Multi-Information Source Optimization of a Dual-Phase Microstructure for Optimum Strength and Strain Hardenability

Seyede Fatemeh Ghoreishi,
Graduate Research Assistant
Dept. of Mechanical Eng.
Texas A&M University
College Station, TX 77843
Email: f.ghoreishi88@tamu.edu

Abhilash Molkeri
Graduate Research Assistant
Dept. of Materials Science and Eng.
Texas A&M University
College Station, TX 77843
Email: abhilashmolkeri@tamu.edu

Ankit Srivastava
Assistant Professor
Dept. of Materials Science and Eng.
Texas A&M University
College Station, TX 77843
Email: ankit.sri@tamu.edu

Raymundo Arroyave
Professor
Dept. of Materials Science and Eng.
Texas A&M University
College Station, TX 77843
Email: rarroyave@tamu.edu

Douglas Allaire*
Assistant Professor
Dept. of Mechanical Eng.
Texas A&M University
College Station, TX 77843
Email: dallaire@tamu.edu

The Materials Genome Initiative (MGI) calls for the acceleration of the materials development cycle through the combination of experiments, simulation, and data. This paper proposes a materials design framework for addressing this call. Specifically, a multi-information source optimization framework is proposed that identifies the next best information source to query and where in the input space to query it by following a novel value-gradient policy. The querying decision takes into account the ability to learn correlations between information sources, the resource cost of querying an information source, and what a query is expected to provide in terms of improvement over the current state. The framework is demonstrated on the optimization of a ground truth quantity of interest, which here is strength normalized strain hardening rate of a ductile dual-phase material. The ground truth is represented by a microstructure-based finite element model and three lower fidelity information sources, including an isostrain, isostress, and isowork model are available for identifying the ground truth optimum.

*Address all correspondence to this author.

1 Introduction

Over the past two decades, there has been considerable interest in the development of frameworks to accelerate the materials development cycle. Back in the late 90s, Greg Olson popularized the concept of materials-as-hierarchical-systems [1, 2], amenable for improvement through the exploitation of *explicit* processing-structure-properties-performance (PSPP) relationships. Olson used this framework to develop (inverse) linkages connecting performance/property requirements to desired (multi-scale) structural features and the latter to the corresponding processing steps. A decade later, the Integrated Computational Materials Engineering (ICME) [3, 4] framework prescribed the combination of theory, experiments and computational tools to streamline and accelerate the materials and manufacturing development cycle. Similarly, the Materials Genome Initiative [5] called for the acceleration of the materials development cycle through the combination of experiments, simulations and data. We would like to point out that ICME and MGI are *aspirational* in that the former does not prescribe the way to carry out the *integration* of multiple tools and the latter does not put forward a feasible strategy to *accelerate* the materials development cycle. On the other hand,

ICME and MGI have motivated considerable development in terms of the sophistication in the tool sets used to carry out the computer-assisted exploration of the materials design space [6–9].

On the integration front, it has long been recognized that in order to establish quantitative PSPP relationships it will be necessary to integrate multiple (computational) tools across multiple scales [9]. Realizing such integration is a necessary (albeit not sufficient) condition to achieving any measure of success when attempting to carry out computationally-assisted materials development exercises. Explicit integration of multiple tools is technically challenging, particularly because of the considerable expense of computational models, the complexity of the input/output interfaces of such models, and the asynchronous nature of the development of such tools.

Approaches that instead use statistical techniques and machine learning tools to better sample the design space have proven to be effective [10], although recently some efforts have emerged that attempt to explicitly integrate models within a single framework for materials design [11–13].

Another strategy for the accelerated discovery of materials (most closely associated to the MGI) has been the use of high-throughput (HT) experimental [14–16] and computational [17] approaches that, while powerful, have important limitations as they tend to be sub-optimal in resource allocation as experimental decisions do not account for the cost and time of experimentation. Resource limitation cannot be overlooked as it is often the case that once a bottle-neck in HT workflows has been eliminated (e.g., synthesis of ever more expansive materials libraries), another one suddenly becomes apparent (e.g., need for high-resolution characterization of materials libraries).

Recently, notions of optimal experimental design, within the overall framework of Bayesian Optimization (BO), have been put forward as a strategy to overcome the limitations of traditional (costly) exploration of the design space. For example, Balachandran et al. [18] recently put forward a framework that balances the need to exploit current knowledge of the design domain with the need to explore it by using a metric that evaluates the value of the next experiment (or simulation) to carry out. BO-based approaches rely on the construction of a response surface of the design space and are typically limited to the use of a single model to carry out the queries. This is an important limitation, as often times, at the beginning of a materials discovery problem, there is not sufficient information to elucidate the feature set (i.e., model) that is most related to the specific performance metric to optimize. Recent work on non-hierarchical fusion for design has led to promising avenues for information source integration [19–21] that we generally build off here. Other recent work on the integration of information sources for design includes a framework that is capable of adaptively selecting competing models connecting materials features to performance metrics through Bayesian Model Averaging (BMA), followed by optimal experimental design [22].

It is clear from the brief discussion above that, while considerable progress has been made recently in the development of novel frameworks for accelerated materials development, several important challenges remain to be solved. Model-based ICME-style frameworks tend to focus on integrating tools at multiple levels under the assumption that there is *a single* model/tool relevant at a specific scale of the problem. This precludes the use of multiple models that may be more/less effective in different regions of the performance space. Data-centric approaches, on the other hand, tend to focus (with some exceptions) on the brute-force exploration of the materials design space, without accounting for the considerable cost associated with such exploration.

In this work, we present a framework that addresses the two outstanding issues listed above in the context of the optimal microstructural design of ductile multi-phase materials, such as advanced high strength steels. Specifically, we expand on the work presented in a companion paper, whereby we carry out the fusion of multiple information sources that connect microstructural descriptors to mechanical performance metrics. This fusion is done in a way that accounts for and exploits the correlations between each individual information source (reduced order model constructed under different simplifying assumptions) and between each information source and the ground truth (represented in this case by a full-field microstructure-based finite element model). In this work, we value the impact a new query to an information source has on the fused model. In particular, we perform the search over the input domain and the information source options concurrently to determine which next query will lead to the most improvement in our objective function. This concurrent approach, to our knowledge, has not been addressed in the literature. In addition, our exploitation of correlations between the discrepancies of the information sources in the fusion process differs significantly from previous work and enables the identification of ground truth optimal points that are not shared by any individual information sources in the analysis.

The remainder of the paper is as follows. In Section 2, we present a brief description of the reduced-order models as well as the (simulated) ground truth. This is followed by the description of the proposed multi-information optimization framework in Section 3. Results are then presented and discussed in Section 4, and finally, conclusions are drawn in Section 5.

