

Application of Bayesian Optimization and Regression Analysis to Ferromagnetic Materials Development

A. R. Will-Cole¹, A. Gilad Kusne^{2,3}, Peter Tonner², Cunzheng Dong¹, Xianfeng Liang¹,
Huaihao Chen¹, and Nian X. Sun¹, *Fellow, IEEE*

¹Department of Electrical Engineering, Northeastern University, Boston, MA 02115 USA

²National Institute of Standards and Technology, Gaithersburg, MD 20899 USA

³Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742 USA

Bayesian optimization (BO) is a well-developed machine learning (ML) field for black-box function optimization. In BO, a surrogate predictive model, here a Gaussian process, is used to approximate the black-box function. The estimated mean and uncertainty of the surrogate model are paired with an acquisition function to decide where to sample next. In this study, we applied this technique to known ferromagnetic thin-film materials such as ferromagnetic $(\text{Fe}_{100-y}\text{Ga}_y)_{1-x}\text{B}_x$ ($x = 0-21$ and $y = 9-17$) and $(\text{Fe}_{100-y}\text{Ga}_y)_{1-x}\text{C}_x$ ($x = 1-26$ and $y = 2-18$) to demonstrate optimization of structure–property relationships, specifically the dopant concentration or stoichiometry effect on magnetostriction and ferromagnetic resonance linewidth. Our results demonstrated that BO can be deployed to optimize structure–property relationships in FeGaB and FeGaC thin films. We have shown through simulation that using BO methods to guide experiments reduced the number of samples required to statistically determine the maximum or minimum by 50% compared to traditional methods. Our results suggest that BO can be used to save time and resources to optimize ferromagnetic films. This method is transferrable to other ferromagnetic material structure–property relationships, providing an accessible implementation of ML to magnetic materials development.

Index Terms—Bayesian optimization (BO), FeGa alloys, ferromagnetic resonance (FMR), magnetostriction, thin films.

I. INTRODUCTION

THE analysis and optimization of structure–property relations is a critical practice in materials engineering. Traditionally, this process is performed via the Edisonian approach of “trial and error” guided by expert knowledge. While previous success has been realized with “trial and error” methods, the cost is expensive due to long temporal scales (many months to years), depleted resources (funding, equipment use/maintenance, and so on), and researcher exhaustion. Typically, by the time new material is fully integrated into an industrial product, the patent protection for the initial material development nears expiration (novel material commercialization can take up to 20 years, while patents last 30 years)—this is not ideal [1]. Experience and intuition can be beneficial during the materials design process and aid in expediting it; however, this adds implicit bias, and a researcher can become overwhelmed and less effective as the optimization involves higher dimensions, such as multiple structure parameters, i.e., compositions, structures, and grain size.

Statistical and machine learning (ML) algorithms have streamlined optimization in other fields, such as the pharmaceutical industry [2]. Interest in using ML in materials science increased significantly with the turn of the century [3]. In particular, the theory community has embraced advances in ML for screening new material compositions. In addition, computational approaches for modeling flexoelectricity have

been combined with deep neural networks to predict optimal design parameters for composite flexoelectric cantilevers [4]. ML has found success in various material applications, such as identifying novel perovskites, high-efficiency photovoltaic materials, and superconductors [5]–[7]. Experimentalists in the rapid-throughput community have also embraced implementation of ML—these focused on closed- and open-loop optimizations. In a closed-loop optimization, there is no external intervention from the bench researcher. One of the first fully closed-loop processes was applied to improving carbon nanotube growth through an autonomous research system—a closed-loop iterative material experimental tool capable of chemical vapor deposition growth, *in situ* Raman spectroscopy, and random forest/genetic algorithm analysis to inform and execute adjustments to synthesis pressure, temperature, and gas partial pressures [8]. Other approaches for closed-loop optimization have focused on the Bayesian optimization (BO) for composition–structure–property relationships rather than processing–structure relationships. These composition–structure–property relationships were efficiently explored via BO—pulsed laser deposition film composition effect on atomic structure collected by synchrotron beam-line data was determined [9], [10], resulting in the discovery of a best-in-class phase change memory material. Open-loop optimizations require intervention and participation of the bench researcher. There are limited instances of ML implemented in more traditional experimental setting for open-loop optimizations—e.g., photovoltaic cell optimization and synthesis of polymer fibers [11], [12]. With respect to magnetic material optimizations, there have been studies that have focused on ML for predictive properties of soft magnetic materials, but not focused on a BO approach [13], [14].

Manuscript received September 24, 2021; accepted October 28, 2021. Date of publication November 2, 2021; date of current version December 22, 2021. Corresponding author: N. X. Sun (e-mail: n.sun@northeastern.edu).

This article has supplementary material provided by the authors and color versions of one or more figures available at <https://doi.org/10.1109/TMAG.2021.3125250>.

Digital Object Identifier 10.1109/TMAG.2021.3125250

0018-9464 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

Here, we present the BO and regression analysis as a method for optimizing structure (i.e., stoichiometry)—magnetostrictive and micromagnetic property relationships in FeGaB [15], [16] and FeGaC [17] alloy thin films. We have selected the FeGaB and FeGaC alloy systems because they are the best magnetostrictive materials for magnetoelectric (ME) antenna devices and ME sensors [18]–[21]. Thin-film material requirements for ME antennas are a low ferromagnetic linewidth in the GHz and simultaneously high saturation magnetostriction, and thus, we have selected these properties as our objectives of interest. Manipulating these properties requires optimizing a two-alloy system, which can prove challenging for co-sputtering deposition. Thus, here, we demonstrate how ML is a viable tool to enable high-performing ME materials. This optimization process is transferrable to a wide range of ferromagnetic materials and is not deposition method specific.

II. OPTIMIZATION APPROACH

The optimization algorithms used here are BO using the Gaussian process surrogate with radial basis function kernel and expected improvement acquisition function. The specific implementation of these algorithms was derived from the open-source, python libraries developed by the Sheffield ML group. We used GPy—a Gaussian processes framework—and GPyOpt—a Gaussian process optimization library [22], [23]. There are advantages to GPy and GPyOpt, specifically the ability to solve problems with nontrivial solutions, to solve problems in parallel, and to implement multiple variable types [23]. Alternative libraries include Spearmint [24]–[27], BayesOpt [28], RoBo [29], and GPflowOpt [30]. The fundamentals of the algorithms implemented in GPy and GPyOpt are defined in the following.

