Cost-effective search for lower-error region in material parameter space using multifidelity Gaussian process modeling

Shion Takeno,[1,2] Yuhki Tsukada,[3,4,*] Hitoshi Fukuoka,[3] Toshiyuki Koyama,[3]
Motoki Shiga ●,[2,4,5] and Masayuki Karasuyama ●[1,4,6,†]

[1]*Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology,*
Gokiso-cho, Showa-ku, Nagoya, 466-8555, Japan
[2]*Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan*
[3]*Department of Materials Design Innovation Engineering, Graduate School of Engineering,*
Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan
[4]*JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan*
[5]*Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Gifu University,*
1-1 Yanagido, Gifu, 501-1193, Japan
[6]*Center for Materials Research by Information Integration, National Institute for Materials Science,*
1-2-1 Sengen, Tsukuba, Ibaraki, 305-0047, Japan

Information regarding precipitate shapes is critical for estimating material parameters. Hence, we considered estimating a region of material parameter space in which a computational model produces precipitates having shapes similar to those observed in the experimental images. This region, called the lower-error region (LER), reflects intrinsic information of the material contained in the precipitate shapes. However, the computational cost of LER estimation can be high because the accurate computation of the model is required many times to better explore parameters. To overcome this difficulty, we used a Gaussian-process-based multifidelity modeling, in which training data can be sampled from multiple computations with different accuracy levels (fidelity). Lower-fidelity samples may have lower accuracy, but the computational cost is lower than that for higher-fidelity samples. Our proposed sampling procedure iteratively determines the most cost-effective pair of a point and a fidelity level for enhancing the accuracy of LER estimation. We demonstrated the efficiency of our method through estimation of the interface energy and lattice mismatch between $MgZn_2$ and α-Mg phases in an Mg-based alloy. The results showed that the sampling cost required to obtain accurate LER estimation could be drastically reduced.

I. INTRODUCTION

Material parameters are often estimated by fitting a theory or model to experimentally observed microstructures. For example, the interface energy between precipitate and matrix phases is estimated by fitting the Ostwald ripening model [1] (theoretical formula) to time-series experimental data of the precipitate radius during the coarsening process. Some recent studies estimated material parameters by comparing data regarding microstructure evolution obtained through experiments and simulations [2–5]. Because a precipitate prefers an energetically favorable shape [6–10], information about precipitate shapes is valuable for estimating material parameters. In Mg-based alloys, rod- or plate-shaped precipitates with various aspect ratios have been observed [11–31]. Moreover, precipitate shapes can be predicted using some advanced computational models if the interface energy and lattice mismatch between the precipitate and matrix phases are given [32–35]. Hence, fitting the computational models to experimental data on precipitate shape enables us to estimate material parameters. However, parameter estimation based on precipitate shapes is time consuming because the computational cost for predicting precipitate shapes is high.

Therefore, to mitigate this problem, we recently introduced a *Gaussian process* (GP)-based selective sampling procedure for material parameter estimation from precipitate shapes [36]. Figure 1 shows a schematic illustration of this approach. When we have a computational model that predicts the energetically favorable shape of the precipitate under given material parameters, we can calculate the discrepancy between the precipitate shape observed in the experiment and that predicted using the computational model. Because experimental data on precipitate shapes are naturally uncertain, the exact minimum of the discrepancy is not necessarily a unique optimal parameter. Instead, the *lower-error region* (LER) of the material parameter space, in which the discrepancy is smaller than a given threshold, is estimated. By determining the threshold from the variance of the precipitate shapes in the experiment, LER estimation can provide a region with reasonable parameters that can be consistent with the current experimental results.

*tsukada.yuhki@material.nagoya-u.ac.jp
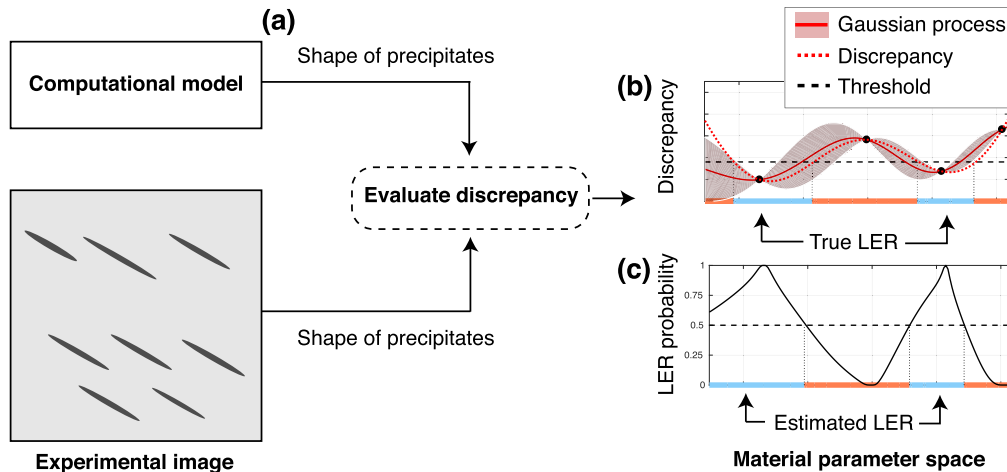†karasuyama@nitech.ac.jp

FIG. 1. LER estimation by GP. (a) Discrepancy between the computational model and experimental image is evaluated through the difference in the precipitate shapes. (b) Using few observed discrepancy values (black circles) in the material parameter space, GP regression approximates discrepancy surface. The red dotted line is the underlying true discrepancy that is unknown beforehand. The solid line and shaded regions represent the GP regression and its predictive variance, respectively. (c) Probability of LER estimated by the GP model. From the GP, the probability that each material parameter has a discrepancy value smaller than the threshold can be estimated. If the probability is higher than 0.5, the region is estimated as LER.

Although GP-based LER identification can be much more efficient than the exhaustive search or naïve random sampling methods, obtaining accurate shapes of the precipitates at every iteration requires a considerably high computational cost. However, by controlling the accuracy of numerical computations, we can also obtain approximate discrepancy values with much lower computational costs. In a computational model for predicting precipitate shapes [35], the total energy (sum of strain and interface energies) of a spheroidal precipitate is formulated as a function of the precipitate aspect ratio r if the material parameters are given. By computing the total energy using different values of r, the equilibrium shape (aspect ratio) of the precipitate that minimizes the total energy can be predicted. If we change the step size of r in the numerical computation, the tradeoff between the computational cost and accuracy can be controlled. In this study, we considered GP-based LER estimation that adaptively incorporates training data from different levels of approximate calculations. The degree of approximation is called *fidelity*. Although lower-fidelity data contain stronger approximations, it is often useful to narrow the candidate region during early-stage screening in our material parameter exploration. We considered efficiently identifying LER by sampling discrepancies not only from the highest-fidelity calculations but also from lower-fidelity calculations that are much easier to perform.

