior predictive information, we use Gaussian processes as intermediate surrogate models that may be temporarily updated with candidate query points [51, 91, 5, 34]. To drive candidate query points toward the Pareto front, we use the expected hypervolume improvement metric presented in [92]. The approach presented in Ref. [92] provides an exact means of calculating the expected hypervolume improvement. This leads to an efficient computational process since a closed-form expression can be used to find the expected hypervolume improvement (EHVI). Overall, our novel Bayesian multi-information source multi-objective optimization framework can exploit multiple non-hierarchical information sources in an efficient manner that produces higher quality Pareto fronts at less computational expense than current available methods. This is achieved via combined use of model reification based information fusion within a Bayesian optimization paradigm over a set of available information sources where querying is directed by an easily computable closed-form acquisition function based on the EHVI. We have chosen a Bayesian optimization paradigm here because the problems we seek to address involve data-driven optimization. That is, our objective function estimates, as computed by available information sources, are analytically unknown and must be learned during search. While there are other optimization strategies, such as model management approaches of Refs. [93] and

[94] and model fusion approaches of Refs. [35] and [95], Bayesian optimization is viewed as a superior computational strategy when tasks of exploration and exploitation must be traded off as discussed in Ref. [96]. Our approach is demonstrated on a test problem with a two-dimensional input space. We then demonstrate our approach on an aerostructural wing design problem involving a 17-dimensional input space. These input space dimensions stress the limits of typical Gaussian process regression modeling, and our approach is still shown to perform well. In each demonstration, we consider two objectives, however, this is not a limitation of the work.

3.3 Background

Our multi-information source optimization approach for multiple objectives employs Gaussian processes as intermediate surrogate models and fuses information using the process of model reification. We describe each of these ingredients in turn in this section. We then conclude this section with background on a general multi-objective optimization formulation based on the Pareto frontier, which is how we approach such problems here. Once we have established the necessary ingredients of our approach, we move to a description of our formal hypervolume indicator based framework for multi-information source multi-objective optimization in Sec. 3.4. Gaussian process regression and information fusion from multiple sources are discussed in detail in chapter 2, therefore, in this section, the main focus is on other parts required in our framework for optimization of multi-objective functions.

3.3.1 Multi-Objective Optimization

A multi-objective optimization problem can be defined as

$$\text{minimize } \{f_1(\mathbf{x}), ..., f_n(\mathbf{x})\}, \mathbf{x} \in \mathcal{X} \tag{3.1}$$

where $f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})$ are the objectives and $\mathcal{X}$ is the feasible design space. Throughout this chapter, we develop unconstrained approaches, however, the inclusion of penalty terms could be considered for constraint handling. Another possibility for constraint handling could involve the construction of a Lagrangian, where the objective is the EHVI and constraints are incorporated in

the usual fashion. This could provide a means for evaluating the Karush-Kuhn-Tucker conditions within the Bayesian optimization framework and open avenue for exciting future work in algorithmic development aimed pursuing the satisfaction of these conditions. For problems such as (5.9), it is usually the case that no single point optimizes each individual objective simultaneously. To deal with this, approaches based on the creation of a scalar objective using utility theory are common, as well as approaches based on finding non-dominated solutions approaching the Pareto frontier. We focus on the latter here. For this case, optimal solutions, $\mathbf{y}$, to a multi-objective problem with n objectives are denoted as $\mathbf{y} \prec \mathbf{y}'$, and are defined as

$$\{\mathbf{y} : \mathbf{y} = (y_1, y_2, \ldots, y_n),\ y_i \leq y_i'\ \ \forall\, i \in \{1, 2, \ldots, n\},\ \ \exists\, j \in \{1, 2, \ldots, n\} : y_j < y_j'\} \tag{3.2}$$

where $\mathbf{y}' = (y_1', y_2', \ldots, y_n')$ denotes any possible objective output. The set of $\mathbf{y} \in \mathcal{Y}$, where $\mathcal{Y}$ is the objective space, is the Pareto front of the problem. This is shown conceptually for a bi-objective problem in Fig. 3.1. All points on the Pareto front are non-dominated. Our approach, which is common in the literature, is thus to find the Pareto front as efficiently as possible for a given multi-objective optimization problem.

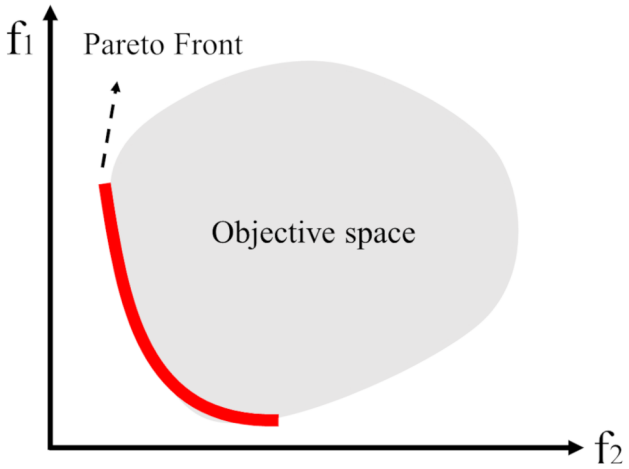

Figure 3.1: All points on the red line are non-dominated and constitute the solution set.

There are many techniques in use for approximating Pareto frontiers for multi-objective optimization problems. Among these are the weighted sum approach [23], the adaptive weighted sum approach [24], normal boundary intersection methods [25], hypervolume indicator methods [26, 27, 28, 29, 30, 31, 32], and others. The hypervolume indicator approach is well-suited to expected improvement based algorithms, which have been shown to work well in a multiple information source setting (see, e.g., Refs. [33, 34, 6, 35]). Thus, our approach proposes the incorporation of hypervolume indicator improvement within a multi-information source querying framework. Hypervolume indicator approaches are based on the concept of a hypervolume in objective space. These hypervolumes are measured relative to a fixed reference point and the enclosed volume between the approximated set of Pareto points and the reference point is computed. The concept is shown notionally in Fig. 3.2. Here, the shaded area is the hypervolume to be computed. In general, if a given set of points has a higher hypervolume than another set, then the given set is a better estimator of the Pareto front. Hypervolume indicator algorithms seek to maximize the hypervolume in objective space so as to best approximate the Pareto front. Thus, the value of new query points can be estimated (using prior predictive distributions if using Gaussian processes) by measuring the expected improvement in the hypervolume that would occur given the query takes place. In Fig. 3.2, the shaded area in blue shows the amount of increase in hypervolume when a new non-dominated point is found and added to the solution set.

3.4 Approach

Bayesian optimization is an optimization technique aimed at learning what is needed about an underlying black-box function to efficiently optimize it (see e.g., Refs. [97, 98, 99, 100]. As such, Bayesian optimization methods seek to trade off the tasks of exploration and exploitation. These methods traditionally employ Gaussian process surrogate models that can be temporarily updated to assess the quality of a candidate query point. This quality is measured by an acquisition function, such as expected improvement, probability of improvement, the knowledge gradient, and others. Our multi-objective optimization approach here treats information sources as black-boxes and uses the EHVI acquisition function. Thus, our proposed method is one of Bayesian optimization.

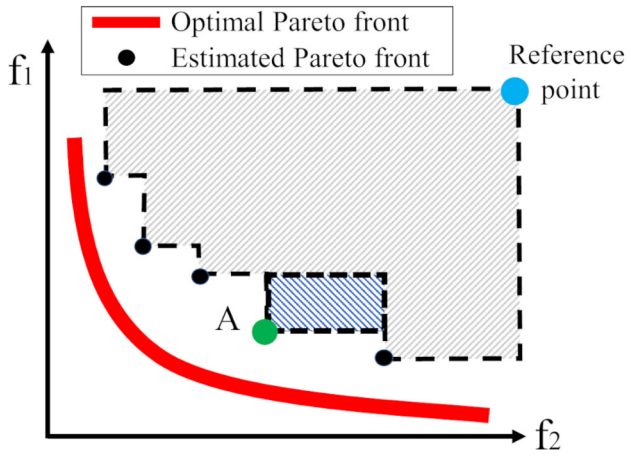

Figure 3.2: The blue shaded region corresponds to the hypervolume improvement due to adding point A to the solution set.

Generally, our approach is based on determining, with available prior information, where to query and what source to query to maximize the hypervolume indicator while being budget aware. To achieve this, we make use of the updatable Gaussian process surrogates described previously for each information source. These surrogates can be used as prior predictive distributions that can be temporarily updated with potential query locations that result in potential changes to the hypervolume indicator. By searching over the space of potential query locations and potential information sources with these prior predictive surrogates, we are able to efficiently identify the next best query to execute. Once this query is executed, all surrogates (including correlation information) may be updated, and then can serve as prior predictive distributions for the next iteration. In this section, we describe in detail our approach to achieving this. We begin with the necessary preliminaries regarding the fast calculation of the expected hypervolume improvement (EHVI) [101] within a multi-information source framework. This discussion follows largely from Ref. [92] where more details can be found if desired. We then describe our algorithm for multi-information source multi-objective optimization. In Sec. 3.5 we demonstrate the use of this framework on a test problem and an aircraft wing design problem.

3.4.1 Preliminaries

Following Ref. [92] for the development of the fast computation of EHVI, we present here our implementation within a multi-information source setting. We begin by considering a current solution set, $\mathcal{S}$, of non-dominated points in objective space at some point during a multi-objective optimization process. The dominated hypervolume, denoted as $\mathcal{H}(\mathcal{S})$, can then be computed given $\mathcal{S}$ and a reference point. Improvement to the hypervolume due to adding a new solution vector $\mathbf{y}$ is then defined as

$$\mathcal{H}_{\mathcal{I}}(\mathbf{y}, \mathcal{S}) = \mathcal{H}(\mathcal{S} \cup \mathbf{y}) - \mathcal{H}(\mathcal{S}) \tag{3.3}$$

If $\mathcal{H}_{\mathcal{I}}(\mathbf{y}, \mathcal{S}) > 0$, then $\mathbf{y}$ is in the non-dominated region of $\mathcal{S}$ and can be used to update the solution set. Otherwise, there is no improvement over $\mathcal{H}(\mathcal{S})$ by adding $\mathbf{y}$ and the query adds no value. In the context of Bayesian optimization, $\mathbf{y}$ is a random output of a probabilistic model related to a potential solution in the design space. Hence, $\mathcal{H}_{\mathcal{I}}(\mathbf{y}, \mathcal{S})$ is also a random variable. Therefore, it is possible to calculate its expected value, which is the EHVI. Comparing EHVI values for different potential solutions in the design space and finding the maximum EHVI leads to an information-economic querying policy that ensures maximum gains are achieved in each successive query. In a multi-information source context, however, the different cost of querying each source should also be taken into account.

The formula for calculating EHVI as outlined in [102] is given as

$$E[\mathcal{H}_{\mathcal{I}}(\mathbf{y})] = \int_{U} P(\mathbf{y} \prec \mathbf{y}') d\mathbf{y}' \tag{3.4}$$

where $P(\mathbf{y} \prec \mathbf{y}')$ is the probability that $\mathbf{y}'$ is dominating $\mathbf{y}$ and U is the dominated hypervolume. In our context, this can be computed in closed-form as will be shown below. Given that we have independent Gaussian process models for every objective for each information source, the posterior predictive output of each model given the data is a random variable identified as $y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ where $i \leq m$ and μ_i, σ_i^2 are the mean and variance of the i^{th} objective accordingly (note, we have not included information source specific indices here for notational clarity). For a new potential

solution in the design space, we have the following equation:

$$P(\mathbf{y} \prec \mathbf{y}') = \prod_{i=1}^{m} \Phi\left(\frac{y_i' - \mu_i}{\sigma_i}\right) \tag{3.5}$$

where Φ is the cumulative distribution function of the standard normal random variable. Details regarding the closed-form expression of Eq. (3.4) along with a fast approach to compute the hypervolume associated with a solution set can be found in Refs. [92, 102, 103, 104, 105].

3.4.2 Multi-information Source Multi-Objective Optimization Framework

Using the Gaussian process as the surrogate model for each objective and EHVI as the acquisition function, we can perform Bayesian optimization to approximate a solution set for a multi-objective optimization problem. It is necessary to notice that the model discrepancies are changing whenever new information is found about the ground truth by querying the information sources. The model discrepancy is defined as the difference between the predicted value of the model built with data from an information source and the model built with the available data from ground truth for a specific design space point. Therefore, model discrepancies should be updated regularly. However, querying the ground truth to update its surrogate model is costly. Thus, we need to define a condition for when to query the ground truth. Such a condition can be, for example, when a certain number of updates have been made to available information sources, or also when a specific amount of the total allotted budget is spent. This method allows the decision maker to query cheap information source more between ground truth queries if it finds the cheap information source is still providing useful knowledge about the ground truth. This is inline with expected intuition regarding the exploitation of cheap information sources given their nearly negligible cost in comparison to expensive sources and ground truth itself.

Algorithm 1 presents our overall framework. Our procedure to optimize a multi-objective function is established assuming the function has m objectives, and there are n information sources of differing fidelity available to provide information about the ground truth. Here, the ground truth is the highest fidelity information source for estimating a quantity of interest. This could, for exam-

Algorithm 1 Multi-Objective Bayesian Optimization

construct GP_1 to GP_m given available data from the ground truth
for i from 1 to n **do**
  for j from 1 to m **do**
    construct $GP_{j,i}$ for objective (j) of the information source (i) given the data
  end for
end for
fuse models and construct the initial Pareto front
while available budget > 0 **do**
  X-sample set $\longleftarrow$ Latin Hypercube Sampling
  for k from 1 to n **do**
    for s from 1 to size(X-sample set) **do**
      Y-sample $\longleftarrow$ query X-samples from GP_{1k} to GP_{mk}
      construct temporary GPs by updating GP_{1k} to GP_{mk} using sample s
      updated_fused_values $\longleftarrow$ fuse other models with the updated one
      generate test_samples using fused_means and fused_variances
      improvement(s,k) $\longleftarrow$ **EHVI**(test_samples,updated_fused_values,Pareto front)
    end for
  end for
  X = sample to be queried , V = Information Source(IS) to query from $\longleftarrow$ Max(improvement)
  $Y = (y_1, ...y_m) = \text{IS}_{(V)}(X)$
  update $GP_{1,V}$ to $GP_{m,V}$ using X and Y
  fuse models
  U $\longleftarrow$ query a randomly generated set of design points from fused model
  find non-dominated vectors in U to update Pareto front
  if requirements to query ground truth is met **then**
    G $\longleftarrow$ a set of design points with arbitrary size distributed along the Pareto front
    $Y_G \longleftarrow$ query set G from the ground truth
    update GP_1 to GP_m
    update model discrepancies
    fuse models and update the Pareto front
  end if
end while
fuse models to construct fusedGP_1 to fusedGP_m
$\mathcal{U} \longleftarrow$ query a randomly generated set of design points from fused models
$\mathcal{S} \longleftarrow$ find non-dominated vectors in $\mathcal{U}$ to update the Pareto front
$\mathcal{X} \longleftarrow$ the design space corresponding to non-dominated set $\mathcal{S}$

ple, be a real world experiment on a realized system, or a validated high fidelity simulation model (often with associated uncertainty, which do not incorporate here without loss of generality). We

assume that the querying of ground truth is the most expensive means of acquiring information about. Expense here could mean runtime, cost, or other resource. While it is possible that ground truth may not be the most expensive information source to query, we do not consider that scenario here. The first step involves the construction of Gaussian processes (GPs) and to create the initial Pareto front. This can be established by finding non-dominated design points of initial data available from the ground truth. Next, the fusion step takes place, which involves the previously described model reification process. This is followed by the generation of candidate query points which are tested for EHVI potential given the current set of GPs. The best candidate (query point and information source) is selected and executed. This is followed by another fusion step given the new information. The budget condition is then checked, which would lead to a ground truth query or a check on whether the budget is exhausted. The term budget refers to the resources available to run a new experiment and limits the total number of evaluations. The budget can be defined, for example, as the computational time in simulations or the total money granted to design experiments in a laboratory. If the budget is exhausted, the process terminates with a final analysis of the estimated Pareto front from the fused GP, which leads to subsequent evaluations of best points from the ground truth. If the budget is not exhausted, the process resamples candidate points and repeats. A ground truth evaluation is triggered after spending a specified amount of budget on evaluating lower fidelity information sources. When this occurs the ground truth is queried, its GP is updated, and then all other GPs are updated (owing to a change in the discrepancies and correlations given new ground truth information). The budget exhaustion condition is then checked and the process proceeds as previously described from this point. A complete flow chart of this process is provided in Fig. 3.3.

Algorithm 2 Querying the Ground Truth

1: divide the most updated Pareto front into N slices
2: construct a smooth Gaussian process for each slice given data points in the slice
3: $P \longleftarrow$ choose the closest point to the Gaussian process mean in each slice
4: $P_Y \longleftarrow$ query the design points corresponding to P from the ground truth

When the decision to query ground truth is made based on the budget condition set by expert opinion, a certain number of points, N, are considered as potential queries. Although choosing larger values of N results in more information gain and higher accuracy to estimate the model discrepancies, it is not necessarily desirable since the ground truth is an expensive to evaluate function or experiment. Hence, a trade-off should be considered in assigning a value to N. For the purposes of the demonstration cases that follow in Sec. 3.5, we have set $N = 10$ for the test function and $N = 4$ for the OpenAeroStruct demonstration, however, the study of this parameter is a topic of future work. Algorithm 2 presents our ground truth querying strategy.

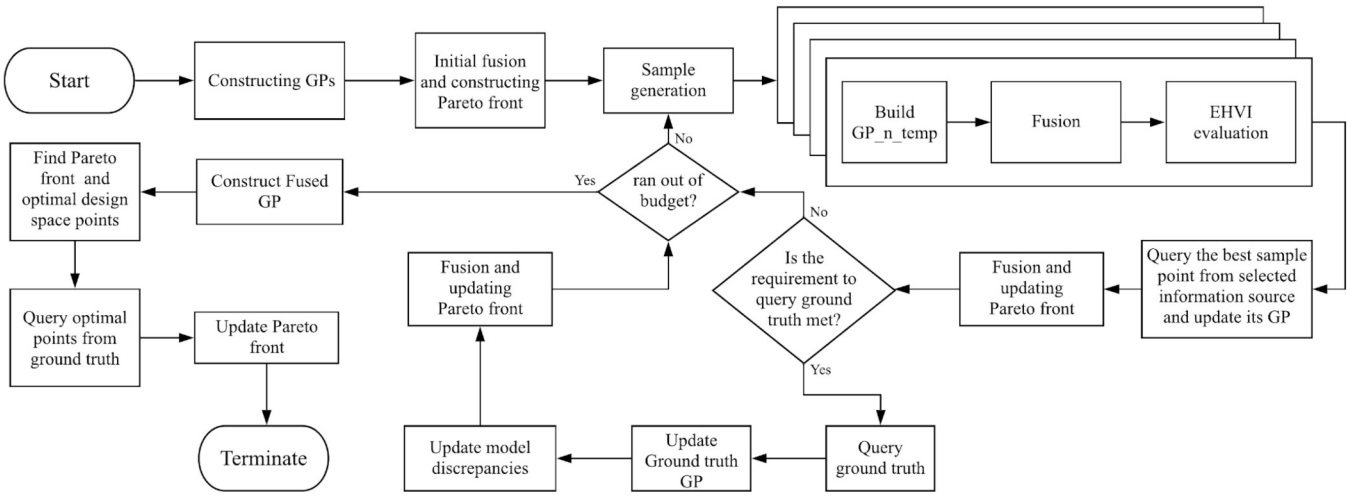

Figure 3.3: Procedure flow of the proposed framework. The ground truth query requirement can be meeting a certain number of iterations or spent budget.

3.5 Application and Results

To evaluate the performance of our proposed algorithm, we have applied it on a test function from Ref. [106]. This is referred to as Poloni's two objective test function, which maps points from a two dimensional design space to a two dimensional objective space. A comparison between the optimal Pareto front associated with the problem and the approximated Pareto front is made to show the effectiveness of the algorithm. To apply the concept of multi-fidelity approach in

optimizing the test function, we have constructed two other functions close to the test function by changing the coefficients and constants. The test function itself is considered ground truth. We follow the demonstration of our framework on the test function with its application to an aircraft wing design problem using OpenAeroStruct [107]. We describe the software and the problem for this application in Sec. 3.5.2.

We note here that our proposed optimization approach is stochastic in nature, and thus, involves uncertainty from a few different sources. The result is that the results are also stochastic in nature. The sources of uncertainty include the use of different training sets to build the initial Gaussian processes, the random nature of how we generate candidate design points to be tested, and the sample-based nature of the fusion process employed. To account for these uncertainties we present the results for several different simulations using different initializations and candidate point locations. We show this uncertainty in the form of 95% empirical confidence intervals in the relevant figures.

3.5.1 Poloni's test function

Poloni's two objective function is a 2-dimensional input test function defined as:

$$\begin{aligned}
&\text{minimize}: \quad f_1(x_1, x_2) = 1 + (\mathrm{A}_1 - B_1(x_1, x_2))^2 + (\mathrm{A}_2 - B_2(x_1, x_2))^2 \\
&\text{minimize}: \quad f_2(x_1, x_2) = (x_1 + 3)^2 + (x_2 + 1)^2,
\end{aligned}$$

where

$$\begin{aligned}
&-\pi \leq x_1, x_2 \leq \pi \\
&\mathrm{A}_1 = 0.5\sin(1) - 2\cos(1) + \sin(2) - 1.5\cos(2) \\
&\mathrm{A}_2 = 1.5\sin(1) - \cos(1) + 2\sin(2) - 0.5\cos(2) \\
&B_1(x_1, x_2) = 0.5\sin(x_1) - 2\cos(x_1) + \sin(x_2) - 1.5\cos(x_2) \\
&B_2(x_1, x_2) = 1.5\sin(x_1) - \cos(x_1) + 2\sin(x_2) - 0.5\cos(x_2).
\end{aligned}$$

Figure 3.4 shows the optimal versus final estimation of the Pareto front and the hypervolume. Since there is no closed form solution for Poloni's test problem, the optimal Pareto front and its

hypervolume are found by exhaustive search. The optimal Pareto front here matches those reported in Refs. [108, 109]. The estimated Pareto front is found using the knowledge of lower fidelity models about the ground truth. Looking at the hypervolume values, it is showing the hypervolume is increasing as a result of improved estimation of the Pareto front admitting that the budget is spent effectively. The budget here is set to 100 to limit the total number of information source evaluations and is used to inform the system when to trigger the ground truth evaluation if needed. The cost of querying the low fidelity and medium fidelity models are set to 1 and 2 units of cost respectively. These values were chosen to ensure adequate use of the information sources. In a practical setting, these values would be computed via resource usage (such as actual runtime). The hypervolume is initially computed using randomly generated data points with a training set of size 20 for every simulation.

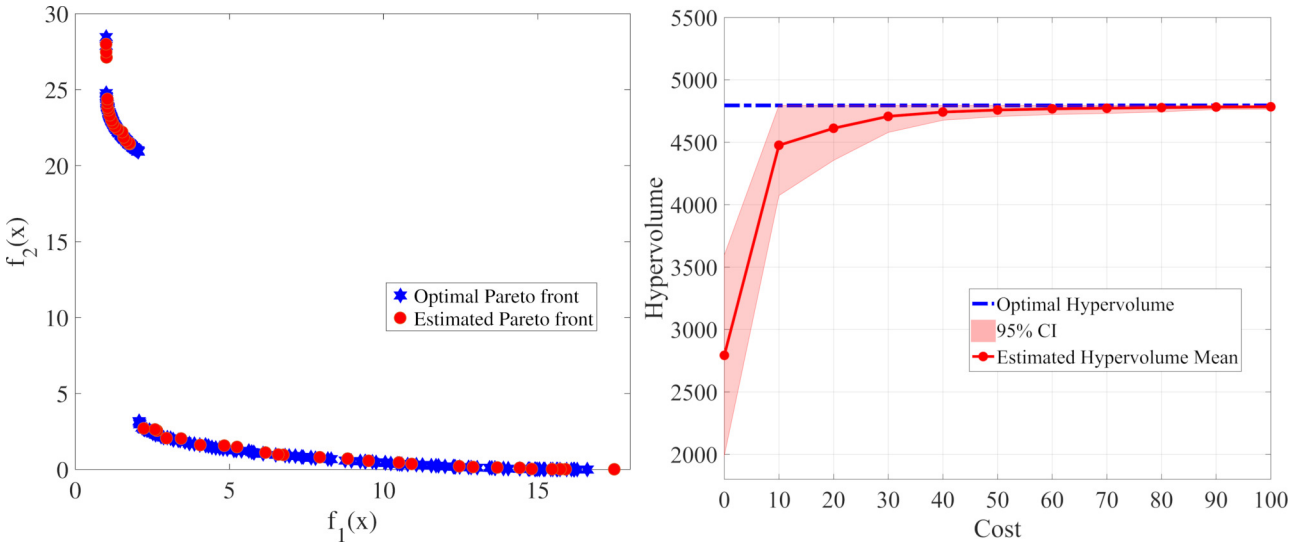

Figure 3.4: The optimal and estimated Pareto front and hypervolume averaged over 30 replications. The reference point is (70,70).

