# Designing staggered platelet composite structure with Gaussian process regression based Bayesian optimization

Kundo Park[a†], Youngsoo Kim[b†], Minki Kim[a,c], Chihyeon Song[d], Jinkyoo Park[d,*], Seunghwa Ryu[a,**]

[a] Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

[b] Department of Nature-Inspired System and Application, Korea Institute of Machinery and Materials (KIMM), Daejeon 34103, Republic of Korea

[c] Department of Integrated Systems Engineering, The Ohio State University, Columbus, OH, 43210, USA

[d] Department of Industrial & Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

[†] These authors have contributed equally to this work.

**Corresponding authors:**

[*] E-mail: jinkyoo.park@kaist.ac.kr

[**] E-mail: ryush@kaist.ac.kr

# Abstract

The staggered platelet composite structure, one of the most well-known examples of biomimetics, is inspired by the microstructure of nacre, where stiff mineral platelets are stacked with a small fraction of soft polymer in a brick-and-mortar style. Significant efforts have been made to establish a framework for designing a staggered platelet pattern that achieves an excellent balance of toughness and stiffness. However, because no analytical formula for accurately predicting its toughness is available because of the complexity of the failure mechanism of realistic composites, existing studies have investigated either idealized composites with simplified material properties or realistic composites designed by heuristics. In the present study, we propose a Bayesian optimization framework to design a staggered platelet structure that renders high toughness. Gaussian process regression (GPR) was adopted to model statistically the complex relationship between the shape of the staggered platelet array and the resultant toughness. The Markov chain Monte Carlo algorithm was used to determine the optimal kernel hyperparameter set for the GPR. Starting with 14 initial training data collected with uniaxial tensile tests, a GPR-based Bayesian optimization using the expected improvement (EI) acquisition function was carried out. As a result, it was possible to design a staggered platelet pattern with a toughness 11% higher than that of the best sample in the initial training set, and this improvement was achieved after only three iterations of our optimization cycle. As this optimization framework does not require any material theories and models, this process can be easily adapted and applied to various other material optimization problems based on a limited set of experiments or computational simulations.



Keywords: Bioinspired composite, Staggered platelet structure, Gaussian process regression, 3D printing, Bayesian optimization, Toughness.

# 1. Introduction

Ever since the emergence of life on Earth, the complex hierarchy in the microstructure of biological materials has slowly but continuously evolved in order to achieve maximal performance in various living environments. The outstanding properties of these natural materials have inspired scientists and engineers to explore the underlying principles of this unprecedented leap in material properties and even fabricate composite materials by mimicking their micro-architecture, the so-called bioinspired materials [1-5]. Among many biomimetic structural materials, staggered platelet structures mimic the microstructure of nacre, where stiff and brittle ceramic platelets are stacked with a small fraction of soft polymers in a brick-and-mortar style [6-9] (**Figure 1.a**). The efficient load transfer mechanism driven by this unique structure allows the overall composite to have an excellent balance of toughness and stiffness, whereas these two properties are mutually exclusive in common structural materials [10-12]. Fueled by the industrial demand for a material that is both stiff and tough, numerous studies have been conducted to understand the load transfer and failure mechanism of staggered platelet structures and to develop an engineering composite by emulating their key material and geometrical characteristics.

Early studies on staggered platelet structures suggested a universal formula able to predict the effective elastic properties of the composite, given the properties of its constituent materials and the geometrical configuration [13, 14]. Ji and Gao proposed a tension–shear chain model to predict the elastic properties by simplifying the geometry to emphasize the shear interaction between the mineral platelets and protein matrix [15]. Wei et al. derived an improved continuum model by rigorously solving the shear-lag model to predict the elastic properties [16]. The development of the multimaterial 3D printing technology has facilitated the production of composites with complex architectures, allowing researchers to examine the deformation behavior of staggered platelet structures experimentally [17-21]. Kim et al. strengthened the analytical model by considering the volume average of elastic energy and then validated the theory by comparing the prediction with the actual mechanical properties of 3D printed samples [22].

Most analytical formulations related to this particular biomimetic architecture, however, are confined to elastic properties, as the prediction of non-linear mechanical properties such as toughness and strength is far more challenging. Multiple possible failure modes exist for staggered platelet structures, among which the active one is determined by the initial geometrical configurations. The complex interplay of local deformation phenomena leads to various forms of failure, making it very difficult to derive analytically a universal formula that can accurately predict the effective plastic properties. For instance, to express analytically the crack-bridging phenomenon, which is a crucial mechanism that contributes to the high toughness of the staggered platelet structure, Shao et al. developed a microstructure-based fracture mechanics model [23]. Rabiei et al. attempted to model the pullout of platelets, which is another important failure morphology frequently observed in tensile tests of staggered platelet structures [24]. A number of other studies have aimed to understand other toughening phenomena at different length scales of this architecture, such as crack blunting, crack branching, and nucleation of microcracks around the crack tip [24-28]. However, a universal formula that can incorporate the complex interplay of all microscopic physical mechanisms and accurately predict the toughness, given the geometrical specifications, remains elusive. Wei et al. attempted to simplify the nonlinear shear behavior of a soft matrix into an ideal elastic–plastic model to establish an analytical formulation of failure [16]. However, owing to the discrepancy between the simple elastic–plastic model and the actual nonlinear behavior of the material, the formula cannot be used as a reliable objective function for designing realistic composites. Despite the tremendous demand for high-toughness composites in diverse industries, the absence of an analytical predictive model hinders the real-life application of nacre-inspired composites, as the optimal design of the staggered pattern for high toughness is difficult to achieve, and a clearly defined cost function is not available.

In the present study, we propose a machine learning-based optimization framework to design a staggered platelet structure that renders high toughness. Gaussian process regression (GPR), a probabilistic regression model, was adopted to model the complex relationship between the input geometrical parameters of the staggered platelet structure and the resultant toughness. This non-parametric regression model allows us to make a probabilistic prediction of the toughness and also provides the confidence interval of the predicted estimations [29-31]. Moreover, the prediction model can be improved by collecting additional experimental data points. In this study, starting with a training data set that included only 14 experimental data points, GPR-based Bayesian optimization was carried out. An optimization cycle that comprises GPR, the highest expected improvement (EI) point estimation, and additional data collection was repeatedly performed to find the staggered platelet pattern that provides improved toughness. As a result, a toughness improvement of 11% was achieved after only three iterations of our optimization cycle. As this optimization framework does not require any material theories and models, this process can be easily adapted and applied to various other material optimization problems.

# 2. Formulation of staggered-platelet design problem for toughness

This study considers a class of constrained optimization problems that have an objective function given by the experimentally measured toughness of a staggered-platelet pattern. Before defining the design variables and constraints of this optimization problem, it is important to determine the geometrical parameters that systematically express a staggered-platelet pattern. As shown in **Figure 1.b**, a staggered platelet pattern can be designed by repeatedly patterning a unit cell that can be defined by four geometrical parameters ($l_a$, $l_b$, h, b). $l_a$ and b are the parameters associated with the length and thickness of the stiff platelets, while $l_b$ and h are the parameters that characterize the vertical and horizontal gaps between adjacent platelets, respectively. By imposing a geometrical constraint in which the total height of a unit cell is 2 mm, we can define three dimensionless parameters, $\phi = \frac{2b}{2b+h}$ (thickness ratio), $\rho = \frac{l_a}{b}$ (length to thickness ratio), and $\xi = \frac{l_b}{l_a}$ (length ratio), based on the interdependency between the four initial geometrical parameters. In this study, for simplicity, the length ratio $\xi$ is regarded as a constraint, thereby having a fixed value of 0.09. The other dimensionless parameters $\phi$ and $\rho$ are the two design variables to be optimized for a staggered platelet structure with high toughness. Still, we note that the suggested framework can be applied when three parameters are considered. The ranges of the two design parameters are chosen as $0.3 < \phi < 0.7$ and $1 < \rho < 8$, which are sufficiently large to explore the different failure mechanisms reported in our previous study [32].