A. Bayesian Optimization

The target, black-box function to be optimized $f(x)$ (i.e., x can be compositional dopant concentration) may be expensive to sample, making it critical to minimize the number of samples x necessary to probe $f(x)$. In this study, x is composition and $f(x)$ is a magnetic property to optimize. This is the case for optimizing thin-film materials, as substrates, targets, deposition tools, and researcher time are expensive. BO describes a set of ML methods for determining the global optimum (maximum or minimum) in the least number of samples, which makes it advantageous for thin-film material optimizations. In these methods, prior belief of the target function is combined with samples of $f(x)$ to form a posterior probability over $f(x)$'s true behavior. Prior beliefs over the functional form of f are encoded by selecting an appropriate surrogate function—here, we implement Gaussian processes as our surrogate function; this is the most widely accepted surrogate function. However, other research has implemented random forests and neural networks [31], [32]. Gaussian processes formalize the assumptions that measurements of the target function are normally distributed and that a collection of measurements can be represented by a multivariate Gaussian random variable with a mean and covariance, allowing one to impose prior assumptions of function behavior through a

covariance or kernel function. The main benefit of implementing a Gaussian process surrogate function is that the uncertainty of the Gaussian process increases away from training data. As previously mentioned, the prior belief of the GP surrogate function can be tuned via the kernel/covariance matrix selection. The most common or default kernel for Gaussian processes is the radial basis function or exponentiated quadratic. The radial basis function kernel is defined by

$$K_{\text{RBF}}(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right) \quad (1)$$

where ℓ is the length scale and σ^2 is the output variance. A large length scale leads to long-range correlation, while a small length scale leads to short-range correlations; thus, the length scale defines how smooth the function is. When selecting a kernel, informed expertise in the application area should be considered [33]. In the context of materials science, the radial basis function kernel may not be appropriate for the first-order phase transitions, and however, these are appropriate for the second-order phase transitions and gradual disordering without a distinct phase transition. In this case study, we know from previous structural characterization that FeGaB and FeGaC exhibit gradual structural disordering from crystalline to amorphous without distinct, sharp phase transition with increase in boron and carbon concentration in the FeGa alloy system, and thus, a radial basis function is an appropriate kernel for the implementation presented here [15], [17]. Based on the application, it is key to use field expertise and decide on an appropriate kernel or employ “kernel engineering” outlined in several texts [33]–[35]. After the surrogate function is updated, the selection of the next sample x is determined through evaluating an acquisition function, which quantifies the utility of potential measurements or experiments. Here, we implement expected improvement as our acquisition function described in Section II-B—this directs sampling toward regions where an improvement is likely.

The BO loop proceeds for iterations $t = 1, 2, \dots$. Subsequent sampling points x_t are determined by finding the maximum of the acquisition function u derived from the surrogate function

$$x_t = \underset{x}{\operatorname{argmax}} u\{x \mid D_{1:t-1}\} \quad (2)$$

where D is defined as the previously sampled points of data $D = \{x_n, y_n\}$, $n \in \{1, \dots, t-1\}$. The target function $f(x)$ is then sampled at the suggested point x_t , and the value of y_t is recorded. Measurements of black-box functions are typically corrupted by noise, ϵ_t , and thus, y_t is defined by the following equation:

$$y_t = f(x_t) + \epsilon_t. \quad (3)$$

Finally, the surrogate function (here a Gaussian process) is updated by adding the point (x_t, y_t) to data D [36]–[38]. It is important to note here that we have assumed a homoscedastic noise model where constant variance is assumed (i.e., all inputs have the same observational noise). It is also possible with BO to assume heteroscedastic noise in GPyOpt [23].

B. Expected Improvement

Another flexible or customizable component of BO is the acquisition function; the acquisition function tunes the desired amount of exploration versus exploitation in the search space. The selection of the next sample point in the search, x_t , is determined via the acquisition function. The GPyOpt library accommodates various acquisition functions, including, and limited to, expected improvement, Gaussian process lower confidence bound (upper confidence bound for maximization problems), and maximum probability of improvement [22], [23]. We investigated the performance of these three aforementioned acquisition functions and found for our application the convergence of the BO to be similar (see the Supplementary Material), thus here, we have just shown our results using expected improvement as our acquisition function. Expected improvement is defined by the following:

$$\text{EI}(x) = \max(\mathbb{E}[(f(x) - f(x^+), 0)]) \quad (4)$$

where $f(x^+)$ is the value of the previous best sample. The location of the best sample is defined as x^+ , where

$$x^+ = \underset{x_i \in x_{1:t-1}}{\operatorname{argmax}} f(x_i). \quad (5)$$

The expected improvement acquisition function can be evaluated analytically as such in the following equation:

$$\text{EI}(x) = \begin{cases} (\mu(x) - f(x^+) - \zeta)\Phi(Z) + \sigma(x)\phi(Z), & \text{given } \sigma > 0 \\ 0, & \text{given } \sigma = 0 \end{cases} \quad (6)$$

where

$$Z = \begin{cases} \frac{(\mu(x) - f(x^+) - \zeta)}{\sigma(x)}, & \text{given } \sigma > 0 \\ 0, & \text{given } \sigma = 0 \end{cases} \quad (7)$$

and $\mu(x)$, $\sigma(x)$, Φ , and ϕ are the mean, standard deviation, cumulative distribution function, and the probability density function, respectively. Note that x_t from (2) and (3) is set to the value of x^+ calculated here. Expected improvement balances the tradeoff between exploitation and exploration. This tradeoff can be tuned by varying the ζ parameter—a higher ζ parameter enables more exploration in the model [36]. More details on expected improvement can be found in [39]–[41]. With the GPy and GPyOpt libraries, it is possible to integrate custom code for knowledge gradient (generalization expected improvement for design problems with noisy evaluations) [42]–[44], Thompson sampling [45], [46], upper confidence bound [47], and so on.