2 Mechanical Behavior of Dual-Phase Microstructures

A class of one of the most technologically sought after structural materials, known as advanced high-strength steels, derive their exceptional properties from complex, heterogeneous microstructures. Of the various advanced high strength steels, dual-phase steels have experienced the fastest growth in the automotive industry [23]. These dual-phase advanced high strength steels primarily consist of hard martensite (phase) islands dispersed in a soft ferrite (phase) matrix [24]. Both these phases of the dual-phase steels undergo non-linear elastic-plastic deformation with strikingly different strength levels and strain hardenability [25, 26]. The

overall mechanical properties of dual-phase steels are thus determined partly by the mechanical properties of the constituent phases, and partly by the microstructural features, such as the volume fraction of the phases. The properties of the phases and the microstructural features can, in principle, be tuned and optimized to achieve a particular performance matrix. The microstructure-property correlation of ductile dual-phase materials can be explored by high-fidelity microstructure-based finite element models. But these come at considerable computational cost that precludes their use to carry out search in the microstructure space for regions of optimal performance. In a companion paper, we carry out the fusion of multiple low-fidelity reduced-order models (termed as information sources) based on isostrain [27], isostress [28] or isowork [29] approximations of the partitioning of the macroscopic strain, stress or work, respectively, among the microstructural constituents. Specifically, it was shown how we exploit statistical correlations between information sources to arrive at a fused model with significantly better fidelity with respect to the ground truth (microstructure-based finite element model) than any individual source (reduced-order model). In this paper, we focus on integrating the multi-information source fusion approach with a Bayesian sequential design optimization framework to arrive at optimal microstructures that maximize the strength normalized strain-hardening rate by identifying and exploiting optimal sequential queries of different information sources. The quantity of interest, strength normalized strain-hardening rate, is an important manufacturing-related attribute as it dictates the ductility and formability of the material. The details of the microstructure-based finite element modeling (ground truth) of dual-phase microstructures and the three lower fidelity reduced-order models (information sources) are described in the companion paper as well as in the following subsections for the sake of completeness.

2.1 Microstructure-based Finite Element Model

Microstructure-based finite element modeling are carried out to calculate the overall mechanical response of the ductile dual-phase microstructures. To this end, we generate 3D representative volume elements (RVEs) of the dual-phase microstructures following the procedure detailed in [30]. The RVE is a composite dual-phase microstructure with two discretely modeled phases: a soft phase representative of the ferrite phase and a hard phase representative of the martensite phase, present in the dual-phase advanced high strength steels. A typical 3D RVE of the dual-phase microstructure is shown in Fig. 1. A typical RVE consists of 27,000 C3D8 brick elements of the ABAQUS/standard element library [31], and has a dimension of $100\mu\text{m} \times 100\mu\text{m} \times 100\mu\text{m}$. The volume fraction of the phases in the RVE is always an integral multiple of the volume of one element, which is $3.7 \times 10^{-5}\mu\text{m}^3$. The RVE is subjected to fully periodic boundary conditions on all six faces and monotonically increasing uniaxial tensile deformation. This allows for the calculation of the overall uniaxial tensile stress-strain response of the composite microstructure.

In the calculations, both the phases are assumed to follow isotropic elastic-plastic stress-strain response. The Young's modulus of both the phases is taken to be $E = 200\text{GPa}$ and the Poisson's ratio is taken to be $\nu = 0.3$. The plastic response of both phases are modeled using the Ludwik power law constitutive relation,

$$\tau = \tau_o + K(\varepsilon_{pl})^n, \tag{1}$$

where τ is the flow stress, ε_{pl} is the equivalent plastic strain, τ_o is the yield strength, K is the strengthening coefficient, and n is the strain hardening exponent. The values of τ_o, K, and n for the constituent phases are given in Tab. 1. The parameters are chosen to represent lower initial yield strength of the ferrite (soft) phase compared to the martensite (hard) phase, and higher strain hardenability of the ferrite phase compared to the martensite phase [26, 30].

Table 1. Constitutive parameters for the constituent phases

Constituent Phase	τ_o **[MPa]**	K **[MPa]**	n
Soft (ferrite)	300	2200	0.5
Hard (martensite)	1500	450	0.06

2.2 Reduced-order Models

We use three low-fidelity reduced-order models as three sources of information. These three reduced-order models are:

i. The Voigt/Taylor isostrain model, where the basic assumption is that the strain field is uniform among the constituent phases [32]. The effective stress is expressed in terms of the local stress average with respect to both phases weighted by their respective volume fractions. That is, for this model we have

$$\varepsilon^T_{pl} = \varepsilon^h_{pl} = \varepsilon^s_{pl}, \quad \tau^T = f_{\text{hard}}\tau^h + (1 - f_{\text{hard}})\tau^s. \tag{2}$$

ii. The Reuss/Sachs isostress model, where the basic assumption is that the stresses among the phases are homogeneous [32]. The effective strain is calculated in terms of the average of the strains in each phase weighted by their respective volume fractions. Thus, for this model we have

$$\tau^T = \tau^h = \tau^s, \quad \varepsilon^T_{pl} = f_{\text{hard}}\varepsilon^h_{pl} + (1 - f_{\text{hard}})\varepsilon^s_{pl}. \tag{3}$$

iii. The isowork model, which is an approximation based on the principle that work of deformation is equally distributed in all the constituent phases in the dual-phase

microstructure at any strain level. That is,

$$\tau^h \varepsilon^h_{pl} = \tau^s \varepsilon^s_{pl}. \quad (4)$$

In Eqs. 2, 3 and 4, ε^T_{pl} is the overall plastic strain, ε^h_{pl} is the plastic strain in the hard (martensite) phase, ε^s_{pl} is the plastic strain in the soft (ferrite) phase, τ^T is the overall stress, τ^h is the stress in the hard (martensite) phase, τ^s is the stress in the soft (ferrite) phase, and f_{hard} is the volume fraction of the hard phase in the microstructure. The stress-strain relations, $\tau = f(\varepsilon_{pl})$, of both phases are assumed to follow, Eq. 1, with the values of the parameters given in Tab. 1.

2.3 Demonstration of Modeling Capabilities

The predicted stress-strain response of dual-phase microstructures with varying volume fraction of the hard phase, f_{hard}, using a high-fidelity microstructure-based finite element model is shown in Fig. 1. As seen in the figure, the flow strength of the dual-phase material increases with increasing volume fraction of the hard phase. But the strain-hardening rate, which is the slope of the stress-strain curve, of the material varies non-monotonically with the volume fraction of the hard phase.

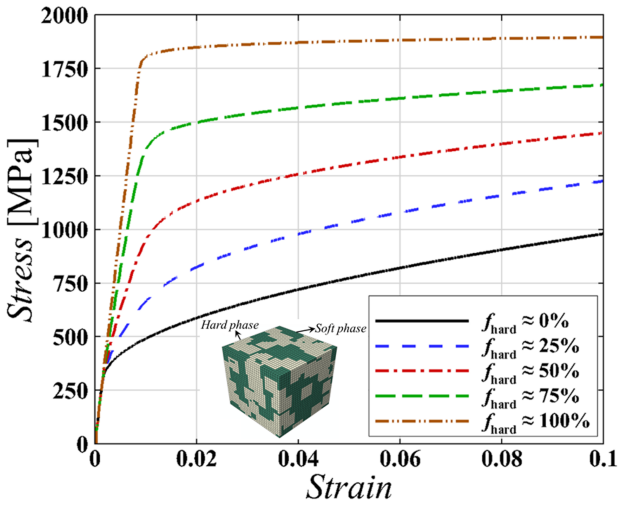

Fig. 1. The stress-strain response of dual-phase phase microstructures with volume fraction of the hard phase, $f_{\text{hard}} = 0\%$, 25%, 50%, 75%, and 100%. A 3D representative volume element of the dual-phase microstructure is shown in the inset.