For the design of a staggered-platelet pattern, another important factor that must be considered is the material selection for the stiff platelets and soft matrix. The toughness of the staggered-platelet composite varies dramatically depending on the mechanical properties of the two constituent materials. As the present research focuses on the optimization of geometrical patterns, the type of materials to be used is regarded as a constraint: Vero PureWhite (RGD837) 3D printing polymer material for the stiff platelets and TangoBlack Plus (FLX980) 3D printing flexible material for the soft matrix. Further details on the material fabrication will be provided in Section 4.1 ‘Preparation of test specimens.’

# 3. Methodology

## 3.1 Gaussian process regression

GPR is a non-parametric (i.e., not limited by a functional form) and probabilistic regression model that can be used to model a complex input–output relationship $y = f(\mathbf{x}) + \epsilon$ based on a limited input data set $\boldsymbol{D} = \{(\mathbf{x}_i, y_i)|i = 1, \dots, n\}$ where the noise $\epsilon$ takes the form of a Gaussian distribution [31, 33]. The advantage of using GPR over other regression models is that it not only makes a reasonable prediction of the output value $f(\mathbf{x}^*)$ with a new input feature $\mathbf{x}^*$, but also provides the confidence interval of the predicted estimation $\epsilon^*$, allowing us to know the accuracy of prediction at different points in the input domain. To predict an output $y^* = f(\mathbf{x}^*) + \epsilon^*$ that corresponds to a new input feature $\mathbf{x}^*$, it is necessary to exploit the property of the Gaussian process in which the outputs $\mathbf{y} = \{y_i|i = 1, \dots, n\}$ and the prediction $\mathrm{y}^*$ will follow a multivariate Gaussian distribution:

$$\begin{bmatrix} \boldsymbol{y} \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix}\right) \qquad (1)$$

Here, $\boldsymbol{K}$ corresponds to a kernel matrix whose components can be expressed as $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, and $\boldsymbol{k}^T$ refers to a kernel vector whose $\mathrm{i}^{\mathrm{th}}$ entry is $k_i = k(\mathbf{x}_i, \mathbf{x}^*)$. The kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$, also called the covariance function, estimates the joint variability of the two input features. In simple terms, it quantifies how the response at point $\mathbf{x}_i$ is affected by the response at other points $\mathbf{x}_j$. In this study, we adopted the Matern 5/2 kernel function to evaluate the covariance of the two input features [34]:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) exp\left(-\frac{\sqrt{5}r}{l}\right) + \delta_{ij}\sigma_{\epsilon,i}^2 \qquad (2)$$

where

$$r = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)} \qquad (3)$$

and $\delta_{ij}$ is a Kronecker delta function.

The value of the kernel function primarily depends on the Euclidean distance $r$ between the two points, meaning that the kernel function value is greater for the two points that are located closer to each other. As can be observed in Eq. (1), the prediction of the output $\mathrm{y}^*$ based on the multivariate Gaussian distribution relationship involves the calculation of the covariance between every possible pair of input features, which allows us to understand that the prediction and its confidence interval are estimated based on all other observation data $\boldsymbol{D}$ with respect to how far the observation data is located from the prediction point $\mathbf{x}^*$.

In the Matern 5/2 kernel function, there exist two positive kernel hyperparameters, which are based on the signal variance $\sigma_f^2$ and characteristic length scale $l$. The signal standard deviation parameter adjusts the magnitude of the overall covariance values, and the characteristic length parameter briefly defines how far the two input values have to be located for their outputs to become mutually uncorrelated. In addition, $\sigma_{\epsilon,i}^2$ accounts for the noise variance of each observation point, which is why the noise variances are added only to the corresponding diagonal of the kernel matrix. If the noise variances of each observation point are not known, the same value of noise variance $\sigma_\epsilon^2$ is given to all the observations, and this value is regarded as an additional kernel hyperparameter to be tuned. The optimal kernel hyperparameter values for proper training of the GPR can be determined by the maximum a posterior (MAP) [35]. In this study, a Markov chain Monte Carlo (MCMC)-based Bayesian inference is carried out to identify the probabilistic distribution of the hyperparameters and determine the best set of hyperparameters that results in maximum posterior.

Based on the standard rules for conditioning Gaussians, the multivariate Gaussian distribution relationship in Eq. (1) can be converted into a predictive distribution of the unknown output $y^*$.

$$p(y^*|\,\mathbf{x}^*, \boldsymbol{D}) = N(y^*|\,\mu^*, \sigma^{2*}) \qquad (4)$$

$$\mu^* = \boldsymbol{k}^T\boldsymbol{K}^{-1}\boldsymbol{y} \qquad (5)$$

$$\sigma^{2*} = k(\mathbf{x}^*, \mathbf{x}^*) - \boldsymbol{k}^T\boldsymbol{K}^{-1}\boldsymbol{k} \qquad (6)$$

As mentioned earlier, the GPR model generally assumes that each of the observation data $\boldsymbol{D} =$

$\{(\mathbf{x}_i, y_i)|i = 1, \dots, n\}$ satisfies the relationship $y_i = f(\mathbf{x}_i) + \epsilon$ and considers that the noise level $\epsilon \sim N(0, \sigma_\epsilon^2)$ is the same for all $n$ observation points. Conversely, a heteroscedastic GPR model can consider different noise variances for the observations, i.e., $\epsilon_i \sim N(0, \sigma_{\epsilon,i}^2)$. The predictive distribution of the unknown output $y^*$ in this case is as follows:

$$p(y^* | \mathbf{x}^*, \boldsymbol{D}) = N(y^* | \mu^*, \sigma^{2*}) \quad (7)$$

$$\mu^* = \boldsymbol{k}^T \boldsymbol{K}_y^{-1} \boldsymbol{y} \quad (8)$$

$$\sigma^{2*} = k(\mathbf{x}^*, \mathbf{x}^*) - \boldsymbol{k}^T \boldsymbol{K}_y^{-1} \boldsymbol{k} \quad (9)$$

where $\boldsymbol{K}_y$ can be obtained by adding each noise variances $\sigma_{\epsilon,i}^2$ to the corresponding diagonal of $\boldsymbol{K}$. The advantage of the heteroscedastic GPR is that the model can assign different weights to each observation with respect to its noise level. The resultant GPR will show a tendency to fit better to observations with smaller noise, while the confidence interval will be large near the observation with significant noise.

### 3.3 Gaussian process regression-based Bayesian optimization framework

Bayesian optimization refers to a sequential design strategy to find the global optimum of an objective function that is expensive to evaluate [36-38]. This optimization scheme is particularly useful for cases where the objective function is a black box with some unknown characteristics, the derivative is not evaluated, and the number of independent variables in the function is small (<10). The overall optimization framework suggested in this study is shown in **Figure 3**. For the preparation of the initial training data set for GPR, we selected 14 different data acquisition points in the rectangular design space, as shown in **Figure 2.a**. To spread the initial data collection points evenly and uniformly over the design space, the locations of 10 points inside the design space were sampled through a uniform Latin hypercube algorithm, and four additional points were placed at the vertices of the design space. Then, the staggered-platelet patterns that correspond to the 14 initial data collection points were built with 3D printing so that the output value, toughness, at each of the input points could be evaluated via uniaxial tensile tests. As the heteroscedastic GPR is a probabilistic regression model that accepts both the mean and variance of the observations, we produced three identical specimens for each of the different patterns.

As we prepared three identical specimens to perform multiple trials of tensile tests for each input point, it was possible to obtain three ln(toughness) values, with which the mean $\mu_i$ and variance $\sigma_i^2$ could be calculated. With the 14 initial training data points $\mathbf{x_i} = (\phi_i, \rho_i)$, $\mathbf{y_i} = (\mu_i, \sigma_i)$ $i = 1, 2, \dots, 14$ prepared, GPR was carried out. The Matern 5/2 kernel function was adopted to evaluate the covariance of the two input features, and the optimal values for the three kernel hyperparameters, $\sigma_f^2$, $l_\phi$, and $l_\rho$, were obtained using the MAP method. To determine each of the hyperparameters that maximizes the posterior, Hamiltonian Monte Carlo (HMC)-based MCMC sampling was carried out, where 18000 steps were sampled based on the above theory with a step size of 0.05, and the first 3000 steps were regarded as a burn-in period [39-41]. The prior assigned for each hyperparameter was a gamma probability function with appropriate shape and rate parameters. Based on the optimal hyperparameter set and the 14 initial training sets, it was possible to characterize a complete GPR, which allowed us to make a reasonable prediction of the toughness at an unknown input feature. The procedures associated with the GPR were implemented in Python, using an open-source Gaussian process framework named GPy developed by the Sheffield machine learning group [42].