C. Pareto Optimal

When the goal is to optimize multiple material properties, each represented by a function f_i , and the relative importance of these properties may change, it is often advisable to identify samples that are Pareto optimal. The Pareto optimal set of samples are those which for any particular target property f_i cannot be improved without reducing another target property. The boundary of the Pareto optimal set is known as the

Pareto optimal front. The front will vary based on whether $f_1(x)$ and $f_2(x)$ should both be maximized, minimized, or one maximized and the other minimized. Pareto optimal is fundamental to multiobjective optimization. There are different ways to approach multiobjective optimization, but a typical approach is to implement scalarization techniques [48], [49]. In scalarization, the objective functions are redefined as constraints and then solved as a single-objective problem, such as with linear weighting, weighted Chebychev, and ϵ -constraint method [50], [51]. For larger numbers of objectives in multiobjective problems, the issue arises in measuring convergence of these Pareto-based algorithms. To determine the quality of the Pareto front approximation, a hypervolume indicator is typically deployed [52], [53]. One multiobjective approach is through the use of a generalization of expected improvement and the hypervolume indicator known as the expected hypervolume improvement. The purpose of employing techniques, such as expected hypervolume improvement, is to assess convergence of the estimation of the Pareto optimal set when there are several objective functions [54], [55]. We recommend the chapter by Emmerich *et al.* [56] for thorough discussion of expected hypervolume improvement for estimation of the Pareto front.

III. APPLIED OPTIMIZATION

We have investigated the FeGaB and FeGaC thin-film material systems using the BO ML model with the goal of predicting which compositional candidates/variants to select in order to minimize the experimental process in terms of time, effort, and cost. Specifically, we utilized historical single-source data from our research group's materials inventory in order to optimize the dopant concentration and stoichiometry influence on the material properties of magnetostriction and ferromagnetic resonance (FMR) linewidth (X-band) for FeGaB and FeGaC thin films [15]–[17]. We present our results for three case studies: single-objective optimization and multiobjective optimization over one stoichiometric variable and single-objective optimization for two stoichiometric variables. Prior to our case study analyses, we examine the experimental data, e.g., magnetostriction and FMR linewidth, for the $(\text{Fe}_{100-y}\text{Ga}_y)_{1-x}\text{B}_x$ ($x = 0\text{--}21$ and $y = 9\text{--}17$) material system, as shown in Fig. 1(a) and (b), respectively. The experimental results reveal the optimized correlation of FMR linewidth (minimization) and magnetostriction constant (maximization) as a function of B-doping to be achieved with an atomic fraction of $\approx 12\%$ B. For ease of use with the ML techniques, the structure data were normalized or rescaled to the maximum of the explored dopant range and shown in Fig 1(c) and (d). The Gaussian process model is fit using the GPy python library [22]. To simulate the process of BO, the Gaussian fit shown in Fig. 1(c) and (d) serves as the function for the true experimentally determined behavior of the FeGaB system.

A. Single-Objective Optimization—One Stoichiometric Variable

The magnetostriction can be maximized in the FeGaB system by implementing a 1-D single-objective optimization.

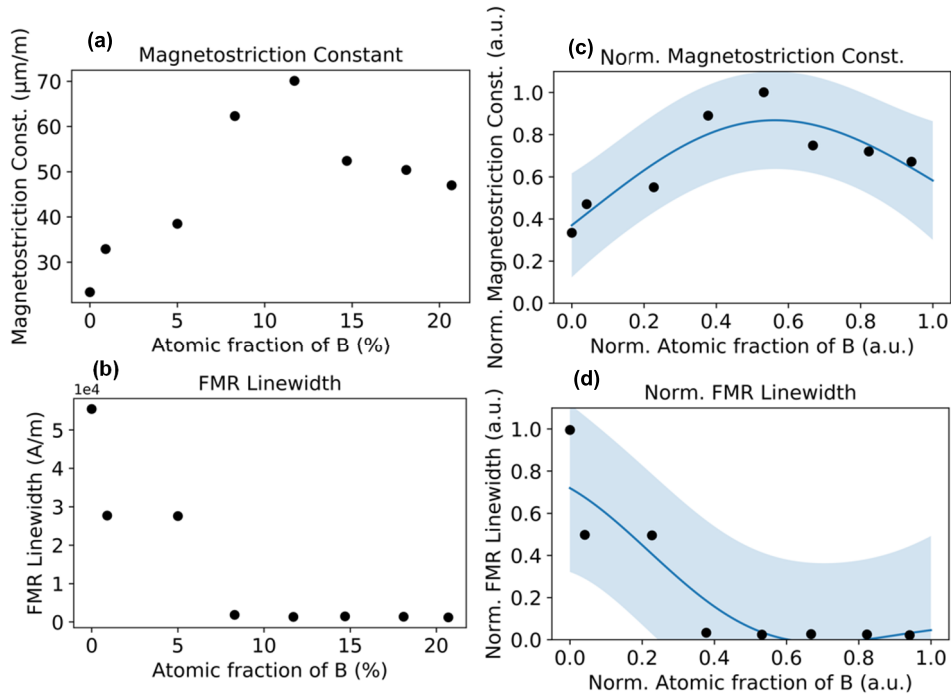


Fig. 1. FeGaB film (a) saturation magnetostriction constant and (b) FMR linewidth as a function of B-doping. (c) Normalized saturation magnetostriction constant. (d) Normalized FMR linewidth as a function of normalized B-doping. Black dots represent known data points, while the blue line is the mean of the Gaussian process fit. $1 \text{ Oe} = (4\pi)^{-1} \times 10^3 \text{ A/m}$.