The variation of the strength normalized strain-hardening rate, $(1/\tau)(d\tau/d\varepsilon_{pl})$, with the volume fraction of the hard phase, f_{hard}, estimated at a strain level, $\varepsilon_{pl} = 1.5\%$, using the microstructure-based finite element calculations, is shown in Fig. 2. As seen in the figure, the value of $(1/\tau)(d\tau/d\varepsilon_{pl})$ at $\varepsilon_{pl} = 1.5\%$, first increases with increasing volume fraction of the hard phase and then starts to decrease. In general a higher value of the quantity $(1/\tau)(d\tau/d\varepsilon_{pl})$ denotes higher formability of the material. Note, in Fig. 2, variation of $(1/\tau)(d\tau/d\varepsilon_{pl})$ with f_{hard} exhibits local perturbations. These perturbations are due to the fact that there are several possible realizations of the RVE of a dual-phase microstructure with a fixed volume fraction of the hard phase. These different realizations result in slightly different values of $(1/\tau)(d\tau/d\varepsilon_{pl})$ for a fixed f_{hard} value. For a few selected volume fractions of the hard phase, seven realizations of the dual-phase microstructures were generated and their mechanical responses were calculated. The standard error on the values of $(1/\tau)(d\tau/d\varepsilon_{pl})$ at $\varepsilon_{pl} = 1.5\%$ due to different realizations of the dual-phase microstructure with fixed volume fraction of the hard phase are also shown in figure as error bars.

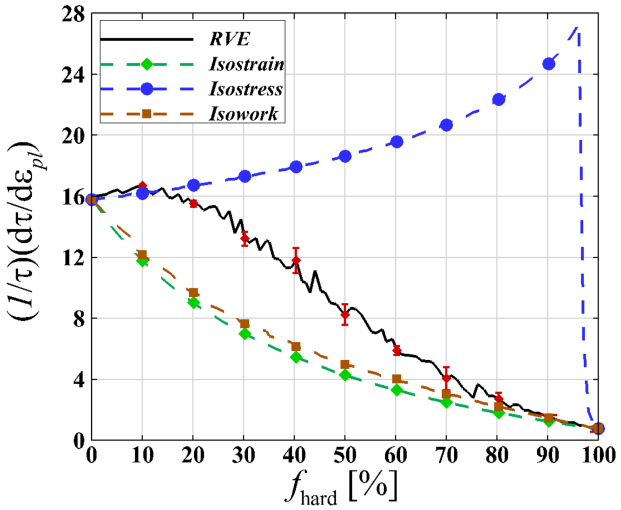

Fig. 2. Comparison of the variation of the strength normalized strain-hardening rate, $(1/\tau)(d\tau/d\varepsilon_{pl})$ at $\varepsilon_{pl} = 1.5\%$, with the volume fraction of the hard phase, f_{hard}, as predicted by the three reduced-order models and the microstructure-based finite element model.

The predictions of $(1/\tau)(d\tau/d\varepsilon_{pl})$ at a strain level, $\varepsilon_{pl} = 1.5\%$, as a function of the volume fraction of the hard phase, f_{hard}, using the three low-fidelity reduced-order models are also shown in Fig. 2. Compared to the finite element model, the isostress model gives a reasonable prediction of $(1/\tau)(d\tau/d\varepsilon_{pl})$ at $\varepsilon_{pl} = 1.5\%$ for low volume fraction of the hard phase but significantly over predicts this quantity for large volume fractions of the hard phase. In contrast, the isostrain and isowork models give reasonable predictions at high volume fraction of the hard phase but under-predict $(1/\tau)(d\tau/d\varepsilon_{pl})$ at $\varepsilon_{pl} = 1.5\%$ at lower volume fractions of the hard phase. It is also important to note here that the maximum values of $(1/\tau)(d\tau/d\varepsilon_{pl})$ at $\varepsilon_{pl} = 1.5\%$ according to each information source are significantly different from the ground truth maximum.

3 Correlation Exploiting Multi-Information Source Optimization

In most materials design tasks, there are always multiple information sources at the disposal of the designer. For example, the forward connections between microstructures and properties/performance can in principle be developed through experiments as well as (computational) models at different levels of fidelity and resolution. Conventional approaches to ICME, on the other hand, often times make the implicit and unrealistic assumption that *there is only one source available* to query the design space—in this work, our framework uses three relatively simple models (under the isostrain, isostress and isowork stress approximations) as representative of multiple information sources while considering a microstructure-sensitive RVE-based simulation as the ground truth.

While information fusion on its own represents a considerable improvement upon the vast majority of ICME-based approaches to materials design currently under development, we posit that an even better approach would necessarily have to account for resource constraints on the exploration of the materials design space (MDS). Specifically, every source used to query the MDS carry a certain (time, monetary, opportunity) cost and thus there are hard limits to the number of queries and the sources used to carry out such queries. Unfortunately, such constraints rarely take a concrete form that can be dealt with using formal constrained optimization approaches. This is due to the often dynamic nature of the materials design and procurement process. As a materials design cycle progresses, the current state of the process may dictate if more resources will be allocated to the process or not. Thus, it is advantageous to tackle such problems in a myopic fashion.

For single information sources and sequential querying, there are two traditional techniques for choosing what to query next in this myopic context [33]. These are efficient global optimization (EGO) [34] and its extensions, such as sequential Kriging optimization (SKO) [35] and value-based global optimization (VGO) [36], and the knowledge gradient (KG) [37–39]. EGO uses a Gaussian process [40] representation of queried information, but assumes no noise [41, 42]. SKO also uses Gaussian processes, but includes a tunable weighting factor to lean towards decisions with higher uncertainty [33]. KG can handle noise and makes its querying selection on the basis of the expected value of the best design after querying. Here, KG does not require that design to have actually been evaluated by an information source.

Recent developments in [19, 20] extend these sequential optimization approaches to the case of multiple information sources. The approach we propose here builds off of these approaches by including and exploiting learned correlations in multi-information source fusion and by defining and implementing a two-step lookahead querying strategy referred to as the value-gradient policy. We describe our formal problem statement, the multi-information source fusion approach, and the value-gradient utility used to guide the querying policy, in the following subsections.

3.1 Problem Formulation

A mathematical statement of the problem is formulated as finding the best design, $\mathbf{x}^*$, such that

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\chi} f(\mathbf{x}), \tag{5}$$

where f is the ground truth objective function, and $\mathbf{x}$ is a set of design variables in the vector space χ. This ground truth objective function is typically very expensive to query. We note that in this formulation, there is a tacit dynamic constraint on resources. While ground truth is impractical to query often in an optimization process, other forms of information are usually available and can be used to approximate the ground truth. These information sources differ in terms of fidelity with respect to the ground truth, as well as resource expenditures required per query. These information sources are also fundamentally related through the fact that they seek to estimate the same quantity of interest. Thus, there *must* exist statistical correlations between these information sources that can potentially be exploited if learned. In this context, the core issue to myopically addressing Eq. 5, is the decision of what information source to query and where in its input domain to execute that query. This decision must balance the cost of the query and what that query is expected to tell us about the solution to Eq. 5.