A significant improvement is achieved before spending the first 10% of the budget. This was achieved largely through the exploitation of the cheaper lower fidelity model. Note that the ground truth is queried each time 10 units of cost is spent to update model discrepancy and hypervolume

on a regular basis.. We present a similar result and the information source query history for the OpenAeroStruct demonstration in Sec. 3.5.2. The diminishing returns in hypervolume are expected given that finding new non-dominated points becomes more difficult as more points are found.

Many previous works have used Poloni's test function to measure the performance of their proposed approaches. For example, Ref. [109] proposes a method using differential evolution. Their result show they have found the optimal Pareto front after 600 function evaluations. Also, in Ref. [108], it is reported that 2500 function evaluations are used to cover the estimated Pareto front close to the optimal Pareto front found using exhaustive search. In Ref. [110], a genetic algorithm approach is taken using populations of more than 500 for 250 generations to obtain the optimal Pareto front. Our proposed method is generally outperforming each of these prior approaches. Here, our method using a total of 100 function evaluations with an additional set of less than 100 evaluations from lower fidelity information sources (which are considered much cheaper than the ground truth, though in general, this would be problem specific). This difference in the necessary number of evaluations emphasizes the efficiency gains achieved by our method while maintaining high quality Pareto front estimates.

3.5.2 OpenAeroStruct Demonstration

OpenAeroStruct is an open-source software developed in NASA's OpenMDAO framework [111], which can be used for fast tightly coupled aerostructural design optimization. The framework implements the coupled adjoint method to compute the aerostructural derivatives used for efficient gradient-based optimization. As noted in Ref. [107], OpenAeroStruct combines a vortex lattice method (VLM) and 1-D finite element analysis using six degrees of freedom 3-D spatial beam elements to model lifting surfaces [112, 107]. A common aerostructural single objective optimization problem is the fuel burn minimization problem using the Breguet range equation. Structural mass minimization of the wing is also frequently considered and thus is used as a second objective in demonstrating our proposed multi-information source multi-objective optimization framework.

The OpenAeroStruct application, as described in Ref. [107], uses the Breguet range equation to compute the fuel burn as a function of structural weight and aerodynamic performance. Design

variables consist of twist distributions, spar thickness distributions, and planform variables such as skin thickness, thickness over cord ratio, and angle of attack. The first four variables are 4-dimensional as four control surfaces were considered for the wing. Hence, the problem has a 17-dimensional design space. Constraints in the standard problem ensure lift equals weight, and that structural failure does not occur.

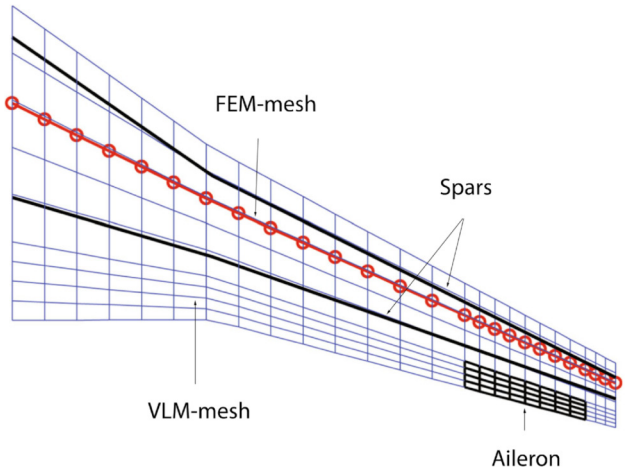

Figure 3.5: A wing with the aerodynamic and structural Meshes [1].

The mesh in the OpenAeroStruct is defined by the number of the spanwise and chordwise points. The fidelity of each model depends on the number of points used to define the lifting surface. A model with a finer mesh is considered to have higher fidelity compared to a model with a coarser mesh. We use three different mesh resolutions in this demonstration to serve as three different multi-fidelity information sources.

The different mesh sizes and costs are shown in Table 3.1, where Num_y is the number of span-wise points and Num_x is the number of chord-wise points. The low fidelity mesh was chosen to ensure meaningful results and the high fidelity mesh was chosen through a mesh refinement analysis that ensured adequate convergence. In Fig. 3.6, the three different meshes are shown. The cost of evaluating each model is based on the computational runtime of a single query.

We applied our approach of multi-information source multi-objective optimization on this two-

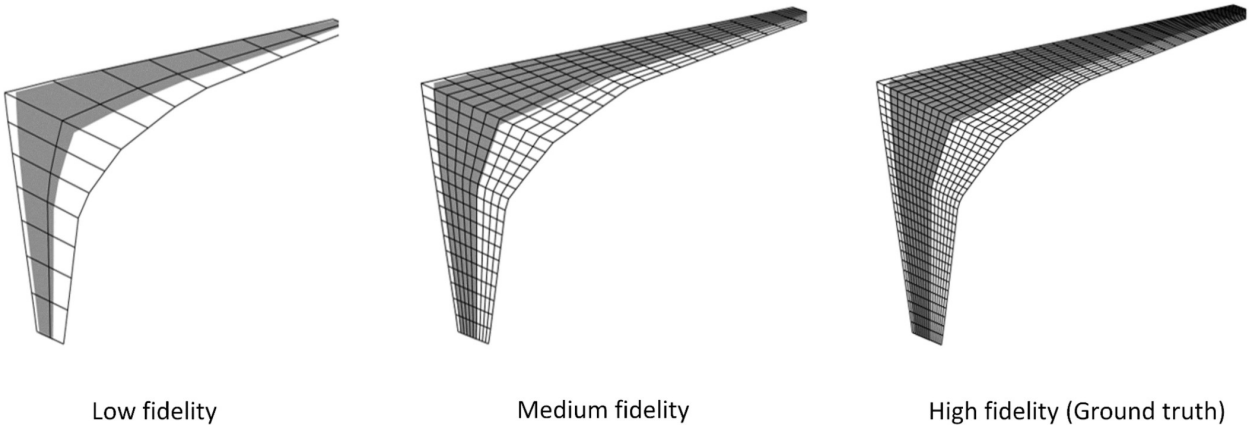

Figure 3.6: Illustration of the meshed wing with different fidelity models. The number of meshes in each model is presented in Table 3.1.

Table 3.1: Mesh sizes and costs for different fidelity models

Fidelity level	Num_y	Num_x	Cost (seconds)
Low	15	3	1.9
Medium	35	11	45.1
High (Ground truth)	55	19	283.9

objective OpenAeroStruct problem with three information sources taking the highest fidelity one as the ground truth. As mentioned earlier, our objective here is to minimize both fuel burn and wing mass by controlling 17 design variables. We assumed a budget of 2000 seconds of computational time on lower fidelity information sources for this demonstration. The results are shown in Fig. 3.7, where random points are shown in blue to show the objective space (these are not part of the algorithm and are for visualization only), the green points are those points selected by our approach with the fused GP, and the red points are the non-dominated green points that have been evaluated with the ground truth (that is, the final step of our algorithm). The figure reveals that our approach has done well in identifying the non-dominated region in the objective space for this 17-dimensional problem. Note that the Pareto front for this problem is not spread along a large region of the objective space. This is expected based on the shape of the objective space as shown

by the randomly queried point.

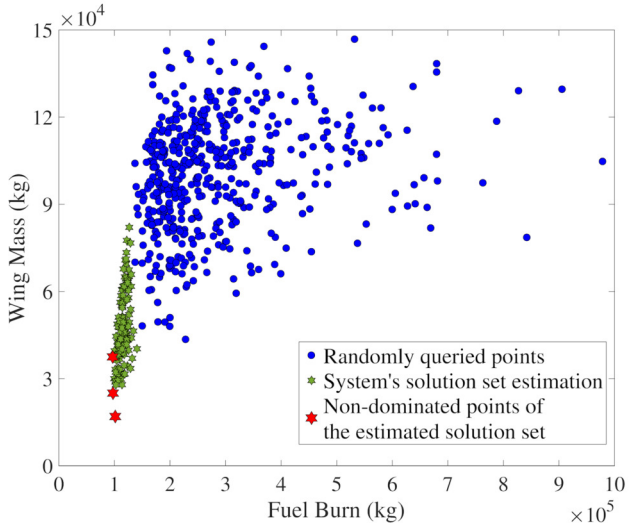

Figure 3.7: Final estimation of the Pareto front from the fused model non-dominated designs shown in red and green stars respectively.

In Fig. 3.8, the hypervolume is updated during the optimization process each time 100 seconds of computational time is spent on evaluating the lower fidelity information sources. This choice results in regular updating of the discrepancy terms. A careful study of the optimal allocation to lower fidelity information sources and ground truth estimates is a topic of future work. We note here that this is not always a clear resource tradeoff since often computation is measured in runtime and physical experiments may need to be measured in monetary units as well as time. The objectives are normalized using the upper limit known for each objective. Since both objectives have large values, differences in hypervolumes might not be sensible and normalization helps to see the changes clearly. The reference point should be dominated by all points in the objective space and here it is fixed as $(1.1, 1.1)$. The most significant changes in the hypervolume quantity is made when spending the first 10% of the budget (as was seen in the previous test case as well). Beyond that, there are improvements in the hypervolume but the returns are diminishing as expected.

Figure 3.9 reveals the cumulative sum of queries from any information source plotted against

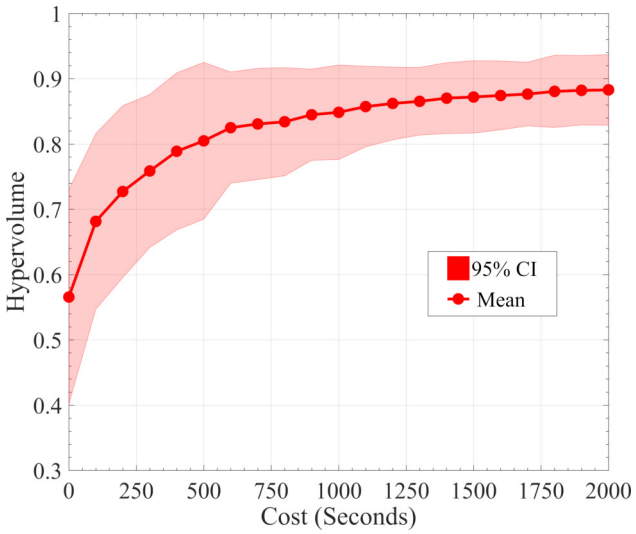

Figure 3.8: Estimated hypervolume with respect to the cost averaged over 30 independent simulations with different starting points.

overall iteration, where iteration is defined as it was previously. As depicted in the figure, initially, the low fidelity information source has been queried much more than the medium fidelity information source. After some number of iterations (about 120 in this case), the value of the low fidelity source has diminished enough that some queries to the higher fidelity source are now necessary. This continues until the budget is exhausted. We see in this case that the low fidelity source continues to be queried as well. This is due to the fact that as higher fidelity information is obtained, the correlations between the low and higher fidelity sources are updated, which results in a possible renewed value in lower fidelity information. This was the case here.

For the above analysis, a possible question is what if there was only one information source available to estimate the quantities of interest in this design problem. To address this question, another experiment is designed to compare the results between the multi-fidelity approach and the single fidelity approach. We have considered the single fidelity optimization task using the medium fidelity information source. The Gaussian process built for the information source is taken as the predictor model and there is no fusion of information and a fused model in the single fidelity optimization case. Fig. 3.10 shows the hypervolume averaged over 30 simulations for each case in a normalized objective space with different starting points. The multi-fidelity approach outperforms

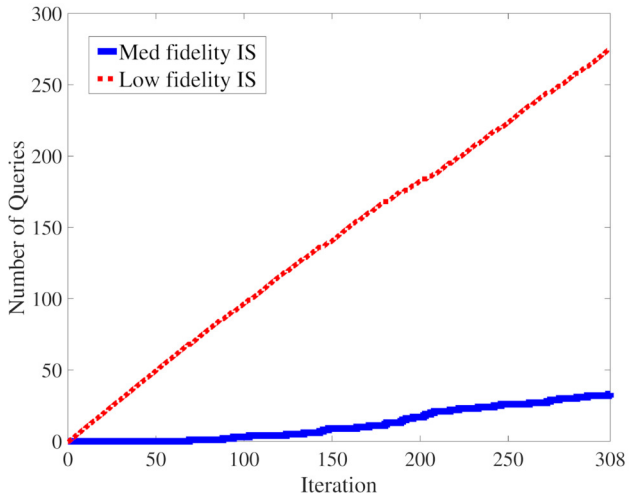

Figure 3.9: The cumulative sum of queries from each information source is shown. We query one information source at each iteration.

the single fidelity approach since it has access to more information about the ground truth. This result shows the low fidelity information source contribution to provide useful information about the quantity of interest in the multi-fidelity configuration.

The next step is to compare the effectiveness of different algorithms proposed to do multi-objective optimization task. Here, we have compared the results of NSGA-II[113, 114, 115], ParEGO[116, 117], and EHVI methods on optimizing the OpenAeroStruct design problem. We will not present the whole algorithms here but interested readers can find details regarding these approaches and implementations in the references above.

In general, the ParEGO algorithm is an extended version of the efficient global optimization (EGO) algorithm initially introduced in [98] for global optimization of single objective expensive black-box functions. The EGO algorithm is a surrogate based method and searches for new solutions using the expected improvement criterion. At each iteration, a set of random solutions are generated in a Latin hypercube or any other space filling design and the solution which maximizes the expected improvement will be queried from the expensive function to update the surrogate model. To extend the method for optimizing multi-objective functions, one approach is to combine all objectives into a single objective using parameterized scalarizing weight vectors [117]. In

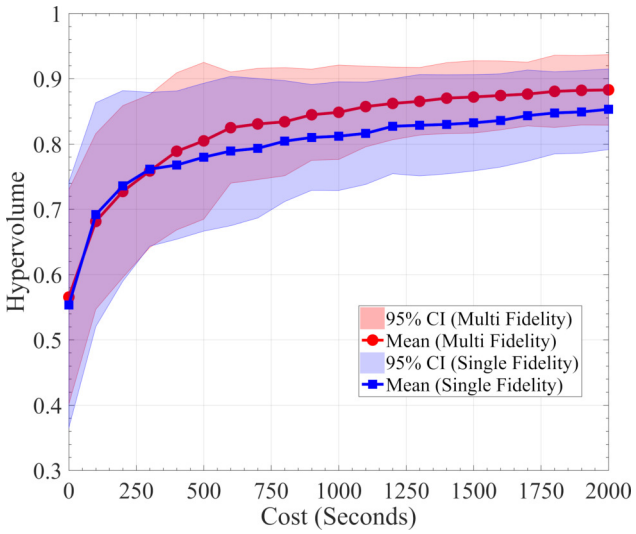

Figure 3.10: Comparing the estimated hypervolume in single fidelity and multi-fidelity approaches averaged over 30 independent replications of simulations.

NSGA-II, a non-dominated sorting is done over the available data or in terms of genetic point of view, population, and they are given a rank according to their non-domination level. New population are generated according to their front rank trying to find new non-dominated solutions. Readers are referred to [113, 114] for more details.

NSGA-II and ParEGO are not set up to take advantage of multiple information sources and they are employed to optimize one function. However, as these algorithms might need to query a large number of points from the function directly, optimizing the ground truth does not make sense with respect to the cost of each query. The goal is to see how they perform if the same amount of resources are available for all methods. Consequently, the optimization is done over the medium fidelity information source to allow reasonable number of queries from the function.

The hypervolume estimations in Fig. 3.11 show the improvement achieved by our EHVI-based approach. Note, the starting point for every simulation is different, thus, we have included the uncertainty region even for the initial hypervolume. The EHVI approach has the advantage to come up with a good estimation of the optimal Pareto front meaning a larger hypervolume much faster than the other approaches. Therefore, in highly budget-constrained experiments, it finds solutions closer to the optimal Pareto front. Although it is seen that the estimated hypervolumes

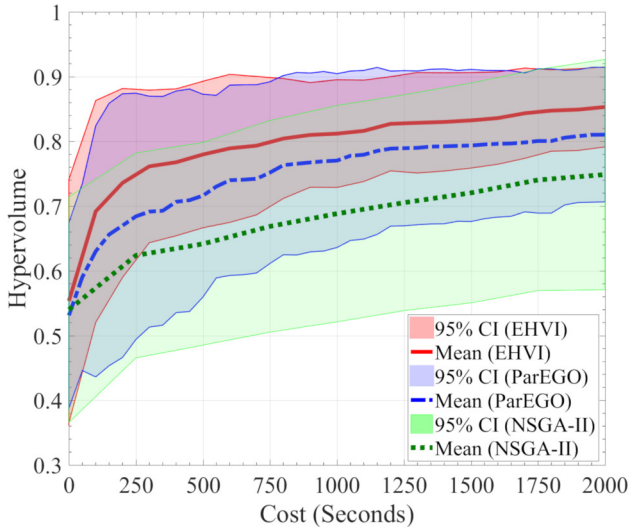

Figure 3.11: Our proposed approach using EHVI has outperformed other methods obtaining larger hypervolume. Values show the averaged hypervolume over 30 independent simulations.

might converge when more resources are available, since they have enough budget to search the space, the EHVI is still suggesting better solutions to the problem. Also, the NSGA-II approach is building up the Pareto front gradually and will likely reach the EHVI and ParEGO estimations of the hypervolume at higher costs.

In Fig. 3.12, we demonstrate the results of using our multi-information source approach versus a single information source approach using EHVI as the acquisition function, as well as the ParEGO and NSGA-II methods. Here, a representative result from the 30 simulations is used to show the results of the different algorithms. We see from the figure the multi-information source approach dominates the other approaches. In some cases, it is possible that the ParEGO and NSGA-II provide a better solution (say in a different choice among the 30 simulations), however, as shown in Fig. 3.11 our method is performing better on average.

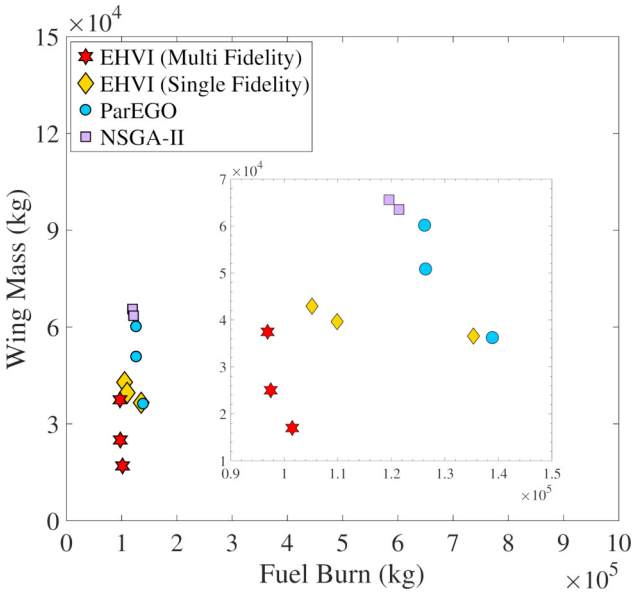

Figure 3.12: Pareto fronts obtained using different methods.

4. INTEGRATION OF ACTIVE SUBSPACE METHOD IN BAYESIAN OPTIMIZATION OF MULTI-INFORMATION SOURCE SYSTEMS*

4.1 Overview

Materials design calls for an optimal exploration and exploitation of the process-structure-property (PSP) relationships to produce materials with targeted properties. In chapter 2, we developed and deployed a closed-loop multi-information source fusion (multi-fidelity) Bayesian Optimization (BO) framework to optimize the mechanical performance of a dual-phase material by adjusting the material composition and processing parameters. While promising, BO frameworks tend to underperform as the dimensionality of the problem increases. Herein, we employ an adaptive active subspace method to efficiently handle the large dimensionality of the design space of a typical PSP-based material design problem within our multi-fidelity BO framework. Our adaptive active subspace method significantly accelerates the design process by prioritizing searches in the important regions of the high-dimensional design space. A detailed discussion of the various components and demonstration of three approaches to implementing the adaptive active subspace method within the multi-fidelity BO framework is presented.

4.2 Introduction

Integrated Computational Materials Engineering (ICME) – based material design [52] relies on solving the inverse problem connecting target properties/performance metrics to material chemistry and processing. This connection is established through (forward) process-structure-property (PSP) relationships [55, 119], which are in turn established through physics- or machine learning-based models [53, 54, 120] and/or experimental data. The solution to this inverse problem entails the exploration and exploitation of PSP relationships to identify the required chemistry-processing combinations that yield desired properties [56]. Practical implementation of ICME frameworks re-

*Reprinted with permission from "Adaptive Active Subspace-based Efficient Multifidelity Materials Design" by D. Khatamsaz, A. Molkeri, R. Couperthwaite, J. James, R. Arróyave, A. Srivastava, and D. Allaire., 2021. Materials and Design, Copyright 2021 Published by Elsevier Ltd.[118]

quires addressing *three major challenges*: the need to exlicitly connect the different models along the PSP chain; the considerable cost associated with the evaluation of each of the models/linkages; and the potentially large dimensionality of the design space.

A significant amount of work has been done to address the *first challenge*, at least in the context of microstructure sensitive materials design, which aims to uncover optimal microstructures that meet specific performance objectives by focusing exclusively on the microstructure-property space [121, 122, 123, 124, 36, 125]. While promising, this approach assumes that the design space consists of a universe of microstructures that are all feasible, potentially, through suitable chemistry-processing combinations. This is an unwarranted assumption as there is no guarantee that an optimal microstructure is feasible, in the sense that it can be attained through an adequate processing protocol. To date, there has been some measure of success in the deployment of fully integrated PSP model chains for materials design[60, 37]. However, this is not a trivial task largely due to the complex, highly coupled, multi-scale nature of the linkages along the PSP chain [58].

To address the *second challenge* associated with the considerable cost of querying the PSP relationships, the materials design community has focused on the development and deployment of closed-loop Bayesian Optimization (BO) frameworks to efficiently explore and exploit the material design space [61, 62, 6, 64, 126, 37]. These frameworks seek a balance between exploration and exploitation in order to efficiently arrive at optimal materials solutions. These approaches are suitable and have been used successfully in both simulation-driven and experiment-centered materials design problems.

The *third challenge*, however, has largely remained unaddressed. This is despite the fact that, more often than not, the design space is large, [36, 37] and BO frameworks tend to underperform as the dimensionality of the problem increases [38]. It is certainly possible to carry out statistical tests to determine the most influential design variables in any optimization task and then to focus exclusively on those degrees of freedom during the design process. However, this requires sufficient data connecting design inputs to design outputs. Thus, there is a need for techniques that can effectively locate the most important (and/or informative) design regions to increase the efficiency

of the materials design process. These efficiency gains can be more pronounced if these potentially productive design regions can be identified with limited data in an adaptive manner, while more information about the design space is gained in the course of exploring/exploiting it [127, 128].

There are techniques to deal with the curse of dimensionality that stems from large design spaces, usually by defining a representative response surface in a lower-dimensional space while maintaining the relationship between the design variables as much as possible. For instance, global sensitivity analysis is used to measure the importance of different design variables in the variation of a quantity of interest [39, 40, 41, 42, 43]. This approach assigns a nominal value to the design variables that have little effect on the objective and perform the optimization over the remaining design variables. Approximating a subspace of the original large design space is also among the most common approaches to dimensionality reduction and can be used to represent data in a lower-dimensional space to ease machine learning objectives, increase the efficiency of optimization tasks [129, 130], aid in model reduction [131], or facilitate optimal control of dynamic systems [132]. Another technique in dimensionality reduction is the Principal Component Analysis (PCA) [44], which linearly projects a large dimensional dataset onto a lower-dimensional space [45, 46, 47] while trying to keep as much information as possible by determining the principal components that capture a majority of the variance in the data. PCA has been employed for microstructure sensitive design to build models to predict (mechanical) properties using a lower-dimensional representation of the complex material microstructure [133, 134, 135, 136, 137, 138, 139, 140]. Note that PCA simply decreases the dimensionality of the design space by considering the correlations among design variables, without accounting for the connection between the design variables and the quantities of interest amenable to optimization. In a PSP-based materials design problem, where the focus is on optimizing the performance metrics by exploring material chemistry and processing options, directly decreasing the dimensionality of the design space without accounting for the design objectives may not be feasible.

Herein, we employ an adaptive active subspace method [48, 49, 50] to efficiently handle the large dimensionality of the design space of a typical PSP-based materials design problem within

our recently developed closed-loop multi-information source fusion (multi-fidelity) BO framework [37]. Specifically, we demonstrate the efficacy of this framework by optimizing the stress, τ, normalized strain hardening rate, $d\tau/d\epsilon_{pl}$, at an arbitrary equivalent plastic strain, $\epsilon_{pl} = 0.9\%$, of a dual-phase material (ferrite-martensite steel) by adjusting the content of the alloying elements C, Mn and Si in the Fe-based alloy, and the processing condition, i.e., the intercritical annealing temperature, T. The normalized strain hardening rate, $(1/\tau)\,(d\tau/d\epsilon_{pl})$, is a useful mechanical performance metric, and a higher value of this parameter indicates better ductility and formability of the material. In the design framework, we utilize the thermodynamic results to predict the chemistry and composition of the constituent phases after the single-stage heat-treatment (intercritical annealing followed by quenching) [126, 37]. This information is then used to predict the mechanical performance of the dual-phase material using a variety of (reduced-order) micromechanical models referred to as Isotress, Isostrain, Isowork, Secant method and Elastic constraint, and a high through-put microstructure-based finite element model that utilizes a three-dimensional representative volume element (RVE) of the material microstructure [64, 37]. All these models, low fidelity micromechanical models, as well as high fidelity microstructure-based finite element models (referred to as RVE and assumed to be the 'ground truth') are treated as information sources. We represent the response of each information source as Gaussian process surrogates and fuse them using standard approaches for the fusion of normally distributed data.