Here, initial three random seed points, x_t , are selected. The algorithm then fits the surrogate Gaussian process function to these initial data points. The maximum expected improvement acquisition function is the next suggested experiment, x_t , sampled. The new data point (x_t, y_t) is used to update the surrogate function. This process iterates a total of three times after the initial seed points. This is generated with the assistance of the GPyOpt library in python [23]. An example run of this process is shown in Fig. 2(a)–(d). The minimum regret and the root-mean-square error (RMSE) as a function of the number observations and iteration of the optimization, respectively, are shown in Fig. 2(e) and (f). The minimum regret is defined by (8) and the RMSE is defined by (9), where f is the predicted value and o is the observation. In (9), the mean is calculated over the samples t

$$\text{regret} = y_{\text{true max}} - y_{\text{predicted max}} \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{t} \sum_{i=1}^t (f_i - o_i)^2} \quad (9)$$

Note that observations 0–2 represent the randomly selected seed points. After four data points, the minimum regret converges and the maximum is identified. The performance of this process over 100 loops of the optimization as a function of initial random seed points was tested to better understand the variances inherent in this algorithm. The minimum regret and the RMSE for one seed point are shown in Fig. 3(a) and (d), two seed points in Fig. 3(b) and (e), and three seed points in Fig. 3(c) and (f). Across all initial seed point possibilities, the convergence appears to be faster for identifying the maximum than capturing the entire behavior; the convergence of the maximum is related to the minimum regret. It is expected

that the global optimum would be captured more quickly than the full behavior of the system for BO. In addition, with the assistance of the optimization algorithm, the minimum regret converges by the third observation or after four data points (seed points and acquisition function determined points)—this simulation reveals that 50% fewer samples are required than the search conducted with traditional methods to identify the maximum in Fig. 1(a), assuming that there were not any “failed” depositions, e.g., depositional mishaps (such as the sputtering gun shutter becoming stuck) or user error (forgetting to sputter etch the substrate). Traditional methods of optimizing structure–property relationships include linearly varying one-variable-at-a-time (OVAT). In the case of the FeGaB and FeGaC films, previous studies in our group have varied the boron or carbon content linearly and consecutively from 0% to approximately 20% [15]–[17]. Then, the balance of the maximum and minimum of the target properties is identified qualitatively. It has previously been shown that this sequential OVAT experimental approach can misidentify the optimum, especially in the case of correlated parameters [11].

A similar process was repeated in the FeGaB system for minimization of FMR linewidth implementing a 1-D single-objective optimization. The optimization process is simulated similar to the optimization in Fig. 2. This optimization used initial three random seed points and three iterations—it is shown in Fig. 4(a)–(d), with the minimum regret and RMSE in Fig. 4(e) and (f). The performance of this process over 100 loops of the optimization as a function of initial random seed points was tested. The minimum regret and the RMSE for one seed point are shown in Fig. 5(a) and (d), two seed points in Fig. 5(b) and (e), and three seed points in

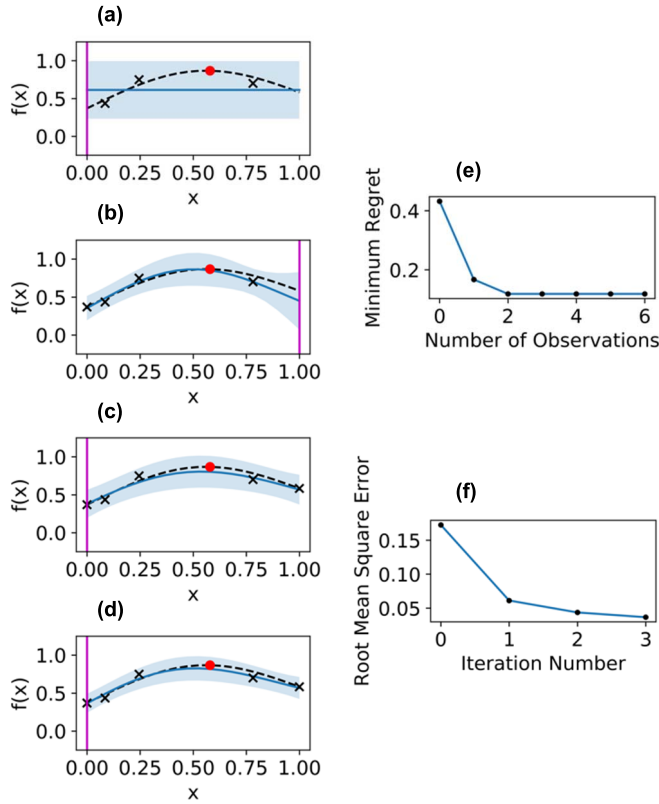


Fig. 2. FeGaB film optimization where $f(x)$ is the normalized magnetostriction and x is the normalized atomic fraction of B. (a) Surrogate function (blue solid line is the mean and shaded region is the 95% confidence interval) and suggested next observation (magenta vertical line), x_t , after the initial random seed points. (b)–(d) Simulated BO process. (e) and (f) Minimum regret and RMSE, respectively.

Fig. 5(c) and (f). Similar to the example simulation for maximization of magnetostriction constant, in the minimization of FMR linewidth case, the minimum regret converges by the third observation or after four data points (seed points and acquisition function determined points)—this is 50% fewer samples than the traditional experiment used to identify the minimum in Fig. 1(b), assuming that there were not any “failed” depositions.

B. Multiobjective Optimization—One Stoichiometric Variable

Often in magnetic film development, there are multiple objectives that need to be optimized based on the specific application. In the application space of high-frequency (1–10 GHz operation) ME mechanical antennas, it is necessary to have high magnetostriction and simultaneously maintain low FMR linewidth [19]. The B-doping concentration in FeGaB films that meet the requirement for high magnetostriction constant and low FMR linewidth can be determined via Gaussian process regression in combination with Pareto analysis. For a 2-D [$f_1(x)$ and $f_2(x)$] Pareto frontier, the front describes the set of options for which you cannot improve on one metric, $f_1(x)$, without diminishing the other metric, $f_2(x)$. Here, we focus on the 2-D case; however, there are analogs for higher dimensions. Here, our $f_1(x)$ and

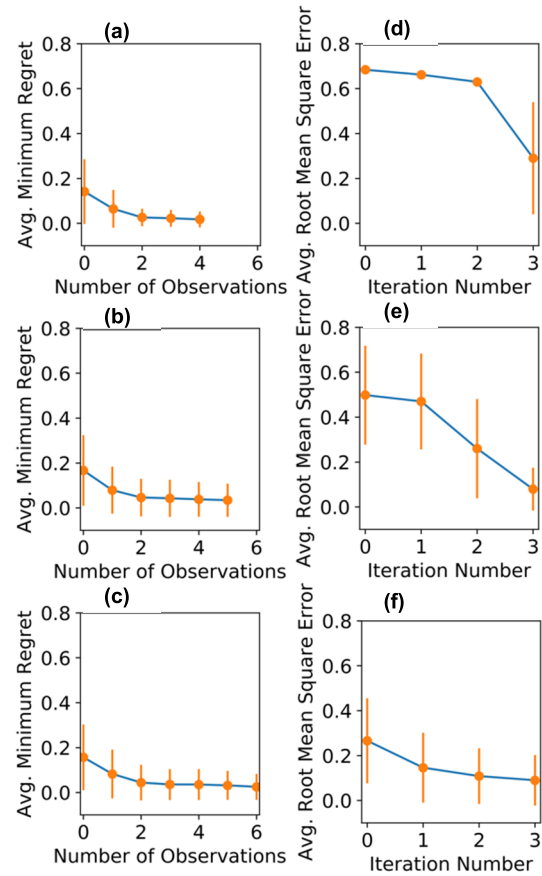


Fig. 3. Optimization of the magnetostriction in the FeGaB system was performed 100 times. The minimum regret was calculated for (a) one seed point, (b) two seed points, and (c) three seed points. The RMSE was calculated for (d) one seed point, (e) two seed points, and (f) three seed points. Note that the mean values and standard deviations are displayed.