To assign a value to each potential query option over the information source space and the domains of the respective information sources, we create intermediate Gaussian process surrogates for each information source learned from previous queries. We assume that we have S information sources, $\bar{f}_i(\mathbf{x})$, where $i \in \{1,2,\ldots,S\}$, available to estimate the ground truth, $f(\mathbf{x})$, at design point $\mathbf{x}$. We further assume that we have N_i previous query results available for information source i. These results are 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 $\mathbf{y}_{N_i}$ represents the corresponding outputs from information source i. Posterior Gaussian process distributions of each $\bar{f}_i$, denoted as $f_{GP,i}(\mathbf{x})$, at any point $\mathbf{x}$ in the input space, are then given as

$$f_{GP,i}(\mathbf{x}) \mid \mathbf{X}_{N_i}, \mathbf{y}_{N_i} \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma^2_{GP,i}(\mathbf{x})\right), \tag{6}$$

where

$$\mu_i(\mathbf{x}) = K_i(\mathbf{X}_{N_i}, \mathbf{x})^T \left[K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma^2_{n,i} I\right]^{-1} \mathbf{y}_{N_i}, \tag{7}$$

and

$$\sigma^2_{GP,i}(\mathbf{x}) = k_i(\mathbf{x},\mathbf{x}) - K_i(\mathbf{X}_{N_i}, \mathbf{x})^T \left[K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma^2_{n,i} I\right]^{-1} K_i(\mathbf{X}_{N_i}, \mathbf{x}). \tag{8}$$

Here, k_i is a real-valued kernel function associated with information source i 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 $k_i(\mathbf{x}_{m,i}, \mathbf{x}_{n,i})$, $K_i(\mathbf{X}_{N_i}, \mathbf{x})$ is the $N_i \times 1$ vector whose m^{th} entry is $k_i(\mathbf{x}_{m,i}, \mathbf{x})$ for information source i, and the term $\sigma^2_{n,i}$ can be used to model observation error for information sources or to guard against numerical ill-conditioning. For the kernel function, we use the commonly used squared exponential kernel function given as

$$k_i(\mathbf{x}, \mathbf{x}') = \sigma_s^2 \exp\left(- \sum_{h=1}^{d} \frac{(x_h - x'_h)^2}{2l_h^2} \right), \tag{9}$$

where d is the dimension 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 between the points within dimension h. The parameters of the Gaussian process (σ_s^2, l_h and σ_n^2) associated with each information source can be estimated via maximum likelihood or Bayesian techniques [40].

To these Gaussian process surrogates, we further quantify the discrepancy of each information source with respect to ground truth. These discrepancies can be estimated from, for example, expert opinion or available ground truth data, and can vary over the input space. We add the estimated uncertainty due to information source discrepancy, $\delta_{f,i}(\mathbf{x})$, to the uncertainty associated with the Gaussian process of information source i, denoted by $\delta_{GP,i}(\mathbf{x})$. Specifically, each of the S available information sources for estimating the ground truth objective can be written as

$$f_i(\mathbf{x}) = \mu_i(\mathbf{x}) + \delta_i(\mathbf{x}), \tag{10}$$

where

$$\delta_i(\mathbf{x}) = \delta_{GP,i}(\mathbf{x}) + \delta_{f,i}(\mathbf{x}). \tag{11}$$

Figure 3 shows a depiction of total uncertainty for an information source, which includes both the uncertainty associated with the Gaussian process and the uncertainty associated with the fidelity of the information source.

3.2 Correlation Exploiting Fusion

Available information sources for estimating a ground truth quantity of interest are necessarily correlated by virtue of their estimation task. If they were not correlated, then presumably they are irrelevant to the estimation task at hand. In our companion paper, we develop and demonstrate a fusion process for learning and exploiting the correlation between the discrepancies of the available information sources. We briefly summarize the approach here, as it is essential to our ability to optimize the ground truth when the available information sources do not share the same optimum value.

Since our information sources are represented by intermediate Gaussian processes, their fusion follows that of normally distributed information. Under the case of known correlations between the discrepancies of information sources,

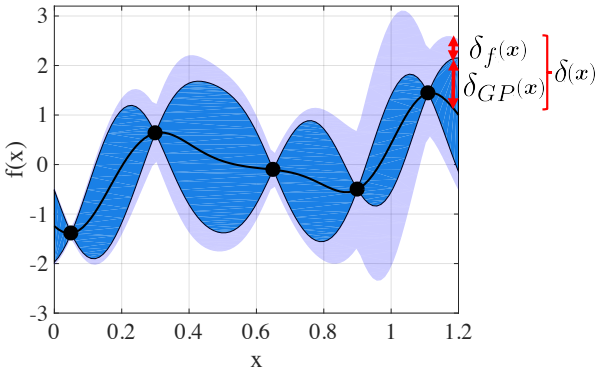

Fig. 3. A depiction of total uncertainty, which includes both the uncertainty associated with the Gaussian process and uncertainty associated with the fidelity of the information source.

the fused mean and variance are shown to be [43]

$$\mathbb{E}[\hat{f}(\mathbf{x})] = \frac{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \boldsymbol{\mu}(\mathbf{x})}{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \tag{12}$$

$$\mathrm{Var}\left(\hat{f}(\mathbf{x})\right) = \frac{1}{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \tag{13}$$

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

The key to the proper use of fusion of normally distributed information is the estimation of the correlation coefficients over the domain. For this, we use the reification process defined in [44, 45]. In this process, to estimate the correlation coefficients between the deviations of information sources i and j, each of the information sources i and j, one at a time, is *reified*, or treated as a ground truth model. Assuming that information source i is reified, the correlation coefficients between the information sources i and j, for $j = 1, \ldots, i-1, i+1, \ldots, S$, are given as

$$\rho_{ij}(\mathbf{x}) = \frac{\sigma_i^2(\mathbf{x})}{\sigma_i(\mathbf{x})\sigma_j(\mathbf{x})} = \frac{\sigma_i(\mathbf{x})}{\sqrt{(\mu_i(\mathbf{x}) - \mu_j(\mathbf{x}))^2 + \sigma_i^2(\mathbf{x})}}, \tag{14}$$

where $\mu_i(\mathbf{x})$ and $\mu_j(\mathbf{x})$ are the mean values of the Gaussian processes of information sources i and j respectively, at design point $\mathbf{x}$, and $\sigma_i^2(\mathbf{x})$ and $\sigma_j^2(\mathbf{x})$ are the total variances at point $\mathbf{x}$. Afterward, information source j is reified to estimate $\rho_{ji}(\mathbf{x})$. Then, the variance weighted average of the two estimated correlation coefficients is used as the estimate of the correlation between the errors as

$$\bar{\rho}_{ij}(\mathbf{x}) = \frac{\sigma_j^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})} \rho_{ij}(\mathbf{x}) + \frac{\sigma_i^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})} \rho_{ji}(\mathbf{x}). \tag{15}$$

These average correlations are then used to estimate the fused mean and variance in Eqs. 12 and 13.