Multifidelity modeling is a machine-learning (ML) framework that combines inexpensive lower-accuracy data and expensive higher-accuracy data to estimate a model with a lower sampling cost of the training data. By multifidelity modeling, GP integrates different fidelity samples through which similarities among different fidelity functions are automatically estimated, and information from the lower-fidelity functions enhance the inference of the highest-fidelity function. Our cost-effective sampling criterion is based on *information entropy*, which evaluates the uncertainty of the probabilistic estimation. At every iteration, the most cost-effective pair of a sampling point and a fidelity level can be selected for

reducing the uncertainty of LER in terms of information entropy (Supplemental Material Sec. A [37] shows an example on an artificial one-dimensional function). By combining information entropy and multifidelity modeling, our proposed method enables us to estimate LER efficiently by sampling only a small number of points compared with the exhaustive search; in particular, we can avoid sampling of higher-fidelity functions many times that results in high computational costs. Although multifidelity modeling is used in materials science applications such as band-gap predictions [38], our study uses a multifidelity-based exploration algorithm involving material parameters. We applied our proposed method called multifidelity LER (MF-LER) estimation to estimate the interface energy and lattice mismatch between $MgZn_2$ (β_1') and α-Mg phases in an Mg-based alloy, in which we have three different fidelity levels requiring 5, 10, and 60 minutes to compute, respectively. We demonstrated that our approach drastically accelerated the material parameter search by efficiently using lower-fidelity samples. Although we focused on an Mg-based alloy in our study, MF-LER is applicable to other material parameter estimation problems because multifidelity calculations are prevalent in computational materials science, in which the computational cost often becomes a severe bottleneck.

II. METHODS

A. Problem setting

Let r_{expt} be the average of aspect ratios of the precipitate obtained from an experimental image and $r_{\boldsymbol{x}_i,\text{comput}}^{(m)}$ be the aspect ratio predicted using a computational model with the material parameter $\boldsymbol{x}_i \in \mathbb{R}^d$ and the fidelity level $m \in \{1, \dots, M\}$. We assumed a set of N candidates $\{\boldsymbol{x}_i\}_{i=1}^N$ in the material parameter space (for example, grid points uniformly taken in the space). If the higher-fidelity level m is calculated, more accurate results can be obtained though it requires a

higher computational cost. Let $\lambda^{(1)} \leqslant \lambda^{(2)} \leqslant \ldots \leqslant \lambda^{(M)}$ be the sampling cost of each fidelity (computational time of the model). The discrepancy between the aspect ratios obtained from the experimental image and through the computational model is defined by

$$y_{\boldsymbol{x}_i}^{(m)} = \frac{1}{2} \sum_{t \in \mathcal{T}} \left(r_{\text{expt}}(t) - r_{\boldsymbol{x}_i, \text{comput}}^{(m)}(t) \right)^2,$$

where t is the time and $\mathcal{T}$ is a set of times when the shapes of precipitates are experimentally measured.

Suppose that the observed discrepancy contains an independent additive noise term given as follows: $y_{\boldsymbol{x}}^{(m)} = f_{\boldsymbol{x}}^{(m)} + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_{\text{noise}}^2)$. Then, the LER, in which the true discrepancy of the highest-fidelity function $f_{\boldsymbol{x}}^{(M)}$ is less than a given threshold h, is defined as

$$LER = \left\{ i \big| f_{\boldsymbol{x}_i}^{(M)} \leqslant h \right\}.$$

If a large set of the highest-fidelity values of $y_{\boldsymbol{x}_i}^{(M)}$ can be obtained for a variety of $\boldsymbol{x}_i$, LER can be identified accurately. However, this leads to prohibitive computational costs because the fidelity level M needs the highest computational cost, and further, the number of candidate material parameters is often high. Our goal is to identify the LER with the smaller total sampling cost (the sum of $\lambda^{(m)}$ over the sampled points).

B. Multifidelity Gaussian process

Suppose we already have the dataset $\mathcal{D}_n = \{(\boldsymbol{x}_i, y_{\boldsymbol{x}_i}^{(m_i)}, m_i)\}_{i=1}^n$ containing a set of triplets consisting of an input $\boldsymbol{x}_i \in \mathbb{R}^d$, fidelity $m_i \in \{1, \ldots, M\}$, and output $y_{\boldsymbol{x}_i}^{(m_i)} \in \mathbb{R}$. To jointly model different fidelity observations with the GP, we used a multifidelity extension of GP regression (MF-GP) [39], which is also known as a co-kriging model. Let $f^{(M)} \sim GP(0, k_M(\boldsymbol{x}, \boldsymbol{x}'))$ be the GP for the highest fidelity $m = M$, in which the prior mean is 0 and the covariance function is $k_M : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ (the covariance function is also called kernel function). Note that we can set the prior mean as 0 without loss of generality [40]. We define the output for the lower fidelity $m = M - 1, \ldots, 1$ recursively from M as follows;

$$f_{\boldsymbol{x}}^{(m)} = f_{\boldsymbol{x}}^{(m+1)} + g_{\boldsymbol{x}}^{(m+1)},$$

where $g^{(m+1)} \sim GP(0, k_g(\boldsymbol{x}, \boldsymbol{x}'))$ with the kernel function $k_g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. The function $g_{\boldsymbol{x}}^{(m+1)}$ represents the difference between $f_{\boldsymbol{x}}^{(m)}$ and $f_{\boldsymbol{x}}^{(m+1)}$. For example, when $M = 3$, we obtain $f_{\boldsymbol{x}}^{(2)} = f_{\boldsymbol{x}}^{(3)} + g_{\boldsymbol{x}}^{(3)}$, and $f_{\boldsymbol{x}}^{(1)} = f_{\boldsymbol{x}}^{(2)} + g_{\boldsymbol{x}}^{(2)} = f_{\boldsymbol{x}}^{(3)} + g_{\boldsymbol{x}}^{(3)} + g_{\boldsymbol{x}}^{(2)}$. The difference between $f_{\boldsymbol{x}}^{(3)}$ and $f_{\boldsymbol{x}}^{(2)}$, which have neighboring fidelity levels, is modeled using the single GP model $g_{\boldsymbol{x}}^{(3)}$. In contrast, the difference between $f_{\boldsymbol{x}}^{(3)}$ and $f_{\boldsymbol{x}}^{(1)}$, whose fidelity levels are more distant from each other, is modeled by the sum of the two GP models $g_{\boldsymbol{x}}^{(3)}$ and $g_{\boldsymbol{x}}^{(2)}$. As a result, in this model, the difference between $f_{\boldsymbol{x}}^{(3)}$ and $f_{\boldsymbol{x}}^{(1)}$ has a larger variance, compared with $f_{\boldsymbol{x}}^{(2)}$ and $f_{\boldsymbol{x}}^{(1)}$.