$f_2(x)$ are magnetostriction and FMR linewidth, respectively. As previously mentioned, we are maximizing magnetostriction and minimizing FMR linewidth. First, we sort our metrics according to FMR linewidth from the lowest to highest value. Now, our magnetostriction values are sorted according to the FMR linewidth. We then start at the lowest FMR linewidth and its corresponding magnetostriction value. When comparing these values to the next ascending FMR linewidth value and corresponding magnetostriction, we determine whether the next ascending measures have an increased magnetostriction value. If the aforementioned is true, then the previous FMR linewidth and corresponding magnetostriction value belong to the Pareto front. If the next ascending FMR linewidth has a corresponding magnetostriction that is lower, then this is a dominant solution and not a part of the Pareto frontier. This process continues until the highest FMR linewidth and the corresponding magnetostriction are reached. Fig. 6 shows the Pareto plot of $f_1(x)$ versus $f_2(x)$, i.e., magnetostriction and FMR linewidth, respectively. The Pareto frontier, otherwise known points where the magnetostriction is maximized, and the FMR linewidth are simultaneously minimized. The Pareto analysis can serve as a method of multiobjective optimization when there is not a clear figure of merit available to optimize.

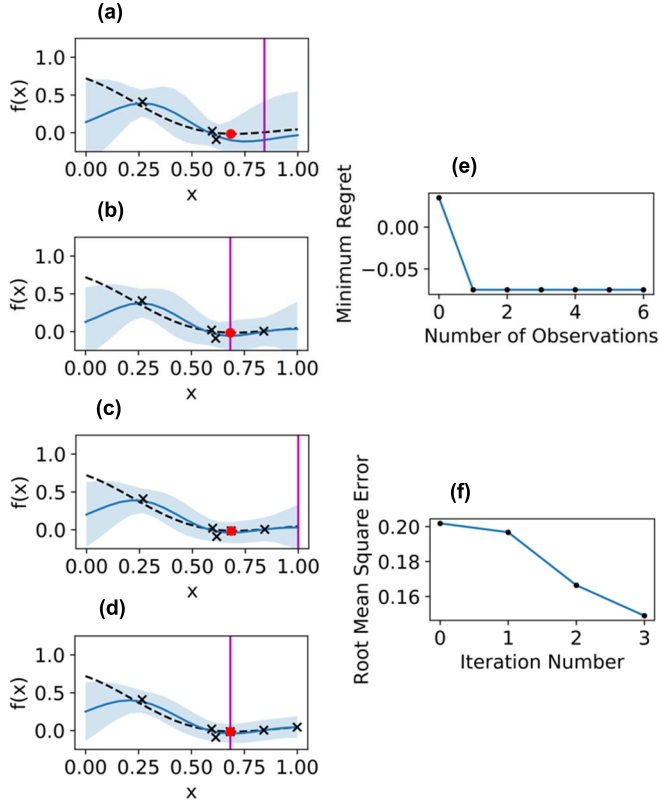


Fig. 4. FeGaB film optimization where $f(x)$ is the normalized FMR linewidth and x is the normalized B-content doping. (a) Surrogate function (blue solid line is the mean and shaded region is the 95% confidence interval) and suggested next observation (magenta vertical line), x_t , after the initial random seed points. (b)–(d) BO process. (e) and (f) Minimum regret and RMSE, respectively.

C. Single-Objective Optimization—Two Stoichiometric Variables

The single-objective optimization in FeGaB system can be expanded for two structure parameters (stoichiometry), $x_{1,i}$ and $x_{2,i}$, i.e., atomic fraction of B and Fe:Ga ratio. Prior to implementation, it is necessary to practice correlation analysis between the two structure parameters. The correlation matrix, r , is determined via the following:

$$r = \frac{\sum (x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2)}{\sqrt{\sum (x_{1,i} - \bar{x}_1)^2 \sum (x_{2,i} - \bar{x}_2)^2}}. \quad (10)$$

The r -value indicates the strength of correlation, whether positive or negative, between the structure parameters. For the data in Fig. 7(a), the off-diagonal terms of the correlation matrix were calculated to be 0.9533, which indicates a strong positive correlation between atomic fraction of B and the Fe/Ga ratio. It is apparent here that the B-doping disproportionately displaces the Ga atoms rather than the Fe atoms in the FeGaB films. In this case with correlated structure parameters, this 2-D single-objective optimization collapses down to a 1-D single-objective optimization problem through implementing linear regression analysis in combination with Gaussian regression. This two-structure single-objective optimization for