To carry out a Bayesian optimization that efficiently explores the design space, expected improvement (EI) was adopted as an acquisition function, and a new data point was added at the location that had the maximum EI value. The EI in general is defined as follows:

$$\mathrm{EI}(\mathbf{x}) = \mathbb{E}\max(f(\mathbf{x}) - f(\mathbf{x}^+), 0) \quad (10)$$

In the case of the Gaussian process model, the EI is analytically defined as follows [43]:

$$\mathrm{EI}(\mathbf{x}) = \begin{cases} w\big(\mu(\mathbf{x}) - f(\mathbf{x}^+)\big)F(Z) + (1-w)\sigma(\mathbf{x})P(Z) & if\ \sigma(\mathbf{x}) > 0 \\ 0 & if\ \sigma(\mathbf{x}) = 0 \end{cases} \quad (11)$$

where

$$Z = \begin{cases} \frac{\mu(\mathbf{x}) - f(\mathbf{x}^+)}{\sigma(\mathbf{x})} & if\ \sigma(\mathbf{x}) > 0 \\ 0 & if\ \sigma(\mathbf{x}) = 0 \end{cases} \quad (12)$$

where $f(\mathbf{x}^+)$ is the toughness value of the best sample thus far, and $\mathbf{x}^+$ is the location of that sample in the design space. $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are the mean and standard deviation of the GPR predicted at $\mathbf{x}$, respectively. $F(Z)$ and $P(Z)$ are the cumulative distribution function and probability distribution function of the standard normal distribution, respectively. As can be understood from the equations, a high EI value can be found at the location that has both high predicted mean $\mu(\mathbf{x})$ and high prediction uncertainty $\sigma(\mathbf{x})$. $w$ is the parameter that allows the user to control the relative importance of the two important aspects of optimization: exploitation and exploration. The exploitation, associated with $\mu(\mathbf{x})$, is a characteristic of the EI function that attempts to suggest the position with the highest predicted toughness. The exploration, associated with $\sigma(\mathbf{x})$, suggests the position where the Gaussian process has high prediction uncertainty, so that augmentation of a data point can effectively improve the quality of the regression. In this research, a $w$ value of 0.5, was used to ensure a good balance of both aspects.

The next design point, $\mathbf{x}'$, can be computed as the point that maximizes the EI acquisition function as follows:

$$\mathbf{x}' = \underset{\mathrm{x}}{\operatorname{argmax}}\, \mathrm{EI}(\mathbf{x}) \qquad (13)$$

To determine the $\mathbf{x}'$ that maximizes the EI, a mesh grid with a spacing of 0.01 in the $\phi$-coordinate and 0.1 in the $\rho$-direction is created over the entire design space, and the EI is evaluated on all the grid points. Among all the grid points, the $\mathbf{x}$ that provides the highest EI value is chosen as the point for the next data acquisition.

Using the suggested design $\mathbf{x}'$ (i.e., the staggered platelet pattern), three new specimens were fabricated, and their toughness was measured via a uniaxial tensile test. Based on the 15 data points obtained thus far (14 initial training data + 1 additional data), a new GPR was performed. Again, the acquisition function was evaluated based on the results of the regression, and the next data acquisition location was determined. Hereafter, this process of repetitively adding a new data point to explore the design space will be called the optimization cycle. The optimization cycle was repeated until a predefined finishing criterion was satisfied. In this study, the optimization cycle was stopped after three repetitions.

# 4. Experimental setup

## 4.1 Preparation of test specimens

To evaluate the toughness of the given patterns, specimens with different staggered-platelet patterns in the middle (**Figure 2.b**) were built with a high-resolution polyjet 3D printing machine (Stratasys J750). We used the Vero PureWhite (RGD837) 3D printing polymer material for the stiff platelets and the TangoBlack Plus (FLX980) 3D printing flexible material for the soft matrix part. The stiff polymer material has an elastic modulus of 2–3 GPa and tensile strength of 50–65 MPa, whereas the soft rubber-like matrix material has a tensile strength of 0.8–1.5 MPa and Shore A hardness of 26–28 [44]. Converting the durometer hardness value into an equivalent elastic modulus based on the empirical correlation suggested by Larson, K., we found that the soft matrix material had an elastic modulus of approximately 0.94–1.04 MPa, which is $1/2000^{th}$ to $1/3000^{th}$ of the elastic modulus of the stiff reinforcement material [45].

As the orientation of 3D printing significantly affects the mechanical properties of the end product, all the specimens used in this study were printed in the same orientation to ensure consistency in the quality of the printed samples [22]. Because of the absence of a tensile test standard associated with polymer–rubber composites with sub-structured patterns, the standard document of ISO26203-2 was employed to design a widely usable dog-bone specimen for various testing conditions [46]. The staggered-platelet pattern at the middle of the dog-bone specimen had a length of 100 mm and width of 20 mm (**Figure 2.b**). The width of the grip part of the specimen was 40 mm, and the total length of the specimen was 200 mm. All the specimens had a thickness of 3 mm. The specimen dimensions were also tested through a series of preliminary experiments to ensure uniform deformation and allow failure to occur at the patterned part of the specimen. In addition, to observe the crack propagation phenomena in the patterned part closely, a single edge pre-crack of 8 mm was formed at the middle of this part.

## 4.2 Uniaxial tensile test

As the heteroscedastic GPR is a probabilistic regression model that accepts both the mean and variance of the observations, we produced three identical specimens for the repeatability of each pattern. To examine the toughness of the 3D printed specimens, uniaxial tensile tests of the specimens were performed using a universal tensile testing machine (Shimadzu AGX-V series, equipped with a 20 kN load cell) (**Figure 2.b**). The specimen protection system of the universal testing machine was utilized to avoid the formation of undesirable damage in the specimen during the mechanical clamping process. The test was performed in the displacement control mode with 6 mm/min crosshead speed and 100 Hz data acquisition frequency. During the tensile tests, the deformation and failure of the printed pattern were captured with a high-resolution digital camera (Sony Alpha a7 II, with resolution of 24.3 megapixels) to examine closely both the failure morphology and crack propagation path. To calculate the toughness, we used a trapezoidal numerical integration function in MATLAB (trapz) to integrate the displacement–load curve obtained from the tensile test.

# 5. Results and discussion

## 5.1 Preparation of initial training data set

As can be observed in **Figure 2.a**, Latin hypercube sampling is adopted to determine the location of 10 data collection points that are spread evenly and uniformly inside the rectangular design space. Four additional data collection points are placed at the vertices of the design space, so that the coverage of the data points over the design space becomes more complete. Fourteen different staggered platelet patterns corresponding to each of the 14 data collection points are also shown in **Figure 2.a**, where the black region indicates the soft matrix material and the white counterpart indicates the stiff platelets. The 3D modeling of specimens with staggered platelet patterns from #1 to #4 is shown as an example in **Figure 2.b**. As can be observed from the photographs taken during the tensile tests (**Figures 2.b** and **4.b**), the patterned region can generally endure a large deformation before the propagation of cracks owing to the flexible nature of the soft matrix material, which has a failure strain of 170%–220% on its own [44]. It can also be noted that, owing to the presence of an initial single-edge precrack, the propagation of the crack is always initiated at the tip of the notch and propagates in the width direction.

The results of all the tensile tests are listed in **Table 1**, and all the load-displacement curves are shown in **Figure S1** in Supporting Information. For each of the 14 different staggered platelet patterns, three trials of tensile tests were carried out, and the mean $\mu$ and variance $\sigma^2$ of ln(toughness) were calculated at the right side of the table. The mean ln(toughness) results are visualized as a 3D scatter plot in **Figure 5.a**.