In MF-GP, the kernel function for a pair of training instances $\{(\boldsymbol{x}_i, y_{\boldsymbol{x}_i}^{(m_i)}, m_i), (\boldsymbol{x}_j, y_{\boldsymbol{x}_j}^{(m_j)}, m_j)\}$ is written as $k((\boldsymbol{x}_i, m_i), (\boldsymbol{x}_j, m_j)) = k_M(\boldsymbol{x}_i, \boldsymbol{x}_j) + (M - \max(m_i, m_j))k_g(\boldsymbol{x}_i, \boldsymbol{x}_j)$ (see Ref. [39] for detail). Using the kernel matrix $\boldsymbol{K} \in \mathbb{R}^{n \times n}$ in which element i, j is defined by $k((\boldsymbol{x}_i, m_i), (\boldsymbol{x}_j, m_j))$, the GP for all fidelities $f^{(1)}, \ldots, f^{(M)}$

can be integrated into one GP in which the predictive mean and variance are obtained as

$$\mu_{\boldsymbol{x}}^{(m)} = \boldsymbol{k}_n^{(m)}(\boldsymbol{x})^\top \left(\boldsymbol{K} + \sigma_{\text{noise}}^2 \boldsymbol{I} \right)^{-1} \boldsymbol{y},$$

$$\sigma_{\boldsymbol{x}}^{2(m)} = k((\boldsymbol{x}, m), (\boldsymbol{x}, m))$$
$$\quad - \boldsymbol{k}_n^{(m)}(\boldsymbol{x})^\top \left(\boldsymbol{K} + \sigma_{\text{noise}}^2 \boldsymbol{I} \right)^{-1} \boldsymbol{k}_n^{(m)}(\boldsymbol{x}),$$

where $\boldsymbol{y} = (y_{\boldsymbol{x}_1}^{(m_1)}, \ldots, y_{\boldsymbol{x}_n}^{(m_n)})^\top$ and $\boldsymbol{k}_n^{(m)}(\boldsymbol{x}) = (k((\boldsymbol{x}, m), (\boldsymbol{x}_1, m_1)), \ldots, k((\boldsymbol{x}, m), (\boldsymbol{x}_n, m_n)))^\top$. By using MF-GP, information from the training data is transferred across different fidelities. As a result, the similarities among different fidelity functions are automatically estimated, and the inference of the highest-fidelity function is enhanced by the lower-fidelity observations.

C. Sampling criterion for LER estimation

Estimating the LER can be considered a classification problem in which each input $\boldsymbol{x}_i$ is classified based on whether it is included in *LER*. Let

$$z_{\boldsymbol{x}} = \begin{cases} 1, & \text{if } f_{\boldsymbol{x}}^{(M)} \leqslant h, \\ 0, & \text{if } f_{\boldsymbol{x}}^{(M)} > h \end{cases}$$

be an indicator variable of the LER classification. From the definition, we obtain $p(z_{\boldsymbol{x}} = 1) = p(f_{\boldsymbol{x}}^{(M)} \leqslant h)$ and $p(z_{\boldsymbol{x}} = 0) = p(f_{\boldsymbol{x}}^{(M)} > h)$. If $p(z_{\boldsymbol{x}} = 0)$ and $p(z_{\boldsymbol{x}} = 1)$ are considerably different, the confidence of the prediction is considered high, while if these two values are close to 0.5, the confidence of the prediction is considered low.

For the dataset $\mathcal{D}_n = \{(\boldsymbol{x}_i, y_{\boldsymbol{x}_i}^{(m_i)}, m_i)\}_{i=1}^n$, the total cost of sampling is $\sum_{i=1}^n \lambda^{(m_i)}$. We considered estimating an accurate z with the small total sampling cost. To evaluate the benefit of sampling from a variety of fidelity levels, we used *information theory* [41]. Let $p(z_{\boldsymbol{x}} | \mathcal{D}_n)$ be the conditional distribution of $z_{\boldsymbol{x}}$, given the training data $\mathcal{D}_n$, and $p(z_{\boldsymbol{x}} | y_{\boldsymbol{x}}^{(m)}, \mathcal{D}_n)$ be the conditional distribution of $z_{\boldsymbol{x}}$, given the training data $\mathcal{D}_n$, and a new observation $y_{\boldsymbol{x}}^{(m)}$. For a pair of m and $\boldsymbol{x}$, the amount of information, called *information gain*, obtained about $z_{\boldsymbol{x}}$ through observing $y_{\boldsymbol{x}}^{(m)}$ is written as

$$I(z_{\boldsymbol{x}}; y_{\boldsymbol{x}}^{(m)}) = H(p(z_{\boldsymbol{x}} | \mathcal{D}_n))$$
$$\quad - \mathbb{E}_{p(y_{\boldsymbol{x}}^{(m)} | \mathcal{D}_n)} \left[H\left(p(z_{\boldsymbol{x}} | y_{\boldsymbol{x}}^{(m)}, \mathcal{D}_n) \right) \right], \quad (1)$$