Our approach to implementing the adaptive active subspace method within the multi-fidelity BO framework is schematically shown in Fig. 4.1. The active subspace method is a technique to look for the directions in the design space for which a function has the largest variability. By forming a subspace using those directions, an approximation of the function is obtained on a lower-dimensional space referred to as the *active subspace* [49]. Thus, increasing the efficiency of the design process by more effectively searching for the optimal solution within the high-dimensional design space [50, 48]. In the context of materials design, the function is the PSP relationship that is being evaluated, and the basic idea is to find the directions in the design space (T, C, Mn, and Si) that give the largest variation in the objective value (normalized strain hardening rate) by

using the available data at every stage of the optimization task. The directions suggesting larger variation than a user-specified value then form the active subspace. Following this, we employ a knowledge gradient acquisition function to determine the 'next best point' to evaluate within the active subspace. In order to obtain the true input values for the 'next best point,' the chosen point in the active subspace must be inversely mapped to the true design space. Since there is no unique solution for this inverse mapping problem, a second BO step is performed to determine the 'next best point' in the true design space. At this stage, a decision about which information source (low fidelity micromechanical model) to query is also made by temporarily updating each information source and comparing their results. Finally, the PSP relationship is evaluated at this 'best point' using the thermodynamic-based model and the selected micromechanical model to estimate the objective value.

4.3 Methods

Here, our objective is to maximize the the stress, τ, normalized strain hardening rate, $d\tau/d\epsilon_{pl}$, i.e. $(1/\tau)\,(d\tau/d\epsilon_{pl})$ at an arbitrary equivalent plastic strain, $\epsilon_{pl} = 0.9\%$, of a dual-phase material. The dual-phase material system considered is a ferritic-martensitic steel which is produced by subjecting the material system composed of Fe, C, Mn and Si to a single-stage intercritical annealing heat treatment followed by rapid quenching. Therefore, our optimization problem aims to find the values of the intercritical annealing temperature, T, and C, Mn and Si content of the Fe-based alloy that correspond to the maximum value of the $(1/\tau)\,(d\tau/d\epsilon_{pl})$.

In chapter 2, we addressed the optimization problem discussed above with a multifidelity BO framework to incorporate the response of different mechanical models, which enabled the collection of information about the optimum design in a less costly manner in comparison to employing a finite element model alone. The multifidelity aspect of the approach was used to exploit the fact that in most materials design problems we have available several different models that can potentially be used to estimate a quantity of interest. These models are usually based on different physics-based and numerical assumptions, which leads to models with varying expense in terms of computational resources required for a query to the model and varying fidelity. The exploitation

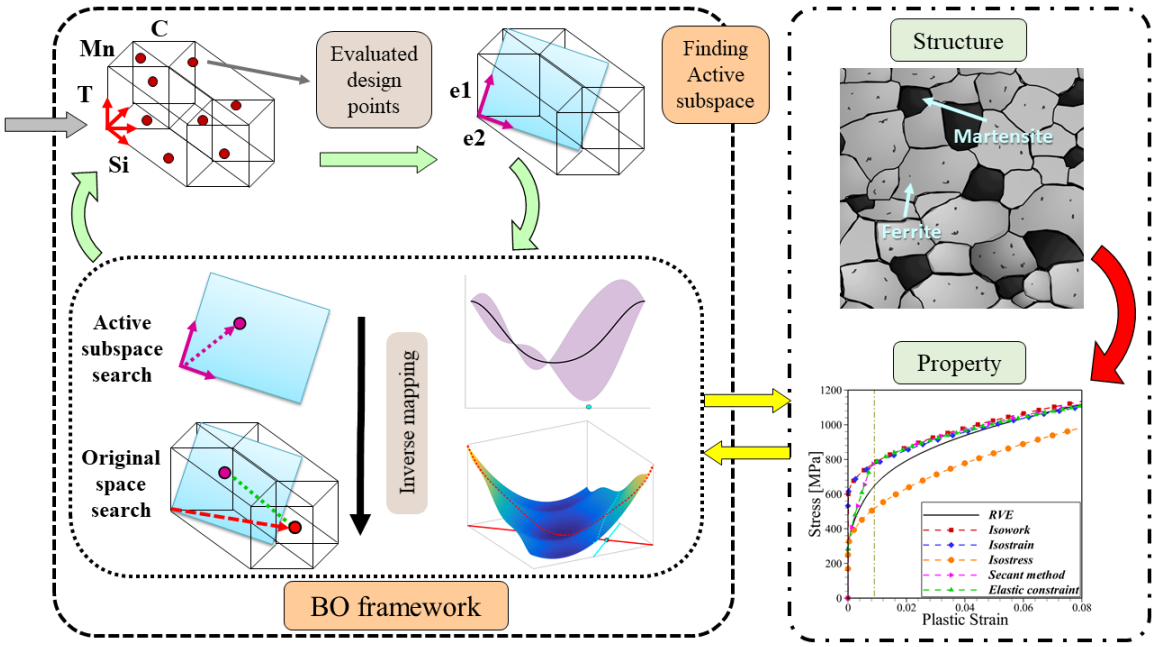

Figure 4.1: Implementation of adaptive active subspace method within a multifidelity Bayesian Optimization (BO) framework. The basic idea is to find the active subspace, i.e., the directions in the material design space – intercritical annealing temperature (T), and alloying elements C, Mn, and Si – that give the largest variation in the mechanical property (normalized strain hardening rate) by using the available data at every stage of the optimization task. Next, the process - structure-property (PSP) relationship is mapped to the active subspace, and the first step of the BO framework is applied to find the 'next best point' to evaluate within the active subspace. The 'next best point' is then mapped back to the original design space by implementing a second BO step. Finally, the PSP relationship is evaluated at this best point using the thermodynamic-based model and the selected micromechanical model to estimate the objective value. This new data is added to the framework for the next iteration.

of each possible model, or information source, was achieved via an information fusion process described in Refs. [8, 5]. While the overall multifidelity BO approach was shown to be more efficient than traditional BO approaches, the process can still be computationally impractical when applied over large design spaces. Here, to address this challenge, we consider the application of adaptive dimensionality reduction in the context of our multifidelity BO framework using the active subspace method, which is described in detail below in Section 4.3.1.

Figure 5.3 illustrates the steps in our proposed adaptive active subspace-based multifidelity BO framework. The framework starts by reducing the dimensionality of the design space and

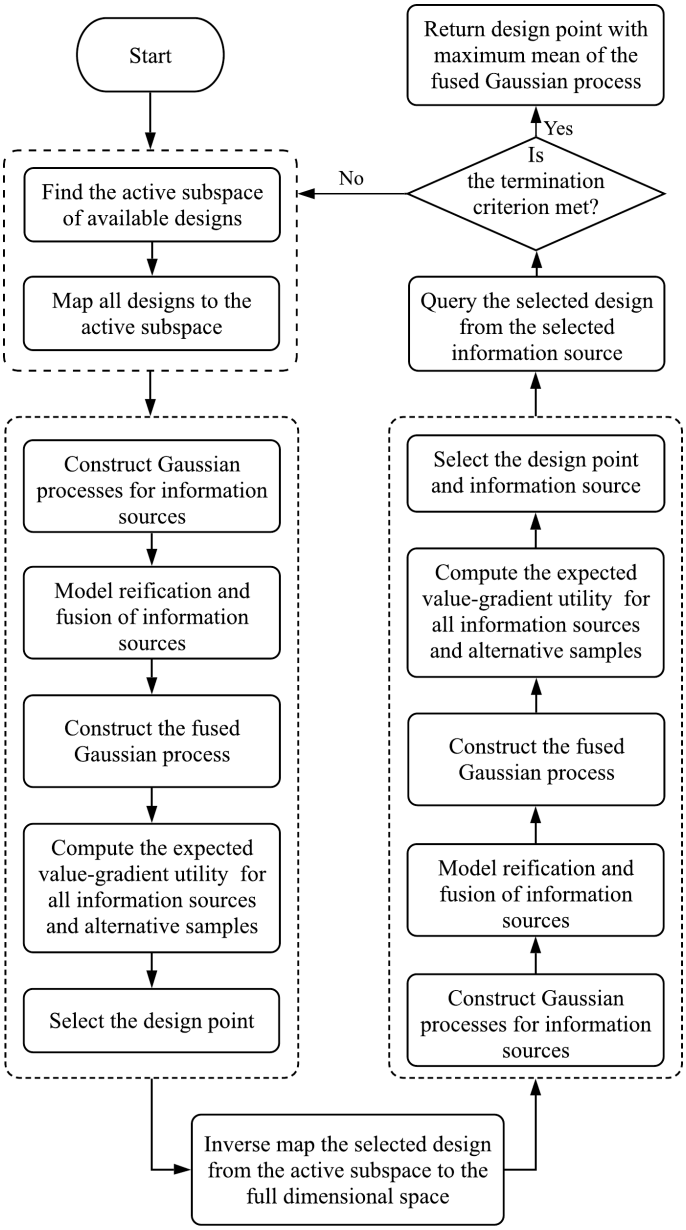

Figure 4.2: Flowchart of the proposed approach. At the start of every iteration, the active subspace is found and all data are projected onto it. Then, the first step of BO is applied over this active subspace. The best design candidate is mapped back to the full dimensional design space, resulting in a solution subspace. The second step of BO then is applied over this subspace to select the best design and information source to query.

projecting all evaluated designs to a lower dimensional design space (the current active subspace), and then two steps of Bayesian optimization are executed. The first step of BO is applied over the active subspace to find a best design candidate to query in the lower dimensional space. Next, the

obtained design is inverse mapped to the full dimensional design space, which results in a subspace of potential solutions of the inverse problem. The second step of BO is applied over this subspace to obtain the next design and information source to query.

In every step of the Bayesian optimization process, we use surrogate models to estimate the expected objective values of design points that have yet to be evaluated. In a multifidelity setting, we therefore have multiple surrogate models to construct (one for each information source) and for use in predicting design points not yet queried. Since every information source contains useful data regarding the expensive objective function to be optimized, we also employ a fusion technique, known as model reification [8, 5, 141] to fuse data from all information sources to obtain a fused predictive model to estimate the expensive objective function. This fused model encompasses our current state of knowledge during the design process. An update to any of the information sources results in a fused model update representing the system's new state of knowledge. During every iteration, we generate a set of potential design points using a space filling technique, for example, Latin hypercube sampling, evaluate them from each information source's surrogate model and temporarily update the fused model. We then use an acquisition function to quantify the expected change in the system's knowledge about the maximum objective value when evaluating that design point from the information source. The next design point to be evaluated is then selected by choosing the information source and design point that resulted in the largest expected change in the system's knowledge of the maximum objective value.

In the following subsections, we describe the active subspace method in detail and then different strategies and implementation options of the framework, which are then used in the presentation of results in Section 4.4. Other concepts and ingredients of the framework such as microstructure space prediction, mechanical response prediction, Gaussian process regression, information fusion, and knowledge gradient are well discussed in chapter 2.

4.3.1 Active Subspace

The active subspace method is a technique to look for the directions in the design space for which a function has the largest variability. By forming a subspace using those directions, an ap-

proximation of a function is obtained on a lower dimensional space called the active subspace. The advantage of constructing a subspace to approximate a function is that learning a subspace of the original high-dimensional design space is easier [50, 48]. This advantage leads to significant efficiency gains, speeding up the optimization in design applications and reducing resource usage [142]. Briefly, the idea is to find the directions in the design space which contain the largest variation in the objective value. In other words, a new coordinate system is built based on eigen vectors of the space with eigenvalues defining how strong the variation of the objective value is when moving toward that direction. The matrix $\mathbf{U}$ has n eigen vectors corresponding to the first n largest eigen vectors and is called the transformation matrix. Other eigen vectors are stored in matrix $\mathbf{V}$ which defines an orthogonal space to the active subspace. Any design point in the original design space can be transformed to the active subspace using the transformation matrix:

$$\mathbf{z} = \mathbf{U}^{\mathrm{T}}\mathbf{x} \tag{4.1}$$

The function g represents the original function f in the active subspace as

$$g(\mathbf{z}) = g(\mathbf{U}^{\mathrm{T}}\mathbf{x}) \approx f(\mathbf{x}) \tag{4.2}$$

Now, we seek to learn the objective function g in the active subspace instead of the original objective function f on the design space $\mathcal{X}$. A detailed discussion on how to compute the active subspace associated with an objective function is presented in Ref. [50]

After performing the Bayesian optimization over the active subspace and once a candidate point, $\mathbf{z}^*$ which is the projection of $\mathbf{x}^*$ in the high-dimensional space, is selected, it needs to be mapped back to the original design space. This allows the second step of the optimization to identify the best point and information source to query. The challenge here is that there are an infinite number of high dimensional vectors that have the same projection to the point in the lower dimensional space. Therefore, we propose a method to overcome this problem.

Using the definition of orthogonality of eigen vectors of a symmetric matrix, which is the

covariance matrix calculated to obtain the eigen vectors, any eigen vector in matrix V is orthogonal to any eigen vector in the transformation matrix U or in general, the active subspace. Consequently, any linear combination of eigen vectors in matrix V is orthogonal to the active subspace. Thus, by writing the equation of vectors created by the linear combination of orthogonal eigen vectors to the active subspace which pass from the design point selected in the first step of optimization, $\mathbf{x}^*$, we are able to generate an infinite number of design points in the higher dimensional space with the same projection to the active subspace.

Assuming the original design space has m dimensions and the active subspace has n dimensions, the matrix V will have $m - n$ eigen vectors. An orthogonal vector to the active subspace is given as

$$\overrightarrow{\mathrm{P}} = \sum_{k=1}^{m-n} a_k \overrightarrow{\mathrm{e}}_k, \tag{4.3}$$

where a_k is a random number that for simplicity, is generated from 0 to 1 and $\overrightarrow{\mathrm{e}}_k$ is an eigen vector in V. Then using the orthogonal vector $\overrightarrow{\mathrm{P}}$ and $\mathbf{x}^*$, the corresponding design point in the high-dimensional space to $\mathbf{z}^*$, the equation of the linear subspace passing through $\mathbf{x}^*$ and orthogonal to the active subspace is given by

$$\frac{x(1) - x^*(1)}{\mathrm{P}(1)} = \frac{x(2) - x^*(2)}{\mathrm{P}(2)} = \ldots = \frac{x(m) - x^*(m)}{\mathrm{P}(m)} = t \tag{4.4}$$

where a design vector in m-dimensional design space, $\mathbf{x} = [x(1), x(2), \ldots, x(m)]^{\mathrm{T}}$, is found by solving m sub-equations for a given t. All design vectors obtained in this step satisfy the relation $\mathbf{z}^* = \mathrm{U}^{\mathrm{T}}\mathbf{x}$. Although all generated design vectors satisfy the relations mathematically, the constraint here is to have all m design variables in the bounds defined by the designer. Therefore, before generating random t values, its range of variability should be specified. First, by replacing the lower bound for every design variable in sub-equations in equation (4.4), m different values are obtained for t

$$t_{low,i} = \frac{\mathrm{lb}(i) - x^*(i)}{\mathrm{P}(i)}, \quad 1 \leq i \leq m \tag{4.5}$$

and the same calculations are done for the design variable upper bounds

$$t_{up,i} = \frac{\text{ub}(i) - x^*(i)}{\text{P}(i)}, \quad 1 \leq i \leq m \tag{4.6}$$

Next, to have all design variables within the bounds, the lower bound for t from $2m$ values found is the closest value to zero between all negative t values and the upper bound is the closest value to zero between all positive t values. This way, it is guaranteed that all the design variables will remain in their bounds when being mapped back to the m-dimensional design space for any random t generated. Note that since a linear subspace can be expanded in any direction starting from $\mathbf{x}^*$, it is ensured that t can take both negative and positive values. Finally, a set of samples in the original design space is generated, $\mathcal{X}_f$, by mapping back the design point $\mathbf{z}^*$. The acquisition function is then employed for the second time to find the best design to be evaluated next.

4.3.2 Strategies and Implementation

There are different strategies available for combining the concept of Bayesian optimization of multifidelity systems and the active subspace method. We consider three such strategies here.

The first approach is to build the active subspace upon the ground truth response surface (GT active subspace). The intention behind this decision is to focus on searching the subspace directly related to the design space of the highest fidelity model. At the beginning of every iteration, the ground truth active subspace is formed and all data from other information sources are projected to this subspace. This results in new models defined on a lower dimensional design space. The Bayesian optimization framework is then exploited to search this lower dimensional design space looking for the best potential design to be evaluated to provide the most information about the optimum design.

The point that maximizes the acquisition function value is then selected as the next-best point to evaluate. This point is then mapped back to the original high-dimensional space. Since there are an infinite number of possible solutions when mapping from a low-dimensional to high-dimensional design space, the Bayesian optimization approach is repeated for this solution set. From this

optimization approach, the next-best design point and an information source are chosen to be queried.

The second approach is to transform all information sources to the active subspace of the temporary updated information source (TUIS active subspace). In this context, instead of transforming to the active subspace at the beginning of every iteration, transformations are done every time an information source is temporarily updated. Therefore, different active subspaces corresponding to each information source are taken into account and the system might find an information source suggesting larger variation in the objective value and by extension the fused model. Thus, it is likely that a more informative point may be identified. By transforming all active subspaces associated with each information source together in turn, the performance of other information sources in different active subspaces is investigated as well. This results in investigating all information sources and fused models in different active subspaces, resulting in a more informative decision making process.

The last approach is to let every information source operate in its own active subspace independently (Independent active subspaces). In this case, when a design point is to be evaluated, it should be transformed to the corresponding active subspace first. Again, all the processes explained earlier remain the same. This approach offers a cheaper framework in comparison to the second approach since the number of transformations and modeling time will be decreased.

4.4 Results and Discussions

We implement an adaptive active subspace method to efficiently handle the large dimensionality of the design space of a typical PSP-based material design problem within our recently developed closed-loop multifidelity BO framework. Here, we demonstrate the obtained results from utilizing three approaches discussed earlier to implement the adaptive active subspace method within the multifidelity BO framework. For comparison purposes, the multifidelity BO framework developed in chapter 2 that does not take advantage of the active subspace method (referred here as the standard approach or Std) is also considered.

Fig. 4.3 shows how quickly each method attained progressively higher objective values as a

function of the number of iterations. All models are initialized with 10 random points in the design space. The results are the average of 5 replicas with different initializations, and the filled region shows the 95% confidence intervals. At the end of every iteration, the system chooses a point and an information source to query. Then, after every 10 iterations, the best estimation of the optimal solution suggested by the fused model is evaluated from the ground truth model.

There are some important conclusions that can be drawn from these results. First, the active subspace approach improves the performance of the optimization significantly as each of these methods leads to faster improvement in the objective and reaches the optimal design region more quickly than traditional multi-information source BO. This is because the active subspace approaches search more effectively over lower dimensional regions in the design space by adaptively locating the regions of largest variation (i.e., the most active regions) in the objective value. Therefore, greater jumps in the objective value are observed when the active subspace method is implemented. Next, using the TUIS active subspace method results in slower improvement rates as compared with the GT and Ind approaches. This is related to the need to construct more active subspaces in this approach and force possible deleterious connections between information sources in these temporary subspaces. Employing the independent active subspaces, or Ind approach for each information source has similar performance to using the Ground Truth active subspace, or GT approach. In both approaches we see superior improvement in the objective as compared with the traditional approach and the TUIS approach. Finally, at iteration 200, all methods have arrived at the optimum design region and there is no advantage of using a particular strategy beyond this point. This is to be expected as each method has acquired enough information at this point in its respective approach to accurately approximate the ground truth objective. Overall, superior improvement rates of the active subspace approaches are associated with the ability of these methods to avoid costly exploration in less important regions of the high-dimensional design space in the early stages of the optimization process.

Although comparing the progression toward the optimum design region based on the number of function evaluations gives a sense of the effectiveness of queries made by employing the active

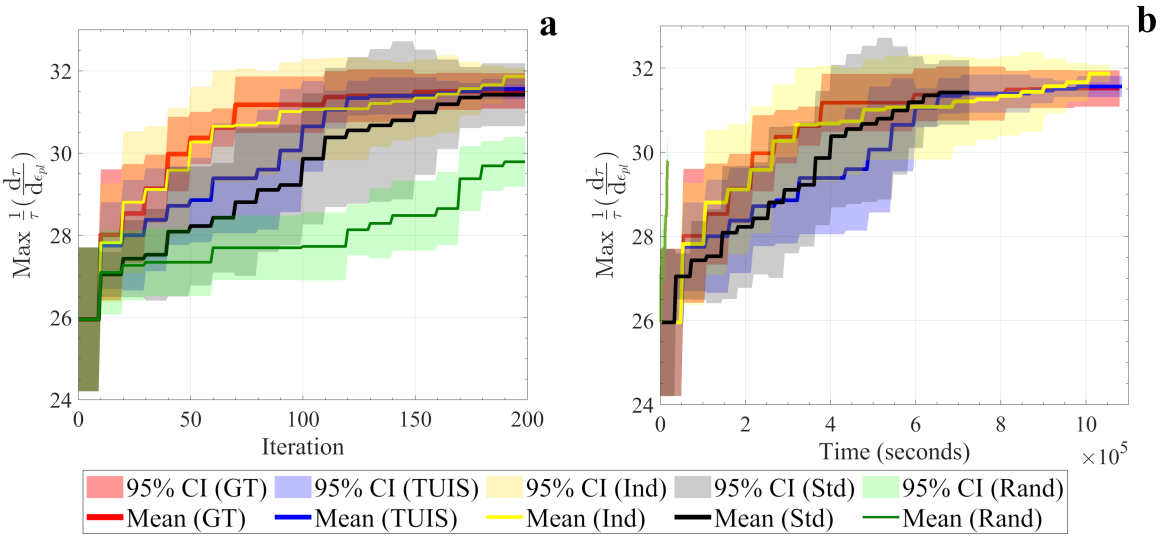

Figure 4.3: Estimated optimum objective value as a function of (a) iteration and (b) time. (a) The active subspace methods using Ground Truth (GT), Temporary Updated Information Source (TUIS), and Independent (Ind) active subspaces have been shown to outperform the Standard (Std) approach without applying the active subspace method. (b) In terms of computational cost, the active subspace approaches are again superior. The computational cost accounts for modeling, active subspace and knowledge gradient calculations in addition to the function evaluations. Using the TUIS active subspace is slightly more expensive due to the larger number of active subspace calculations and transformations required. The results are obtained from 5 different initializations and the mean and 95% confidence intervals are shown.

subspace method, a more thoughtful comparison is to compare the time required for each approach to reach a target value. In this case, we consider the cost of modeling, including updating and evaluating the Gaussian processes, calculation of the active subspaces and knowledge gradient in addition to the function evaluations done during the optimization process. In Fig. 4.3(b), we have illustrated the objective value attained and time required for each approach. These results suggests that using the active subspace method, in particular the GT active subspace or Ind active subspace approaches, results in higher objective values in less computational time in comparison to the standard multifidelity or multi-information source optimization. On the other hand, using the TUIS active subspace approach shows little to no improvement, which is related to the number of active subspace computations and transformations made in a single iteration. While these results show that the conventional multifidelity approach performs as well as the active subspace approach

after a certain amount of time, the active subspace approaches provide significant improvement in the results at early stages of the process. In addition to this, the active subspace approach also shows lower variability in the results—this last aspect is very important as low uncertainty is a desired attribute of any design framework.

We note here that the uncertainty (or variance) in the results stem from several sources. First of all, we have used different training sets for initialization of the models, so each iteration starts from different initial conditions. Second, we have modeled the information sources using a stochastic process, namely Gaussian process models. These probabilistic models predict the objective value with normally distributed uncertainty in the prediction. In addition, at every stage of the optimization, we generate random samples using Latin hypercube sampling. Therefore, for each different run, there will be different samples to evaluate as candidate design points. Finally, the ground truth function in this particular design application, RVE, is noisy and can provide different objective values for the same design input [37]. The confidence intervals in Fig. 4.3 show the total uncertainty since differentiating between each source of uncertainty was not practical.

We are also interested in knowing which are the active subspaces that are preferred at every stage of the design process and what design variables are contributing most in the active subspace formation. The different algorithms used in the current work lead to diverse active subspace configurations and subsequently selections of information sources. In Fig. 4.4, the cumulative sum of the times every active subspace is chosen has been plotted. The labels show the primary design variables that the active subspace is composed of. As such, the labels indicate the dimensionality of an active subspace and which design variables will be searched preferentially. In other words, all design variables participate and are searched in a particular active subspace, but to different degrees. In Fig. 4.4 only the main participants of each active subspace are shown and are included in the labels. Note that, the ground truth is queried every 10^{th} iteration and the corresponding active subspace will be updated accordingly. Figure 4.4(a) shows that when using the GT active subspace, the system initially starts searching the carbon space, then it searches the magnesium space, and finally, the temperature space.