Depending on the initial staggered platelet pattern, there were two different ways in which crack propagation occurred in the staggered platelet patterns: crack propagation through the stiff platelets, and around the stiff platelets. The two different modes of crack propagation can be easily distinguished by observing the crack path in the fractured specimens (**Figure S2** in Supporting Information). The first mode is the crack propagating straight through the stiff platelets on its way, which is a phenomenon observed in the specimens with high $\rho$ and $\phi$ values: samples #1, #2, #6, #7, and #9. This failure mode generally results in brittle failure of the sample, in which the sample exhibits high failure strength and limited elongation. The black load–displacement curve of sample #2 shown in **Figure 4.a** is an example of the brittle failure behavior of specimens in this category. The second failure mode is characterized by the crack detouring around the stiff platelets in its propagating direction and traveling only in the soft matrix. Such crack propagation occurred in the specimens with low $\rho$ and $\phi$ values: samples #3, #4, #5, #8, #10, #11, #12, #13, and #14. The samples in this category exhibited more ductile behavior during their tensile tests, resulting in a relatively higher failure strain and lower strength. The red load–displacement curve of sample #3 shown in **Figure 4.a** is an example of the ductile failure behavior of samples in this category.

The blue curve in **Figure 4.a** is the load–displacement curve of sample #7, the pattern that had the highest toughness among all 14 initial patterns. The graph shows an excellent balance of strength and ductility, allowing the graph itself to have a larger area under the curve compared to the two other graphs in the same grid. The reason for this high toughness will be discussed in detail in Section 5.4 'Physical understanding of the improved toughness.'

## 5.2 Gaussian process regression and maximum expected improvement geometry

To define a GPR model over the 14 initial observation data, an optimum set of kernel hyperparameters $\sigma_f^2$, $l_\phi$, and $l_\rho$ was obtained by determining the hyperparameter values that maximize the posterior function. The record of 15000 steps of the HMC-based MCMC sampling is shown in **Figure 6**. As the distribution of sampled points statistically represents the probability density of the likelihood function, we determined the location where the sampled points were most densely concentrated and selected the coordinate of that position as our optimal hyperparameter $(\sigma_f^2, l_\phi, l_\rho) = (3.65, 31.78, 5.89)$. With the optimal hyperparameter values determined, we defined the kernel functions to be used for our GPR.

$$k_\phi(\phi_i, \phi_j) = 3.65^2 \left(1 + \frac{\sqrt{5}r_\phi}{31.78} + \frac{5r_\phi^2}{3 \cdot 31.78^2}\right) exp\left(-\frac{\sqrt{5}r_\phi}{31.78}\right) \quad \text{for } \phi \qquad (14)$$

and

$$k_\rho(\rho_i, \rho_j) = 3.65^2 \left(1 + \frac{\sqrt{5}r_\rho}{5.89} + \frac{5r_\rho^2}{3 \cdot 5.89^2}\right) exp\left(-\frac{\sqrt{5}r_\rho}{5.89}\right) \quad \text{for } \rho \qquad (15)$$

where

$$r_\phi = \sqrt{(\phi_i - \phi_j)^2} \text{ and } r_\rho = \sqrt{(\rho_i - \rho_j)^2} \quad (16)$$

Based on the 14 initial observations and the kernel functions with optimal hyperparameters, it was possible to define the GPR, with which we can make a reasonable prediction of toughness (regression mean $\mu$) of any pattern within our design space, as well as the uncertainty of the prediction (regression variance $\sigma^2$). The regression mean is visualized as a 3D surface plot over the entire design space, as shown in **Figure 5.b**, where the height and color of the surface indicate the predicted ln(toughness) value. The regression mean plot allows us to understand that a high-toughness staggered-platelet pattern is likely to exist in the upper right quadrant of the design space (high $\phi$ and $\rho$). Indeed, the highest value of regression mean was located at $(\phi, \rho) = (0.65, 6.5)$ having the value of ln(toughness) = 7.624.

A major advantage of using a Gaussian process-based statistical regression model is that it is possible to predict both the mean $\mu$ and variance $\sigma^2$ of the toughness. While the regression mean helps us to predict the magnitude of ln(toughness), the regression variance allows us to understand the uncertainty of the prediction made at each point. Unlike the regression mean, which can be readily visualized as a surface plot, it is more difficult to visualize the variance of regression in a 3D space. To visualize both the mean and variance of the GPR model, we selected one length-to-thickness ratio value of $\rho = 6.7$ and plotted the 2D GPR graph with respect to the thickness ratio $\phi$ in **Figure 7.a**. The $\rho$ value of 6.7 is intentionally chosen to display the observation data of sample #7 in the graph, which has a $(\phi, \rho)$ value of (0.6, 6.7) and is the sample that had the highest toughness value among all 14 of the initial observation points. In the graph, the solid curve indicates the GPR mean, whereas the yellow region indicates the standard deviation of the GPR. As illustrated in the graph, the GPR model, formed based on the observation data points and optimal hyperparameter set, allows us to determine both the statistically predicted value of the toughness and the uncertainty of the prediction at any point in the design space.

To determine the next data acquisition point that allows us to explore the design space most efficiently, the EI is calculated over the entire design space, and the position with the highest EI value is identified. As shown in **Figure 7.a**, the best observation data obtained thus far is sample #7, which has $f(\mathbf{x}^+)$ of 7.59, at $\mathbf{x}^+$ position of (0.6, 6.7). With this reference value, we calculated the EI over the entire domain and visualized the result as a heat map in **Figure 7.b**. As the GPR mean value $f(\mathbf{x})$ in a large portion of the design space is lower than the ln(toughness) value of the best sample, the EI value of 0 is dominant in this plot. The region with positive EI values was observed in the upper right quadrant of the domain, and the highest EI value among them was 0.0162 at $(\phi, \rho) = (0.65, 6.4)$. This maximum EI point has a high GPR mean ($\mu = 7.623$) and an average uncertainty of prediction ($\sigma^2 = 0.0532$), which indicates a high probability of finding a staggered platelet pattern that has improved toughness. This new data acquisition point suggested by our GPR seems intuitive, as the two nearest observation data are samples #7 and #9, which have the highest (7.59) and second highest (7.48) toughness values among all the initial training sets.

## 5.3 Optimization cycles

Based on the next data acquisition point suggested by the GPR, a new tensile test specimen with a staggered platelet pattern of $(\phi, \rho) = (0.65, 6.4)$ is modeled and shown in **Figure 7.c**. To ensure consistency in manufacturing conditions, the new specimen was fabricated with the same type of 3D printer, the same materials, and the same printing orientation setup that we used for the printing of 14 types of samples in the initial training set. Uniaxial tensile tests were carried out with the three newly fabricated samples, and their resultant toughness values are listed in **Table 1** in the Cycle 1 row. The resultant toughness value of 7.448 was indeed lower than the highest toughness in the initial training data, indicating that the toughness improvement was not observed during the first cycle of optimization. However, this new staggered platelet pattern suggested by our GPR model had the 3rd highest toughness value among all the 15 different patterns we have tested so far, which is an optimistic result indicating that our optimization model is capable of making reasonable suggestions.

Based on the 15 observation data we have obtained thus far (14 initial training data and 1 data from Cycle 1), another optimization cycle was implemented by carrying out the same series of procedures, including identification of the optimal kernel hyperparameter, building of the GP-based regression model, and characterization of the next data acquisition point (**Figure 3**). The results of GPR (regression mean, regression

variance, and EI) are shown in **Figure S3** in Supporting Information. As presented in both **Table 1** and **Figure 8**, the new data acquisition point for Cycle 2 is located at $(\phi, \rho) = (0.60, 8.0)$, which is a location where we can hardly expect to find a high toughness as it is distant from the observation data points that have high toughness values. The reason for this location having the highest EI value is not because it has a high GPR mean (the chance of finding a high toughness), but because this position had a good balance of GPR mean and GPR variance (uncertainty in prediction). In the EI-based optimization method, the purpose of obtaining a new data point in such a position is not to find the optimal value, but to improve the quality of the GPR in the most efficient way by adding a reliable data point at a location of high uncertainty. Indeed, the actual toughness value evaluated at this position was only 7.248, but the regression variance at this point was significantly reduced from 0.0471 to 0.0108 by adding this new data point.