3.3 Value-Gradient Querying

By computing the fused means and variances in the input design space χ, we construct a fused Gaussian process over this space. This fused information source contains all of our current knowledge about the ground truth objective function. Our goal is to optimize ground truth by leveraging new queries to the less resource-expensive information sources. Once our resources for information source querying have been exhausted, the predicted ground truth optimal design can be synthesized or produced in an experiment. This information can of course then be fed back to the multi-information source optimization framework and used to update information source discrepancy and correlation information.

The task then, is to determine what information source to query and where to query it, concurrently, so as to produce the most value in terms of addressing Eq. 5, with the tacit resource constraint in mind. For this decision, we propose a utility, which we refer to as the value-gradient utility, which takes into account both the immediate improvement in one step and expected improvement in two steps. The idea here being that we seek to produce rapid improvement, with the knowledge that every resource expenditure could be the last, but we also seek to position ourselves best for the next resource expenditure. In this sense, we are equally weighting next step value with next step (knowledge) gradient information, hence the term value-gradient.

The immediate improvement can be quantified by the maximum mean function value of the fused Gaussian process, μ^*_{fused}. Since the best estimate of the objective function is represented by the fused Gaussian process, which is the probabilistic representation of the ground truth objective function, there is uncertainty in the value of the predicted ground truth objective function upon querying the next sample. Thus, we compute the expected value of improvement using the posterior predictive distribution of the fused model. Letting $(i_{1:N}, \mathbf{x}_{1:N}, y_{1:N})$ be the information sources, design points, and the corresponding objective values used for the first N queries and $\hat{f}$ denote the posterior distribution of the fused model, the expected improvement (EI) at design point $\mathbf{x}$ is defined as

$$\begin{aligned} EI(\mathbf{x}) &= \mathbb{E}\left[\max_{\mathbf{x}'\in\chi} \mathbb{E}[\hat{f}(\mathbf{x}') \mid i_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_{N+1} = \mathbf{x}, y_{1:N}] \right. \\ &\qquad \left. - \max_{\mathbf{x}'\in\chi} \mathbb{E}[\hat{f}(\mathbf{x}') \mid i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]\right] \\ &= \mathbb{E}\left[\max_{\mathbf{x}'\in\chi} \mathbb{E}[\hat{f}(\mathbf{x}') \mid i_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_{N+1} = \mathbf{x}, y_{1:N}]\right] \\ &\qquad - \max_{\mathbf{x}'\in\chi} \mathbb{E}[\hat{f}(\mathbf{x}') \mid i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}], \end{aligned} \tag{16}$$

where the last expression comes out of the expectation operator as it is a known value when conditioned on the first N queries.

The KG policy of [37, 46, 47] takes an information-economic approach to maximize this expectation. Letting $S^N = \mathbb{E}[\hat{f}(\mathbf{x}) \mid i_{1:N}, \mathbf{x}_{1:N}, y_{1:N}]$ be the knowledge state, the value of being at state S^N is defined as $V^N(S^N) = \max_{\mathbf{x}\in\chi} S^N$. The knowledge gradient, which is a measure of expected improvement, is defined as

$$\nu^{KG}(\mathbf{x}) = \mathbb{E}[V^{N+1}\left(S^{N+1}(\mathbf{x})\right) - V^N(S^N) \mid S^N]. \tag{17}$$

The KG policy for sequentially choosing the next query is then given as

$$\mathbf{x}^{KG} = \arg\max_{\mathbf{x}\in\chi} \nu^{KG}(\mathbf{x}). \tag{18}$$

Calculation of the knowledge gradient is discussed in detail in two algorithms presented in [47].

Given both immediate and expected improvement, our proposed value-gradient utility is given as

$$U = \mu^*_{\text{fused}} + \max_{\mathbf{x}\in\chi} \nu^{KG}(\mathbf{x}), \tag{19}$$

where the first term is the maximum value of the mean function of the current fused model and the second term is the maximum expected improvement that can be obtained with another query as measured by the knowledge gradient over the fused model. We can then define a value-gradient policy as the policy that selects the next query such that the value-gradient utility is maximized. By considering the immediate gain in the next step and the expected gain in the step that follows, the value-gradient is a two-step look-ahead policy.

To determine the next design point and information source to query, we generate Latin hypercube samples in the input design space as alternatives denoted as $\mathbf{X}_f$. Among these alternatives, we seek to find the query that maximizes the value-gradient utility of Eq. 19. According to Eq. 6, an evaluation of information source i, at design point $\mathbf{x}$, is distributed normally with mean $\mu_i(\mathbf{x})$ and variance $\sigma^2_{GP,i}(\mathbf{x})$. For a given alternative, $\mathbf{x}$, we draw N_q independent samples from the distribution at that point as

$$f_i^q(\mathbf{x}) \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma^2_{GP,i}(\mathbf{x})\right),\ i = 1, \ldots, S \text{ and } q = 1, \ldots, N_q. \tag{20}$$

In order to predict the impact of querying each alternative on the utility function, we temporarily augment the design point, $\mathbf{x}$, and the sampled information source output value, $f_i^q(\mathbf{x})$, one at a time, to the available samples of information source i. By adding this sample, the Gaussian process of information source i and as a result, the fused Gaussian process, are temporarily updated. Then, the maximum mean function value and the maximum knowledge gradient of the temporarily updated fused Gaussian process are evaluated. These quantities can then be used to compute the value-

gradient utility that would result if the sample, $\left(\mathbf{x}\,,\, f_i^q(\mathbf{x})\right)$, was realized from information source i. This is given as

$$U^q_{\mathbf{x},i} = \mu^{*,\text{temp}}_{\text{fused}} + \max_{\mathbf{x}' \in \chi} \nu^{KG}(\mathbf{x}'). \tag{21}$$

This process is repeated for all N_q samples by removing the previously added sample and augmenting with the next new sample. The expected value-gradient utility obtained from adding alternative $\mathbf{x}$ to information source i is then computed as

$$EU_{\mathbf{x},i} = \frac{1}{N_q} \sum_{q=1}^{N_q} U^q_{\mathbf{x},i}. \tag{22}$$

This expected utility is evaluated for all the alternatives and all the information sources. By denoting $C_{\mathbf{x},i}$ as the cost of querying information source i at design $\mathbf{x}$, we find the query $(i_{N+1}, \mathbf{x}_{N+1})$ that maximizes the expected value-gradient utility per unit cost, given by

$$(i_{N+1}, \mathbf{x}_{N+1}) = \underset{i \in \{1,\ldots,S\}\,,\, \mathbf{x} \in \mathbf{X}_f}{\arg\max} \frac{EU_{\mathbf{x},i}}{C_{\mathbf{x},i}}. \tag{23}$$

After querying the design point $\mathbf{x}_{N+1}$ from the selected information source, i_{N+1}, the corresponding Gaussian process and afterward, the fused Gaussian process, are updated. This process repeats until a termination criterion, such as exhaustion of the querying budget, is met. Then, the optimum solution of Eq. 5 is found based on the mean function of the fused Gaussian process. This design is then to be created at ground truth. Information from this creation can then be fed back into the framework if more resources are allocated.