where H is the information entropy and $\mathbb{E}_{p(y_{\boldsymbol{x}}^{(m)} | \mathcal{D}_t)}$ is the expectation over $y_{\boldsymbol{x}}^{(m)}$. Information entropy H is a standard uncertainty measure of a random variable in information theory, which is defined as $H[p(X)] = \mathbb{E}_{p(X)}[-\log_2(p(X))]$ for a random variable X. In our case, both $p(z_{\boldsymbol{x}} | \mathcal{D}_n)$ and $p(z_{\boldsymbol{x}} | y_{\boldsymbol{x}}^{(m)}, \mathcal{D}_n)$ are Bernoulli distributions. In general, for a Bernoulli distribution $p(z)$ $(z \in \{0, 1\})$, the information entropy is $H(p(z)) = -p(1)\log_2 p(1) - p(0)\log_2 p(0)$ that takes the maximum value 1 when $p(1) = p(0) = 1/2$ (most uncertain) and the minimum value 0 when $p(1) = 0$ and $p(0) = 1$, or $p(1) = 1$ and $p(0) = 0$. Thus, the first term of (1) is the uncertainty of the current $z_{\boldsymbol{x}}$, and the second term is the expected uncertainty after adding the candidate $y_{\boldsymbol{x}}^{(m)}$ into the training data. Because $y_{\boldsymbol{x}}^{(m)}$ is not observed yet, the expectation is taken over the current GP estimation of $p(y_{\boldsymbol{x}}^{(m)} | \mathcal{D}_n)$. In other words, (1) can be seen as the expected uncertainty reduction
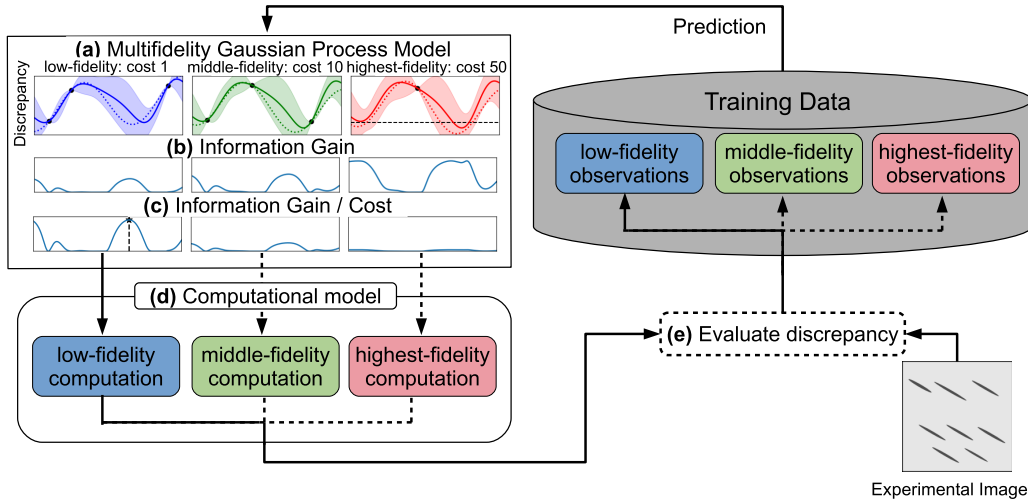
FIG. 2. Schematic illustration of MF-LER. (a) MF-GP provides predictions using all the observations across different fidelities. The cost values for the low-, middle-, and highest- fidelity functions in this illustration are 1, 10, and 50, respectively. (b) Information gain for identifying the LER in the highest-fidelity function is evaluated through the MF-GP model. (c) Information gain is divided by the sampling cost, which enables us to evaluate the cost effectiveness of sampling. (d) The computational model is calculated with the selected fidelity and material parameters (in this illustration, the low-fidelity function is selected). (e) Discrepancy between the computational model and the experimental image is evaluated (through the precipitate shapes), and the result is added to the training data.

after sampling $y_{\boldsymbol{x}}^{(m)}$. Computational details of the information gain are provided in Supplemental Material Sec. B [37].

For higher m, a larger amount of information about $z_{\boldsymbol{x}}$ can be obtained. However, this requires a higher sampling cost. We selected a pair of m and $\boldsymbol{x}_i$ that maximizes the following cost-effectiveness criterion $a(\boldsymbol{x}_i, m)$, in which information gain is divided by the sampling cost of $y_{\boldsymbol{x}_i}^{(m)}$:

$$a(\boldsymbol{x}_i, m) = \frac{I\big(z_{\boldsymbol{x}_i}; y_{\boldsymbol{x}_i}^{(m)}\big)}{\lambda^{(m)}}.$$

Because this criterion represents the amount of information per unit sampling cost, our sampling process can be efficient in terms of the actual computational cost rather than the number of iterations. Figure 2 shows the entire procedure of our method, called MF-LER, in which the most cost-effective pair of a sampling point and a fidelity level is iteratively selected. Further, a demonstration using a simple one-dimensional function is shown in Supplemental Material Sec. A [37].

III. RESULTS

A. Computational model for predicting precipitate shape

We assumed a rod- or plate-shaped precipitate as a spheroid $(x^2/a^2 + y^2/b^2 + z^2/c^2 = 1, a = b, r = c/a)$. The total energy (sum of strain energy and interface energy) of the spheroidal precipitate is formulated as

$$E_{\text{total}}(r) \equiv E_{\text{strain}}(r) + E_{\text{interface}}(r)$$
$$= \frac{V_0}{2} C_{ijkl} \varepsilon_{kl}^0 \big(\varepsilon_{ij}^0 - S_{ijmn}(r)\varepsilon_{mn}^0\big) + A(r)\gamma,$$

where V_0 is the precipitate volume (in m^3), C_{ijkl} is the elastic modulus tensor (in Pa), ε_{ij}^0 is the crystal lattice mismatch between the precipitate and matrix phases, $S_{ijmn}(r)$ is Eshelby's tensor [42], $A(r)$ is the interface area (in m^2), and γ

is the interface energy (in Jm^{-2}). The formulas for computing $S_{ijmn}(r)$ and $A(r)$ can be found in Ref. [35]. When the values of the material parameters are given, we can compute E_{total} as a function of the aspect ratio of the spheroid r, which is changed from 1 to 100 with a step size Δr. Then, the aspect ratio that minimizes E_{total} is r_{comput} that corresponds to the equilibrium precipitate shape. In this study, we computed the aspect ratio of the MgZn$_2$ phase in the α-Mg phase. The ε_{ij}^0 is given by

$$\varepsilon_{ij}^0 = \begin{pmatrix} \varepsilon_{11}^0 & 0 & 0 \\ 0 & \varepsilon_{22}^0 & 0 \\ 0 & 0 & 0.00182 \end{pmatrix},$$

where $\varepsilon_{11}^0 = \varepsilon_{22}^0$ [35,36]. The elastic modulus tensor for the α-Mg phase [43] was used for the computation. Computational models with low-, middle- and highest-fidelity functions were prepared by setting Δr as $10^{-3}, 10^{-4}$, and 10^{-5}, respectively. We considered estimating the interface energy γ and lattice mismatch ε_{11}^0 between the MgZn$_2$ and α-Mg phases, using the experimental data on the changes in the aspect ratio of the rod-shaped MgZn$_2$ phase in an Mg-based alloy aged at 160 °C for 2, 8, and 24 hours [28].