Based on the 16 observation data (14 initial training data and 2 additional data) obtained thus far, another optimization cycle was carried out. The highest EI point for Cycle 3 was characterized at $(\phi, \rho) = (0.59, 5.7)$, a position that was predicted to have a high toughness of 7.504 based on the GPR. By fabricating the 3D printed specimen and performing uniaxial tensile tests, it was possible to evaluate the actual ln(toughness) of the newly suggested staggered platelet pattern, which was 7.688. **Figure 8** illustrates the result of the three optimization cycles based on GPR, and it allows us to understand that it is possible to determine a higher toughness staggered-platelet geometry in only three optimization cycles. Although the toughness result in the second optimization cycle was not satisfactory, the introduction of this data point to the GPR model significantly improved the predictability of the model and allowed us to make an excellent prediction in the following optimization cycle.

In summary, to optimize the staggered platelet pattern for higher toughness, we adopted GPR model that uses EI as an acquisition function to perform Bayesian optimization. The proposed optimization strategy made it possible to explore the design space efficiently and determine an improved staggered platelet pattern with the addition of only three more data points. Such efficient optimization steps could be made because each data acquisition point suggested by the EI function is not overly obsessed with finding the highest toughness point in its given regression; instead, it attempts to improve the predictability of the regression model when it is necessary. The second data point added to our GPR model is an example of data augmentation oriented toward improving the predictability of the regression. As a result, owing to the above-mentioned strength of our optimization strategy, we could determine a staggered platelet pattern that had improved toughness in only three optimization cycles.

The strength of our GPR based optimization framework is that it incorporates and reflects the 'reality' in various ways. First, our optimization is performed using data obtained from an actual experiment. As the acquisition of toughness data through actual experiments is both time consuming and expensive, existing studies have attempted to use numerical simulations or simplified analytical models to estimate the toughness of staggered platelet structures. Although the optimization of the structure may be possible with such data acquisition methods, the optimized result has a significant credibility problem, as neither simulation nor analytical formula can perfectly model the actual deformation behavior of a staggered platelet structure. In contrast, our optimization framework uses the data points obtained through an actual tensile test of 3D printed samples and, subsequently, the reliability of the data points is ensured. Furthermore, in reality, the toughness values evaluated through the tensile test show slight variations from sample to sample, as the testing environment cannot be maintained constant. The heteroscedastic GPR model adopted in our optimization framework is capable of considering different noise variances for each of the observation data, effectively incorporating the noise in real-life measurements.

## 5.4 Physical understanding of the improved toughness

The resultant toughness of a staggered platelet pattern depends predominantly on its failure mechanism. As previously discussed in Section 5.1 'Preparation of the initial training dataset,' there are two different failure modes. In the first mode, the crack propagates straight through the stiff platelets on its way, generally exhibiting a high failure strength but low ductility. This behavior was observed in the specimens with high $\rho$ and $\phi$ values, which correspond to a pattern that has long and thick stiff platelets stacked with a narrow soft matrix layer between them. The second failure mode is characterized by the crack detouring around the stiff platelets in its propagating direction and traveling only in the soft matrix. Such crack propagation occurred in the specimens with low $\rho$ and $\phi$ values, which correspond to a pattern that has short and thin stiff platelets with sufficient interlayer of soft matrix between each other. The samples in this category exhibited more ductile behavior during their tensile tests, resulting in a relatively higher failure strain, but lower strength and elastic modulus.

Under a uniaxial tensile load, two major types of stress arise in a staggered platelet composite structure: the nominal stress in stiff platelets and the shear stress that occurs in the soft matrix between the platelets [22, 32]. If

either of the two stresses reaches its threshold earlier than the other and the corresponding load-bearing capacity is lost, crack propagation is initiated, and the entire composite structure starts to fail. Therefore, a high toughness is guaranteed if the staggered platelet pattern is designed such that both stress-bearing components can endure the load as much as possible and, ideally, their failure threshold stresses are both reached at the moment of failure.

In the first failure mode, where the crack propagates straight through the stiff platelets, the nominal stress in the stiff platelets reaches its failure threshold earlier than the shear stress in the soft interlayer, and thus this mode does not efficiently exploit the stress bearing capacity and superior ductility of the soft matrix. In the second failure mode, where the crack detours around the stiff platelets, the shear stress in the soft matrix material that fills the gap between the stiff platelets reaches its failure threshold earlier than the nominal stress in the stiff platelets. In this case, the load-bearing capacity of the stiff platelets is not fully exploited. Therefore, a high-toughness staggered platelet pattern must be able to disperse the external tensile load efficiently into the stiff platelets and soft matrix, so that the stress threshold of both components can be reached as quickly as possible.

Another important factor that determines the toughness of a staggered platelet composite is the volume ratio between the stiff platelets and the soft matrix. In accordance with an experimental study conducted by Kim et al., the ductility of a staggered platelet composite depends strongly on the volume content of the soft matrix material, as its failure strain is generally much greater than that of the stiff platelet material [32]. In their research, straight crack propagation was observed in staggered platelet samples with a high volume fraction of stiff platelet material, exhibiting a brittle failure mode. In contrast, in samples with a lower platelet volume fraction, the propagation of cracks followed a zigzag pattern, showing improved ductility. Therefore, a high-toughness staggered platelet pattern must have a sufficient amount of soft matrix material to ensure the ductility of the overall structure.

The complex interplay of various other mechanisms associated with the toughness of staggered platelet composites makes it difficult to optimize the structure analytically. The GPR-based structure optimization strategy introduced in the present study allowed us to determine a well-balanced staggered platelet pattern with high toughness. The staggered platelet pattern $(\phi, \rho) = (0.59, 5.7)$ determined in the last cycle of optimization was capable of dispersing the external tensile load into both the stiff platelets and soft matrix, almost fully exploiting the load-bearing capacity of both components. As can be observed from the load–displacement curves shown in **Figure 8.d** (and **Figure S1** in Supporting Information), the staggered platelet pattern after three optimization cycles had an excellent balance of strength and ductility, as this pattern made a near-optimal use of both the stiff and ductile components for the bearing of external load. In addition, this pattern had an adequate volume fraction of soft material, which improved the failure strain of the tested specimen.

## 5.5 Significance of heteroscedastic Gaussian process regression as an experimental data-based regression model

As presented in **Table 1**, the observation data (staggered-platelet patterns) have experimental noise variance values $\sigma^2$ that are quite different from each other. Among all the observations made, the staggered platelet pattern with $(\phi, \rho) = (0.33, 4.1)$ resulted in the highest noise variance (0.06569), which was approximately 200 times greater than the smallest noise variance (0.00029) shown in the specimen $(\phi, \rho) = (0.7, 8.0)$. In the case of the staggered-platelet pattern that showed the highest noise variance, the volume content of the soft matrix material was high and the size of the stiff platelets was very small, allowing the crack to propagate more randomly within the soft matrix. Consequently, although the three tested samples had identical staggered-platelet patterns, the resultant crack paths were different in the three trials and subsequently led to a high noise variance. In contrast, the staggered platelet pattern with the smallest noise variance had a high volume content of stiff material, thereby allowing the crack to propagate only in a straight manner. Subsequently, the failure morphology of the tested samples was exactly the same in all three trials, and this staggered-platelet pattern led to a very low noise variance.

For such an experiment-based optimization problem where the noise variance is input-dependent, heteroscedastic GPR is an excellent regression model to adopt. While the standard GPR assumes that all observation data have the same noise variance, the heteroscedastic GPR model is capable of adapting to different noise variance values for each observation point and embedding the noise variance trend into the resultant GPR. In the present study, the resultant heteroscedastic regression showed a tendency to fit better to observations with smaller noise, while the confidence interval was large near the observation with significant noise. Therefore, by employing a heteroscedastic GPR model, we could effectively make a statistical regression to an experimental dataset, and the resultant regression could not only make a reasonable prediction of mean $\mu$, but also effectively model the input-dependent noise variance $\sigma^2$.

# 6. Conclusion

In the present study, a machine learning-based Bayesian optimization framework is introduced for the design of a staggered platelet structure with high toughness. Gaussian process regression (GPR), a non-parametric regression model, was adopted to statistically model the complex relationship between the shape of the staggered platelet array and the resultant toughness. For the regression model, the Matern 5/2 kernel function was adopted, and the optimal kernel hyperparameters were obtained via the maximum a posterior (MAP) method. The Hamiltonian Monte Carlo (HMC)-based Markov chain Monte Carlo (MCMC) method was chosen for the determination of hyperparameter values. After the regression of the data points, the expected improvement (EI) was chosen as the acquisition function to determine the next data acquisition point.