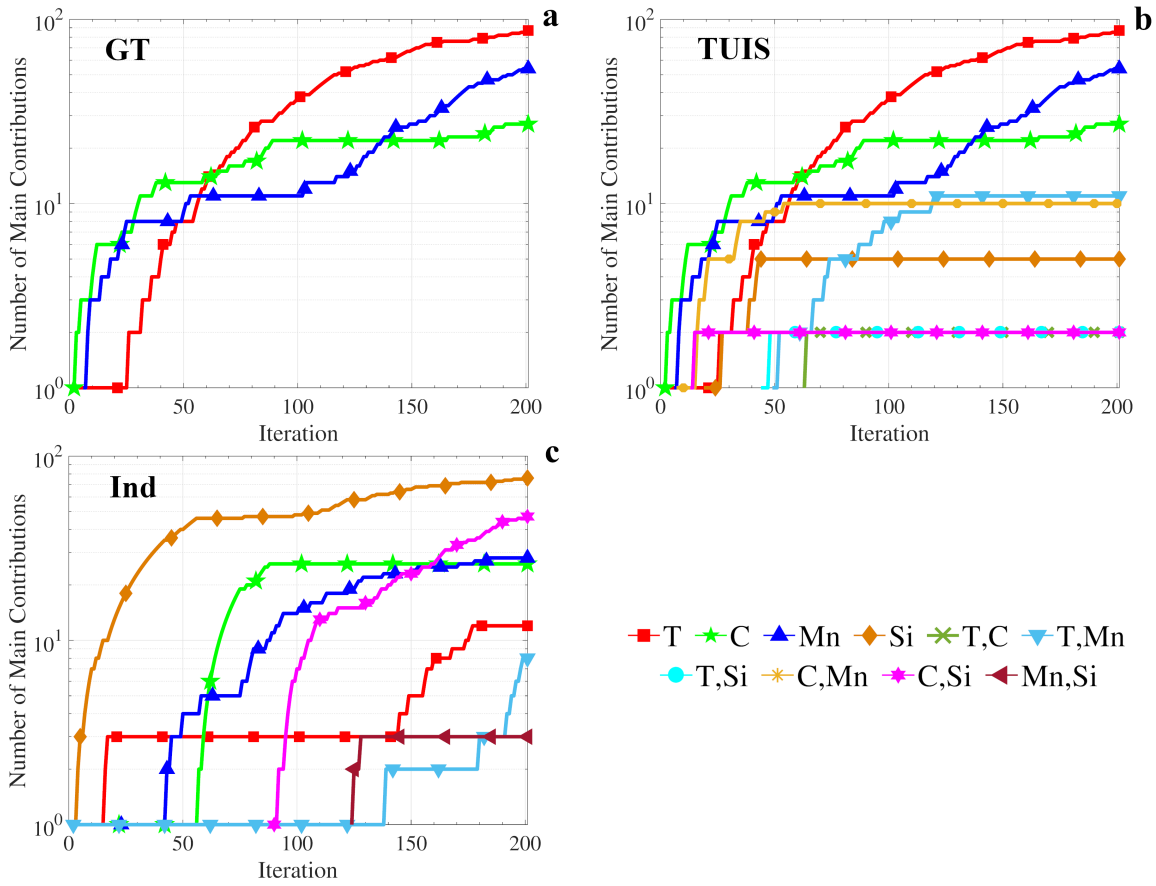

Figure 4.4: Number of times a design variable has contributed to form the active subspace. (a) using Ground Truth (GT) active subspace (b) using Temporary Updated Information Source (TUIS) active subspace (c) using Independent (Ind) active subspaces. While Si is not showing any contribution in forming the ground truth active subspace, in the other cases, a variety of design variable combinations are participating to build the active subspace.

The same results when using the TUIS and Independent active subspace approaches are shown in Fig. 4.4(b) and 4.4(c). The point here is that the one-dimensional active subspaces are preferred over the higher dimensional active subspaces. This shows that the system finds more value in searching active subspaces mainly composed of one important design variable at a time, and, once the information from the single design variables is exhausted, the system starts searching subspaces with main contributions from a combination of design variables. However, these active subspaces are still smaller than the original design space.

While Fig. 4.4 provides useful information about important design variables at every iteration

in different case studies, it is beneficial to look at how much other variables are participating in the active subspace formation. As mentioned earlier, active subspaces are built upon the most effective directions in the design space that considers the change in all design variables but in different degrees. These directions are the eigen vectors of the covariance matrix defined as

$$\mathbf{C} \approx \frac{1}{M} \sum_{i=1}^{M} \nabla_{\mathbf{x}} f(\mathbf{x}_i) \nabla_{\mathbf{x}} f(\mathbf{x}_i)^{\mathrm{T}} \tag{4.7}$$

assuming M samples are evaluated from the function f previously and the gradient is calculated numerically using a finite difference method. This is done since the function is a 'black-box' and there is no closed-form expression for the gradient. Once the eigen vectors and the associated eigenvalues are found, a single vector is formed using the linear combination of all eigen vectors, each multiplied by their eigenvalue to emphasize the strength of each variable toward a particular direction. At every iteration, we will have a 4-dimensional vector showing the effective participation of every variable based on a scalar value. Since we are not able to show a 4-dimensional space, a simple mathematical projection of the 4-dimensional space to 2-dimensional space was used. This projection allows us to graphically show the distribution of active subspaces within the design space. These results are shown in Figure 4.5.

Every point on these projection plots corresponds to a point in the 4-dimensional space. The projections allow us to show the locations of each of the pure 1D, 2D and 3D subspaces (labeled on the figure). The actual subspaces that are used in the calculations for the three approaches are plotted in relation to these. The labels of each of these subspaces is determined by the magnitude of the eigenvalues, where an eigenvalue greater than 0.5 assigns that input dimension to the label. As illustrated in Fig. 4.5, this approach to labeling allows the points to deviate quite significantly from the pure subspaces, however, the clusters are still visible. We can also quite easily observe that all approaches mostly use 1D and 2D subspaces in the calculations. This is a promising result since it shows that the active subspace approach is operating as expected. We also observe that very few unique subspaces are used in the GT-subspace approach which is likely a result of only using the active subspace of the ground truth model. Both the Individual and TUIS active subspace

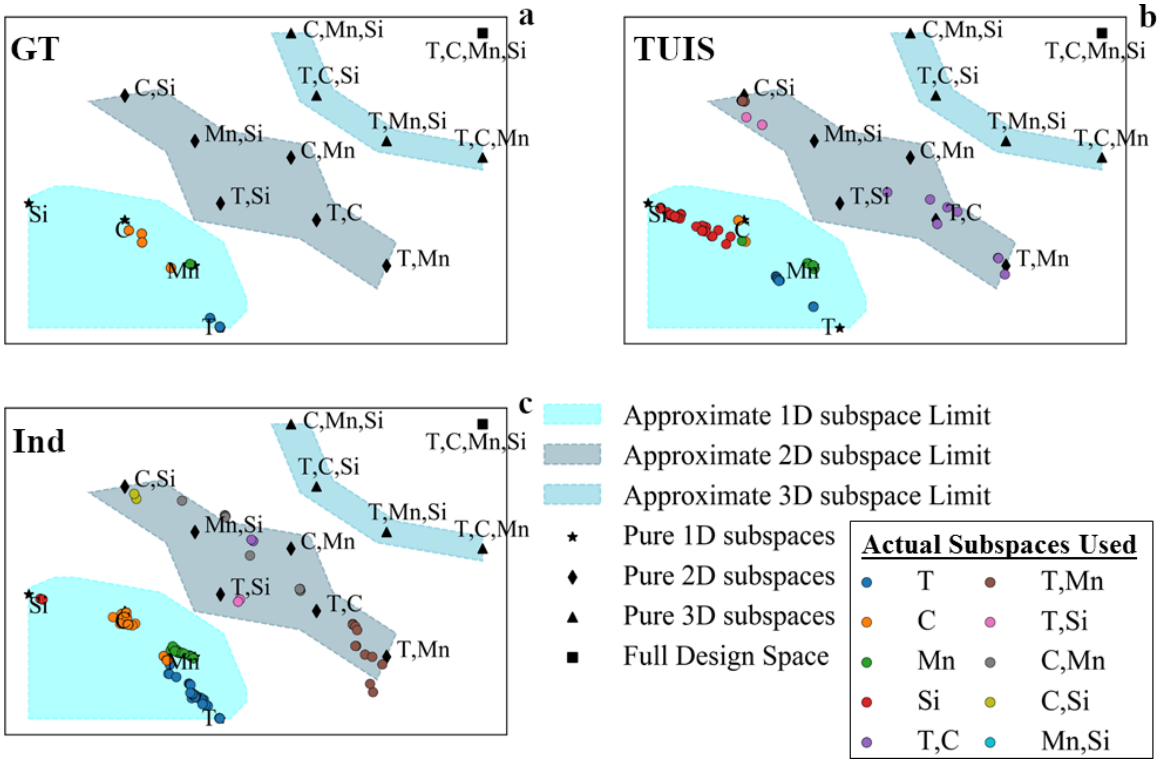

Figure 4.5: Distribution of active subspaces within the design space. A 2-D projection over the 4-D design space (a) using Ground Truth (GT) active subspace (b) using Temporary Updated Information Source (TUIS) active subspace (c) using Independent (Ind) active subspaces.

approaches show a much broader selection of active-subspaces, with many 2D subspaces also being utilized. As a final note, these differences in the active-subspaces used in the optimization does not appear to significantly affect the optimization process, as shown in the results above. The importance of this result is that the use of an active subspace approach is not dependent on the active-subspaces that are used in the optimization.

The contribution of the information sources in the optimization process and the number of queries made from each information source can show which lower fidelity models are providing more valuable knowledge about the optimum design in different case studies. Additionally, since every information source can have a different active subspace, that changes over time as more data is added to the model, the selection of the information source to query is directly affecting results in Fig. 4.5. It is thus interesting to know about the participation of the information sources as well.

Figure 4.6 shows the cumulative number of times an information source is queried based on the iterations. At every iteration, only a single information source is queried.

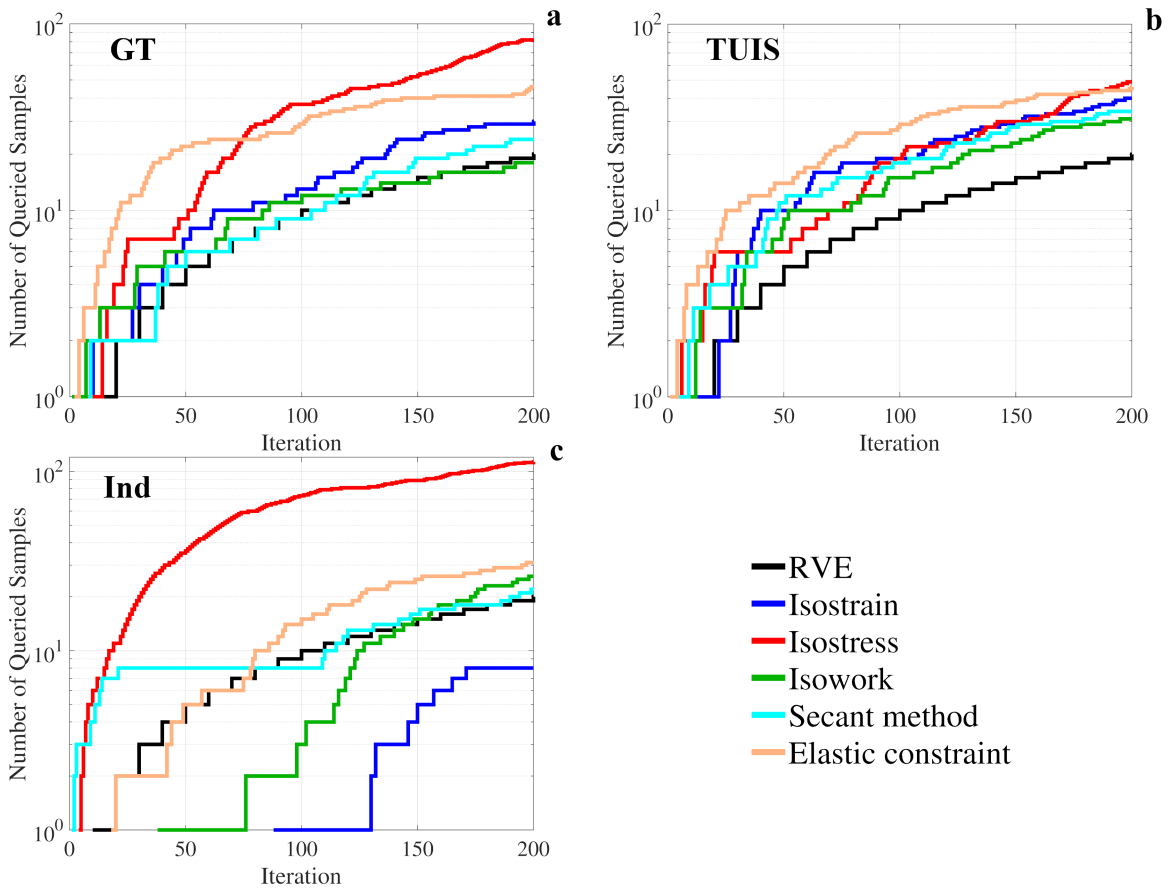

Figure 4.6: Number of queries from different information sources and the ground truth (RVE). (a) using Ground Truth (GT) active subspace (b) using Temporary Updated Information Source (TUIS) active subspace (c) using Independent (Ind) active subspaces. the isostress and elastic constraint models highly contributing to provide valuable information regarding the optimum design in all approaches.

Figure 4.6 suggests that for all cases, the isostress and elastic constraint information sources have been selected more than any other information source. It shows that these two models have a smaller discrepancy with the ground truth model around the optimum design point in comparison to the other information sources. In this chapter, the focus is on how using active subspace approach results in a more efficient and faster search for the optimal designs in the initial iterations.

Therefore, at every iteration, we let the system query the best information source providing the highest information value to the system about the optimal design regardless of how much such queries costs.

5. MULTI-OBJECTIVE MULTI-INFORMATION SOURCE CONSTRAINED BAYESIAN OPTIMIZATION*

5.1 Overview

Bayesian Optimization (BO) has emerged as a powerful framework to efficiently explore and exploit materials design spaces. To date, most BO approaches to materials design have focused on the materials discovery problem as if it were a single expensive-to-query 'black box' in which the target is to optimize a single objective (i.e., material property or performance metric). Also, such approaches tend to be constraint agnostic. Here, we present a novel multi-information BO framework capable of actively learning materials design as a multiple objectives and constraints problem. We demonstrate this framework by optimally exploring a Refractory Multi-Principal-Element Alloy (MPEA) space, here specifically, the system Mo-Nb-Ti-V-W. The MPEAs are explored to optimize two density-functional theory (DFT) derived ductility indicators (Pugh's Ratio and Cauchy pressure) while learning design constraints relevant to the manufacturing of high-temperature gas-turbine components. Alloys in the BO Pareto-front are analyzed using DFT to gain an insight into fundamental atomic and electronic underpinning for their superior performance, as evaluated within this framework.

5.2 Introduction

Improved gas-turbine engine (GTE) technology requires the continued development of high-temperature materials with higher strength and creep resistance at operational temperatures. The current materials of choice for use in such extreme operating conditions are Ni-based superalloys [144]. As gas-turbine efficiency can be improved by increasing the inlet temperature to the engine [145], there is a motivation to operate gas turbine engines at increasingly elevated temperatures. Currently, Ni-based superalloys operate at temperatures approaching their melting

*Reprinted with permission from "Multi-Objective Materials Bayesian Optimization with Active Learning of Design Constraints: Design of Ductile Refractory Multi-Principal-Element Alloys" by D. Khatamsaz, B. Vela, P. Singh, D. Johnson, D. Allaire, and R. Arróyave., 2022. Acta Materialia, Copyright 2022 Acta Materialia Inc. Published by Elsevier Ltd.[143]

temperatures($\sim 0.9\, T_m$) [146]. Novel gas turbine technologies require metallic materials that can perform at temperatures exceeding 1150°C, beyond which nickel-based superalloys are unable to perform due to inherent limitations from their melting temperatures [144]. While sophisticated cooling systems, such as cooling channels [147], thermal barrier coatings [148], and engineering of thermal conductivity [148], have been employed in the design of modern jet turbine blades, Ni-based superalloys are quickly approaching their operational limit, spurring exploration for novel, ultrahigh-temperature materials.

Recently, refractory multi-principal-element alloys (MPEAs) have garnered much attention as an emerging class of high-temperature materials. MPEAs consist of several alloying components (typically 4 or more) with concentrations ranging from 5 to 35 at.%, whereas conventional alloys rely on a single predominant constituent. Refractory MPEAs generally form body-centered-cubic (bcc) solid solutions [149] that have been shown to possess high-temperature properties comparable to those of the current Ni-based superalloys [150]. The compositional complexity of these refractory MPEAs creates opportunities to design alloys with unique properties, such as high-temperature yield strength [150], low density [151], creep resistance [152], and oxidation resistance [153]. However, while refractory MPEAs may satisfy performance constraints at elevated temperatures, due to the ductile-to-brittle-transition temperature of these alloys, many are brittle at room temperature and thus not machinable. The extreme difficultly associated with processing refractory alloys has historically limited their development [154]. In fact, room temperature ductility is a significant bottleneck in the development of refractory MPEAs [155]. For this reason we proposed a novel framework capable of multi-objective Bayesian optimization and active learning of multiple constraint boundaries in order to optimize for ductility in refractory MPEAs while under GTE relevant design constraints.

With the advent of Integrated Computational Materials Engineering (ICME), it has become possible solve the inverse problem and design alloys with tailored properties [52]. ICME relies on simulation in tandem with experiments to build linkages along the process-structure-property performance (PSPP) chain. Performance constraints are often defined in terms of materials properties.

In turn, these materials properties are dictated by the structure of the material. Finally, the structure of the material is determined by the processing conditions used to realize the material. When inverted, these linkages can guide the search for alloys that meet certain performance constraints. Thus within an ICME framework, refractory MPEAs can be designed, in principle, with ductility in mind.

Other works where Bayesian optimization is used within the ICME paradigm include examples of mutli-objective optimization, optimization under unknown constraints, and active learning to reduce the design space. Regarding multi-objective optimization, Solomou et al. [62] demonstrated tri-objective Bayesian optimization for the design of precipitation hardened shape memory alloys, simultaneously optimizing for austenitic finish temperature, specific thermal hysteresis (defined by the difference of austenitic finish temperature and martensitic start temperature), and the maximum transformation strain. Regarding Bayesian optimization under constraints, Griffiths et al. [156] were able to optimize drug-like molecules while under the constraint that said molecules must be valid molecular structures. The Bayesian optimization was performed over the latent space of a variational autoencoder that encoded the molecular structure of candidate designs. The authors used Bayesian Neural Network (BNN) classifiers that would output the probability of a point in the latent space being mapped to a valid molecular structure or not. Optimization was then carried out in feasible regions of the design space. Regarding active learning to reduce the design space, using the e-PAL framework, Jablonka et al. [157] were able to use active learning to efficiently estimate the Pareto-front during the multi-objective design of polymers for dispersant applications. The polymers were designed for optimal adsorption free energy, dimer free energy barrier, and radius of gyration. The polymer design space was iteratively reduced as e-PAL classified points as either likely dominated or likely Pareto optimal, actively learning the Pareto-front. The framework converges when all remaining points are classified as dominated (disregarded) or Pareto optimal.

In this chapter, we present a novel framework to perform multi-objective Bayesian optimization under unknown constraints. The framework is capable of actively learning the constraint boundaries as well as iteratively reducing the design space by discerning between feasible and infeasible

design regions. We seek to computationally link structure to property in refractory MPEAs by optimizing well-known ductility indicators, i.e., Pugh's ratio and Cauchy's pressure. The framework is benchmarked by designing ductile refractory MPEAs while under two constraints (density and solidus temperature) relevant to gas turbine application as a case study. Furthermore, a detailed DFT calculations is done on predicted MPEAs to assess underlying features driving ductility and its origin. While this work is limited to ductility, there exists opportunity to account for more objectives and constraints.

5.2.1 Design Objectives and Constraints

A known ductility indicator in refractory MPEA design is the valence electron concentration (VEC), as shown to be true theoretically [158] and experimentally [155]. A low VEC will promote shear failure and suppress cleavage failure in bcc-based alloys due to shear instability introduced by decreasing VEC [158]. It is a common MPEA design rule that that refractory alloys with low VECs are more ductile than those with higher VECs. For example, driven by the fact that equimolar HfNbTiZr and HfNbTaTiZr alloys are known to be ductile, Sheikh et al.[155] minimized the VEC in the HfNbTaTiZr alloy space under the constraint that the alloy be single-phase bcc. This constraint was encoded by the enthalpy of mixing and the atomic size mismatch, which are known indicators of the stability of the bcc solid-solution phase. The authors identified and synthesized $Hf_{0.5}Nb_{0.5}Ta_{0.5}Ti_{1.5}Zr$ which had an elongation at fracture of 18.8%. While the VEC is useful in identifying ductile alloys in a HTP manner, it does not *explicitly* account for elasticity in the crystal structure of the alloy. This points toward the inability of such metrics to find the *most* ductile alloys, therefore, indicators that capture the crystalline elasticity of an alloy are needed.

The ductility/brittleness of MPEAs can also be encoded by metrics derived from the elastic properties of alloys, such as the Pugh's ratio and the Cauchy pressure. These two indicators of ductility have been used extensively in the design of ductile MPEAs [159, 160, 161]. For pure crystalline metals, the Pugh's ratio is defined as the ratio of the bulk modulus over the shear modulus (B/G). This ratio encodes the competition between resistance to plastic deformation (G) and the fracture strength (B); Thus B/G captures the extent of the plastic range without fracture [162].

Pettifor [163] proposed Cauchy pressure as an indicator of intrinsic ductility/brittleness, which is the difference of two elastic constants C_{12} and C_{44}. A positive Cauchy pressure indicates non-directional metallic bonds resulting in intrinsic ductility of the crystal, whereas a negative Cauchy pressure corresponds to directional bonds and results in an intrinsically brittle crystal structure. Both indicators can be estimated with high-fidelity DFT frameworks at great computational cost. However, as the MPEA composition space is combinatorically vast, sufficient exploration of the space is intractable using conventional (computational or experimental) approaches.

Furthermore, while a given alloy may be ductile, its other properties may not be appropriate for GTE applications. For example, a ductile alloy may be too dense ($\rho < 11$ g/cc) for use in aviation. Likewise, an alloy that is optimized for ductility may have too low of a solidus temperature ($T_s <$ 2000°C) for use inside the hot-zones of GTEs. While we do not wish to optimize for these two properties, we still must classify alloys based on whether they meet said constraints. Therefore, to explore effectively this vast design space for ductile alloys while under design constraints related to GTE application, and under resource constraints due to the high cost of the DFT truth-model, intelligent optimization schemes capable of balancing resources between optimization *and* classification are needed. This classification step essentially aims to discover the feasible alloy space amenable for further optimization.

5.2.2 Multi-Information Source, Constraint-Aware Bayesian Optimization

Limitations on computational resources is a bottleneck in solving optimization problems in engineering applications. Many of engineering systems are in the form of black-box objective functions that require numerical approaches to search the input space for designs corresponding to optimum values of the quantities of interest. Among the proposed approaches for such design applications are Bayesian optimization (BO) techniques. Bayesian techniques offer a more efficient optimization by employing a heuristic-based search and, more importantly, the ability to update the system's state of knowledge continuously as new observations are introduced to the system.

Furthermore, in many instances in materials engineering, several models are available that represent the same system of interest. These models are based on different assumptions and/or

simplifications, and thus differ in fidelity and cost of evaluation. These models are treated as sources that provide useful information about a quantity of interest and are thus called *information sources*. In Refs. [37, 118, 141, 6, 33], it has been shown that employing multi-information sources in multi-fidelity BO frameworks offers a more robust and efficient approach to implement in design applications in comparison to single model optimization techniques.

In multi-information source BO, the assumption is that every source contains useful information regarding optimum design; thus, accurately fusing these sources results in a fused model that can mimic the response of the highest-fidelity model, known as the ground truth, enabling the search of the design space for the optimum design at no considerable computational cost. As later developments of the works [37, 118, 141, 6, 33], in Ref. [84], a novel framework is proposed that is capable of optimizing multiple objectives in multi-fidelity settings. In many engineering applications, there are several quantities of interest to be optimized simultaneously that urges the need for multi-objective optimization techniques. In Ref. [84], it has been shown that single fidelity approaches either using a BO framework or other techniques such as ParEGO and genetic algorithms are outperformed computationally when multiple sources of information contribute to provide information about the quantities of interest.

