Computational Materials Science 129 (2017) 311–322

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Optimal experimental design for materials discovery

Roozbeh Dehghannasiri [a,b,*], Dezhen Xue [c,d], Prasanna V. Balachandran [c], Mohammadmahdi R. Yousefi [e], Lori A. Dalton [e,f], Turab Lookman [c,*], Edward R. Dougherty [a,b,*]

[a] Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA
[b] TEES-AgriLife Center for Bioinformatics and Genomic Systems Engineering, Texas A&M University, College Station, TX 77843, USA
[c] Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[d] State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, China
[e] Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA
[f] Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 6 July 2016
Received in revised form 7 November 2016
Accepted 23 November 2016

Keywords:
Optimal experimental design
Materials design
Mean objective cost of uncertainty (MOCU)

ABSTRACT

In this paper, we propose a general experimental design framework for optimally guiding new experiments or simulations in search of new materials with desired properties. The method uses the knowledge of previously completed experiments or simulations to recommend the next experiment which can effectively reduce the pertinent model uncertainty affecting the materials properties. To illustrate the utility of the proposed framework, we focus on a computational problem that utilizes time-dependent Ginzburg-Landau (TDGL) theory for shape memory alloys to calculate the stress-strain profiles for a particular dopant at a given concentration. Our objective is to design materials with the lowest energy dissipation at a specific temperature. The aim of experimental design is to suggest the best dopant and its concentration for the next TDGL simulation. Our experimental design utilizes the mean objective cost of uncertainty (MOCU), which is an objective-based uncertainty quantification scheme that measures uncertainty based upon the increased operational cost it induces. We analyze the performance of the proposed method and compare it with other experimental design approaches, namely random selection and pure exploitation.

1. Introduction

The Materials Genome Initiative (MGI) in the U.S. [1] has catalyzed much recent interest in accelerating materials discovery. One of the outstanding challenges in materials science is to reduce the number of costly and time-consuming trial and error experiments required to find new materials with desired properties. This is not a trivial problem because the materials search space is vast due to the complex interplay of structural, chemical and microstructural degrees of freedom and only a small fraction has been experimentally investigated [2]. High-throughput efforts, including high-throughput calculations and combinatorial experiments, have largely been the approaches of choice to narrow the combinatorial search space [3–5]. Recently, there has been much interest in using data-driven machine learning tools for optimally guiding experiments or calculations towards materials with desired properties [2,6–16]. Such methods have met considerable success in fields such as game theory, pattern recognition, artificial intelligence, and event forecasting. However, the application of pure data-driven approaches to materials science can be biased and yield suboptimal results, as the available training data are quite limited compared to the number of features (or material descriptors) and size of the search space [17–19]. A distinct advantage of materials science is that knowledge in the form of constitutive or scaling relations and various constraint equations is often available from theory or known empirically. Such prior knowledge can be used in conjunction with data to quantify uncertainty and construct operators that are optimal relative to that uncertainty. Moreover, these operators on average outperform those designed solely using data. The efficacy of this approach has been successfully demonstrated in the biological sciences. For instance, in genomics there is a large body of knowledge regarding gene/protein signaling pathways. This knowledge can be transformed in such a way as to be useful for constructing biomarkers [20,21] and then

* Corresponding authors at: Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA (R. Dehghannasiri, E.R. Dougherty), Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA (T. Lookman).

E-mail addresses: roozbehdn@tamu.edu (R. Dehghannasiri), xuedezhen@mail.xjtu.edu.cn (D. Xue), pbalachandran@lanl.gov (P.V. Balachandran), yousefi@ece.osu.edu (M.R. Yousefi), dalton@ece.osu.edu (L.A. Dalton), txl@lanl.gov (T. Lookman), edward@ece.tamu.edu (E.R. Dougherty).

http://dx.doi.org/10.1016/j.commatsci.2016.11.041
0927-0256/

incorporated into a Bayesian framework to design optimal classifiers for decisions involving patient diagnosis, prognosis, and therapy [22–25]. How these methods can be utilized to predict properties and guide new experiments is therefore of significant importance and has not been demonstrated in materials science.

Our focus in this work is on experimental design. Experimental design has a long and varied history in science and engineering, the reason being that a properly designed experimental procedure provides much greater efficiency than simply making random probes. Indeed, Francis Bacon's call for experimental design in the *New Organon* in 1620 is often taken to be the beginning of modern science. In particular, in biomedicine, where the processes are very complex, one can gather a virtually limitless amount of information without getting to the crux of the matter. In view of the large number of measurements often needed in materials science, we deal with this problem as a multi-dimensional optimization problem, which typically requires training data in order to be solved [26,27]. Prior knowledge regarding parameters and features affecting the desired properties of materials is also crucial. However, it is often the case that prior knowledge is insufficient and the presence of large uncertainties degrades the experimental design. Therefore, one needs to improve the predictability of the model with respect to the ultimate objective by making additional measurements, which in turn requires synthesizing new materials. Therefore, it is necessary to target experimental efforts where the material with the desirable properties may be found by minimizing the number of experiments. This can be done via *experimental design* that distinguishes between different experiments based upon information they can provide.