Starting with 14 initial training data collected with uniaxial tensile tests, the above-described optimization process was carried out. After obtaining a new data point through the experiment, another iteration of the optimization cycle was performed until an appropriate finishing criterion was satisfied. In this study, three iterations of the optimization cycle were carried out, and it was possible to design a staggered platelet pattern with a toughness 11% higher than the best sample in the initial training set. As this optimization framework does not require any material theories and models, this process can be easily adapted and applied to various other material optimization problems.

The resultant staggered platelet pattern achieved superior toughness compared to the other data points because the pattern could efficiently disperse the external tensile load into both stiff platelets and the soft matrix, so that the nominal stress threshold of platelets and shear stress threshold of the matrix could be reached almost instantaneously at the moment of crack propagation. As a result, this balanced pattern design efficiently exploited the load-bearing capacity of both the stiff reinforcement and soft matrix. In addition, the final staggered platelet pattern had an appropriate volume ratio between stiff and soft material, so that the overall structure improved the failure strain.

# Acknowledgement

This research was supported by Basic Science Research Program (2019R1A2C4070690) and Creative Materials Discovery Program (2016M3D1A1900038) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) of the Republic of Korea, as well as the KAIST-funded Global Singularity Research Program for 2019 (N11190118).

# 7. Reference


1. Wegst, U.G., et al., *Bioinspired structural materials.* Nature materials, 2015. **14**(1): p. 23-36.
2. Suresh Kumar, N., et al., *A review on biological and biomimetic materials and their applications.* Applied Physics A, 2020. **126**(6): p. 445.
3. Zhang, C., D.A. Mcadams II, and J.C. Grunlan, *Nano/Micro-Manufacturing of Bioinspired Materials: a Review of Methods to Mimic Natural Structures.* Advanced Materials, 2016. **28**(30): p. 6292-6321.
4. Zhao, N., et al., *Bioinspired materials: from low to high dimensional structure.* Advanced Materials, 2014. **26**(41): p. 6994-7017.
5. Suresh Kumar, N., et al., *A review on biological and biomimetic materials and their applications.* Applied Physics A, 2020. **126**: p. 1-18.
6. Katti, K.S. and D.R. Katti, *Why is nacre so tough and strong?* Materials Science and Engineering: C, 2006. **26**(8): p. 1317-1324.
7. Wang, J., Q. Cheng, and Z. Tang, *Layered nanocomposites inspired by the structure and mechanical properties of nacre.* Chemical Society Reviews, 2012. **41**(3): p. 1111-1129.
8. Zhao, H., Z. Yang, and L. Guo, *Nacre-inspired composites with different macroscopic dimensions: strategies for improved mechanical performance and applications.* NPG Asia Materials, 2018. **10**(4): p. 1-22.
9. Yang, W., et al., *Strength analysis of bio-inspired composites reinforced by regularly and randomly staggered platelets.* Composite Structures, 2019. **216**: p. 415-426.
10. Li, J., et al., *Hybrid Hydrogels with Extremely High Stiffness and Toughness.* ACS Macro Letters, 2014. **3**(6): p. 520-523.
11. Wegst, U.G.K., et al., *Bioinspired structural materials.* Nature Materials, 2015. **14**(1): p. 23-36.
12. Liu, Y., et al., *Design of super-tough co-continuous PLA/NR/SiO2 TPVs with balanced stiffness-toughness based on reinforced rubber and interfacial compatibilization.* Composites Science and Technology, 2018. **165**: p. 231-239.
13. Begley, M.R., et al., *Micromechanical models to guide the development of synthetic 'brick and mortar' composites.* Journal of the Mechanics and Physics of Solids, 2012. **60**(8): p. 1545-1560.
14. Farzanian, S. and R. Shahsavari, *Mapping the coupled role of structure and materials in mechanics of platelet-matrix composites.* Journal of the Mechanics and Physics of Solids, 2018. **112**: p. 169-186.
15. Ji, B. and H. Gao, *Mechanical properties of nanostructure of biological materials.* Journal of the Mechanics and Physics of Solids, 2004. **52**(9): p. 1963-1990.
16. Wei, X., M. Naraghi, and H.D. Espinosa, *Optimal Length Scales Emerging from Shear Load Transfer in Natural Materials: Application to Carbon-Based Nanocomposite Design.* ACS Nano, 2012. **6**(3): p. 2333-2344.
17. Libonati, F., et al., *Bone-Inspired Materials by Design: Toughness Amplification Observed*

*Using 3D Printing and Testing* Advanced Engineering Materials, 2016. **18**(8): p. 1354-1363.
18. Mirzaeifar, R., et al., *Defect-Tolerant Bioinspired Hierarchical Composites: Simulation and Experiment.* ACS Biomaterials Science & Engineering, 2015. **1**(5): p. 295-304.
19. Dimas, L.S., et al., *Tough Composites Inspired by Mineralized Natural Materials: Computation, 3D printing, and Testing*. Advanced Functional Materials, 2013. **23**(36): p. 4629-4638.
20. Dimas, L.S. and M.J. Buehler, *Modeling and additive manufacturing of bio-inspired composites with tunable fracture mechanical properties.* Soft Matter, 2014. **10**(25): p. 4436-4442.
21. Sakhavand, N. and R. Shahsavari, *Universal composition–structure–property maps for natural and biomimetic platelet–matrix composites and stacked heterostructures.* Nature Communications, 2015. **6**(1): p. 6523.
22. Kim, Y., et al., *An extended analytic model for the elastic properties of platelet-staggered composites and its application to 3D printed structures.* Composite Structures, 2018. **189**: p. 27-36.
23. Shao, Y., et al., *Discontinuous crack-bridging model for fracture toughness analysis of nacre*. Journal of the Mechanics and Physics of Solids, 2012. **60**(8): p. 1400-1419.
24. Rabiei, R., S. Bekah, and F. Barthelat, *Failure mode transition in nacre and bone-like materials.* Acta Biomaterialia, 2010. **6**(10): p. 4081-4089.
25. Lin, A. and M.A. Meyers, *Growth and structure in abalone shell.* Materials Science and Engineering: A, 2005. **390**(1): p. 27-41.
26. Mayer, G., *Rigid Biological Systems as Models for Synthetic Composites*. Science, 2005. **310**(5751): p. 1144-1147.
27. Espinosa, H.D., et al., *Merger of structure and material in nacre and bone – Perspectives on de novo biomimetic materials.* Progress in Materials Science, 2009. **54**(8): p. 1059-1100.
28. Song, F. and Y.L. Bai, *Effects of nanostructures on the fracture strength of the interfaces in nacre.* Journal of Materials Research, 2003. **18**(8): p. 1741-1744.
29. Cao, Y., et al., *Efficient Optimization for Sparse Gaussian Process Regression.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015. **37**(12): p. 2415-2427.
30. Yuan, J., et al., *Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression.* International Journal of Machine Tools and Manufacture, 2008. **48**(1): p. 47-60.
31. Denzel, A. and J. Kästner, *Gaussian process regression for geometry optimization.* The Journal of Chemical Physics, 2018. **148**(9): p. 094114.
32. Kim, Y., et al., *A three-dimensional fracture pattern diagram of staggered platelet structures.* Composite Structures, 2019. **220**: p. 769-775.
33. Lizotte, D., et al. *Gaussian process regression for optimization*. in *NIPS Workshop on Value of Information*. 2005. Citeseer.
34. Williams, C.K. and C.E. Rasmussen, *Gaussian processes for regression.* 1996.
35. Blum, M. and M.A. Riedmiller. *Optimization of Gaussian process hyperparameters using*

*Rprop*. in *ESANN*. 2013. Citeseer.