Regarding classification (or identification of a materials feasible space), a challenge in constrained optimization problems is correctly recognizing the feasible regions and their boundaries. Although sometimes checking the feasibility of a design input is done at no considerable costs, for example, by simply inserting the design variables into a analytical equation, there exist cases that constraints are defined by computationally expensive models which makes it impractical to verify the feasibility of every single design by querying their respective models. Therefore, constructing cheaper machine learning models to represent the constraints can reduce the overall cost of solving a constrained optimization problem.

A natural choice is to use surrogate models and information-theoretic approaches to learn the constraint models and accurately estimate their value at different locations in the input space [164]. However, it might be unnecessary to model the constraint over the entire input space as the bound-

ary separating the feasible and infeasible regions is what is truly of importance. Thus, in this study, we propose a Bayesian classification (BC) framework that uses classifiers and an active learning technique to effectively learn the constraint boundaries and recognize the feasible regions via checking the class memberships of any locations in the input space. Similar to multi-fidelity BO frameworks, in classification problems, there may exist several information sources that model the same constraint. Thus, we introduce our classification framework in form of a multi-fidelity BC configuration. Then, by coupling this multi-fidelity BC framework with the multi-objective multi-fidelity Bayesian framework introduced in [84], we create a design framework that actively learns the constraint boundaries and guides the search toward the optimum design by recognizing the feasible regions.

Here, we make further developments to the multi-objective, multi-fidelity BO framework introduced in Ref. [84] and propose a novel approach to solve constrained-design problems. Specifically, we deploy this framework in the Mo-Nb-Ti-V-W system, an exemplary MPEA system. By balancing the need to learn the constraint boundaries more accurately with improving the system's knowledge about the optimum values of quantities of interest, our proposed framework is able to make decisions about the best action to take at every stage of the process.

5.3 Methods

5.3.1 Gaussian Process Regression

In the context of BO, surrogate models are employed to model the behavior of objective functions and to represent a cheaper source to estimate the objective values associated to different design inputs and then calculate the expected gains regarding a potential design evaluation at much lower computational costs without the need to call the objective function directly.

We have implemented Gaussian process regression (GPR) to model objective functions [51]. Gaussian process models are powerful tools for probabilistic modeling purposes. They are easy to manipulate and simple to update as new observations are made available. In multi-fidelity BO frameworks, there are several sources to estimate the same quantity of interest at different fidelity

levels, each providing key piece of information about the ground-truth quantity of interest. In presence of multi-information sources, several Gaussian processes are constructed to represent the response surface of these information sources.

Following Refs. [33] and [6], we assume we have available some set of information sources, $f_i(\mathbf{x})$, where $i \in \{1, 2, \ldots, S\}$, that can be used to estimate the quantity of interest, $f(\mathbf{x})$, at design point $\mathbf{x}$. These surrogates are indicated by $f_{\text{GP},i}(\mathbf{x})$. Assuming there are N_i evaluations of information source i denoted by $\{\mathbf{X}_{N_i}, \mathbf{y}_{N_i}\}$, where $\mathbf{X}_{N_i} = (\mathbf{x}_{1,i}, \ldots, \mathbf{x}_{N_i,i})$ represents the N_i input samples to information source i and $\mathbf{y}_{N_i} = (f_i(\mathbf{x}_{1,i}), \ldots, f_i(\mathbf{x}_{N_i,i}))$ represents the corresponding outputs from information source i, then the posterior distribution of information source i at design point $\mathbf{x}$ is given as

$$f_{\text{GP},i}(\mathbf{x}) \mid \mathbf{X}_{N_i}, \mathbf{y}_{N_i} \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma^2_{\text{GP},i}(\mathbf{x})\right) \tag{5.1}$$

where

$$\begin{aligned} \mu_i(\mathbf{x}) &= K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma^2_{n,i} I]^{-1} \mathbf{y}_{N_i} \\ \sigma^2_{\text{GP},i}(\mathbf{x}) &= k_i(\mathbf{x}, \mathbf{x}) - K_i(\mathbf{X}_{N_i}, \mathbf{x})^T \\ &\qquad [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma^2_{n,i} I]^{-1} K_i(\mathbf{X}_{N_i}, \mathbf{x}) \end{aligned} \tag{5.2}$$

where k_i is a real-valued kernel function over the input space, $K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i})$ is the $N_i \times N_i$ matrix whose m, n entry is $k_i(\mathbf{x}_{m,i}, \mathbf{x}_{n,i})$, and $K_i(\mathbf{X}_{N_i}, \mathbf{x})$ is the $N_i \times 1$ vector whose m^{th} entry is $k_i(\mathbf{x}_{m,i}, \mathbf{x})$ for information source i. We have also included the term $\sigma^2_{n,i}$, which is used to model observation error for information sources based on experiments. Without loss of generality, we employ the squared exponential covariance function as the kernel function specified as

$$k_i(\mathbf{x}, \mathbf{x}') = \sigma_s^2 \exp\left(-\sum_{h=1}^{d} \frac{(x_h - x'_h)^2}{2l_h^2}\right) \tag{5.3}$$

where d is the dimensionality of the input space, σ_s^2 is the signal variance, and l_h, where $h = 1, 2, \ldots, d$, is the characteristic length-scale that indicates the correlation strength between the points within the dimension h. The parameters σ_s^2 and l_h associated with each information source can be estimated by maximizing the log marginal likelihood.

When using multiple information sources to estimate a ground truth quantity of interest, it is important to quantify the uncertainty in the response of the information sources with respect to the ground truth, which is defined as the discrepancy term to compensate for the lower fidelity estimation of the ground truth quantity of interest. We quantify the total variance that captures both the variance associated with the Gaussian process representation and the quantified variance associated with the fidelity of the information source over the input space, as

$$\sigma_i^2(\mathbf{x}) = \sigma_{\text{GP},i}^2(\mathbf{x}) + \sigma_{f,i}^2(\mathbf{x}) \tag{5.4}$$

where $\sigma_{f,i}^2(\mathbf{x})$ is the variance related to the fidelity of information source i that can be estimated from, for example, expert opinion or available real-world data.

5.3.2 Gaussian Process Classification

Similar to optimization problems, Bayes' theorem can be employed to calculate the joint probability $p(y,\mathbf{x})$, where y is the class label, in classification problems:

$$p(y|\mathbf{x}) = \frac{p(y)p(\mathbf{x}|y)}{\sum_{c=1}^{C} p(C_c)p(\mathbf{x}|C_c)} \tag{5.5}$$

A challenge of Gaussian process classification (GPC) that is not present in Gaussian process regression is that of non-Gaussian likelihoods in GPC. To overcome this, following Ref. [51], we use a discriminative GPC approach that seeks to model $p(y|\mathbf{x})$ directly, which avoids the need to specify prior distributions over $p(y)$ and the specification of class-conditional densities, $p(\mathbf{x}|C_c)$. Discriminative GPCs are probabilistic classifiers that predict the probability of belonging to a class by placing a Gaussian process prior over a latent function $f(\mathbf{X})$ and computing the posterior distribution at a desired location $\mathbf{x}$ [51, 165]. Here, we are not interested in the values of the latent function, but it is used to conveniently formulate the classifier. This latent function is responsible to connect the input to the output, where the output is class membership probability. Consequently, we need the latent function posterior distribution respectively. The Laplace approximation algo-

rithm is employed that utilizes Gaussian approximation to the posterior of the latent variables. The approximated posterior is then used to obtain the class membership distribution, where Monte Carlo sampling is done to estimate the class membership probability. Similar to the construction of Gaussian process regressions, assuming we have available some set of information sources $i \in \{1, 2, \ldots, S\}$ with N_i labeled samples, the latent function $f_i(\mathbf{x})$ has a multivariate normal distribution defined by

$$\begin{aligned} \mu_i(\mathbf{x}) &= K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i})]^{-1} f(\mathbf{X}) \\ \Sigma_i(\mathbf{x}) &= k_i(\mathbf{x}, \mathbf{x}) - K_i(\mathbf{X}_{N_i}, \mathbf{x})^T \\ &\qquad [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i})]^{-1} K_i(\mathbf{X}_{N_i}, \mathbf{x}) \end{aligned} \tag{5.6}$$

The class label predictions are obtained by sampling from the calculated posterior distribution and passing the samples through a sigmoid function σ, for example, the logistic sigmoid, to ensure the output is bounded to [0,1]. Then the mean of the obtained distribution is the class membership probabilities.

In the context of BC and learning purposes, the uncertainty associated to the predictions are essential in calculation of an expected utility value. Note that this is the distinguishing characteristic of the Gaussian process classification as a probabilistic model in comparison to other classification techniques that makes GPC well-suited for probabilistic frameworks and learning purposes. A more detailed discussion is presented in [51].

5.3.3 Information Fusion of Multiple Sources

Assuming that every information source participating in the optimization process contains some useful information regarding the ground truth quantity of interest, the goal is to accurately fuse the information provided by these information sources to approximate the quantity of interest as accurately as possible at much lower costs in comparison to evaluating the ground truth objective function [5, 6, 7, 8].

Several approaches exist for fusing multiple sources of information, such as Bayesian model-

ing averaging [9, 10, 11, 12, 13, 14], the use of adjustment factors [15, 16, 17, 18], covariance intersection methods [19], and fusion under known correlation [20, 21, 22].

Our assumption is that every information source contains useful information regarding the ground truth quantity of interest and as more information sources are incorporated into a fusion process, the expectation is to have the variance of the quantity of interest estimates decreased. This is not necessarily the case for all of the aforementioned fusion techniques with the exception of fusion under known correlation. Unlike most traditional multi-fidelity approaches [93, 91, 166, 167, 168, 169, 170, 171], in our approach, we do not assume a hierarchy of information sources and our goal is optimization with respect to ground truth and not optimization with the highest fidelity source. Therefore, determining correlations prior to fusion is essentially important. To estimate the correlation coefficients between information sources, we use the reification process introduced in Refs. [5, 8]. In reification process, a pair of information sources are selected each time and they are reified ('made real') in turn, which means one information source is assumed to be the true model and the deviation of the second information source with respect to the reified model is calculated. These calculated deviations are used to obtain the correlation between the mean squared errors of the information sources. The covariance matrix is formed after the reification is done over each pair of information sources. Readers are encouraged to check Refs. [5, 6, 7, 8] for detailed discussion on how the correlation estimation is performed. In case of known correlations between the discrepancies of information sources, the fused mean and variance at a particular design point $\mathbf{x}$ are defined as [22]

$$E[\hat{f}(\mathbf{x})] = \frac{\mathbf{e}^{\mathrm{T}}\tilde{\Sigma}(\mathbf{x})^{-1}\boldsymbol{\mu}(\mathbf{x})}{\mathbf{e}^{\mathrm{T}}\tilde{\Sigma}(\mathbf{x})^{-1}\mathbf{e}} \tag{5.7}$$

$$\mathrm{Var}\left(\hat{f}(\mathbf{x})\right) = \frac{1}{\mathbf{e}^{\mathrm{T}}\tilde{\Sigma}(\mathbf{x})^{-1}\mathbf{e}} \tag{5.8}$$

where $\mathbf{e} = [1, \ldots, 1]^{\mathrm{T}}$ is a S dimensional column vector of ones, $\boldsymbol{\mu}(\mathbf{x}) = [\mu_1(\mathbf{x}), \ldots, \mu_S(\mathbf{x})]^{\mathrm{T}}$ given S models, and $\tilde{\Sigma}(\mathbf{x})^{-1}$ is the inverse of the covariance matrix between the information sources. A more detailed discussion on this fusion technique and some examples of its implementation are presented in Refs. [77, 172, 5, 141, 64, 50, 33, 37].

5.3.4 Multi-Objective Optimization

A multi-objective optimization problem can be defined as

$$\text{minimize } \{f_1(\mathbf{x}), ..., f_n(\mathbf{x})\}, \mathbf{x} \in \mathcal{X} \tag{5.9}$$

where $f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})$ are the objectives and $\mathcal{X}$ is the feasible design space. In multi-objective optimization problems, it is usually the case that there is no single solution that optimizes all objectives simultaneously. Thus, the solution to such design problems is a set of non-dominated designs that are not superior to each other, forming the Pareto-front in the objective space. In this case, optimal solutions, $\mathbf{y}$, to a multi-objective optimization problem with n objectives are denoted as $\mathbf{y} \prec \mathbf{y}'$, and are defined by

$$\{\mathbf{y} : \mathbf{y} = (y_1, y_2, \ldots, y_n),\ y_i \leq y_i'\ \forall\, i \in \{1, 2, \ldots, n\},\ \ \exists\, j \in \{1, 2, \ldots, n\} : y_j < y_j'\} \tag{5.10}$$

where $\mathbf{y}' = (y_1', y_2', \ldots, y_n')$ denotes any possible objective output. The set of $\mathbf{y} \in \mathcal{Y}$, where $\mathcal{Y}$ is the objective space, is the Pareto-front of the problem.

There are several techniques to employ to estimate the Pareto-front in multi-objective optimization problems such as weighted sum approach [23], the adaptive weighted sum approach [24], normal boundary intersection methods [25], hypervolume indicator methods [26, 27, 28, 29, 30, 31, 32], and others. In the context of BO, and working with expected improvement- based algorithms, the hypervolume indicator approaches are well-suited as they allow for introducing a single random variable, combining multiple random variables (here, objective values), to consider the uncertainty of all other variables as a whole. Thus, in a BO framework, a hypervolume indicator can be the target variable to be optimized. The idea is to define the volume between the estimated Pareto-front and a fixed point in the objective space as the hypervolume and relate the expected improvement associated to each objective value to the total hypervolume and calculate the expected hypervolume improvement. We follow the work proposed in Refs. [84, 105] for BO of multi-objective

functions in presence of multiple information sources. An in-depth discussion on the calculation of the expected hypervolume improvement can be found in the mentioned references and [92].

5.3.5 Active Learning in Bayesian Classification

Using GPCs to label the samples, there are uncertainty associated to the label predictions that also shows how uncertain is the classifier about the label at a particular location. The class membership Y is a random variable indicated by a distribution P. In general, we look for a measure that shows the degree of uncertainty based on the given distribution for class memberships. A convenient option is using discrete entropy to calculate the uncertainty in the label predictions:

$$H(Y) = -\sum_{i=1}^{k} p_i \log(p_i) \tag{5.11}$$

where there are k classes and p_i is the probability of belonging to class i. Higher values of entropy shows the larger uncertainty of the classifier about the label of a sample. Thus, we look forward to query samples and update the classifiers at locations where the classifier is highly uncertain about the true label.

An issue when using entropy as a measure of uncertainty is that in the case of multi-class classification, it is possible that the classifier is only uncertain about the memberships in, for example, two classes and is highly certain about the other class memberships that may be unimportant. This is the case when the entropy is still high, and it does not provide any information about the confidence on some class memberships. To address this issue, following Ref. [173], we can take the "Best versus second Best" approach. This technique only considers the top two uncertain class memberships and measures the uncertainty based on the difference between the two probabilities.

For our constrained BO, we only perform binary classification since we are using classifiers to separate the feasible and unfeasible regions. In this case, our problem reduces to finding the samples closest to the classifier's predicted constraint boundary. In other words, the samples with the smallest difference between class membership probabilities are chosen to be queried from the constraints to update the classifier and learn about the true constraint boundary.

5.3.6 Truth model – Density Functional Theory

The truth model in the multi-objective optimization side framework was queried through the DFT-based KKR (Korringa-Kohn-Rostoker Green's function) method, in which the coherent-potential approximation (CPA) accounts properly for direct configurational average over chemical disorder [174], concomitantly with the charge self-consistency required within the standard DFT. One of the objectives, i.e., bulk moduli, used by the framework were calculated by employing a gradient-corrected exchange-correlation functional (PBE) [175]. Additionally, the Warren-Cowley short-range order (SRO, given as $\alpha_{\mu\nu}(k;T)$) of the designed compositions were calculated as implemented within DFT-KKR-CPA linear-response theory to analyze the ordering tendencies found in the final compositions [176, 177, 178].

blackThe structural optimization to estimate local-lattice distortion (LLD), a ductility metric of designed compositions in Table 5.1, were performed using the DFT method as implemented within the Vienna Ab initio Simulation Package [179, 180, 181, 175]. The Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) functional [175] was employed for geometrical relaxations with total-energy and force convergence criteria of 10^6 eV and 0.01 eV/Å, respectively. To mimc MPEAs within a finite cell, supercells were designed using SCRAPs [182] – SuperCell Random APproximateS. We chose two SCRAP sizes, (i) 128 atoms for design 1, and (ii) 160 atoms for design #2 to #7 in Table 1. The largest possible supercells were used to avoid size effects [183]. The Brillouin zone integration in charge self-consistency and ionic relaxtion were performed on $1\times1\times1$ using Monkhorst-Pack method [184] with a plane-wave cutoff energy of 520 eV, where the effect of the core electrons and interaction between the nuclei and the valence was treated by the projector-augmented wave (PAW) [185, 186].

The DFT-KKR-CPA is expensive to query. blackTherefore, computationally inexpensive alternatives, e.g., rule-of-mixtures approximation, are needed for both the Pugh's ratio and the Cauchy pressure to explore the objective space. In this work, to estimate the Pugh's ratio cheaply, we express the bulk and shear modulus in terms of the Poisson ratio. We then calculate the weighted average of the elemental Poisson ratio to evaluate Equation 5.12, where i iterates along all N el-

ements in the design space, ν_i and x_i is the Poisson ratio of the ith element, respectively. For the Cauchy pressure, we calculate the weighted average of C_{12} and C_{44} elastic constants and find their difference according to Equation 5.13.

$$\overline{B/G} = \frac{2(1 + \sum_{i=1}^{N} \nu_i x_i)}{3(1 - \sum_{i=1}^{N} \nu_i x_i)} \tag{5.12}$$

$$\overline{C}_{pres} = \sum_{i=1}^{N} C_{12} x_i \; - \sum_{i=1}^{N} C_{44} x_i \tag{5.13}$$

5.3.7 Thermodynamic Simulation

The truth model in the classification side of the framework was queried through a high fidelity CALculation of PHase Diagrams (CALPHAD) based simulation scheme. Equipped with the MPEA specific TCHEA5 thermodynamic database, Thermo-Calc's equilibrium simulation was used to query both the density and the solidus temperature. The integration of these models within this automated framework was achieved using the Thermo-Calc API, TC-Python. These thermodynamic equilibrium simulations are relatively expensive to query, As such, cheaper alternatives are required to explore the constraint space. The rule-of-mixtures was used as a cheap alternative to the CALPHAD truth model for both density and solidus temperature.

5.4 Results

5.4.1 Multi-Fidelity Bayesian classification

In this work, we use Gaussian process classification (GPC) to model the constraint boundaries to distinguish the feasible and unfeasible regions in the design space (binary classification). A BC framework employs entropy measure to determine the uncertainty and search for the best next experiment to query the constraints and update corresponding GPCs to decrease the labeling uncertainty the most. Similar to multi-fidelity BO frameworks introduced in Refs. [84, 141, 6, 37], a BC framework can be designed in multi-fidelity settings. There may exist several models to represent the same constraint with different fidelity levels and evaluation costs. Using the reification process

followed by fusion of multiple sources introduced in Refs. [5, 22, 77], a fused classifier can be constructed for each constraint that accurately models the constraint boundary using information gained from different sources. To show how a multi-fidelity BC approach is able to determine the constraint boundaries more accurately and efficiently, a test problem is designed and shown in Fig. 5.1. The highest fidelity model represents the true boundary while there are also two lower fidelity models to estimate the boundary between regions 'A' and 'B'.

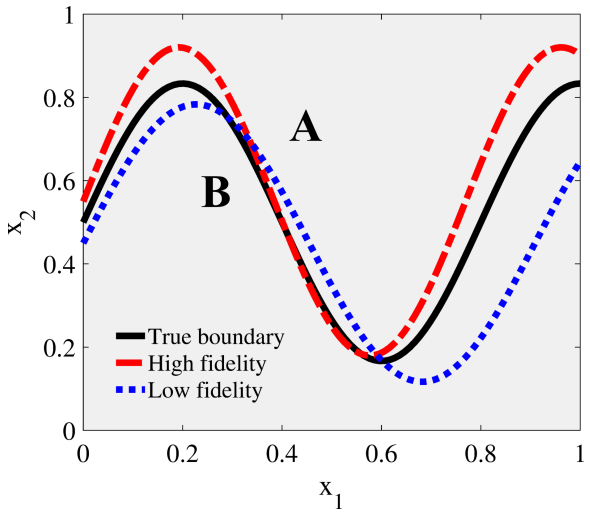

Figure 5.1: 2 Dimensional classification test problem. The space is divided into regions 'A' and 'B'. Two lower fidelity models estimate the true boundary.

In Fig. 5.2, the results of employing the BC framework in single- and multi-fidelity settings have been shown. While the black curve is the true boundary that is desired to be recognized, the blue curve shows the classifier's estimation of the boundary with 95 % confidence intervals. Red dots are the locations in the input space queried in the BC process to update the classifiers and decrease the uncertainty of classification. They show the estimated constraint boundary after 20, 50, and 100 queries from the highest fidelity model. After every 10 queries to lower fidelity models, a fused model is constructed and the design closest to the boundary is chosen to be queried from the highest fidelity model. The closer a design is to the estimated boundary, the larger the

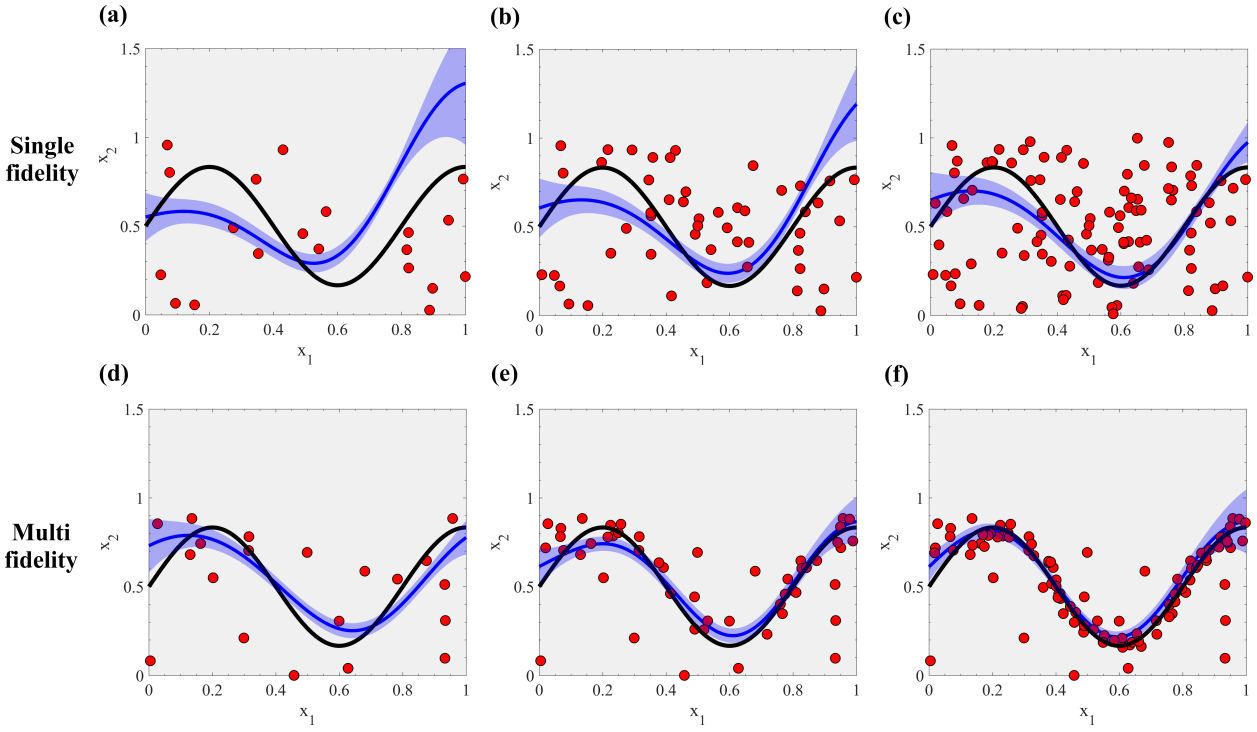

Figure 5.2: Boundary estimation results using single and multi fidelity Bayesian classification approaches. True constraint boundary (in black) versus estimated constraint boundary with 95 % confidence intervals (in blue). Red dots show where the framework has chosen to query. (a) and (d): 20 queries. (b) and (e): 50 queries. (c) and (f): 100 queries.

classifier uncertainty is about its class membership. Note that the boundary is where the classifier assigns the probability of class memberships very close to 50 % in case of binary classification.

As seen in Fig. 5.2, employing multiple sources to represent the same constraint has improved the performance of the classifier significantly. Not only it has estimated the true boundary more accurately with narrower confidence intervals, but also it has done so with a smaller number of queries from the highest fidelity model. Plus, the queried locations show how more effectively the framework is able to make highly informative queries to help in determining the constraint boundary.