4 Framework Demonstration

In this section, we demonstrate the application of our framework to the optimization of the ground truth strength normalized strain hardening rate for the dual-phase steel application. We stress here that the purpose of our framework is the optimization of ground truth. That is, our motivation is the creation of a myopic multi-information source optimization framework for addressing Eq. 5 in the context of materials design. Thus, we seek to identify the best candidate for a ground truth experiment with whatever resources we have available. Once those resources are exhausted, a ground truth experiment takes place based on the recommendation of our framework. The result of that experiment can then be fed back into the framework. If more resources are then allocated, perhaps on the basis of promising results, then the framework can be employed again.

The specific demonstration consists of the use of the three reduced-order models (isostrain, isostress, and isowork) to query the impact of quantifiable microstructural attributes on the mechanical response of a composite microstructure—in this case a dual-phase steel. The *ground truth* in this case is the finite element model of the dual-phase material. The objective is the maximization of the (ground truth) normalized strain hardening rate at $\varepsilon_{pl} = 1.5\%$. The design variable is the percentage of the hard phase, f_{hard}, in the dual-phase material. We assume that our resources limit us to five total queries to (any of) the information sources before we must make a recommendation for a ground truth experiment. Given promising ground truth results, five more queries can be allocated to the information sources. The framework is initialized with one query from each information source and one query from the ground truth. This information is used to construct the initial intermediate Gaussian process surrogates.

The value-gradient policy of our framework was used to select the next information source and the location of the query in the input space for each iteration of the process. For comparison purposes, the KG policy operating directly on the ground truth was also used to reveal the gains that can be had by considering all available information sources. The convergence results of our proposed approach using all information sources and the KG policy on the ground truth are shown in Fig. 4. On the figure, the dashed line represents

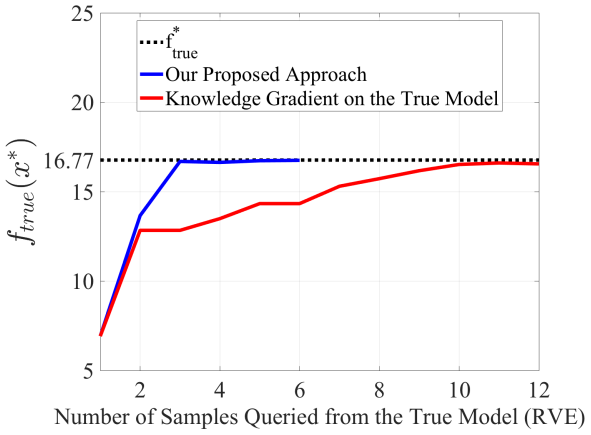

Fig. 4. The optimal solution obtained by our proposed approach and by applying the knowledge gradient on a GP of only the true data (RVE) for different number of samples queried from the true model.

the optimal value of the ground truth quantity of interest. It is clear from this figure that our approach outperformed the knowledge gradient, and in doing so, saved considerable expense by reducing the number of needed ground truth experiments. Table 2 presents the results of each ground truth experiment conducted according to the recommendation of our framework. From the table we see that the third recommendation for a ground truth experiment produces a nearly optimal design. We note that the original sample from ground truth used for initialization was taken at $f_{\text{hard}} = 95\%$, which is far away from the true optimal. This can be seen below in Fig. 5 in the left column. Thus, the framework, by leverag-

Table 2. The optimal solution obtained by the fused model, and the true value at the obtained optimum design point. The true optimal solution by the RVE model is $(\mathbf{x}^*, f^*) = (8.5427, 16.7685)$.

Experiment	$\mathbf{x}^*_{fused}$	f^*_{fused}	$f_{true}(\mathbf{x}^*_{fused})$
2	28.6432	7.5037	13.6638
3	10.0503	11.9833	16.6923
4	10.5528	13.9217	16.6389
5	9.5477	15.4655	16.7325
6	8.7963	16.7089	16.7528

ing the three *inexpensive* available information sources, was able to quickly direct the ground truth experiment to a higher quality region of the design space. The final three experiments shown in Tab. 2 show that little more is gained in terms of ground truth objective and that the fused model has learned more about the ground truth in that region. At this point, it is likely that more resources would not be allocated to this design problem and the framework was able to successfully find the best design.

Updates to each information source Gaussian process surrogate model and the fused model representing our knowledge of ground truth are shown in Fig. 5 for iterations 1, 15, and 30 of the information source querying process. As can be seen from the left column, the first experiment from ground truth and the first query from each information source told us little about the location of the true objective. However, on iteration 15, the fused model, shown by the smooth red curve, has identified the best region of the design space, although it under-predicts the ground truth at this point. We note that at this point, only 3 expensive ground truth experiments have been conducted. By iteration 30, the fused model is very accurate in the region surrounding the optimal design for ground truth. At this point, six ground truth experiments have been conducted. From the figure, and also from Fig. 2, it is clear that none of the information sources share the ground truth optimum. The *ability of the framework to find this optimum rested upon the use of correlation exploiting fusion*, and would not have been possible using traditional methods.

To conclude this demonstration we present the history of the queries to each information source and the ground truth. This information is provided in Fig. 6. Note, that the iteration now counts queries to each information source as well as ground truth experiments. From the figure, it is clear that all three information sources are exploited to find the ground truth optimal design, implying that, however imperfect, *all* sources available to the designer must be used, in an optimal manner, in order to identify the optimal ground truth.

5 Conclusions and Future Work

In this paper we presented and demonstrated a myopic multi-information source optimization framework. The framework focused on determining the next information source to query and where in the input domain to query it by trading off resource expense and gains expected in ground truth objective function quality. To value each next potential query we presented a novel value-gradient policy, which seeks to maximize a two-step lookahead utility based on immediate value and the knowledge gradient for a potential next step. The framework was demonstrated on the optimization of ground truth strength normalized strain hardening rate for a dual-phase material. The results of the demonstration revealed the promise of this framework as a suitable methodology for answering the MGI call for accelerating the materials development cycle.

In future work, the framework developed here will be extended to handle multiple objectives and studied for scalability to high dimensional input spaces. Additionally, we will explore the possibility of carrying out optimal sequential queries in which the sources of information are not input/output aligned. A specific example, for example, would be combining sources that establish relationships between processing parameters/conditions and microstructure with sources that connect microstructures to properties/performance. Much remains to be done, but this work presents a plausible research program towards the realization of the promise of ICME, which in the end rests on tool (or information source) *integration*.

Acknowledgements

The authors would like to acknowledge the support of the National Science Foundation through grant No. NSF-CMMI-1663130, *DEMS: Multi-Information Source Value of Information Based Design of Multiphase Structural Materials*. Arroyave would also like to acknowledge the support of the National Science Foundation through grant No. NSF-CMMI-1534534, *DMREF: Accelerating the Development of Phase-Transforming Heterogeneous Materials: Application to High Temperature Shape Memory Alloys*. Allaire and Arroyave would also like to acknowledge the support of the National Science Foundation through grant No. NSF-DGE-1545403, *NRT-DESE: Data-Enabled Discovery and Design of Energy Materials (D^3EM)*.