36. Snoek, J., H. Larochelle, and R.P. Adams, *Practical bayesian optimization of machine learning algorithms.* arXiv preprint arXiv:1206.2944, 2012.
37. Gardner, J.R., et al. *Bayesian Optimization with Inequality Constraints*. in *ICML*. 2014.
38. Pelikan, M., *Bayesian optimization algorithm*, in *Hierarchical Bayesian optimization algorithm*. 2005, Springer. p. 31-48.
39. Hanson, K.M. *Markov Chain Monte Carlo posterior sampling with the Hamiltonian method*. in *Medical Imaging 2001: Image Processing*. 2001. International Society for Optics and Photonics.
40. Gu, M.G. and H.T. Zhu, *Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2001. **63**(2): p. 339-355.
41. Geyer, C.J., *Practical markov chain monte carlo.* Statistical science, 1992: p. 473-483.
42. SheffieldML. *GPy: A Gaussian process framework in Python*. 2020; Available from: https://github.com/SheffieldML/GPy.
43. Feng, Z., et al., *A multiobjective optimization based framework to balance the global exploration and local exploitation in expensive optimization.* J Glob Optim, 2014. **61**: p. 1-18.
44. Stratasys, *PolyJet Materials Data Sheet*. 2017.
45. Larson, K., *Can you estimate modulus from durometer hardness for silicones.* Dow Corning Corporation, 2016: p. 1-6.
46. ISO, *ISO 26203-2 Metallic materials — Tensile testing at high strain rates — Part 2: Servo-hydraulic and other test systems*. 2011.

# Figure set

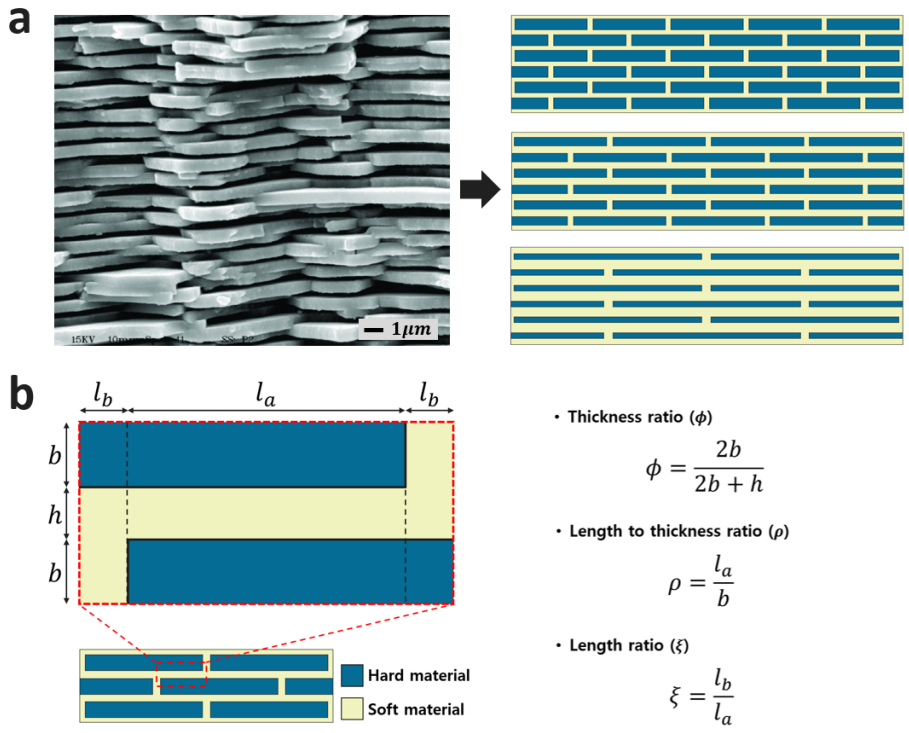



**Figure 1. Schematic of 2D staggered platelet structure**

(a) Staggered platelet structures inspired by the microstructure of natural nacre.

(b) Graphical representation of geometrical variables in a unit cell of the staggered platelet pattern, and the three resultant dimensionless geometrical parameters.

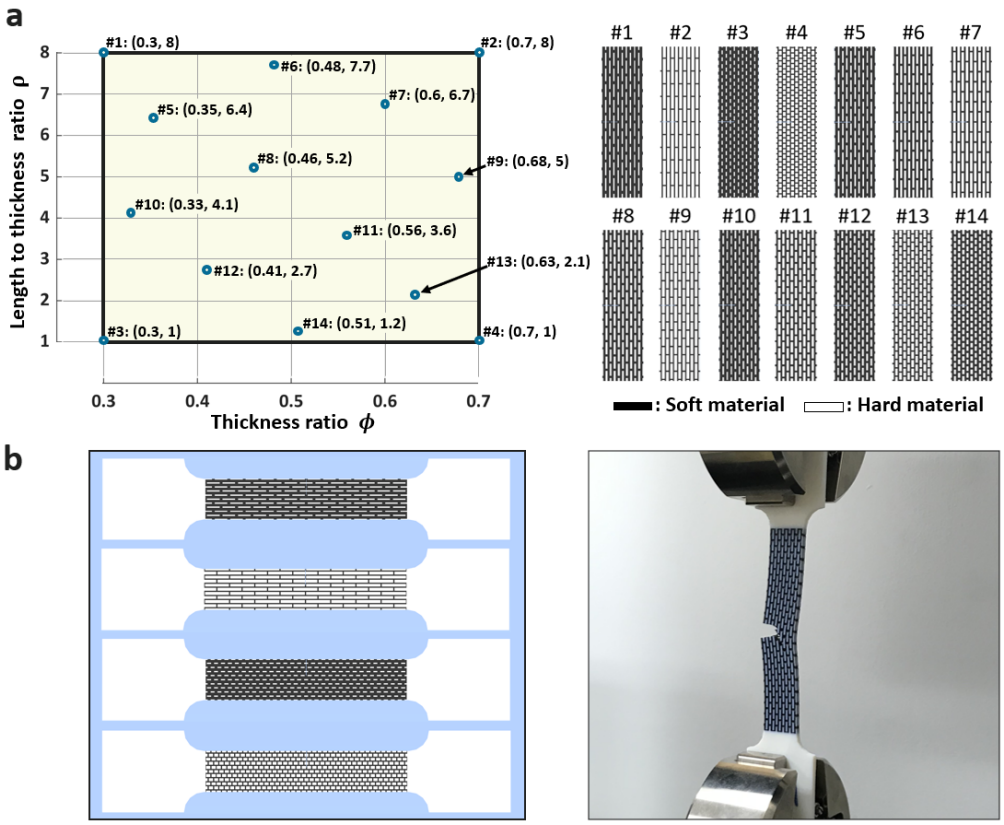



**Figure 2. Preparation of initial training data set**

(a) Fourteen initial data collection points were represented in the rectangular design space and the corresponding staggered platelet patterns. The uniform Latin hypercube sampling method was used to identify the data collection points that were evenly distributed in the domain.

(b) 3D modeling of dog-bone specimens with staggered platelet patterns (samples #1–#4) and uniaxial tensile testing of the 3D printed specimens.

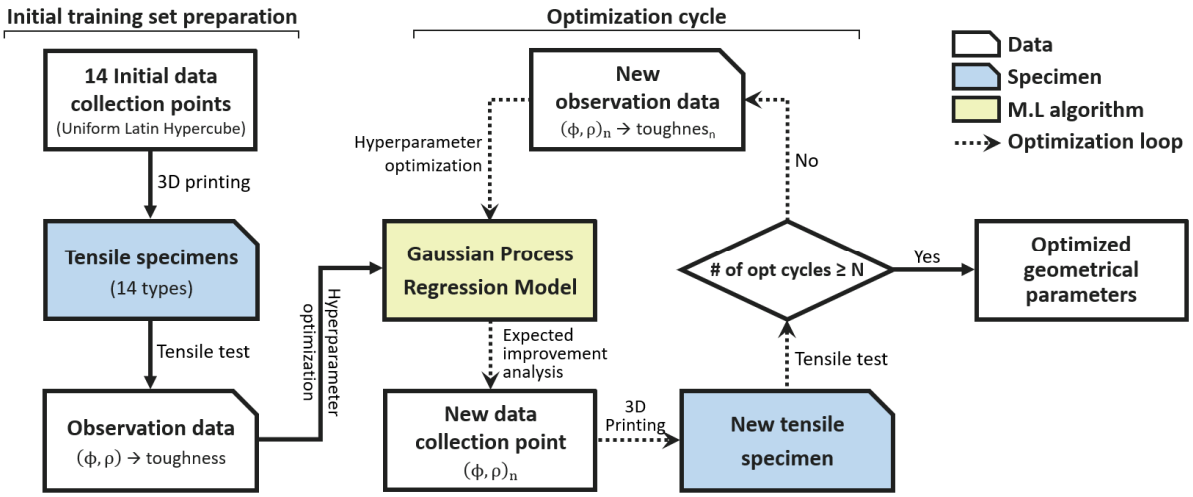



**Figure 3. Overall workflow chart of Gaussian process regression (GPR)-based geometry optimization**

The procedures related to the initial training set preparation are represented by solid arrow lines, and the optimization cycle is represented by dashed arrow lines.