5.4.2 Proposed design framework

Fig. 5.3 illustrates the flowchart of the proposed design framework to solve constrained optimization problems. Here, the BO and BC frameworks are coupled to build a larger configuration

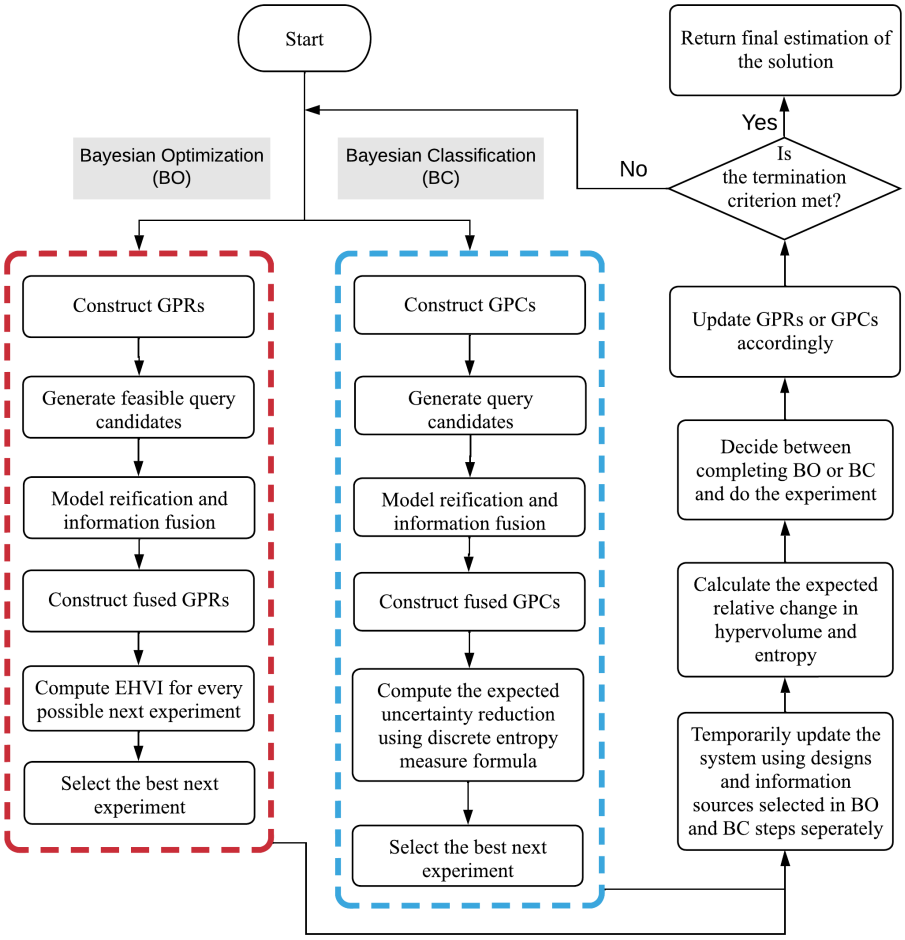

Figure 5.3: Flowchart of design framework. This flowchart shows the main steps in the design framework. Gaussian process regressions (GPRs) model the objective function and Gaussian process classifiers (GPCs) model the constraint boundaries.

that wisely and optimally makes decisions about improving the system's knowledge regarding the optimum design while learning the constraint boundaries to recognize the feasible design region.

In the BO part of this design framework, the most up-to-date fused classifiers are used to determine the feasible regions and prevent the framework to search unfeasible regions. Therefore, before spending any computational resources to test a set of generated samples to search for the best next experiment, fused classifiers determine if all samples are feasible. If not, the unfeasible samples will be removed and new samples are generated until we find the desired number of feasible samples to test. To make decision on feasibility of a design, we used $\mu - 2\sigma > 0.5$ to

guarantee at least 95 percent confidence in predictions. One can change this confidence interval accordingly, depending on how hard (or difficult) is a constraint. Then, Expected HyperVolume Improvement (EHVI) is employed as the utility function to search for promising designs to query and increase the hypervolume the most and improve the estimation of the Pareto-front [92].

In the BC component of the framework, to actively learn the constraint boundaries and increase the accuracy of the fused classifiers, an entropy measure is used to gauge the uncertainty in labeling the samples. Since we do binary classification to label samples as either feasible or unfeasible, the problem is reduced to finding the closest samples to the predicted constraint boundary by classifiers. At every iteration of the framework, a decision has to be made between querying the objective functions and thus improving the optimum value of a quantity of interest (hypervolume in case of multi objective optimization) or querying the constraints and updating classifiers to increase the accuracy of fused classifiers.

To balance the decision between these two options, the expected relative change to the optimum value of an objective (i.e. the expected hypervolume improvment) is calculated and is compared to the relative change in entropy if we update the constraints. This is done by temporarily augmenting the samples determined as the best next experiment and observing the changes in the corresponding values. By comparing the observed relative changes in each case, the framework chooses the action that adds the most value to the system. Note that in both the BO and BC components, multiple information sources may have been employed to represent every objective function and every constraint. Thus, the selected samples are augmented to the GPR or the GPC corresponding to the selected information source.

5.4.3 Material Design Process

The design framework depicted in Fig. 5.3 has been employed to solve the aforementioned alloy design problem. There are two objectives defined in the problem to be maximized: Pugh's ratio and Cauchy pressure. These objectives encode the intrinsic ductility of candidate alloys. There are two GTE-relevant constraints that must be satisfied: solidus temperature and density of the candidate alloys. The solidus temperature for any composition should be greater than 2000°C

such that the alloy can withstand the hot-zone inside GTEs. Furthermore, the density should be less than 11 g/cm^3 such that the alloy is light enough for application in aviation. The design space is 5-dimensional (including temperature). Candidate alloys can consist of permutations and combinations of the following five elements: Mo, Nb, Ti, V, and W. The Thermo-Calc thermodynamic equilibrium model was used as the truth-model that represents the solidus and density constraints. For each constraint, there exists a low order rule-of-mixture approximation. Regarding the objective functions, there are also two models available, one as the lower fidelity information source that is inexpensive to query and responds in a few seconds. Then, a high fidelity KKR model was used as the ground truth (see **Methods** section). Being computationally demanding, the KKR-model was queried after every 10 queries from the lower fidelity information source to update the discrepancy of the low fidelity model and correct its estimation of the objective values.

First remark: We would like to note that the design constraints (solidus temperature and density) are relatively easy to estimate using conventional computational thermodynamic techniques. It would thus be possible to simply carry out a high-throughput exploration of the materials space [187] and simply select the feasible region without any active learning step. The purpose of this exercise is to demonstrate the integration of the discovery of the feasible space within a materials optimization framework. This capability would be significant in the case of 'real world' closed-loop materials discovery tasks in which the discovery of the feasible region in a materials design space is highly non-trivial and dependent on exhaustive experimental characterization. For example, the discovery of alloys with acceptable oxidation resistance *and optimal* mechanical performance may depend on actual experimental oxidation studies. *Our framework is agnostic regarding the nature of the information sources used and thus can easily be adapted to fully experimental or hybrid experimental/computational settings.*

The framework was run over 2,000 iterations. Regardless of how much is learned about the constraint boundaries, the last 500 iterations are allocated to perform only optimization to improve the estimated Pareto-front. In later works, an adaptive technique will be taken to increase the weights over optimization than classification as the number of iterations increase.

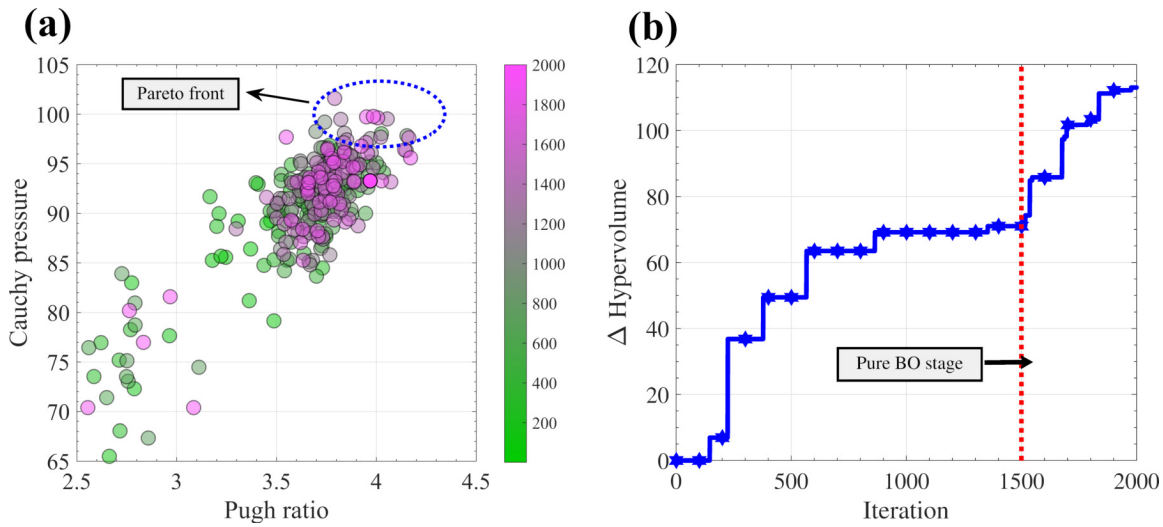

Figure 5.4: Final results of introduced material design problem (a) Queried samples from the KKR model in the objective space. Final estimation of the Pareto-front shows 7 non-dominated designs. (b) Improvement in hypervolume of the estimated Pareto-front as a function of iteration. Note that the hypervolume value is depended on the chosen reference point in the objective space, thus, the change in hypervolume is a better indication of making improvements to the Pareto-front estimation.

The results are depicted in Figure 5.4. In part (a), queried designs from the KKR model are shown in the objective space and the color map shows the order that queries are made. There are total of 284 queries made from the KKR model, where 229 are made in the last 500 iterations. The final estimation of the Pareto-front consists of 7 non-dominated designs. In Table 5.1, the design values corresponding to the estimated Pareto-front are specified. In part (b), the change in hypervolume of the estimate Pareto-front as number of iteration is depicted. As the number of iteration increases, the change in hypervolume value gets smaller. Furthermore, it takes more number of iterations to see further improvements in the hypervolume value, however, by allocating the last 500 iteration for optimization purposes, more improvements are made. This essentially shows that as we move toward the final iterations, it is worth to invest on optimization and make the final improvements possible to the solutions. An important point here is that all queried samples are satisfying both solidus temperature and density constraints, which shows how by actively learning the constraint boundaries and using classifiers, querying unfeasible designs is prevented. Tests

show that about 20 % of the randomly sampled designs violate at least one of the constraints.

Table 5.1: Non-dominated designs corresponding to the Pareto-front in Fig. 5.4. Compositions reported in atomic percentage.

Composition elements	Mo	Nb	Ti	V	W
Design 1	0.014	0.738	0	0.233	0.015
Design 2	0.0330	0.8080	0.0090	0.0890	0.0610
Design 3	0.0100	0.7460	0.0050	0.0910	0.1480
Design 4	0.0110	0.6840	0.0150	0.2740	0.0160
Design 5	0.1140	0.7350	0.0010	0.1490	0.0010
Design 6	0.0150	0.8410	0.0550	0.0010	0.0880
Design 7	0.0140	0.7550	0.0160	0.2020	0.0130

Fig. 5.5 illustrates the cumulative number of actions taken at every iteration. Out of 2,000 iterations, 634 iterations are dedicated to perform BO which means 634 queries are made from lower fidelity model. The KKR model is queried in batch of maximum of 5 designs after every 10 queries to the lower fidelity model. To find the batch of designs to be queried from the KKR model, first, a set of query candidates are generated and then their expected hypervolume improvement is calculated. Note that the objective values at every location in the design space are estimated by GPRs that provide normal distributions as the estimation. Then, samples with positive expected hypervolume improvement are kept and clustered by solving a k-medoid problem. Finally, the medoids (i.e. actual points belonging to a cluster in the design space that are maximally separated from other clusters) of these clusters are queried from the KKR model. This approach assures a good distribution of the queried samples all over the space that helps to fully discover the non-dominated region.

Out of 1,366 queries from the constraints, 1,189 queries are made from the density function. The reason that the framework struggles to identify the density constraint could be the complex shape of feasible and unfeasible regions, making it challenging to recognize the boundary. However, the results show that the framework has learned enough to confidently recognize the feasible regions.

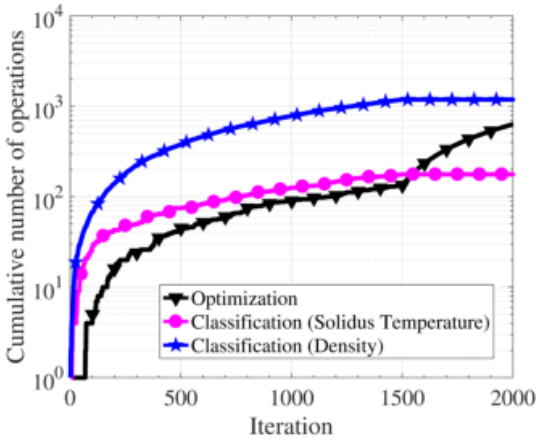

Figure 5.5: Cumulative number of operations as a function of iteration. For most number of iterations, the framework has decided to query the density model and update its classifier. It is an indication of the complexity of the density constraint boundary and the framework struggles to learn it and reduce the classifier's uncertainty in labeling the design inputs.

5.5 Discussion

5.5.1 HTP Brute Force Analysis to Evaluate Framework Classification

In order to assess the utility of using such an classification framework, we benchmarked the BC of refractory MPEA space against a traditional HTP 'brute-force' approach. In this brute-force approach the density and solidus temperature constraints were queried at increments of 5 at.% considering unary to quinary systems, resulting in 10,626 queries in total. However, with the BC framework, only 1,366 queries were needed to find the boundary in the constraint space, dramatically improving the tractability of the problem as the total number of queries over the potentially feasible space was reduced by close to a factor of nine. Querying these expensive information sources is beneficial as it can better elucidate how well the classification aspect of the framework actually performs in this refractory MPEA space.

In order to visualize how composition affects the constrained properties (solidus temperature and density) in this 5-dimensional refractory MPEA space, we rely on a dimensionality reduction technique known as t-distributed stochastic neighbor embedding (tSNE). Each point represents an alloy with a distinct composition. In tSNE embeddings, points that are close to each other in

high dimensional space are plotted close to each other in 2-dimensional space. Points colored in Fig. 5.8c are alloys that contain 50% or more of a particular element. Points closer to the corners of this "pentagonal" shape approach unary compositions. For example, points near the red corner of the tSNE are rich in vanadium; The point on the "corner" represents pure vanadium. Points along the edge connecting the red and blue regions are Mo-V binaries. The inner regions of the tSNE represent the refractory MPEA space. We would like to point out that this representation is mostly qualitative as tSNE embeddings only preserve the local structure of the dataset, while being agnostic with regard to its global structure.

The light blue stars represent the 7 alloys that lie on the Pugh's Ratio-Cauchy Pressure Pareto-front. These alloys are in the Niobium-rich region of the tSNE. Their location in this tSNE embedding provides a visualization of where in the refractory MPEA space the Pareto-front lies. The location of Pareto-front in the refractory MPEA space is not completely unexpected, as Nb has a relatively low VEC of 5. However, according the prevailing VEC theory of ductility in refractory MPEAs [155], Ti-rich alloys are expected to be the most ductile in this region as Ti has a VEC of 4, the lowest in the alloy space. However, in Figs. 5.6a and 5.6b, Ti-rich regions have solidus temperatures below 2273 K, violating the solidus constraint. The reason the framework converges on Nb-rich alloys instead of Ti-rich alloys is further demonstrated in Fig. 5.6 where the property space is plotted. In Fig. 5.6a the VEC is plotted against the solidus. Many of the Ti-rich alloys (depicted as squares) fall beneath the 2273 K constraint. In Fig. 5.6b, the Ti-rich alloys that do pass the solidus constraint do not have a large a Pugh's ratio as the Nb-Rich alloys (depicted as stars). While not as ductile as Ti-rich alloys, most Nb-rich alloys pass the density and solidus constraints. The classification side of the proposed framework was able to recognize the Ti-rich region of the design space as infeasible by querying the solidus information source only 177 times.

Second remark: We would also like to point out that our design setting is truly agnostic with regard to the complexity of the alloy space. Rather than focusing on high-complexity compositions, our framework optimized a set of property targets, subject to application specific design constraints. Whether such compositions are located within the "high entropy" alloy space is immaterial when

trying to discover optimal alloys with a target application in mind. It is the view of the authors that property/constraint-aware and "entropy" agnostic exploration of High Entropy Alloy spaces is a more productive research program as compared to approaches that focus on alloy complexity as the sole objective of the alloy design task.

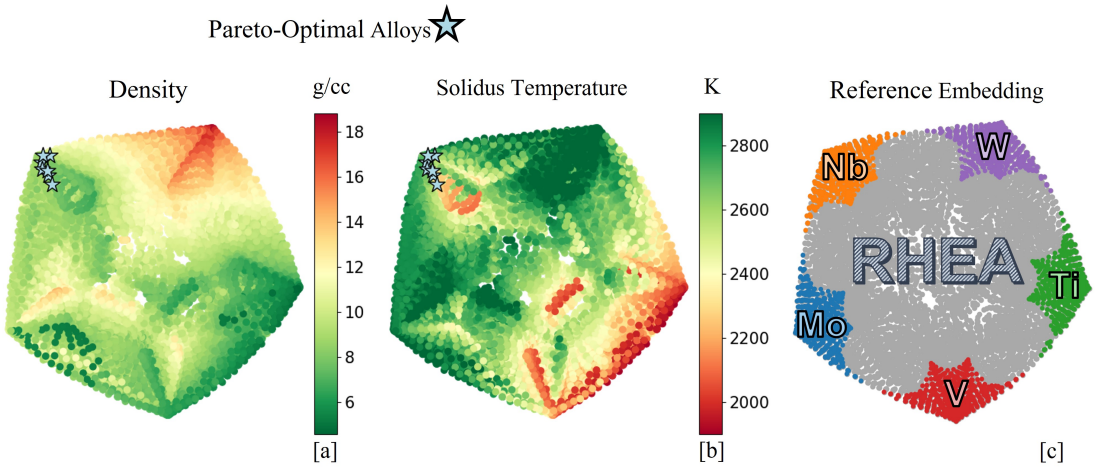

Figure 5.6: t-SNE projection of the MoNbTiVW alloy space. Colored points in the reference embedding (right) denote alloys that contain alloys that contain 45% or more of a particular element. The solidus temperature is plotted on the same embedding (right). Likewise, the density is plotted on the embedding (left).

5.5.2 DFT analysis of Pareto-front-selected Refractory MPEAs

The seven Nb-V rich alloys in Table 5.1 with superior mechanical properties that comprise the Pareto-front were selected for further analysis using DFT. Singh *et al.* [188] has shown that the ability of an alloy to resist local distortion and compositional changes have direct impact on ductility. However, the understanding of these features controlling distortion and local compositional changes is currently not discussed well. We performed detailed DFT calculations [174, 176, 179] of stability (formation energy), distortion factor and local distortion (atomic displacements), and short-range order (local chemical fluctuations) on design compositions.

Generally, refractory materials have very high moduli in the bcc phase because of their low

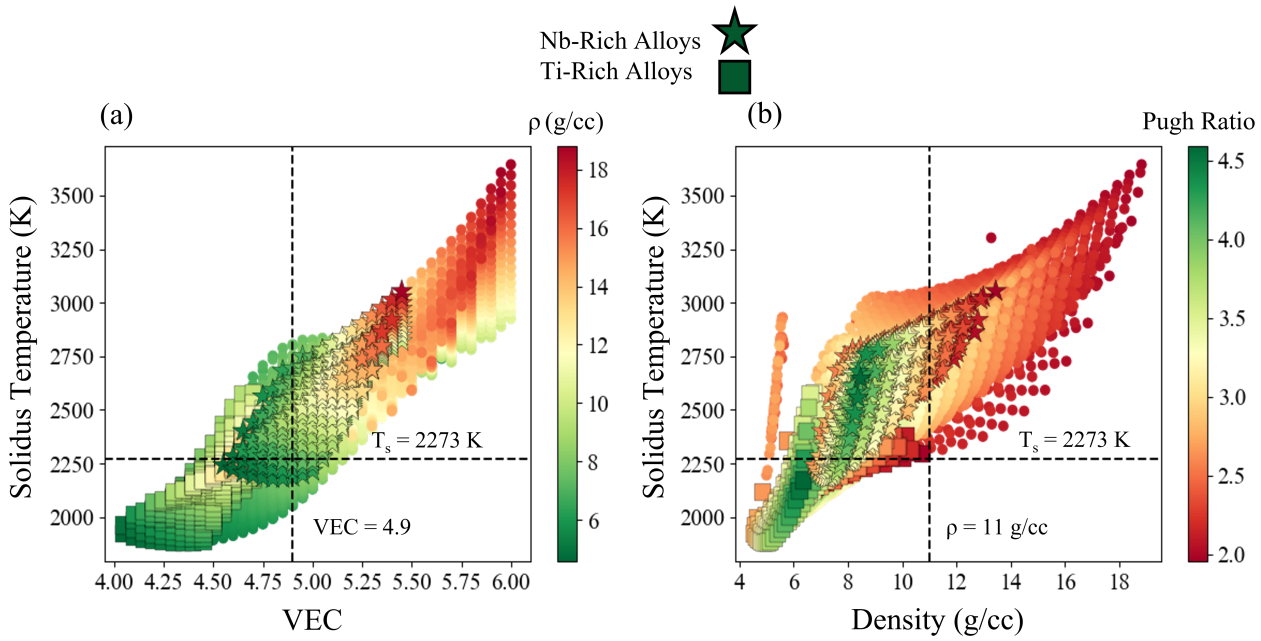

Figure 5.7: t-SNE projection of the MoNbTiVW alloy space. Colored points in the reference embedding (right) denote alloys that contain alloys that contain 45% or more of a particular element. The solidus temperature is plotted on the same embedding (right). Likewise, the density is plotted on the embedding (left).

compactness. The interaction between these elements are weaker than other crystal structures such as in fcc materials. Therefore, the ductility is definitely an issue, which helps a material to sustain against permanent (large) deformation under a tensile loading at room temperature without fracturing. In Fig. 5.8a-b, we plot BM of design compositions with respect to Nb-V composition (in atomic-fraction) and phase stability. The calculated BM was found in the moderate range in Fig. 5.8a, similar to highly ductile fcc-based materials; moreover, each compositions were in the desired MPEA stability range (-150 meV-atom^{-1} <E$_{form}$ <65 meV-atom^{-1}) [178].

In Fig. 5.8c, we plotted distortion metric with respect to scalar displacement, where most compositions show minimal local distortion except #7 that has both large distortion metric and local atomic displacement. Our findings suggest that the distortion factor may further get increased with temperature due to non-uniform local contraction and expansion in different regions with different atomic sizes, which may lead to large thermal strain in local lattices. In Fig. 5.8d, we found

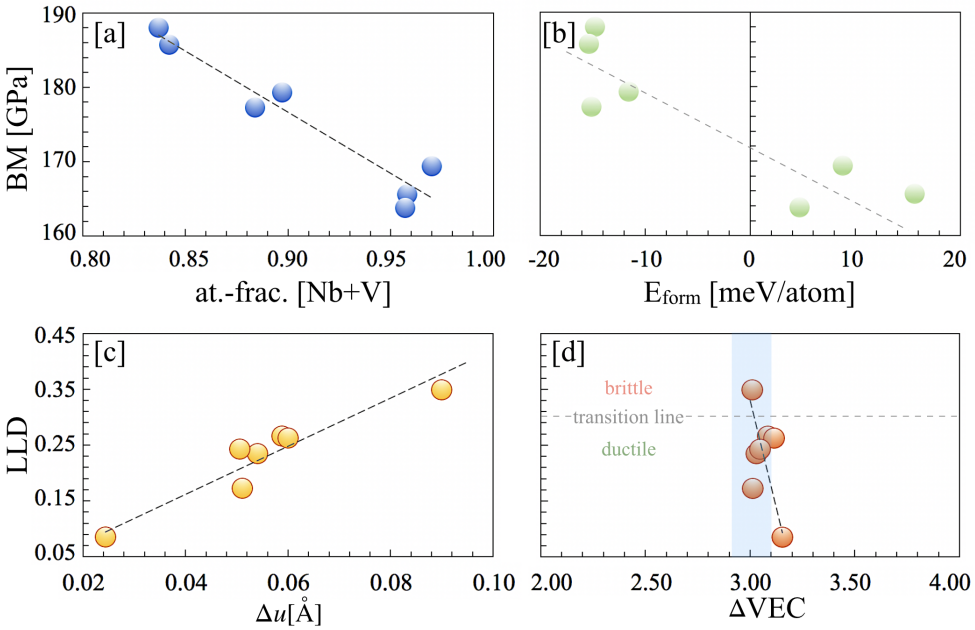

Figure 5.8: DFT analysis of thermodynamics and structural properties. (a,b) We plotted bulk-moduli with respect to atomic-fraction (Nb+V) and formation-enthalpy (E_{form}).The structural analysis of key mechanical properties responsible for ductility in Refractory MPEAs, (a) local-lattice distortion with respect to (c) static-displacement, and (d) valence electron count difference in bcc alloys.

elemental compositions can be tuned to alter local atomic displacement that correlates well with ductility in refractory MPEAs [188]. To provide more detail, we plot distortion metric [188] with respect to difference of VEC (ΔVEC) of bcc alloys in Fig. 5.8d. The dashed line horizontal to x-axis shows the separation between ductile and brittle materials based on the atomic distortion criterion. Clearly, highly ductile materials are expected to have low distortion as found in fcc alloys.