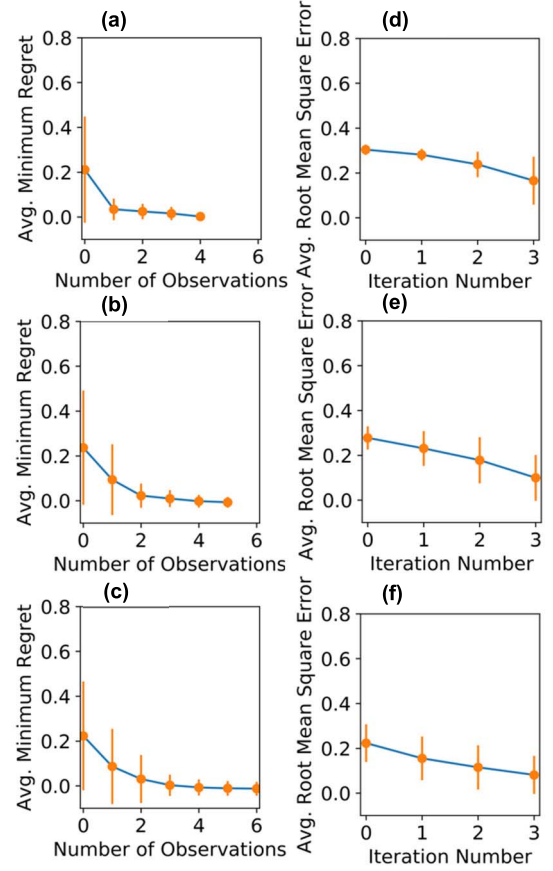


Fig. 5. Optimization of the FMR linewidth in the FeGaB system was performed 100 times. The minimum regret was calculated for (a) one seed point, (b) two seed points, and (c) three seed points. The RMSE was calculated for (d) one seed point, (e) two seed points, and (f) three seed points. Note that the mean values and standard deviations are displayed.

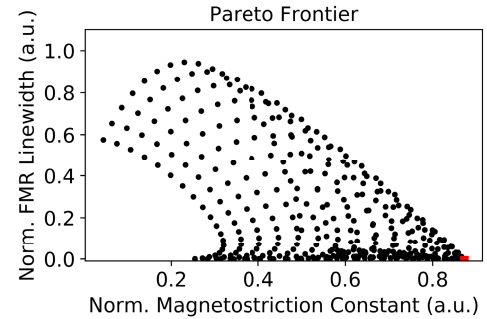


Fig. 6. Pareto frontier for the FeGaB films. The black dots represent the criterion space and the red dots represent the Pareto optimal.

identifying the global maximum of magnetostriction in FeGaB is shown in Fig. 7(b).

In addition, the single-objective optimization in the FeGaC system can be expanded for two structure parameters, $x_{1,i}$ and $x_{2,i}$, i.e., atomic fraction of C and Fe/Ga ratio. For the data in Fig. 7(c), the off-diagonal terms of the correlation matrix were calculated to be 0.0532, which indicates that there is no correlation present. Therefore, the two structure parameters can be treated as independent variables. In Fig. 7(d), the

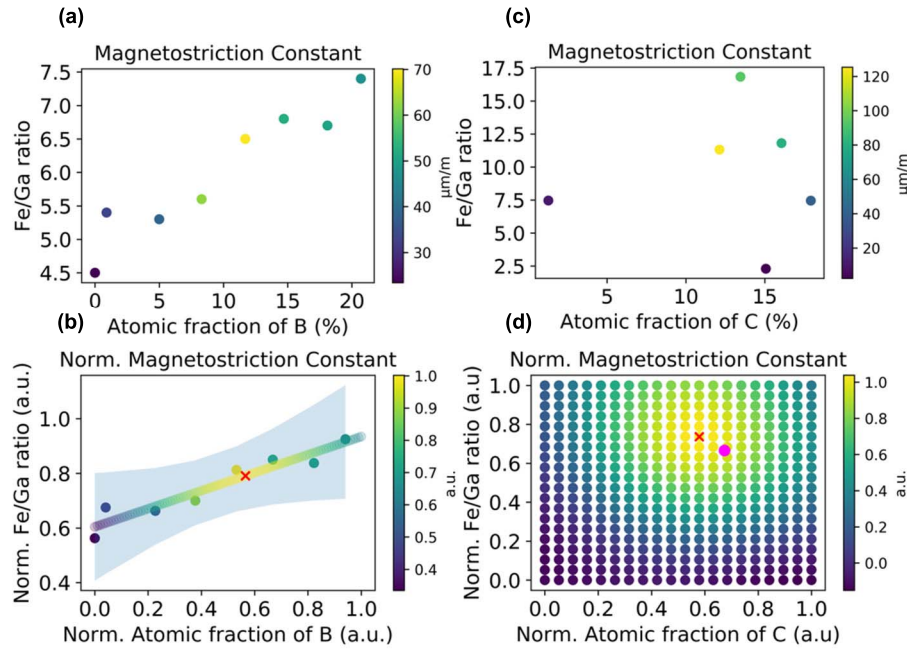


Fig. 7. (a) Atomic fraction of B (%) and Fe/Ga ratio versus the magnetostriction constant. (b) 1-D description of the Gaussian process fit for the magnetostriction as a function of atomic fraction of B considering a linear correlation between atomic fraction of B and Fe/Ga ratio. The normalized data as well as the Gaussian process are displayed. The shaded region represents the 95% confidence bound for the linear regression fit between the normalized atomic fraction of B and the normalized Fe/Ga ratio. (c) Atomic fraction of C (%) and Fe/Ga ratio versus the magnetostriction constant. (d) 2-D description of the Gaussian process fit for magnetostriction as a function of both atomic fraction of C and Fe/Ga ratio. The red “x” denotes the predicted maximum, and the magenta dot represents the next suggested experiment to run based on the evaluation of the acquisition function.

data were normalized and fit to a surrogate function with both the structure parameters as input variables. The predicted maximum is indicated with a red “x.” The next suggested experiment ($x_{1,t}$ and $x_{2,t}$) is determined with the acquisition function; it is shown with a magenta dot. This is an example how to implement BO for higher order structure parameters to optimize material properties.

IV. CONCLUSION

Active learning has been shown highly effective in materials science optimization problems [57], [58]. Here, we have shown how BO and regression analysis (i.e., Gaussian processes and linear regressions) can be implemented to optimize structure–property relationships in FeGaB and FeGaC thin films. With simulating BO methods to guide experiments, the number of samples required to statistically determine the maximum or minimum was reduced by 50% compared with traditional methods. Therefore, BO can be used to save time and resources to optimize ferromagnetic films. The Pareto analysis was used for multiobjective cases and shown to be a viable replacement for a figure of merit. Here, we demonstrated a multiobjective case of low FMR linewidth and high magnetostriction, which is ideal for ME antennas. In addition, we showed some caveats regarding correlated structure parameters—with correlated parameters 2-D or higher dimension problem can be simplified to lower dimension problem. These methods are accessible and can be implemented in a laboratory setting without the use of specialty high throughput tools or a degree in computer science. Through user friendly,

Jupyter notebooks in Python these algorithms can be implemented via open access GPy and GPyOpt libraries [21], [22].