B. Performance evaluation

In our study, we demonstrated the performance of MF-LER using the Mg-based alloy data. On the basis of the analysis in our previous study [36], we set $h = 5$ that was empirically inferred from the standard deviation of the aspect ratio in the experimental image. Figure 3 shows the heat map of discrepancy and LER. We have three fidelity levels $M = 3$, and the sampling costs of the low-, middle-, and highest-fidelity functions are $\lambda^{(1)} = 5$, $\lambda^{(2)} = 10$, and $\lambda^{(3)} = 60$ minutes, respectively. For the candidate parameter $\boldsymbol{x} = (\gamma, \varepsilon_{11}^0)^\top$, we used 250 equally spaced grids in $\varepsilon_{11}^0 \in [-0.250, -0.001]$ and $\gamma \in [0.001, 0.250]$ (Jm^{-2}). Thus, we have a total of $N = 62\,500$
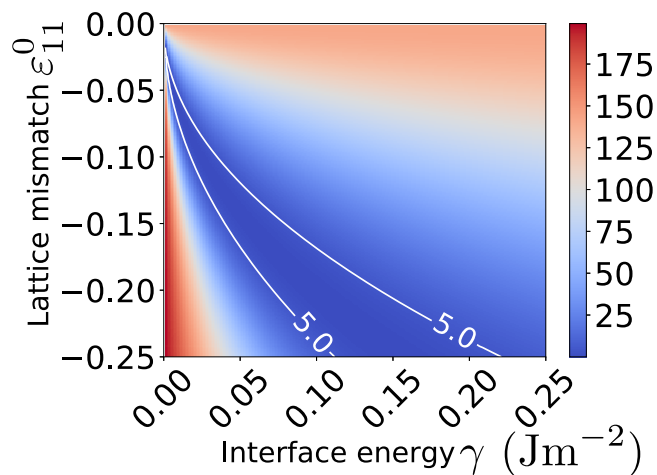
FIG. 3. Heat map of highest fidelity discrepancy and LER defined by threshold $h = 5$.

candidates that require $3\,750\,000 (= 62\,500 \times 60)$ minutes to compute all the points in the highest-fidelity function. To evaluate the usefulness of the low-fidelity observations, we compared MF-LER with two strategies that take samples only from the highest-fidelity function. The first approach is to use information gain (1) as the sampling criterion [36], called single-fidelity LER estimation (SF-LER), and the second approach is single-fidelity GP with random sampling (SF-Random). Note that applying SF-LER and SF-Random only to lower-fidelity functions does not make any prediction about the highest fidelity, because they cannot provide any estimation about differences of multiple fidelities. Further, it also should be noted that since differences among discrepancy surfaces were unknown beforehand, using the predicted LER of the lower-fidelity functions (i.e., $f_{\boldsymbol{x}}^{(m)} < h$ for $m < M$) for the LER estimation of the highest-fidelity function cannot be justified. Therefore, sampling from the highest-fidelity function is indispensable in our problem setting. For the initial points, SF-Random and SF-LER randomly selected the five highest-fidelity points, and MF-LER randomly selected the ten lowest fidelity points. For all approaches, a candidate $\boldsymbol{x}$ is classified as LER if $p(z_{\boldsymbol{x}} = 1) \geqslant 0.5$. Detailed explanations of the GP are given in Supplemental Material Sec. C [37].

Figure 4 shows the sampling processes of MF-LER, SF-LER, and SF-Random. At "Cost 500 (minutes)," MF-LER determines LER approximately, while SF-LER and SF-Random do not estimate LER accurately. At "Cost 1000 (minutes)," MF-LER starts sampling from the highest-fidelity function and identifying LER in more detail. However, the predicted LER of SF-LER and SF-Random are still largely different from the true region. At "Cost 2000 (minutes)," SF-LER starts identifying the rough shape of the LER, and the prediction of SF-Random is not still stable. At this cost, MF-LER provides almost precise LER estimation. The total cost 2000 (minutes) is only approximately 0.05% ($\approx 2000/3\,750\,000 \times 100$) of that used for exhaustive search on the highest-fidelity surface. This suggests that MF-LER is effective to accelerate the search process by reducing the sampling from expensive computations.

Figure 5 shows the quantitative performance evaluation. We evaluated the accuracy of LER estimation through the predicted binary label $z_{\boldsymbol{x}}$. Note that our objective was only to identify LER and not to approximate the entire discrepancy surface accurately that would require a higher number of samples. We used standard evaluation measures of the classification problem called *recall*, *precision*, and *F score*. Because MF-LER sampled only from low-fidelity function values as initial points, unlike the other two methods, the initial cost values of MF-LER in the plot are different from those of SF-LER and SF-Random. All the results are the averages of 10 runs with random initial points.

The left plot in Fig. 5 shows recall, defined by

$$\frac{\text{TP}}{|\text{LER}|},$$

where TP (true positive) is the number of points $i \in \text{LER}$ which has $p(z_{\boldsymbol{x}_i} = 1) \geqslant 0.5$. This is the ratio of the number of LER points that are correctly identified over the number of points in the true LER. This evaluates how many LER points are correctly identified. At the beginning, recall was approximately 0.1 for all sampling strategies owing to the absence of sampled points. However, MF-LER rapidly increased recall substantially faster than SF-LER and SF-Random. The middle plot in Fig. 5 shows precision, defined by

$$\frac{\text{TP}}{\text{TP} + \text{FP}},$$

where FP (false positive) is the number of points $i \notin \text{LER}$ which has $p(z_{\boldsymbol{x}_i} = 1) \geqslant 0.5$. Precision has the same numerator as recall, but the denominator is equal to the number of points predicted as the LER. This evaluates the specificity of prediction, which cannot be considered by recall. The precision values were higher than the recall values in the beginning, indicating that the large part of predicted LER was actually $y_{\boldsymbol{x}_i}^{(M)} \leqslant h$. Similar to recall, MF-LER was better than SF-LER and SF-Random at all the cost values. Because recall and precision have a tradeoff relationship, their harmonic mean, referred to as the F score, is often used as a comprehensive evaluation criterion. The right plot of Fig. 5 shows the superior performance of MF-LER in terms of the F score.