References

[1] Olson, G. B., 1997. "Computational design of hierarchically structured materials". *Science,* **277**(5330), pp. 1237–1242.

[2] Olson, G. B., 2000. "Designing a new material world". *Science,* **288**(5468), pp. 993–998.

[3] Allison, J., 2011. "Integrated computational materials engineering: A perspective on progress and future steps". *JOM,* **63**(4), pp. 15–18.

[4] Council, N. R., et al., 2008. *Integrated computational materials engineering: a transformational discipline for improved competitiveness and national security*. National Academies Press.

[5] Holdren, J. P., et al., 2011. "Materials genome initia-

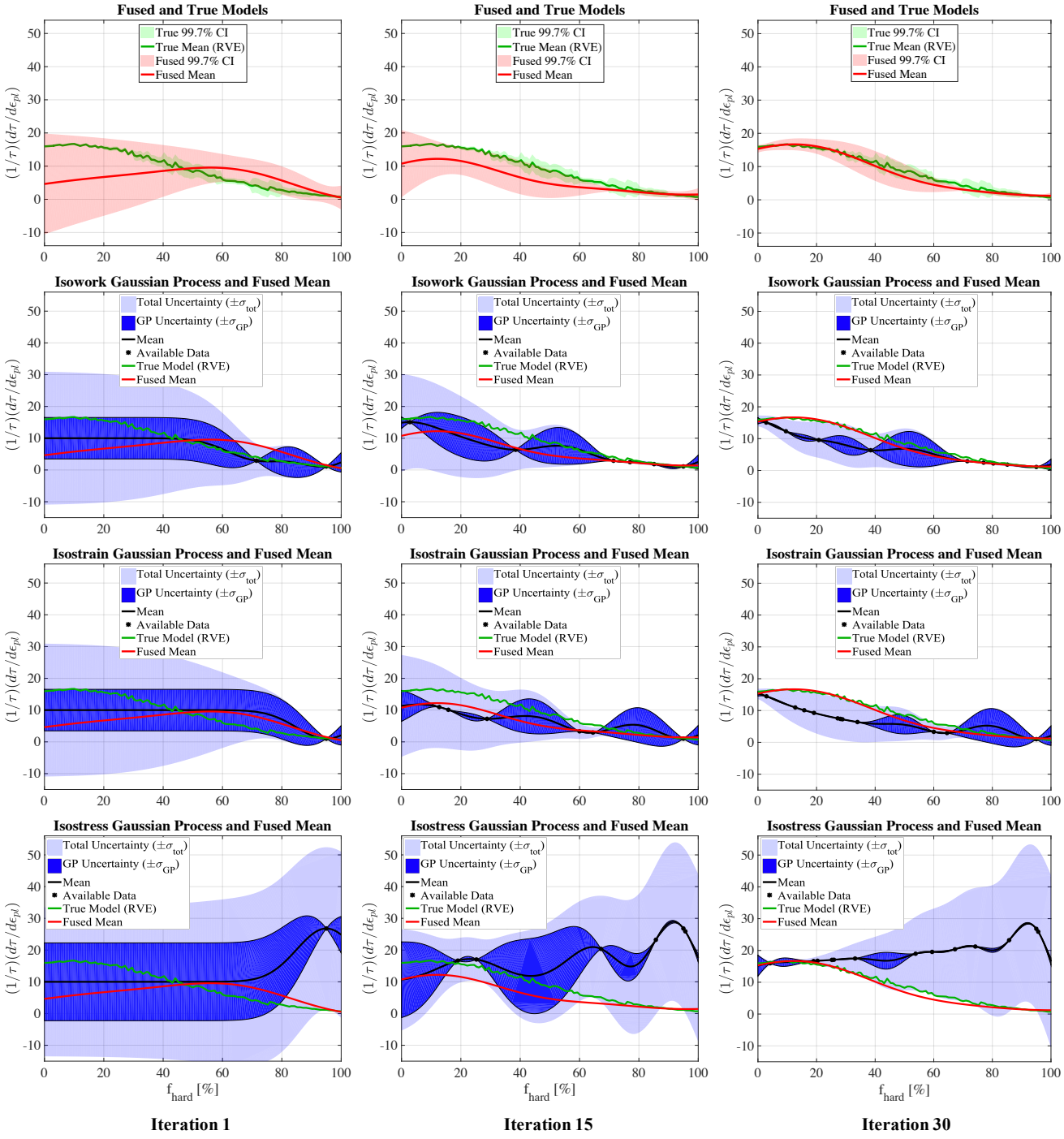

Fig. 5. The fused model and Gaussian processes of the isowork, isostrain and isostress models in comparison with the true (RVE) model in iterations 1, 15 and 30.

tive for global competitiveness". *National Science and Technology Council OSTP. Washington, USA*.

[6] Madison, J. D., 2016. "Integrated computational materials engineering: Tools, simulations and new applications". *JOM,* **68**(5), pp. 1376–1377.

[7] Agrawal, A., and Choudhary, A., 2016. "Perspective: materials informatics and big data: realization of the fourth paradigm of science in materials science". *APL Materials,* **4**(5), p. 053208.

[8] Kalidindi, S. R., and De Graef, M., 2015. "Materials data science: current status and future outlook". *Annual Review of Materials Research,* **45**, pp. 171–193.

[9] Voorhees, P., Spanos, G., et al., 2015. Modeling across scales: a roadmapping study for connecting materials models and simulations across length and time scales. Tech. rep., Tech. rep., The Minerals, Metals & Materi-

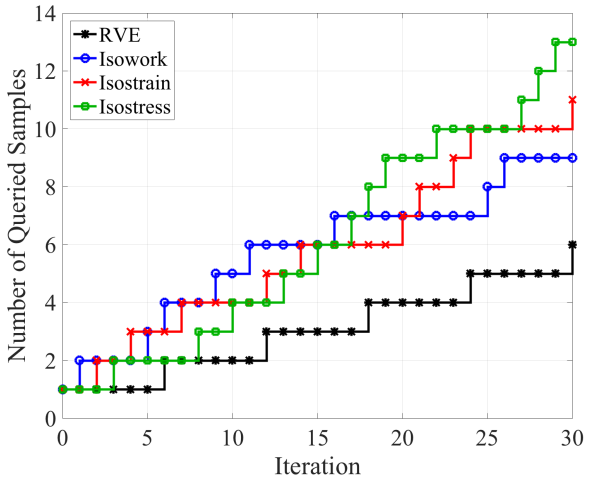

Fig. 6. Number of samples queried from the true model (RVE) and the information sources in each iteration.

als Society (TMS).

[10] Bessa, M., Bostanabad, R., Liu, Z., Hu, A., Apley, D. W., Brinson, C., Chen, W., and Liu, W. K., 2017. "A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality". *Computer Methods in Applied Mechanics and Engineering,* **320**, pp. 633–667.