Furthermore, defects are very common to refractory alloys, which are detrimental for ductility. Notably, it has been reported that the presence of tungsten up to 5 at.% in refractory-based MPEAs [183] either slows down or stops the formation of new defects, moreover, high-vacancy migration energy of heavier elements also slows down the diffusion of defects. Although our framework was not optimized for defect related properties, the presence of small to moderate at.% molybdenum/tungsten further affirms that new designed compositions in Table 5.1 satisfy critical feature

requirements for ductility. This understanding of low defect concentration in tungsten based alloys can be helpful as the defect creation and propagation into crystal phase make them more fragile at higher strains, which is not desired for high-temperature applications.

The disorder has been shown to induce change in the DOS at/near electronic Fermi energy [189], which can be interpreted as the effect of disorder and electron correlation. The transition-metal-based refractory MPEAs are among the candidate alloys due to the presence of their partially filled d-bands. In Fig. 5.9a, we plot the value of total DOS (TDOS) at the Fermi-level. We found an interesting trend where #2 and #7, despite having the same TDOS values at the Fermi-level, show opposite trends in energy stability. To understand this contrast, we plot total density of states for both the alloys in full energy range (-6 eV to 3 eV). Despite being energetically stable #2 shows an unstable peak at the Fermi-level while #7 shows valley in density of states, and this feature in DOS has been found to correlate well with the alloy stability [190]. But we do not know if this small but finite change in DOS structure at the Fermi-level has any severe impact on thermodynamic behavior or charge fluctuations at finite temperature. The SRO analysis can give us useful information related to local change of chemical compositions, which is a critical aspect of ductile materials.

The local chemical fluctuations [191], if significantly large, may affect the synthesis of single phase alloy and mechanical properties, therefore, we believe understanding change in local response, i.e., short-range behavior (SRO), can give useful guideline regarding minimizing pronounced changes of chemical composition. The SRO theory [176] uses local atomic interaction to predict local chemical behavior, this is important as changes in chemical composition has a strong connection with interaction of alloying elements.

In Fig. 5.9c,d, we plot the SRO for #2 and #7 MPEAs to reflect on temperature dependent changes in their chemical behavior. Both the alloys in Fig. 5.9c-d show clustering trend below spinodal temperature (T_{sp}) driven by Mo-W pairs. An absolute instability to k_o mode [176] occurs below T_{sp}, where $[\alpha^{(-1)}(k_o; T_{sp})]^{IJ}_{\mu\nu}$=0. For k_o=(000), the alloy is unstable to segregation (atomic or vacancy clustering), rather than local ordering. Interestingly, we found that the #7 shows com-

peting ordering (B2 (H=111) type) mode while clustering (Γ) mode is slightly stronger, i.e., Γ = (000) peak compete with finite k_o peak. The most unstable SRO mode for both the alloys with large chemical fluctuation has the largest peak in $\alpha_{\mu\nu}(k_o; T > T_{sp})$ at wavevector k_o for a specific Mo-W pairs in the solid-solution phase.

The DOS values at Fermi energy for #2 and #7 MPEAs are almost same but their electronic-structure in Fig. 5.9b at the Fermi energy are quite different, where #2 has has small peaks. The finite peak at Fermi energy is the reason for thermodynamic instability towards clustering despite its energy stability (E_{form} (#2)=-12 meV-atom^{-1}). The prediction of clustering indicates weak interaction among alloying elements both at higher temperature and below T_{sp}, i.e., small or no charge fluctuation in disorder phase.

Except for design composition #7 in Table 5.1, the DFT analysis indicates the ductility for all other design compositions. The small lattice distortion, weak charge fluctuation, and possibility of low defect concentration in W based alloys [183] of predicted compositions is strong indication that the design framework is able to capture the useful structure-property trend in refractory MPEAs critical for GTEs application.

The machinability of refractory MPEAs is a major bottleneck in their development and deployment as structural materials in GTEs. However, with ICME approaches, GTE-amenable refractory MPEAs can be designed with ductility in mind. This can be achieved by optimizing for ductility indicators, such as the Pugh's Ratio and Cauchy Pressure, while under GTE relevant constraints such as the density and the solidus temperature. The Pugh's Ratio and Cauchy Pressure can be estimate with DFT methods; likewise, the density and solidus constraints can be queried by Thermo-Calc's thermodynamic equilibrium simulations, classifying points as feasible based on whether they satisfy both constraints. However, due to the vast nature of the MPEA design space and the high computational cost of these models, computational resources must be allocated in such a way that optimization of objectives is balanced with classification of meeting constraints.

The results from our framework show significant promise in the use of multi-objective and classification frameworks within an ICME methodology for materials design. Most notably, the

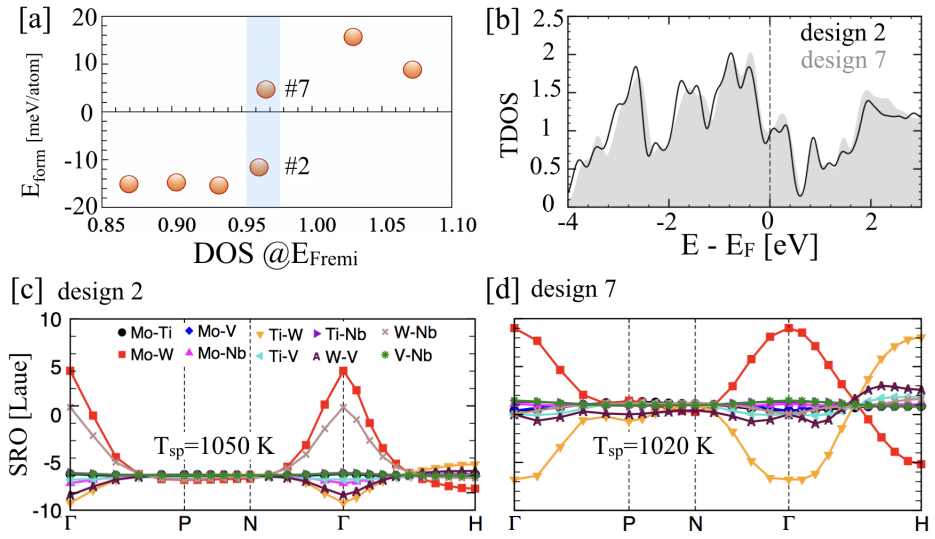

Figure 5.9: Electronic-structure and short-range order and analysis. (a) Total DOS at Fermi energy for MPEAs in Table 1. (b) The total density of states (DOS) plot for two key design compositions from the shaded region in (a), i.e., #2 and #7 . (b) The total density of states, and (c,d) short-range order for #2 and #7.

results showed that the proposed multi-information source BO framework is capable of efficiently exploring high-dimensional materials design spaces under multiple objective targets. Regarding classification, our novel framework (with 1,366 queries) is around 8 times faster than the traditional HTP "brute-force" approaches (with 10,626 queries), a huge improvement in terms of resource allocation.

While the proposed framework has been implemented and deployed in an *in silico* platform, the overall principle is truly source-agnostic. Each information source for both objectives and constraints is transformed into a Gaussian Process representation. Therefore, this approach can potentially be deployed for the efficient exploration and exploitation of materials spaces in physical closed-loop materials discovery platforms. With the exponential increase in interest of materials community towards autonomous materials discovery platforms, we believe that the frameworks like one proposed here will provide a useful direction to develop novel material discovery platforms.

6. CONCLUSIONS AND FUTURE WORK

In this dissertation, different frameworks have been developed and proposed to increase the efficiency of Bayesian optimization frameworks.

In chapter 2, a multi-information source Bayesian optimization is deployed to solve engineering design problems in presence of multiple sources of information. These sources are lower in accuracy or fidelity, but cheaper to evaluate with respect to the ground truth model. Since all these models are estimating the same quantity of interest, it is assumed that there exists some degree of correlation between these models and the ground truth. Using the reification process, the correlation between the information sources is estimated, then, by fusing the information obtained from different sources, a fused Gaussian process is obtained that enables accurate prediction of the ground truth quantity of interest. Knowledge gradient is employed as the utility function to evaluate possible knowledge gains of designs yet to be queried from a variety of information sources. In any iteration of the framework, an information source and a design that maximizes the utility function are chosen to be queried. The method was shown to be robust when subjected to various resource constraints that imposed restrictions on how the framework would query the expensive high-fidelity ground truth. It was found that these constraints played an important role in how different information sources were queried by the framework. The results demonstrate that in the case where we do not consider the actual computational cost of each information source and do not impose any budgetary constraint, the framework has a tendency to exhaustively explore the available information sources. On the other hand, when we impose tight budget constraints, the framework extensively extracts the information out of cheaper information sources and is still able to provide a reasonable prediction of the optimum design space.

In chapter 3, the multi-information source Bayesian optimization is extended to enable the optimization of multi-objective design problems. In realistic engineering problems, it is often the case that a set of objectives to be optimized simultaneously. However, the challenge is these objectives usually compete against each other and there is not a single solution to such design problems

optimizing all objectives at the same time instead, a region is considered as the solution set with no design superior to others. The methodology proposed in this chapter seeks to exploit all available information sources for efficiently identifying non-dominated points in the objective space as a means of estimating the true Pareto front. The approach was based on the fast evaluation of the expected hypervolume improvement through the use of temporarily updated Gaussian process surrogate models of each information source. The process also incorporates model reification to fuse new information rigorously as it becomes available through proper accounting for correlation between the sources. This study concludes that multi-information source Bayesian optimization approaches to directing efficient querying when the budget is constrained can be effective ways of estimating the Pareto front of a multi-objective problem. In particular, the ability to rapidly query lower-fidelity sources while accounting for their correlation with higher-fidelity sources and ground truth has enabled efficient (less than 10% of the budget for the problems studied here) identification of promising regions for non-dominated point searching. Then, improvement over the Pareto front estimation is shown when more information sources are available. Any information source can provide useful information about the quantity of interest that is not accessible from other information sources. Finally, the performance of the proposed approach is compared to two other well-known multi-objective optimization approaches, called ParEGO and NSGA-II. The results demonstrate the effectiveness of the current approach in budget-constrained situations.

Chapter 4 is discussing a dimensionality reduction technique to be employed within multi-information source Bayesian optimization frameworks. There, it is suggested strategies to equip a multifidelity Bayesian optimization framework with a gradient-based active subspace approach to address the issue of underperforming Bayesian optimization frameworks when the dimensionality of the design space increases. We have demonstrated this framework on a microstructure-sensitive design problem. Although employing a multifidelity Bayesian optimization approach alone results in less costly design procedures, exploring a high-dimensional design space could still be costly and thus, efficiency gains are desired in this regard. The results suggest that by taking advantage of incorporating the active subspace method in the multifidelity Bayesian optimization frameworks,

with fewer function evaluations, it is possible to obtain a better estimation of the optimal design faster. This is due to the active subspace method prioritizing searches in the important regions in the high-dimensional design space and representing the data on the lower dimensional active subspace that ease the curse of dimensionality problem. By investigating different strategies to use the active subspace method in a multifidelity Bayesian optimization framework, it has been shown that the Ground Truth and Individual Active Subspace approach performed better than the conventional multifidelity Bayesian optimization approach. Therefore, these two methods can be beneficial to reach a target objective value faster with larger initial steps toward the optimum design. While we have applied this approach to a very specific class of materials design problems, the framework has wider applicability, as it is often the case that in materials design problems only a small fraction of the degrees of freedom are active at any one time.

Finally, in chapter 5, a framework is developed to handle constrained optimization problems. In previous chapters, different frameworks are proposed suggesting more efficient approaches to designing complex systems. However, there were no constraints imposed in any of the engineering problems solved. In presence of design constraints, it is vital to recognize the boundaries to correctly determine the feasible regions and spend the computational resource more efficiently. This is addressed by developing a multifidelity framework to be implemented within optimization frameworks that can handle the constraints by learning the feasibility boundaries efficiently. Similar to a Bayesian optimization framework, a Bayesian classification framework is proposed that uses Gaussian process classification to model the constraint boundaries. Entropy measure is used as the utility function to search for the best information source and design to be queried for a more accurate boundary estimation. The decrease in entropy can be a sign of getting more accurate predictions of feasibility. Similar to multi-information source Bayesian optimization, the reification technique is used to fuse multiple information sources and obtain a fused classifier for a highly efficient estimation of constraint boundaries. The results show significant promise in the use of multi-objective and classification frameworks within an ICME methodology for materials design. Most notably, the results showed that the proposed multi-information source BO framework is ca-

pable of efficiently exploring high-dimensional materials design spaces under multiple objective targets.

In conclusion, in this study, it is shown how our multifidelity approach is capable of efficiently optimizing single and multi-objective functions. In the case of high-dimensional design space, where the Bayesian optimization frameworks tend to underperform, and a larger number of function evaluations is needed in search of optimal designs, we used the active subspace method to identify more important subspaces in the design space based on the objective function variation to suggest easier learning and searching process in machine learning tasks. Finally, we proposed a multifidelity Bayesian classification/optimization framework to handle constrained design problems by efficiently learning the constraint boundaries. Future works in this area can be extending the optimization frameworks to optimize over the problem space, for example, by evolving information sources as we learn more about the entire system and each information source. Another possibility is to develop multi-objective optimization frameworks to employ the active subspace method to quickly recognize the regions highly contributing to obtaining larger hypervolumes so that better Pareto fronts are discovered with less computational expenses. Finally, in constructing the Gaussian process for Bayesian optimization or Bayesian classification, a fixed kernel is used to estimate the correlation between data in the space, however, exploring kernel spaces makes it possible to choose a specific kernel considering the nature of the objective function at particular locations in the space.

REFERENCES

[1] A. Elham and M. J. van Tooren, "Coupled adjoint aerostructural wing optimization using quasi-three-dimensional aerodynamic analysis," *Structural and Multidisciplinary Optimization*, vol. 54, no. 4, pp. 889–906, 2016.

[2] R. Lam, K. Willcox, and D. H. Wolpert, "Bayesian optimization with a finite budget: An approximate dynamic programming approach," in *Advances in Neural Information Processing Systems*, pp. 883–891, 2016.

[3] W. B. Powell and I. O. Ryzhov, *Optimal learning*, vol. 841. John Wiley & Sons, 2012.

[4] P. Frazier, W. Powell, and S. Dayanik, "The knowledge-gradient policy for correlated normal beliefs," *INFORMS journal on Computing*, vol. 21, no. 4, pp. 599–613, 2009.

[5] D. Allaire and K. Willcox, "Fusing information from multifidelity computer models of physical systems," in *2012 15th international conference on information fusion*, pp. 2458–2465, IEEE, 2012.

[6] S. F. Ghoreishi, A. Molkeri, A. Srivastava, R. Arroyave, and D. Allaire, "Multi-information source fusion and optimization to realize icme: Application to dual-phase materials," *Journal of Mechanical Design*, vol. 140, no. 11, 2018.

[7] S. F. Ghoreishi, W. D. Thomison, and D. Allaire, "Sequential information-theoretic and reification-based approach for querying multi-information sources," *Journal of Aerospace Information Systems*, vol. 16, no. 12, pp. 575–587, 2019.

[8] W. D. Thomison and D. L. Allaire, "A model reification approach to fusing information from multifidelity information sources," in *19th AIAA non-deterministic approaches conference*, p. 1949, 2017.

[9] M. Clyde, "Model averaging." In Subjective and Objective Bayesian Statistics, 2nd ed., Chapter 13, Wiley-Interscience, 2003.

[10] M. Clyde and E. George, "Model uncertainty," *Statistical Science*, vol. 19, pp. 81–94, 2004.

[11] D. Draper, "Assessment and propagation of model uncertainty," *Journal of the Royal Statistical Society Series B*, vol. 57, no. 1, pp. 45–97, 1995.

[12] J. Hoeting, D. Madigan, A. Raftery, and C. Volinsky, "Bayesian model averaging: A tutorial," *Statistical Science*, vol. 14, no. 4, pp. 382–417, 1999.

[13] E. Leamer, *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York, NY: John Wiley & Sons, 1978.

[14] D. Madigan and A. Raftery, "Model selection and accounting for model uncertainty in graphical models using occam's window," *American Statistical Association*, vol. 89, no. 428, pp. 1535–1546, 1994.

[15] A. Mosleh and G. Apostolakis, "The assessment of probability distributions from expert opinions with an application to seismic fragility curves," *Risk Analysis*, vol. 6, no. 4, pp. 447–461, 1986.

[16] J. Reinert and G. Apostolakis, "Including model uncertainty in risk-informed decision making," *Annals of Nuclear Energy*, vol. 33, no. 4, pp. 354–369, 2006.

[17] M. Riley and R. Grandhi, "Quantification of modeling uncertainty in aeroelastic analyses," *Journal of Aircraft*, vol. 48, no. 3, pp. 866–873, 2011.

[18] E. Zio and G. Apostolakis, "Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories," *Reliability Engineering & System Safety*, vol. 54, no. 2-3, pp. 225–241, 1996.

[19] S. Julier and J. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations." In proceedings of the American Control Conference, pp. 2369-2373, 1997.

[20] S. Geisser, "A bayes approach for combining correlated estimates," *Journal of the American Statistical Association*, vol. 60, pp. 602–607, 1965.

[21] P. Morris, "Combining expert judgments: A bayesian approach," *Management Science*, vol. 23, pp. 679–693, 1977.

[22] R. Winkler, "Combining probability distributions from dependent information sources," *Management Science*, vol. 27, no. 4, pp. 479–488, 1981.

[23] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: new insights," *Structural and multidisciplinary optimization*, vol. 41, no. 6, pp. 853–862, 2010.

[24] I. Y. Kim and O. L. de Weck, "Adaptive weighted-sum method for bi-objective optimization: Pareto front generation," *Structural and multidisciplinary optimization*, vol. 29, no. 2, pp. 149–158, 2005.

[25] I. Das and J. E. Dennis, "Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems," *SIAM journal on optimization*, vol. 8, no. 3, pp. 631–657, 1998.

[26] N. Beume, "S-metric calculation by considering dominated hypervolume as klee's measure problem," *Evolutionary Computation*, vol. 17, no. 4, pp. 477–492, 2009.

[27] L. Bradstreet, L. While, and L. Barone, "A fast many-objective hypervolume algorithm using iterated incremental calculations," in *IEEE Congress on Evolutionary Computation*, pp. 1–8, IEEE, 2010.

[28] M. T. Emmerich, A. H. Deutz, and J. W. Klinkenberg, "Hypervolume-based expected improvement: Monotonicity properties and exact computation," in *2011 IEEE Congress of Evolutionary Computation (CEC)*, pp. 2147–2154, IEEE, 2011.

[29] C. M. Fonseca, L. Paquete, and M. López-Ibánez, "An improved dimension-sweep algorithm for the hypervolume indicator," in *2006 IEEE international conference on evolutionary computation*, pp. 1157–1163, IEEE, 2006.

[30] L. M. Russo and A. P. Francisco, "Quick hypervolume," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 481–502, 2013.

[31] Q. Yang and S. Ding, "Novel algorithm to calculate hypervolume indicator of pareto approximation set," in *International Conference on Intelligent Computing*, pp. 235–244, Springer, 2007.

[32] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach," *IEEE transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.

[33] S. F. Ghoreishi and D. Allaire, "Multi-information source constrained bayesian optimization," *Structural and Multidisciplinary Optimization*, vol. 59, no. 3, pp. 977–991, 2019.

[34] S. F. Ghoreishi and D. L. Allaire, "Gaussian process regression for bayesian fusion of multi-fidelity information sources," in *2018 Multidisciplinary Analysis and Optimization Conference*, p. 4176, 2018.

[35] R. Lam, D. L. Allaire, and K. E. Willcox, "Multifidelity optimization using statistical surrogate modeling for non-hierarchical information sources," in *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, p. 0143, 2015.

[36] D. T. Fullwood, S. R. Niezgoda, B. L. Adams, and S. R. Kalidindi, "Microstructure sensitive design for performance optimization," *Progress in Materials Science*, vol. 55, pp. 477–562, Aug. 2010.

[37] D. Khatamsaz, A. Molkeri, R. Couperthwaite, J. James, R. Arróyave, D. Allaire, and A. Srivastava, "Efficiently exploiting process-structure-property relationships in material design by multi-information source fusion," *Acta Materialia*, p. 116619, 2021.

[38] E. Raponi, H. Wang, M. Bujny, S. Boria, and C. Doerr, "High dimensional bayesian optimization assisted by principal component analysis," in *International Conference on Parallel Problem Solving from Nature*, pp. 169–183, Springer, 2020.

[39] A. Saltelli, "Sensitivity analysis for importance assessment," *Risk analysis*, vol. 22, no. 3, pp. 579–590, 2002.

[40] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity analysis in practice: a guide to assessing scientific models*, vol. 1. Wiley Online Library, 2004.

[41] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[42] H. Christopher Frey and S. R. Patil, "Identification and review of sensitivity analysis methods," *Risk analysis*, vol. 22, no. 3, pp. 553–578, 2002.

[43] D. L. Allaire and K. E. Willcox, "A variance-based sensitivity index function for factor prioritization," *Reliability Engineering & System Safety*, vol. 107, pp. 107–114, 2012.

[44] G. H. Dunteman, *Principal components analysis*. No. 69, Sage, 1989.

[45] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[46] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.

[47] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[48] T. M. Russi, *Uncertainty quantification with experimental data and complex system models*. PhD thesis, UC Berkeley, 2010.

[49] P. G. Constantine, E. Dow, and Q. Wang, "Active subspace methods in theory and practice: applications to kriging surfaces," *SIAM Journal on Scientific Computing*, vol. 36, no. 4, pp. A1500–A1524, 2014.

[50] S. F. Ghoreishi, S. Friedman, and D. L. Allaire, "Adaptive dimensionality reduction for fast sequential optimization with gaussian processes," *Journal of Mechanical Design*, vol. 141, no. 7, 2019.

[51] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[52] J. Allison, "Integrated computational materials engineering: A perspective on progress and future steps," *JOM*, vol. 63, no. 4, pp. 15–18, 2011.

[53] W. Y. Wang, B. Tang, D. Lin, C. Zou, Y. Zhang, S.-L. Shang, Q. Guan, J. Gao, L. Fan, H. Kou, *et al.*, "A brief review of data-driven icme for intelligently discovering advanced structural metal materials: Insight into atomic and electronic building blocks," *Journal of Materials Research*, vol. 35, no. 8, pp. 872–889, 2020.

[54] B. Gautham, R. Kumar, S. Bothra, G. Mohapatra, N. Kulkarni, and K. Padmanabhan, "More efficient icme through materials informatics and process modeling," in *Proceedings of the 1st World Congress on Integrated Computational Materials Engineering (ICME)*, p. 35, Wiley Online Library, 2011.

[55] M. F. Horstemeyer, *Integrated Computational Materials Engineering (ICME) for metals: using multiscale modeling to invigorate engineering design with science*. John Wiley & Sons, 2012.

[56] R. Arróyave and D. L. McDowell, "Systems approaches to materials design: past, present, and future," *Annual Review of Materials Research*, vol. 49, pp. 103–126, 2019.

[57] D. L. McDowell and S. R. Kalidindi, "The materials innovation ecosystem: a key enabler for the materials genome initiative," *Mrs Bulletin*, vol. 41, no. 4, p. 326, 2016.

[58] P. Voorhees, G. Spanos, *et al.*, "Modeling across scales: a roadmapping study for connecting materials models and simulations across length and time scales," *TMS, Warrendale, PA*, vol. 14, 2015.

[59] V. Savic, L. Hector, U. Basu, A. Basudhar, I. Gandikota, N. Stander, T. Park, F. Pourboghrat, K. S. Choi, X. Sun, *et al.*, "Integrated computational materials engineering (ICME) multi-scale model development for advanced high strength steels," tech. rep., SAE Technical Paper, 2017.

[60] L. Lin and W. Ren, "An implementation of icme in materials information exchanging interfaces," *Materials discovery*, vol. 12, pp. 9–19, 2018.

[61] A. Talapatra, S. Boluki, T. Duong, X. Qian, E. Dougherty, and R. Arróyave, "Autonomous efficient experiment design for materials discovery with bayesian model averaging," *Physical Review Materials*, vol. 2, no. 11, p. 113803, 2018.

[62] A. Solomou, G. Zhao, S. Boluki, J. K. Joy, X. Qian, I. Karaman, R. Arróyave, and D. C. Lagoudas, "Multi-objective bayesian materials discovery: Application on the discovery of precipitation strengthened niti shape memory alloys through micromechanical modeling," *Materials & Design*, vol. 160, pp. 810–827, 2018.

[63] B. Peherstorfer, K. Willcox, and M. Gunzburger, "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *Siam Review*, vol. 60, no. 3, pp. 550–591, 2018.

[64] S. F. Ghoreishi, A. Molkeri, R. Arróyave, D. Allaire, and A. Srivastava, "Efficient use of multiple information sources in material design," *Acta Materialia*, vol. 180, pp. 260–271, 2019.