ACKNOWLEDGMENT

A. R. Will-Cole would like to acknowledge the Machine Learning for Materials Research Bootcamp 2020/2021 co-organized by the University of Maryland and the National Institute of Standards and Technology and Dr. M. W. Cole and Dr. Robert Cole for fruitful discussions. The work of A. R. Will-Cole was supported by the National Defense Science and Engineering Graduate Fellowship of the Office of Naval Research. This work was supported in part by the NIH under Award UF1NS107694 and in part by the NSF TANMS ERC under Award 1160504.

REFERENCES

- [1] T. W. Eagar, “Bringing new materials to market,” *Technol. Rev.*, vol. 98, no. 2, p. 42, Feb. 1995.
- [2] D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, and R. K. Tekade, “Artificial intelligence in drug discovery and development,” *Drug Discovery Today*, vol. 26, no. 1, pp. 80–93, Jan. 2021.
- [3] I. Takeuchi, H. Koinuma, E. J. Amis, J. M. Newsam, L. T. Wille, and C. Buelens, “SYMPOSIUM S: Combinatorial and artificial intelligence methods in materials science,” in *Proc. Mater. Res. Soc. Symp.*, Dec. 2002, pp. 358–371.
- [4] K. M. Hamdia, H. Ghasemi, Y. Bazi, H. AlHichri, N. Alajlan, and T. Rabczuk, “A novel deep learning based method for the computational material design of flexoelectric nanostructures with topology optimization,” *Finite Elements Anal. Des.*, vol. 165, pp. 21–30, Nov. 2019.
- [5] G. Pilania, P. V. Balachandran, C. Kim, and T. Lookman, “Finding new perovskite halides via machine learning,” *Frontiers Mater.*, vol. 3, p. 19, Apr. 2016.