IV. DISCUSSION

A. Uncertainty of material parameters

The lattice mismatch between matrix and precipitate phases is often calculated based on the information of crystallographic lattice correspondence and lattice parameter. However, it is challenging to determine the lattice correspondence between the α-Mg and nanometer-size $MgZn_2$ phases by transmission electron microscopy. Prior to material parameter estimation, we minimized the number of parameters to be estimated. The lattice mismatch along $[0001]_{\alpha}$ could be calculated as $\varepsilon_{33}^0 = 0.00182$, since the lattice correspondence along $[0001]_{\alpha}$ was presumed from the crystallographic orientation relationship between the α-Mg and $MgZn_2$ phases [12]. Furthermore, on the basis of the formation mechanism of $[0001]_{\alpha}$ rod-shaped precipitate [35], the lattice mismatch along the directions perpendicular to $[0001]_{\alpha}$ was assumed as equal to each other $\varepsilon_{11}^0 = \varepsilon_{22}^0$.
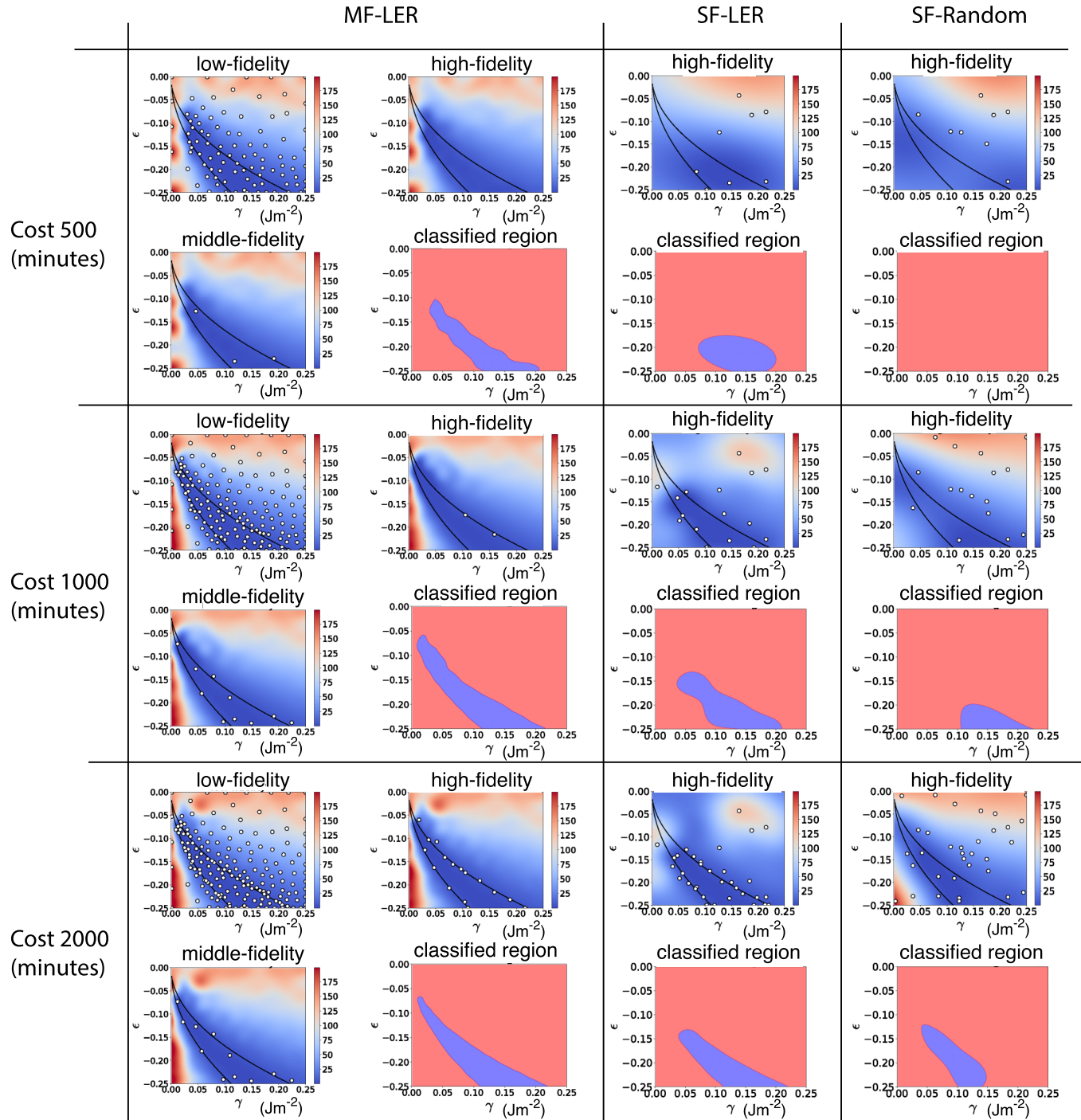
FIG. 4. Illustrative comparison of MF-LER, SF-LER, and SF-Random. For each sampling strategy, results obtained using three different total sampling costs namely 500, 1000, and 2000 minutes are shown. In each sampling cost of MF-LER, three heat maps represent the predictive mean $\mu_x^{(m)}$ for $m = 1$ (upper left), $m = 2$ (lower left), and $m = 3$ (upper right). The right-bottom binary image in MF-LER is the predicted LER [the blue region is $p(z_{x_i} = 1) \geqslant 0.5$]. For SF-LER and SF-Random, the heat map for the predictive mean of the highest-fidelity function and binary image of the predicted LER are shown. In each heat map, the white points represent the observed samples, and the black lines represent the boundary of the LER for $h = 5$.

As described in Sec. III A, the total energy of the precipitate is formulated as the sum of the strain energy and interface energy. The precipitate shape is determined by the interplay between the strain energy and interface energy [44]. The balance between the strain energy and interface energy

can be calculated as

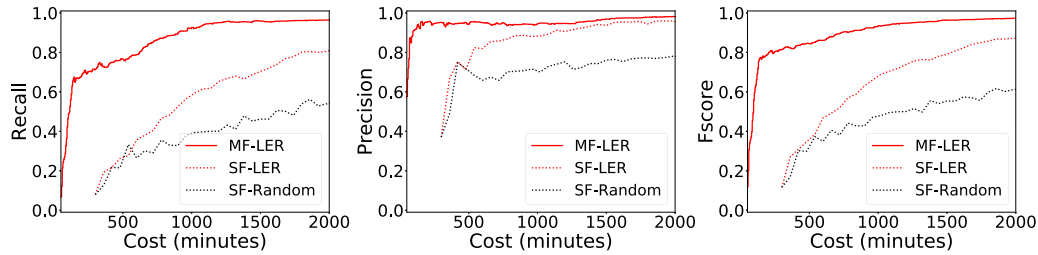$$K = \frac{E_{\text{strain}}}{E_{\text{interface}}} = \frac{BV_0}{A\gamma} = \frac{LB}{\gamma},$$

FIG. 5. Performance evaluation with $h = 5$. The left, middle, and right plots indicate the recall, precision, and F score (see the main text for definition), respectively. The horizontal axis denotes the accumulated sampling cost.

where

$$B = \tfrac{1}{2}C_{ijkl}\varepsilon_{kl}^0\big(\varepsilon_{ij}^0 - S_{ijmn}(r)\varepsilon_{mn}^0\big).$$

Note that V_0 is the precipitate volume, A is the interface area, and $L = V_0/A$ is a typical precipitate size. The equilibrium shape of an isolated precipitate is determined by the value of K [44]. Hence, at a given precipitate volume, a unique parameter set [the lattice mismatch (ε_{ij}^0) and interface energy (γ)] cannot be obtained from experimental data on precipitate aspect ratio. Meanwhile, we conceived that we might obtain a unique parameter set (ε_{ij}^0 and γ) by using experimental data on the changes in the aspect ratio of $MgZn_2$ phase during aging, where L increases with time. This is why we used experimental data in an Mg-based alloy aged at 160 °C for 2, 8, and 24 hours [28]. However, since experimental data on precipitate shapes are naturally uncertain, we estimated LER instead of a unique parameter set (ε_{ij}^0 and γ) that minimizes the discrepancy ($y_{x_i}^{(M)}$) defined in Sec. II A.