In the present study, we demonstrate experimental design by formulating a computational problem based on the time-dependent Ginzburg-Landau (TDGL) theory for shape memory alloys (SMAs) that has been shown to capture reasonably well the underlying physics of the shape memory effect (SME) and superelasticity (SE) [28]. SMAs are a subclass of martensitic structural phase transitions that display SME and SE. The SE effect arises when an alloy above a certain temperature is deformed such that on unloading it returns to the original strain state. They are important because of their high tensile strength and ability to recover [29–31]. When the high symmetry austenite structure is exposed to stress beyond a critical value, it transitions to the low symmetry martensite phase and when the stress is removed, the martensite reverts back to the parent austenite phase [32]. Typically, such first-order phase transitions are accompanied by large hysteresis in the stress-strain curve, as shown in Fig. 1(a). Therefore, the whole loading-unloading process results in a hysteresis stress-strain loop. The closed area inside the curve is a measure of the amount of energy dissipated during the stress-strain cycle. For practical applications, large energy dissipation or hysteresis is undesirable because it affects the fatigue properties of SMAs in devices (such as cardiovascular stents) that require high sensitivity, precision, and durability. Therefore, lowering the energy dissipation accompanying SE is critical for realizing SMAs in practical applications and serves as our computational *design target* in this work. More specifically, we integrate TDGL theory (which computes the stress-strain curve for prototypical SMAs such as FePd) with our experimental design to rapidly select the material-specific model parameters that minimize the energy dissipation associated with SE. In real experiments, chemical modification of SMAs, for example by doping alloying elements in the host alloy, can affect the shape of the stress-strain SE response and consequently, the energy dissipation. In our simulation, we "mimic" chemical doping by varying the model parameters (additional details discussed in the next section).

The outline of the paper is as follows. In Section 2, we discuss the computational model for SMAs based on mesoscale Landau theory. This serves as the "oracle" to calculate properties and we also use it as a source of data to fit an empirical model that serves as input to our design. In Section 3, we discuss and develop our experimental design strategy which acts on the results from the computational model of Section 2. Section 4 discusses the results from our design and evaluates its performance compared to using a random selection strategy for guiding the next experiment or using the best predicted model value. In Section 5, we distill the key ideas and state the general framework for materials design where model parameters or features are unknown, and illustrate the algorithm in Section 5.1 with a worked example using a polynomial cost function. In Section 5.2, we provide a second example but in the context of a network problem with discrete states, which again explains the mathematics underlying our approach in a simple form. In this example, the cost function is defined as the probability of undesirable states.

2. Ginzburg-Landau theory

We discuss here the TDGL theory for SMAs. Note that throughout this paper we denote vectors by bold letters. We also use uppercase and lowercase letters to represent random variables and their realizations, respectively. Our model for the alloy is a two-dimensional version of the cubic to tetragonal martensitic transformation appropriate for materials such as FePd or InTl [28]. The symmetry-adapted strains e_1, e_2, and e_3 represent hydrostatic, deviatoric, and shear modes, respectively, where $e_2 = (1/2)(\epsilon_{xx} - \epsilon_{yy})$ is the strain responsible for the transition (order parameter, OP), and $e_1 = (1/2)(\epsilon_{xx} + \epsilon_{yy})$ and $e_3 = \epsilon_{xy}$. These strains are defined in terms of the linear strain components of elasticity in 2D, $\epsilon_{ij} = 1/2[(\partial u_i/\partial x_j) + (\partial u_j/\partial x_i)]$, where u_i is the i-th

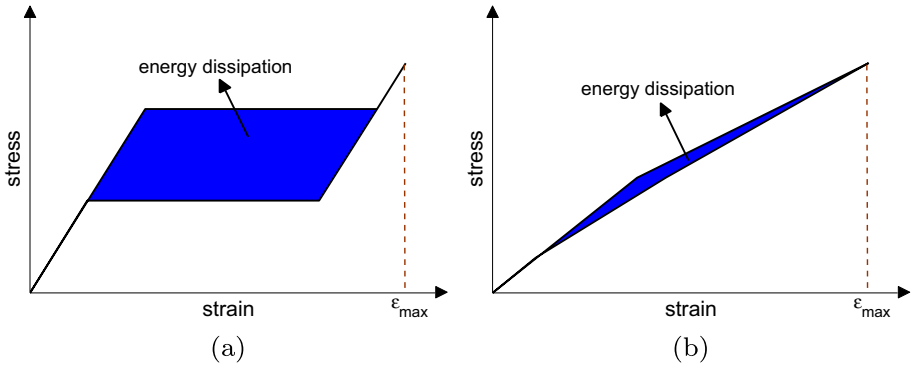

Fig. 1. Stress-strain curves before design and that showing targeted response. (a) Typical stress-strain curve for a shape memory alloy with typically large hysteresis giving rise to large dissipation. (b) The targeted stress-strain response with small hysteresis and dissipation. Our proposed experimental design guides the "next experiments" towards (b).