- [6] E. O. Pyzer-Knapp, K. Li, and A. Aspuru-Guzik, "Learning from the Harvard clean energy project: The use of neural networks to accelerate materials discovery," *Adv. Funct. Mater.*, vol. 25, no. 41, pp. 6495–6502, Sep. 2015.
- [7] V. Stanev *et al.*, "Machine learning modeling of superconducting critical temperature," *NPJ Comput. Mater.*, vol. 4, no. 1, pp. 1–4, Jun. 2018.
- [8] P. Nikolaev *et al.*, "Autonomy in materials research: A case study in carbon nanotube growth," *NPJ Comput. Mater.*, vol. 2, no. 1, pp. 1–6, Oct. 2016.
- [9] A. G. Kusne *et al.*, "On-the-fly machine-learning for high-throughput experiments: Search for rare-earth-free permanent magnets," *Sci. Rep.*, vol. 4, no. 1, pp. 1–7, Sep. 2014.
- [10] A. G. Kusne *et al.*, "On-the-fly closed-loop materials discovery via Bayesian active learning," *Nature Commun.*, vol. 11, no. 1, pp. 1–11, Nov. 2020.
- [11] B. Cao *et al.*, "How to optimize materials and devices via design of experiments and machine learning: Demonstration using organic photovoltaics," *ACS Nano*, vol. 12, no. 8, pp. 7434–7444, Jul. 2018.
- [12] C. Li *et al.*, "Rapid Bayesian optimisation for synthesis of short polymer fiber materials," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, Jul. 2017.
- [13] R. Jha, N. Chakraborti, D. R. Diercks, A. P. Stebner, and C. V. Ciobanu, "Combined machine learning and CALPHAD approach for discovering processing—Structure relationships in soft magnetic alloys," *Comput. Mater. Sci.*, vol. 150, pp. 202–211, Jul. 2018.
- [14] Y. Wang *et al.*, "Accelerated design of Fe-based soft magnetic materials using machine learning and stochastic optimization," *Acta Mater.*, vol. 194, pp. 144–150, May 2020.
- [15] J. Lou, R. E. Insignares, Z. Cai, K. S. Ziemer, M. Liu, and N. X. Sun, "Soft magnetism, magnetostriction, and microwave properties of FeGaB thin films," *Appl. Phys. Lett.*, vol. 91, no. 18, Oct. 2007, Art. no. 182504.
- [16] C. Dong *et al.*, "Characterization of magnetomechanical properties in FeGaB thin films," *Appl. Phys. Lett.*, vol. 113, no. 26, Dec. 2018, Art. no. 262401.
- [17] X. Liang *et al.*, "Soft magnetism, magnetostriction, and microwave properties of Fe-Ga-C alloy films," *IEEE Magn. Lett.*, vol. 10, pp. 1–5, 2019.
- [18] J. Lou *et al.*, "Giant microwave tunability in FeGaB/lead magnesium niobate-lead titanate multiferroic composites," *Appl. Phys. Lett.*, vol. 92, no. 26, Jun. 2008, Art. no. 262502.
- [19] T. Nan *et al.*, "Acoustically actuated ultra-compact NEMS magnetoelectric antennas," *Nature Commun.*, vol. 8, no. 1, pp. 1–8, Aug. 2017.
- [20] Y. Hui, T. Nan, N. X. Sun, and M. Rinaldi, "High resolution magnetometer based on a high frequency magnetoelectric MEMS-CMOS oscillator," *J. Microelectromech. Syst.*, vol. 24, no. 1, pp. 134–143, Feb. 2015.
- [21] M. Li *et al.*, "Ultra-sensitive NEMS magnetoelectric sensor for picotesla DC magnetic field detection," *Appl. Phys. Lett.*, vol. 110, no. 14, Apr. 2017, Art. no. 143510.
- [22] (2012). *GPY: An Open-Source Framework for Gaussian Processes Written in Python*. [Online]. Available: <http://github.com/SheffieldML/GPY>
- [23] (2016). *GPYO: An Open-Source Library for Bayesian Optimization Using GPY*. [Online]. Available: <http://github.com/SheffieldML/GPYO>
- [24] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 2960–2968.
- [25] K. Swersky, J. Snoek, and R. P. Adams, "Multi-task Bayesian optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 1–10.
- [26] J. Snoek, K. Swersky, R. S. Zemel, and R. P. Adams, "Input warping for Bayesian optimization of non-stationary functions," *J. Mach. Learn. Res.*, vol. 32, pp. 1674–1682, Jun. 2014.
- [27] J. Snoek, "Bayesian optimization and semiparametric models with applications to assistive technology," Ph.D. dissertation, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013.
- [28] R. Martinez-Cantin, "BayesOpt: A Bayesian optimization library for nonlinear optimization, experimental design, and bandits," *J. Mach. Learn. Res.*, vol. 15, pp. 3735–3739, Nov. 2014.
- [29] A. Klien, S. Falkner, N. Mansur, and F. Hutter, "RoBo: A flexible and robust Bayesian optimization framework in Python," in *Proc. NIPS Bayesian Optim. Workshop*, Dec. 2017, pp. 1–5.
- [30] N. Knudde, J. van der Herten, T. Dhaene, and I. Couckuyt, "GPflowOpt: A Bayesian optimization library using TensorFlow," 2017, *arXiv:1711.03845*.
- [31] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [32] J. Snoek *et al.*, "Scalable Bayesian optimization using deep neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 2171–2180.
- [33] J. Shawe-Taylor and N. Christianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [34] P. I. Frazier, "A tutorial on Bayesian optimization," 2018, *arXiv:1807.02811*.
- [35] M. Krasser, (2018). *Bayesian Optimization*. [Online]. Available: <http://krasserm.github.io/>
- [36] F. Archetti and A. Candelieri, *Bayesian Optimization and Data Science*. Cham, Switzerland: Springer, 2019.
- [37] A. Zheng and A. Casari, *Feature Engineering for Machine Learning*. Newton, MA, USA: O'Reilly Media, 2018.
- [38] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [39] J. Mockus, "Application of Bayesian approach to numerical methods of global and stochastic optimization," *J. Global Optim.*, vol. 4, no. 4, pp. 347–365, 1994.
- [40] H. J. Kushner, "A new method of locating the maximum point of an arbitrary multiplex curve in the presence of noise," *J. Basic Eng.*, vol. 86, no. 1, pp. 97–106, Mar. 1964.
- [41] J. Mockus, V. Tiesis, and A. Zilinskas, "The application of Bayesian methods for seeking the extremum," *Towards Global Optim.*, vol. 2, pp. 117–129, Dec. 1978.
- [42] P. Frazier, W. Powell, and S. Dayanik, "The knowledge-gradient policy for correlated normal beliefs," *INFORMS J. Comput.*, vol. 21, no. 4, pp. 599–613, Nov. 2009.
- [43] W. Scott, P. Frazier, and W. Powell, "The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression," *SIAM J. Optim.*, vol. 21, no. 3, pp. 996–1026, Jul. 2011.
- [44] P. I. Frazier and J. Wang, "Bayesian optimization for materials design," in *Information Science for Materials Discovery and Design*, vol. 225. Cham, Switzerland: Springer, 2016.
- [45] S. Agrawal and N. Goyal, "Analysis of Thompson sampling for the multi-armed bandit problem," in *Proc. 25th Annu. Conf. Learn. Theory*, 2012, pp. 1–26.
- [46] O. Chapelle and L. Li, "An empirical evaluation of Thompson sampling," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, Dec. 2011, pp. 2249–2257.
- [47] P. Auer, "Using confidence bounds for exploitation-exploration trade-offs," *J. Mach. Learn. Res.*, vol. 3, pp. 397–422, Nov. 2003.
- [48] F. Ruiz, M. Luque, F. Miguel, and M. Del Mar Muñoz, "An additive achievement scalarizing function for multiobjective programming problems," *Eur. J. Oper. Res.*, vol. 188, no. 3, pp. 683–694, Aug. 2008.
- [49] Y. Nikulin, K. Miettinen, and M. M. Mäkelä, "A new achievement scalarizing function based on parameterization in multiobjective optimization," *OR Spectr.*, vol. 34, no. 1, pp. 69–87, Jan. 2012.
- [50] P. Vamplew, R. Dazeley, A. Berry, R. Issabekov, and E. Dekker, "Empirical evaluation methods for multiobjective reinforcement learning algorithms," *Mach. Learn.*, vol. 84, nos. 1–2, pp. 51–80, Jul. 2011.
- [51] K. Van Moffaert, M. M. Drugan, and A. Nowe, "Scalarized multi-objective reinforcement learning: Novel design techniques," in *Proc. IEEE Symp. Adapt. Dyn. Program. Reinforcement Learn. (ADPRL)*, Apr. 2013, pp. 191–199.
- [52] G. Chiandussi, M. Codegone, S. Ferrero, and F. E. Varesio, "Comparison of multiobjective optimization methodologies for engineering problems," *Comput. Math. Appl.*, vol. 63, pp. 912–942, Mar. 2012.
- [53] K. van Moffaert, M. Drugan, and A. Nowe, "Hypervolume-based multi-objective reinforcement learning," in *Proc. Int. Conf. Evol. Multi-Criterion Optim.* Cham, Switzerland: Springer, 2013, pp. 352–366.
- [54] M. Emmerich, "Single and multi-objective evolutionary design optimization assisted by Gaussian random field metamodels," Ph.D. dissertation, Dept. Comput. Sci., Univ. Dortmund, Dortmund, Germany, Oct. 2005.
- [55] K. Yang, M. Emmerich, A. Deutz, and T. Bäck, "Multi-objective Bayesian global optimization using expected hypervolume improvement gradient," *Swarm Evol. Comput.*, vol. 44, pp. 945–956, Feb. 2019.
- [56] M. Emmerich, K. Yang-André, A. Deutz, H. Wang, and C. M. Fonseca, "A multicriteria generalization of Bayesian global optimization," in *Advances in Stochastic and Deterministic Global Optimization* (Springer Optimization and Its Applications). Cham, Switzerland: Springer, 2016, pp. 229–242.
- [57] T. Lookman, P. V. Balachandran, D. Xue, and R. Yuan, "Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design," *NPJ Comput. Mater.*, vol. 5, no. 1, pp. 1–7, Feb. 2019.
- [58] E. J. Braham, R. D. Davidson, M. Al-Hashimi, R. Arróyave, and S. Banerjee, "Navigating the design space of inorganic materials synthesis using statistical methods and machine learning," *Dalton Trans.*, vol. 49, no. 33, pp. 11480–11488, Aug. 2020.