The uncertainty (standard deviation) of r_{expt} was used to infer the threshold h in the LER estimation; the LER becomes narrower with decreasing the uncertainty of the experimental measurement and thus lowering the value of h. Although the statistical values (average and standard deviation) of r_{expt} were used to estimate the LER in our proposed method, it might be possible to narrow the LER by using "raw data" of r_{expt} values for individual precipitates measured in an experimental image, which seems to be a critical issue to be addressed in the future.

B. Numerical accuracy of computational model

In the LER estimation, we evaluated the discrepancy ($y_{x_i}^{(m)}$) between the precipitate aspect ratios obtained from the experimental image (r_{expt}) and through the computational model [$r_{x_i,comput}^{(m)}$]. Note that r_{expt} is the average of precipitate aspect ratios measured in an experimental image, and $y_{x_i}^{(m)}$ is evaluated independently of the uncertainty of r_{expt}. Meanwhile, to obtain $r_{x_i,comput}^{(m)}$, the total energy of a spheroidal precipitate (E_{total}) is computed as a function of the aspect ratio r (Sec. III A); the aspect ratio that minimizes E_{total} is $r_{x_i,comput}^{(m)}$. For obtaining accurate values of $r_{x_i,comput}^{(m)}$, the computational error was minimized by setting Δr as 10^{-5}. Then, $\Delta r = 10^{-5}$ was used in the highest-fidelity function to obtain the value of $y_{x_i}^{(M)}$. The fidelity level was controlled by the value of Δr, which was changed independently of the uncertainty of r_{expt}.

We demonstrated the performance of MF-LER using the Mg-based alloy data. The low-, middle-, and highest-fidelity functions were prepared by setting Δr as 10^{-3}, 10^{-4}, and

10^{-5}, respectively. The values of $r_{x_i,comput}^{(m)}$ computed with $\Delta r = 10^{-5}$ were more accurate than those computed with $\Delta r = 10^{-3}$ or $\Delta r = 10^{-4}$. The histograms of the difference between discrepancy surfaces $y_{x_i}^{(1)}$ and $y_{x_i}^{(3)}$ or $y_{x_i}^{(2)}$ and $y_{x_i}^{(3)}$ are shown in Supplemental Material Sec. D [37]. There were non-negligible differences between the discrepancy surfaces computed with the low-, middle-, and highest-fidelity functions. Hence, there is a tradeoff between the computational cost and accuracy in this case. Note, however, that we are unable to know in advance the effect of the value of Δr on the discrepancy surface $y_{x_i}^{(m)}$. Therefore, a high accuracy computation (in this case, $\Delta r = 10^{-5}$) is often required even when the improvement of accuracy is moderate.

In the computational model for predicting the precipitate shape (Sec. III A), we assume that the $MgZn_2$ phase is spheroidal and the anisotropy in the interface energy can be ignored. This assumption is effective for rod- or plate-shaped precipitates observed in Mg-based alloys [11–31]. However, some faceted or lenticular precipitates have been observed [45–48], which cannot be assumed as spheroidal, and their formation mechanism is closely related to the anisotropy in the interface energy [33]. For such precipitates, further advanced computational models for predicting precipitate shape should be used in material parameter estimation. One of the best choices would be to employ a phase-field model [32–34], which can predict the precipitate shape under the influence of anisotropy in both strain energy and interface energy. If different levels of approximate calculations are available, our proposed MF-LER estimation would be useful to efficiently estimate material parameters in the phase-field model using experimental data of precipitate shape.

C. Considerations on machine-learning modeling

In this study, we used three fidelities. As shown in Sec. II B, MF-LER can deal with an arbitrary number of fidelities. However, improvement in efficiency would saturate with too many fidelities, because it may only increase functions which have very similar surfaces and similar costs, though computations of MF-LER linearly increases with the number of fidelities. We empirically conjectured that, at most, five fidelities would be sufficient for most practical problems.

In addition to the number of fidelities, it is also important to determine the lowest fidelity level that is used as the training data. Theoretically, an arbitrary fidelity level can be incorporated, because our method adaptively estimates similarity between different fidelity levels, by which the most

cost-effective fidelity can be selected at every iteration. However, the similarity between fidelities needs to be estimated from the sampled data, and it may cause additional samplings that are not worth the costs if the fidelity is not effective for the LER identification. Therefore, a safer approach is to confine to fidelities that are strongly correlated with the highest-fidelity function (typically, it would be inferred by prior knowledge or preliminary inspection of the data). Although the superior performance of MF-LER has been confirmed in Figs. 4 and 5, it might be advantageous to prepare the low- and middle-fidelity functions in relation to the uncertainty of experimental measurements. As for the Mg-based alloy data used in this study, the threshold h was assumed to have a value between 1 and 5 based on the standard deviation of r_{expt} [36]. In this case, the low- and middle-fidelity functions prepared with 10^{-3} and 10^{-4} might not be necessarily the best setting for maximizing the efficiency of the MF-LER exploration. A different choice of low- and middle-fidelity functions could further increase the performance of MF-LER, which should be examined in the future.

The discrepancy surface in this study is highly smooth (as shown in Fig. 3), and the dimension is not high (two-dimensional parameter space). On the other hand, when the parameters are in a high dimensional space or the surface is discontinuous, estimating the accurate GP regression model becomes difficult. These are open problems of Bayesian optimization in the machine learning community, and we consider that these issues are also important future work for a wide range of the materials informatics researches.

V. CONCLUSIONS

We proposed an ML-based selective sampling procedure for estimating the LER of the material parameter space. The LER is defined using the discrepancy in the precipitate shapes between the computational model and experimental image. To efficiently explore the material parameter space, we introduced multifidelity modeling that can incorporate several levels of approximate samples. Based on the information entropy measure, our sampling method, called MF-LER, can determine the most cost-effective pair of a sample point and a fidelity level at every iteration. We demonstrated the effectiveness of our method by estimating the interface energy and lattice mismatch between $MgZn_2$ and α-Mg phases in an Mg-based alloy. The results show that lower-fidelity data are highly useful for accelerating the LER estimation drastically. Although we used a computational model for predicting precipitate shape, there are various computational models and approaches in materials science, from atomistic length scale to macroscopic length scale; molecular dynamics, phase field, cellular automata, crystal plasticity, finite-element code, where the computational cost often becomes a severe bottleneck. When experimental data are combined with these computational calculations at different levels of fidelity, MF-LER should be useful to estimate model parameters efficiently.