[65] P. I. Frazier, W. B. Powell, and S. Dayanik, "A knowledge-gradient policy for sequential information collection," *SIAM Journal on Control and Optimization*, vol. 47, no. 5, pp. 2410–2439, 2008.

[66] R. Couperthwaite, D. Allaire, and R. Arróyave, "Utilizing gaussian processes to fit high dimension thermodynamic data that includes estimated variability," *Computational Materials Science*, p. 110133, 2020.

[67] D. Koistinen and R. Marburger, "A general equation prescribing the extent of the austenite-martensite transformation in pure iron-carbon alloys and plain carbon steels," *Acta Metallurgica*, vol. 7, pp. 59–60, Jan. 1959.

[68] A. Srivastava, A. Bower, L. Hector Jr, J. Carsley, L. Zhang, and F. Abu-Farha, "A multiscale approach to modeling formability of dual-phase steels," *Modelling and Simulation in Materials Science and Engineering*, vol. 24, no. 2, p. 025011, 2016.

[69] A. Srivastava, H. Ghassemi-Armaki, H. Sung, P. Chen, S. Kumar, and A. F. Bower, "Micromechanics of plastic deformation and phase transformation in a three-phase trip-assisted advanced high strength steel: Experiments and modeling," *Journal of the Mechanics and Physics of Solids*, vol. 78, pp. 46–69, 2015.

[70] D. Gerbig, A. Srivastava, S. Osovski, L. G. Hector, and A. Bower, "Analysis and design of dual-phase steel microstructure for enhanced ductile fracture resistance," *International Journal of Fracture*, pp. 1–24, 2017.

[71] W. Voigt, "On the relation between the elasticity constants of isotropic bodies," *Ann. Phys. Chem*, vol. 274, pp. 573–587, 1889.

[72] A. Reuss, "Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle.," *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 9, no. 1, pp. 49–58, 1929.

[73] O. Bouaziz and P. Buessler, "Mechanical behaviour of multiphase materials: an intermediate mixture law without fitting parameter," *Revue de Métallurgie–International Journal of Metallurgy*, vol. 99, no. 1, pp. 71–77, 2002.

[74] G. Weng, "The overall elastoplastic stress-strain relations of dual-phase metals," *Journal of the Mechanics and Physics of Solids*, vol. 38, no. 3, pp. 419–441, 1990.

[75] D. M. Saylor, J. Fridy, B. S. El-Dasher, K.-Y. Jung, and A. D. Rollett, "Statistically representative three-dimensional microstructures based on orthogonal observation sections," *Metallurgical and Materials Transactions A*, vol. 35, no. 7, pp. 1969–1979, 2004.

[76] Dassault Systemes, Vélizy-Villacoublay, France, *ABAQUS user's manual*, 2017.

[77] R. L. Winkler, "Combining probability distributions from dependent information sources," *Management Science*, vol. 27, no. 4, pp. 479–488, 1981.

[78] C. Villani, "The Wasserstein distances," in *Optimal Transport: Old and New* (C. Villani, ed.), pp. 93–111, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[79] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, pp. 34–41, Aug. 2001.

[80] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth Mover's Distance as a Metric for Image Retrieval," *International Journal of Computer Vision*, vol. 40, pp. 99–121, Nov. 2000.

[81] M.-F. Chen, "Trilogy of Couplings and General Formulas for Lower Bound of Spectral Gap," in *Probability Towards 2000* (L. Accardi and C. C. Heyde, eds.), pp. 123–136, New York, NY: Springer New York, 1998.

[82] Y. Ollivier, "Ricci curvature of markov chains on metric spaces," 2007.

[83] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks," in *Proceedings of Machine Learning Research* (D. Precup and Y. W. Teh, eds.), vol. 70, (Proceedings of Machine Learning Research), pp. 214–223, PMLR, 2017.

[84] D. Khatamsaz, L. Peddareddygari, S. Friedman, and D. Allaire, "Bayesian optimization of multiobjective functions using multiple information sources," *AIAA Journal*, pp. 1–11, 2021.

[85] A. Ariyarit and M. Kanazaki, "Multi-fidelity multi-objective efficient global optimization applied to airfoil design problems," *Applied Sciences*, vol. 7, no. 12, p. 1318, 2017.

[86] A. Amrit and L. Leifsson, "Applications of surrogate-assisted and multi-fidelity multi-objective optimization algorithms to simulation-based aerodynamic design," *Engineering Computations*, 2019.

[87] L. Leifsson, S. Koziel, and Y. A. Tesfahunegn, "Multiobjective aerodynamic optimization by variable-fidelity models and response surface surrogates," *AIAA Journal*, vol. 54, no. 2, pp. 531–541, 2016.

[88] C. A. C. Coello and G. B. Lamont, *Applications of multi-objective evolutionary algorithms*, vol. 1. World Scientific, 2004.

[89] C. M. M. d. Fonseca, *Multiobjective genetic algorithms with application to control engineering problems.* PhD thesis, University of Sheffield, 1995.

[90] A. Amrit, L. Leifsson, and S. Koziel, "Design strategies for multi-objective optimization of aerodynamic surfaces," *Engineering Computations*, 2017.

[91] D. Allaire and K. Willcox, "Surrogate modeling for uncertainty assessment with application to aviation environmental system models," *AIAA journal*, vol. 48, no. 8, pp. 1791–1803, 2010.

[92] G. Zhao, R. Arroyave, and X. Qian, "Fast exact computation of expected hypervolume improvement," 2018.

[93] N. Alexandrov, R. Lewis, C. Gumbert, L. Green, and P. Newman, "Approximation and model management in aerodynamic optimization with variable-fidelity models," *AIAA Journal*, vol. 38, pp. 1093–1101, November-December 2001.

[94] A. J. Booker, J. E. Dennis, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset, "A rigorous framework for optimization of expensive functions by surrogates," *Structural optimization*, vol. 17, no. 1, pp. 1–13, 1999.

[95] D. Allaire and K. Willcox, "A mathematical and computational framework for multifidelity design and analysis with computer models," *International Journal for Uncertainty Quantification*, vol. 4, no. 1, 2014.

[96] F. Archetti and A. Candelieri, *Bayesian Optimization and Data Science*. Springer, 2019.

[97] J. Močkus, "On bayesian methods for seeking the extremum," in *Optimization techniques IFIP technical conference*, pp. 400–404, Springer, 1975.

[98] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[99] P. I. Frazier, "A tutorial on bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

[100] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, pp. 2951–2959, 2012.

[101] I. Couckuyt, D. Deschrijver, and T. Dhaene, "Fast calculation of multiobjective probability of improvement and expected improvement criteria for pareto optimization," *Journal of Global Optimization*, vol. 60, no. 3, pp. 575–594, 2014.

[102] P. Feliot, J. Bect, and E. Vazquez, "A bayesian approach to constrained single-and multi-objective optimization," *Journal of Global Optimization*, vol. 67, no. 1-2, pp. 97–133, 2017.

[103] L. While, L. Bradstreet, and L. Barone, "A fast way of calculating exact hypervolumes," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 1, pp. 86–95, 2011.

[104] A. Jaszkiewicz, "Improved quick hypervolume algorithm," *Computers & Operations Research*, vol. 90, pp. 72–83, 2018.

[105] D. Khatamsaz, L. Peddareddygari, S. Friedman, and D. L. Allaire, "Efficient multi-information source multiobjective bayesian optimization," in *AIAA Scitech 2020 Forum*, p. 2127, 2020.

[106] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable multi-objective optimization test problems," in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*, vol. 1, pp. 825–830, IEEE, 2002.

[107] J. P. Jasa, J. T. Hwang, and J. R. Martins, "Open-source coupled aerostructural optimization using python," *Structural and Multidisciplinary Optimization*, vol. 57, no. 4, pp. 1815–1827, 2018.

[108] C. Poloni, A. Giurgevich, L. Onesti, and V. Pediroda, "Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimizer for a complex design problem in fluid dynamics," *Computer Methods in Applied Mechanics and Engineering*, vol. 186, no. 2-4, pp. 403–420, 2000.

[109] M. J. Riley, T. Peachey, D. Abramson, K. W. Jenkins, *et al.*, "Multi-objective engineering shape optimization using differential evolution interfaced to the nimrod/o tool," in *IOP Conf. Ser. Mater. Sci. Eng*, vol. 10, p. 012189, 2010.

[110] V. Kelner and O. Léonard, "Application of genetic algorithms to lubrication pump stacking design," *Journal of Computational and Applied Mathematics*, vol. 168, no. 1-2, pp. 255–265, 2004.

[111] J. S. Gray, J. T. Hwang, J. R. R. A. Martins, K. T. Moore, and B. A. Naylor, "Openmdao: an open-source framework for multidisciplinary design, analysis, and optimization," *Structural and Multidisciplinary Optimization*, vol. 59, pp. 1075–1104, Apr 2019.

[112] S. S. Chauhan and J. R. Martins, "Low-fidelity aerostructural optimization of aircraft wings with a simplified wingbox model using openaerostruct," in *International Conference on Engineering Optimization*, pp. 418–431, Springer, 2018.

[113] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii," in *International conference on parallel problem solving from nature*, pp. 849–858, Springer, 2000.

[114] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[115] Dr.S.Baskar, S. Tamilselvi and P.R.Varshini, "Matlab code for constrained nsga ii," 2020. `https://www.mathworks.com/matlabcentral/fileexchange\/49806-matlab-code-for-constrained-nsga-ii-dr-s-baskar-s-tamilselvi-an`

[116] J. Davins-Valldaura, S. Moussaoui, G. Pita-Gil, and F. Plestan, "Parego extensions for multi-objective optimization of expensive evaluation functions," *Journal of Global Optimization*, vol. 67, no. 1-2, pp. 79–96, 2017.

[117] J. Knowles, "Parego: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 50–66, 2006.

[118] D. Khatamsaz, A. Molkeri, R. Couperthwaite, J. James, R. Arróyave, A. Srivastava, and D. Allaire, "Adaptive active subspace-based efficient multifidelity materials design," *Materials & Design*, vol. 209, p. 110001, 2021.

[119] L. Johnson and R. Arróyave, "An inverse design framework for prescribing precipitation heat treatments from a target microstructure," *Materials & Design*, vol. 107, pp. 7–17, Oct. 2016.

[120] Q. Zhao, H. Yang, J. Liu, H. Zhou, H. Wang, and W. Yang, "Machine learning-assisted discovery of strong and conductive Cu alloys: Data mining from discarded experiments and physical features," *Materials & Design*, vol. 197, p. 109248, Jan. 2021.

[121] B. Adams, A. Henrie, B. Henrie, M. Lyon, S. Kalidindi, and H. Garmestani, "Microstructure-sensitive design of a compliant beam," *Journal of the Mechanics and Physics of Solids*, vol. 49, pp. 1639–1663, Aug. 2001.

[122] S. R. Kalidindi, J. R. Houskamp, M. Lyons, and B. L. Adams, "Microstructure sensitive design of an orthotropic plate subjected to tensile load," *International Journal of Plasticity*, vol. 20, pp. 1561–1575, Aug. 2004.

[123] G. Saheli, H. Garmestani, and B. L. Adams, "Microstructure design of a two phase composite using two-point correlation functions," *Journal of Computer-Aided Materials Design*, vol. 11, pp. 103–115, Jan. 2004.

[124] T. Fast, M. Knezevic, and S. R. Kalidindi, "Application of microstructure sensitive design to structural components produced from hexagonal polycrystalline metals," *Computational Materials Science*, vol. 43, pp. 374–383, Aug. 2008.

[125] S. Allain, O. Bouaziz, I. Pushkareva, and C. Scott, "Towards the microstructure design of DP steels: A generic size-sensitive mean-field mechanical model," *Materials Science and Engineering: A*, vol. 637, pp. 222–234, June 2015.

[126] R. Couperthwaite, A. Molkeri, D. Khatamsaz, A. Srivastava, D. Allaire, and R. Arròyave, "Materials design through batch bayesian optimization with multisource information fusion," *JOM*, vol. 72, no. 12, pp. 4431–4443, 2020.

[127] W. Jiang, W. Liao, T. Liu, X. Shi, C. Wang, J. Qi, Y. Chen, Z. Wang, and C. Zhang, "A voxel-based method of multiscale mechanical property optimization for the design of graded tpms structures," *Materials & Design*, p. 109655, 2021.

[128] X. Chen, H. Zhou, and Y. Li, "Effective design space exploration of gradient nanostructured materials using active learning based surrogate models," *Materials & Design*, vol. 183, p. 108085, 2019.

[129] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*. SIAM, 2019.

[130] J. Zhang, S. Bi, and G. Zhang, "A directional gaussian smoothing optimization method for computational inverse design in nanophotonics," *Materials & Design*, vol. 197, p. 109213, 2021.

[131] A. C. Antoulas, *Approximation of large-scale dynamical systems*. SIAM, 2005.

[132] K. Zhou, J. C. Doyle, K. Glover, *et al.*, *Robust and optimal control*, vol. 40. Prentice hall New Jersey, 1996.

[133] C. Kunselman, S. Sheikh, M. Mikkelsen, V. Attari, and R. Arróyave, "Microstructure classification in the unsupervised context," *Acta Materalia*, 2020.

[134] S. R. Kalidindi, S. R. Niezgoda, and A. A. Salem, "Microstructure informatics using higher-order statistics and efficient data-mining protocols," *Jom*, vol. 63, no. 4, pp. 34–41, 2011.

[135] A. Cecen, H. Dai, Y. C. Yabansu, S. R. Kalidindi, and L. Song, "Material structure-property linkages using three-dimensional convolutional neural networks," *Acta Materialia*, vol. 146, pp. 76 – 84, 2018.

[136] J. H. Panchal, S. R. Kalidindi, and D. L. McDowell, "Key computational modeling issues in integrated computational materials engineering," *Computer-Aided Design*, vol. 45, no. 1, pp. 4 – 25, 2013. Computer-aided multi-scale materials and product design.

[137] A. Paul, P. Acar, W. keng Liao, A. Choudhary, V. Sundararaghavan, and A. Agrawal, "Microstructure optimization with constrained design objectives using machine learning-based feedback-aware data-generation," *Computational Materials Science*, vol. 160, pp. 334 – 351, 2019.

[138] J. Jung, J. I. Yoon, H. K. Park, J. Y. Kim, and H. S. Kim, "Bayesian approach in predicting mechanical properties of materials: Application to dual phase steels," *Materials Science and Engineering: A*, vol. 743, pp. 382 – 390, 2019.

[139] Z. Li, B. Wen, and N. Zabaras, "Computing mechanical response variability of polycrystalline microstructures through dimensionality reduction techniques," *Computational Materials Science*, vol. 49, no. 3, pp. 568 – 581, 2010.

[140] X. Hu, J. Li, Z. Wang, and J. Wang, "A microstructure-informatic strategy for Vickers hardness forecast of austenitic steels from experimental data," *Materials & Design*, vol. 201, p. 109497, Mar. 2021.

[141] S. F. Ghoreishi and D. L. Allaire, "A fusion-based multi-information source optimization approach using knowledge gradient policies," in *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, p. 1159, 2018.

[142] P. Liu, Y. Yan, X. Zhang, Y. Luo, and Z. Kang, "Topological design of microstructures using periodic material-field series-expansion and gradient-free optimization algorithm," *Materials & Design*, vol. 199, p. 109437, 2021.

[143] D. Khatamsaz, B. Vela, P. Singh, D. D. Johnson, D. Allaire, and R. Arróyave, "Multi-objective materials bayesian optimization with active learning of design constraints: Design of ductile refractory multi-principal-element alloys," *Acta Materialia*, p. 118133, 2022.

[144] T. M. Pollock, "Alloy design for aircraft engines," *Nature materials*, vol. 15, no. 8, pp. 809–815, 2016.

[145] G. Andrews, "16 - ultra-low nitrogen oxides (nox) emissions combustion in gas turbine systems," in *Modern Gas Turbine Systems* (P. Jansohn, ed.), Woodhead Publishing Series in Energy, pp. 715–790, Woodhead Publishing, 2013.

[146] R. C. Reed, *The superalloys: fundamentals and applications*. Cambridge university press, 2008.

[147] F. N. Nourin and R. S. Amano, "Review of Gas Turbine Internal Cooling Improvement Technology," *Journal of Energy Resources Technology*, vol. 143, 11 2020. 080801.

[148] S. Wee, J. Do, K. Kim, C. Lee, C. Seok, B.-G. Choi, Y. Choi, and W. Kim, "Review on mechanical thermal properties of superalloys and thermal barrier coating used in gas turbines," *Applied Sciences*, vol. 10, no. 16, 2020.

[149] O. N. Senkov, D. B. Miracle, K. J. Chaput, and J.-P. Couzinie, "Development and exploration of refractory high entropy alloys—a review," *Journal of Materials Research*, vol. 33, no. 19, p. 3092–3128, 2018.

[150] O. N. Senkov, D. Isheim, D. N. Seidman, and A. L. Pilchak, "Development of a refractory high entropy superalloy," *Entropy*, vol. 18, no. 3, 2016.

[151] U. Bhandari, C. Zhang, and S. Yang, "Mechanical and thermal properties of low-density al20+xcr20-xmo20-yti20v20+y alloys," *Crystals*, vol. 10, no. 4, 2020.

[152] D. Wang, J. Tan, C. Li, X. Qin, and S. Guo, "Enhanced creep resistance of ti30al25zr25nb20 high-entropy alloy at room temperature," *Journal of Alloys and Compounds*, vol. 885, p. 161038, 2021.

[153] K.-C. Lo, Y.-J. Chang, H. Murakami, J.-W. Yeh, and A.-C. Yeh, "An oxidation resistant refractory high entropy alloy protected by crtao4-based oxide," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[154] N. Philips, M. Carl, and N. Cunningham, "New opportunities in refractory alloys," *Metallurgical and Materials Transactions A*, vol. 51, no. 7, pp. 3299–3310, 2020.

[155] S. Sheikh, S. Shafeie, Q. Hu, J. Ahlström, C. Persson, J. Veselỳ, J. Zỳka, U. Klement, and S. Guo, "Alloy design for intrinsically ductile refractory high-entropy alloys," *Journal of applied physics*, vol. 120, no. 16, p. 164902, 2016.

[156] R.-R. Griffiths and J. M. Hernández-Lobato, "Constrained bayesian optimization for automatic chemical design using variational autoencoders," *Chemical science*, vol. 11, no. 2, pp. 577–586, 2020.

[157] K. M. Jablonka, G. M. Jothiappan, S. Wang, B. Smit, and B. Yoo, "Bias free multiobjective active learning for materials design and discovery," *Nature communications*, vol. 12, no. 1, pp. 1–10, 2021.

[158] L. Qi and D. Chrzan, "Tuning ideal tensile strengths and intrinsic ductility of bcc refractory alloys," *Physical review letters*, vol. 112, no. 11, p. 115503, 2014.

[159] S. M. Shaikh, V. Hariharan, S. K. Yadav, and B. Murty, "Calphad and rule-of-mixtures: A comparative study for refractory high entropy alloys," *Intermetallics*, vol. 127, p. 106926, 2020.

[160] L. Chen, X. Zhang, Y. Wang, X. Hao, and H. Liu, "Microstructure and elastic constants of AlTiVMoNb refractory high-entropy alloy coating on ti6al4v by laser cladding," *Materials Research Express*, vol. 6, p. 116571, oct 2019.

[161] Y. Ye, B. Musico, Z. Lu, L. Xu, Z. Lei, V. Keppens, H. Xu, and T. Nieh, "Evaluating elastic properties of a body-centered cubic nbhfzrti high-entropy alloy – a direct comparison between experiments and ab initio calculations," *Intermetallics*, vol. 109, pp. 167–173, 2019.

[162] S. Pugh, "Xcii. relations between the elastic moduli and the plastic properties of polycrystalline pure metals," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 45, no. 367, pp. 823–843, 1954.

[163] D. G. Pettifor, "Theoretical predictions of structure and related properties of intermetallics," *Materials Science and Technology*, vol. 8, no. 4, pp. 345–349, 1992.

[164] S. Ghoreishi, W. Thomison, and D. Allaire, "Sequential information-theoretic and reification-based approach for querying multi-information sources," *Journal of Aerospace Information Systems*, vol. 16, no. 12, pp. 575–587, 2019.

[165] F. S. Costabal, P. Perdikaris, E. Kuhl, and D. E. Hurtado, "Multi-fidelity classification using gaussian processes: accelerating the prediction of large-scale computational models," *Computer Methods in Applied Mechanics and Engineering*, vol. 357, p. 112602, 2019.

[166] V. Balabanov, R. Haftka, B. Grossman, W. Mason, and L. Watson, "Multifidelity response surface model for HSCT wing bending material weight," in *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, (St. Louis, MO), September 2-4 1998.

[167] V. Balabanov and G. Venter, "Multi-fidelity optimization with high-fidelity analysis and low-fidelity gradients," in *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, (Albany, New York), August 30-September 1, 2004.

[168] S. Choi, J. J. Alonso, and I. M. Kroo, "Two-level multifidelity design optimization studies for supersonic jets," *Journal of Aircraft*, vol. 46, no. 3, pp. 776–790, 2009.

[169] M. Eldred, A. Giunta, and S. Collis, "Second-order corrections for surrogate-based optimization with model hierarchies." 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2004.

[170] A. March and K. Willcox, "Convergent multifidelity optimization using Bayesian model calibration," *Structural and Multidisciplinary Optimization*, vol. 46, no. 1, pp. 93–109, 2012.

[171] A. March and K. Willcox, "Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives," *AIAA Journal*, vol. 50, no. 5, pp. 1079–1089, 2012.

[172] D. Khatamsaz and D. L. Allaire, "A comparison of reification and cokriging for sequential multi-information source fusion," in *AIAA Scitech 2021 Forum*, p. 1477, 2021.

[173] A. J. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2372–2379, IEEE, 2009.

[174] D. D. Johnson, D. M. Nicholson, F. J. Pinski, B. L. Gyorffy, and G. M. Stocks, "Density-functional theory for random alloys: Total energy within the coherent-potential approximation," *Phys. Rev. Lett.*, vol. 56, pp. 2088–2091, May 1986.

[175] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, Oct 1996.

[176] P. Singh, A. V. Smirnov, and D. D. Johnson, "Atomic short-range order and incipient long-range order in high-entropy alloys," *Phys. Rev. B*, vol. 91, p. 224204, Jun 2015.

[177] P. Singh, S. Picak, A. Sharma, Y. I. Chumlyakov, R. Arroyave, I. Karaman, and D. D. Johnson, "Martensitic transformation in $\mathrm{fe}_x\mathrm{mn}_{80-x}\mathrm{co}_{10}\mathrm{cr}_{10}$ high-entropy alloy," *Phys. Rev. Lett.*, vol. 127, p. 115704, Sep 2021.

[178] P. Singh, A. Sharma, A. V. Smirnov, M. S. Diallo, P. K. Ray, G. Balasubramanian, and D. D. Johnson, "Design of high-strength refractory complex solid-solution alloys," *npj Computational Materials*, vol. 4, p. 16, Mar 2018.

[179] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Phys. Rev. B*, vol. 54, pp. 11169–11186, Oct 1996.

[180] G. Kresse and J. Hafner, "Ab initio molecular dynamics for liquid metals," *Phys. Rev. B*, vol. 47, pp. 558–561, Jan 1993.

[181] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," *Computational Materials Science*, vol. 6, no. 1, pp. 15–50, 1996.

[182] R. Singh, A. Sharma, P. Singh, G. Balasubramanian, and D. D. Johnson, "Accelerating computational modeling and design of high-entropy alloys," *Nature Computational Science*, vol. 1, pp. 54–61, Jan 2021.

[183] A. Roy, P. Singh, G. Balasubramanian, and D. D. Johnson, "Vacancy formation energies and migration barriers in multi-principal element alloys," *Acta Materialia*, vol. 226, p. 117611, 2022.

[184] H. J. Monkhorst and J. D. Pack, "Special points for brillouin-zone integrations," *Phys. Rev. B*, vol. 13, pp. 5188–5192, Jun 1976.

[185] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Phys. Rev. B*, vol. 59, pp. 1758–1775, Jan 1999.

[186] P. E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B*, vol. 50, pp. 17953–17979, Dec 1994.

[187] A. Abu-Odeh, E. Galvan, T. Kirk, H. Mao, Q. Chen, P. Mason, R. Malak, and R. Arróyave, "Efficient exploration of the high entropy alloy composition-phase space," *Acta Materialia*, vol. 152, pp. 41–57, 2018.

[188] P. Singh, R. Arroyave, and D. Johnson, "A structural signature for ductility in chemically complex alloys," *in review (submitted)*, 2022.

[189] B. Altshuler and A. Aronov, "Zero bias anomaly in tunnel resistance and electron-electron interaction," *Solid State Communications*, vol. 30, no. 3, pp. 115–117, 1979.

[190] P. Singh, A. V. Smirnov, and D. D. Johnson, "Ta-nb-mo-w refractory high-entropy alloys: Anomalous ordering behavior and its intriguing electronic origin," *Phys. Rev. Materials*, vol. 2, p. 055004, May 2018.

[191] T. Mukherjee, J. S. Zuback, A. De, and T. DebRoy, "Printability of alloys for additive manufacturing," *Scientific Reports*, vol. 6, p. 19717, Jan 2016.