[11] Reddy, S., Gautham, B., Das, P., Yeddula, R. R., Vale, S., and Malhotra, C., 2017. "An ontological framework for integrated computational materials engineering". In Proceedings of the 4th World Congress on Integrated Computational Materials Engineering (ICME 2017), Springer, pp. 69–77.

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

[13] Diehl, M., Groeber, M., Haase, C., Molodov, D. A., Roters, F., and Raabe, D., 2017. "Identifying structure–property relationships through dream. 3D representative volume elements and DAMASK crystal plasticity simulations: An integrated computational materials engineering approach". *JOM,* **69**(5), pp. 848–855.

[14] Potyrailo, R., Rajan, K., Stoewe, K., Takeuchi, I., Chisholm, B., and Lam, H., 2011. "Combinatorial and high-throughput screening of materials libraries: review of state of the art". *ACS combinatorial science,* **13**(6), pp. 579–633.

[15] Suram, S. K., Haber, J. A., Jin, J., and Gregoire, J. M., 2015. "Generating information-rich high-throughput experimental materials genomes using functional clustering via multitree genetic programming and information theory". *ACS combinatorial science,* **17**(4), pp. 224–233.

[16] Green, M. L., Choi, C., Hattrick-Simpers, J., Joshi, A., Takeuchi, I., Barron, S., Campo, E., Chiang, T., Empedocles, S., Gregoire, J., et al., 2017. "Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies". *Applied Physics Reviews,* **4**(1), p. 011105.

[17] Curtarolo, S., Hart, G. L., Nardelli, M. B., Mingo, N., Sanvito, S., and Levy, O., 2013. "The high-throughput highway to computational materials design". *Nature materials,* **12**(3), pp. 191–201.

[18] Balachandran, P. V., Xue, D., Theiler, J., Hogden, J., and Lookman, T., 2016. "Adaptive strategies for materials design using uncertainties". *Scientific reports,* **6**.

[19] Chen, S., Jiang, Z., Yang, S., and Chen, W., 2016. "Multimodel fusion based sequential optimization". *AIAA Journal,* **55**(1), pp. 241–254.

[20] Lam, R., Allaire, D. L., and Willcox, K. E., 2015. "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.

[21] Allaire, D., and Willcox, K., 2014. "A mathematical and computational framework for multifidelity design and analysis with computer models". *International Journal for Uncertainty Quantification,* **4**(1).

[22] Talapatra, A., Boluki, S., Duong, T., Qian, X., Dougherty, E., and Arroyave, R., 2018. "Towards an autonomous efficient materials discovery framework: An example of optimal experiment design under model uncertainty". *arXiv preprint arXiv:1803.05460*.

[23] Bhattacharya, D., 2011. "Metallurgical perspectives on advanced sheet steels for automotive applications". In *Advanced steels*. Springer, pp. 163–175.

[24] Rashid, M., 1981. "Dual phase steels". *Annual Review of Materials Science,* **11**(1), pp. 245–266.

[25] Chen, P., Ghassemi-Armaki, H., Kumar, S., Bower, A., Bhat, S., and Sadagopan, S., 2014. "Microscale-calibrated modeling of the deformation response of dual-phase steels". *Acta Materialia,* **65**, pp. 133–149.

[26] Srivastava, A., Bower, A., Hector Jr, L., Carsley, J., Zhang, L., and Abu-Farha, F., 2016. "A multiscale approach to modeling formability of dual-phase steels". *Modelling and Simulation in Materials Science and Engineering,* **24**(2), p. 025011.

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

[28] Reuss, A., 1929. "Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle.". *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik,* **9**(1), pp. 49–58.

[29] Bouaziz, O., and Buessler, P., 2002. "Mechanical behaviour of multiphase materials: an intermediate mixture law without fitting parameter". *Revue de Métallurgie–International Journal of Metallurgy,* **99**(1), pp. 71–77.

[30] Gerbig, D., Srivastava, A., Osovski, S., Hector, L. G.,

and Bower, A., 2017. "Analysis and design of dual-phase steel microstructure for enhanced ductile fracture resistance". *International Journal of Fracture*, pp. 1–24.

[31] Documentation, A., and Manual, U., 2010. "Version 6.10". *Dassault Systemes*.

[32] Nemat-Nasser, S., and Hori, M., 2013. *Micromechanics: overall properties of heterogeneous materials*, Vol. 37. Elsevier.

[33] Scott, W., Frazier, P., and Powell, W., 2011. "The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression". *SIAM Journal on Optimization,* **21**(3), pp. 996–1026.

[34] Jones, D. R., Schonlau, M., and Welch, W. J., 1998. "Efficient global optimization of expensive black-box functions". *Journal of Global Optimization,* **13**(4), pp. 455–492.

[35] Huang, D., Allen, T. T., Notz, W. I., and Miller, R. A., 2006. "Sequential kriging optimization using multiple-fidelity evaluations". *Structural and Multidisciplinary Optimization,* **32**(5), pp. 369–382.

[36] Moore, R. A., Romero, D. A., and Paredis, C. J., 2014. "Value-based global optimization". *Journal of Mechanical Design,* **136**(4), p. 041003.

[37] Frazier, P. I., Powell, W. B., and Dayanik, S., 2008. "A knowledge-gradient policy for sequential information collection". *SIAM Journal on Control and Optimization,* **47**(5), pp. 2410–2439.

[38] Gupta, S. S., and Miescke, K. J., 1994. "Bayesian look ahead one stage sampling allocations for selecting the largest normal mean". *Statistical Papers,* **35**(1), pp. 169–177.

[39] Gupta, S. S., and Miescke, K. J., 1996. "Bayesian look ahead one-stage sampling allocations for selection of the best population". *Journal of Statistical Planning and Inference,* **54**(2), pp. 229–244.

[40] Williams, C. K., and Rasmussen, C. E., 2006. "Gaussian processes for machine learning". *The MIT Press*.

[41] Schonlau, M., Welch, W. J., and Jones, D., 1996. "Global optimization with nonparametric function fitting". *Proceedings of the ASA, Section on Physical and Engineering Sciences*, pp. 183–186.

[42] Schonlau, M., Welch, W. J., and Jones, D. R., 1998. "Global versus local search in constrained optimization of computer models". *Lecture Notes-Monograph Series*, pp. 11–25.

[43] Winkler, R. L., 1981. "Combining probability distributions from dependent information sources". *Management Science,* **27**(4), pp. 479–488.

[44] Allaire, D., and Willcox, K., 2012. "Fusing information from multifidelity computer models of physical systems". In Information Fusion (FUSION), 2012 15th International Conference on, IEEE, pp. 2458–2465.

[45] Thomison, W. D., and Allaire, D. L., 2017. "A model reification approach to fusing information from multifidelity information sources". In 19th AIAA Non-Deterministic Approaches Conference, p. 1949.

[46] Powell, W. B., and Ryzhov, I. O., 2012. *Optimal Learning*, Vol. 841. John Wiley & Sons.

[47] Frazier, P., Powell, W., and Dayanik, S., 2009. "The knowledge-gradient policy for correlated normal beliefs". *INFORMS Journal on Computing,* **21**(4), pp. 599–613.