[1] M. Kahlweit, Adv. Colloid Interface Sci. **5**, 1 (1975).

[2] S.-i. Ito, H. Nagao, A. Yamanaka, Y. Tsukada, T. Koyama, M. Kano, and J. Inoue, Phys. Rev. E **94**, 043307 (2016).

[3] S. Ito, H. Nagao, T. Kasuya, and J. Inoue, Sci. Technol. Adv. Mater. **18**, 857 (2017).

[4] J. Zhang, S. O. Poulsen, J. W. Gibbs, P. W. Voorhees, and H. F. Poulsen, Acta Mater. **129**, 229 (2017).

[5] K. Sasaki, A. Yamanaka, S. Ito, and H. Nagao, Comput. Mater. Sci. **141**, 141 (2018).

[6] M. E. Thompson, C. S. Su, and P. W. Voorhees, Acta Metall. Mater. **42**, 2107 (1994).

[7] I. Schmidt and D. Gross, J. Mech. Phys. Solids **45**, 1521 (1997).

[8] I. Schmidt, R. Mueller, and D. Gross, Mech. Mater. **30**, 181 (1998).

[9] A. G. Khachaturyan, *Theory of Structural Transformations in Solids* (Dover, New York, 2008).

[10] D. A. Porter, K. E. Easterling, and M. Y. Sherif, *Phase Transformations in Metals and Alloys 3rd edition* (CRC Press, Boca Raton, 2009).

[11] J. B. Clark, Acta Metall. **13**, 1281 (1965).

[12] J. S. Chun and J. G. Byrne, J. Mater. Sci. **4**, 861 (1969).

[13] J. F. Nie and B. C. Muddle, Scr. Mater. **37**, 1475 (1997).

[14] S. Celotto, Acta Mater. **48**, 1775 (2000).

[15] B. Smola, I. Stulíková, F. Buch, and B. L. Mordike, Mater. Sci. Eng. A **324**, 113 (2002).

[16] D. H. Ping, K. Hono, and J. F. Nie, Scr. Mater. **48**, 1017 (2003).

[17] J. C. Oh, T. Ohkubo, T. Mukai, and K. Hono, Scr. Mater. **53**, 675 (2005).

[18] J. F. Nie, X. Gao, and S. M. Zhu, Scr. Mater. **53**, 1049 (2005).

[19] T. T. Sasaki, K. Oh-ishi, T. Ohkubo, and K. Hono, Scr. Mater. **55**, 251 (2006).

[20] C. L. Mendis, C. J. Bettles, M. A. Gibson, and C. R. Hutchinson, Mater. Sci. Eng. A **435-436**, 163 (2006).

[21] C. L. Mendis, K. Oh-ishi, and K. Hono, Scr. Mater. **57**, 485 (2007).

[22] T. T. Sasaki, T. Ohkubo, and K. Hono, Scr. Mater. **61**, 72 (2009).

[23] K. Oh-ishi, R. Watanabe, C. L. Mendis, and K. Hono, Mater. Sci. Eng. A **526**, 177 (2009).

[24] T. T. Sasaki, K. Oh-ishi, T. Ohkubo, and K. Hono, Mater. Sci. Eng. A **530**, 1 (2011).

[25] C. L. Mendis, K. Oh-ishi, T. Ohkubo, and K. Hono, Scr. Mater. **64**, 137 (2011).

[26] C. L. Mendis, K. Oh-ishi, and K. Hono, Metall. Mater. Trans. A **43**, 3978 (2012).

[27] F. R. Elsayed, T. T. Sasaki, C. L. Mendis, T. Ohkubo, and K. Hono, Mater. Sci. Eng. A **566**, 22 (2013).

[28] T. Bhattacharjee, C. L. Mendis, K. Oh-ishi, T. Ohkubo, and K. Hono, Mater. Sci. Eng. A **575**, 231 (2013).

[29] T. Bhattacharjee, T. Nakata, T. T. Sasaki, S. Kamado, and K. Hono, Scr. Mater. **90-91**, 37 (2014).

[30] T. T. Sasaki, F. R. Elsayed, T. Nakata, T. Ohkubo, S. Kamado, and K. Hono, Acta Mater. **99**, 176 (2015).

[31] T. Nakata, C. Xu, R. Ajima, K. Shimizu, S. Hanaki, T. T. Sasaki, L. Ma, K. Hono, and S. Kamado, Acta Mater. **130**, 261 (2017).

[32] Y. Gao, H. Liu, R. Shi, N. Zhou, Z. Xu, Y. M. Zhu, J. F. Nie, and Y. Wang, Acta Mater. **60**, 4819 (2012).

[33] H. Liu, Y. Gao, J. Z. Liu, Y. M. Zhu, Y. Wang, and J. F. Nie, Acta Mater. **61**, 453 (2013).

[34] Y. Z. Ji, A. Issa, T. W. Heo, J. E. Saal, C. Wolverton, and L.-Q. Chen, Acta Mater. **76**, 259 (2014).

[35] Y. Tsukada, Y. Beniya, and T. Koyama, J. Alloy. Compd. **603**, 65 (2014).

[36] Y. Tsukada, S. Takeno, M. Karasuyama, H. Fukuoka, M. Shiga, and T. Koyama, Sci. Rep. **9**, 15794 (2019).

[37] See Supplemental Material at http://link.aps.org/supplemental/ 10.1103/PhysRevMaterials.4.083802 for details of the machine learning method and settings and data of our case study.

[38] G. Pilania, J. E. Gubernatis, and T. Lookman, Comput. Mater. Sci. **129**, 156 (2017).

[39] M. C. Kennedy and A. O'Hagan, Biometrika **87**, 1 (2000).

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

[41] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, 2003).

[42] T. Mura, *Micromechanics of Defects in Solids 2nd rev. edition* (Martinus Nijhoff, Dordrecht, 1987).

[43] The Japan Institute of Metals and Materials (Ed.), *Kinzoku Data Book 4th rev. edition* (Maruzen, Tokyo, 2004).

[44] Y. Wang, L. Q. Chen, and A. G. Khachaturyan, Scr. Metall. Mater. **25**, 1387 (1991).

[45] P. Vostrý, B. Smola, I. Stulíková, F. Buch, and B. L. Mordike, Phys. Status Solidi A **175**, 491 (1999).

[46] M. Nishijima, K. Hiraga, M. Yamasaki, and Y. Kawamura, Mater. Trans. **47**, 2109 (2006).

[47] M. Nishijima and K. Hiraga, Mater. Trans. **48**, 10 (2007).

[48] M. Nishijima, K. Yubuta, and K. Hiraga, Mater. Trans. **48**, 84 (2007).