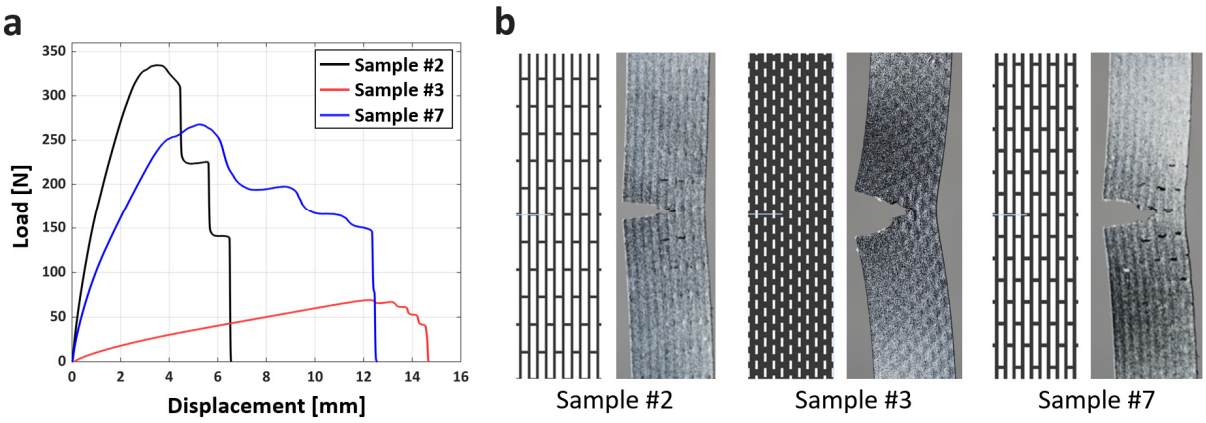



**Figure 4. Result of the tensile tests on the 3D printed samples**

(a) Results of the uniaxial tensile test on 3D printed specimens with three different staggered platelet patterns.

(b) Snapshots of three samples taken at the instant of crack propagation. The surface of samples was spray coated in black and white as we initially intended to perform a digital image correlation (DIC).

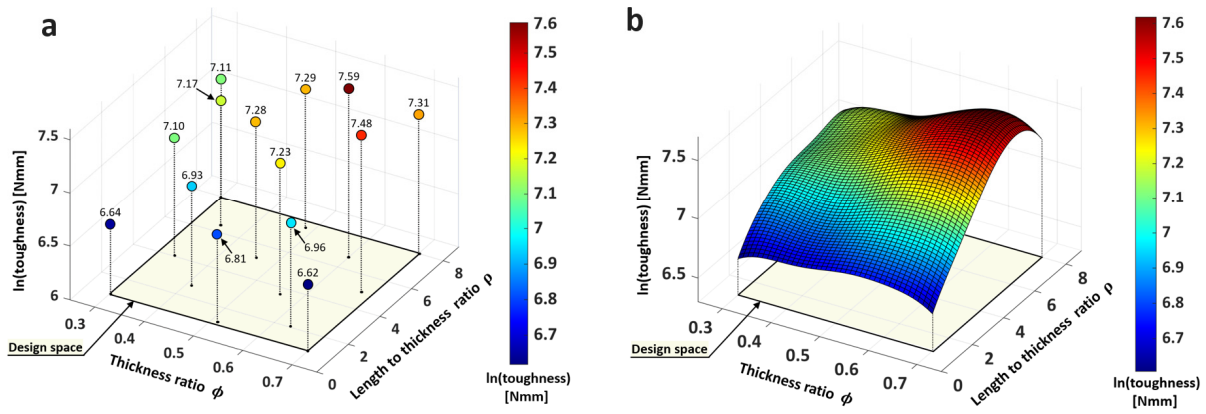



**Figure 5. The 14 initial observations and resultant GPR model**

(a) 3D representation of the initial training dataset consisting of 14 observations. The yellow rectangular region on the φ–ρ plane represents the design space. The ln(toughness) value of each data point is represented by both the color and z-coordinate values.

(b) GPR of 14 initial data points. The regression mean was represented as a 3D surface plot.

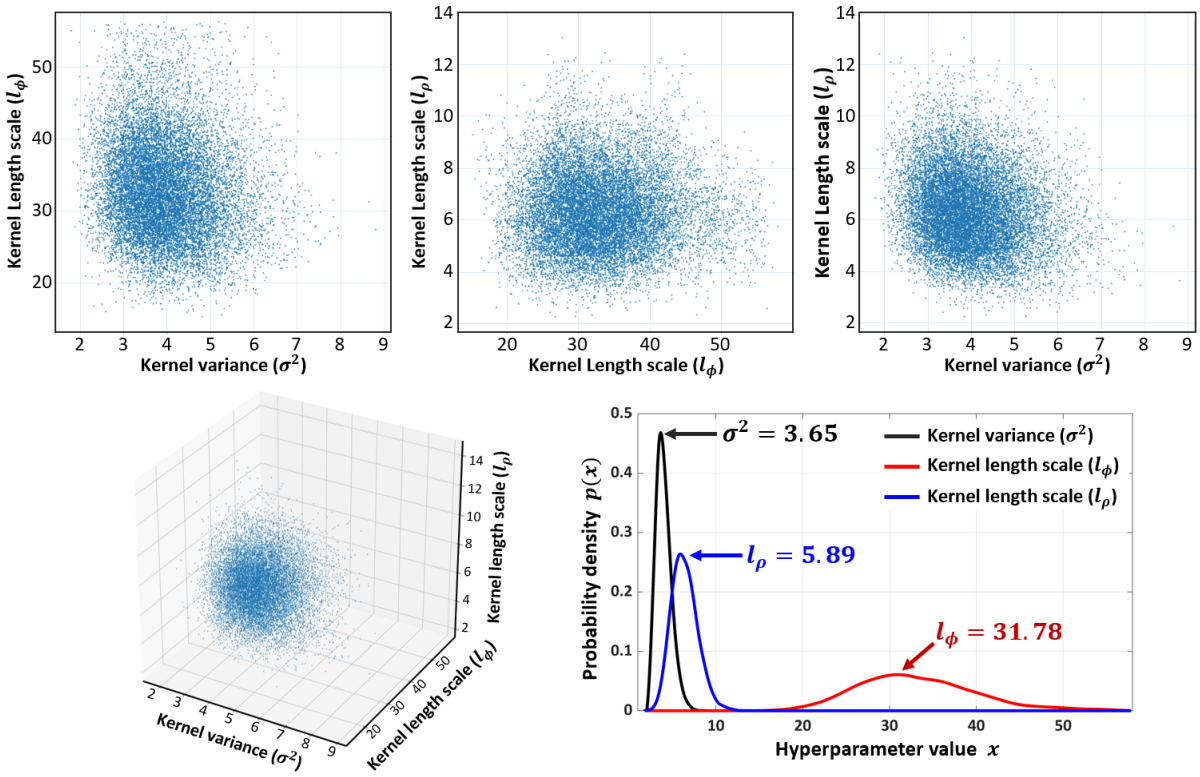



**Figure 6. Optimal hyperparameter estimation via Markov chain Monte Carlo (MCMC) method**

The record of 15000 steps of the MCMC sampling is based on the Hamiltonian Monte Carlo (HMC) algorithm. The sampled points are represented in the 3D space, and their 2D projections are also shown. The probability density graphs clearly indicate the optimal values for each hyperparameter.

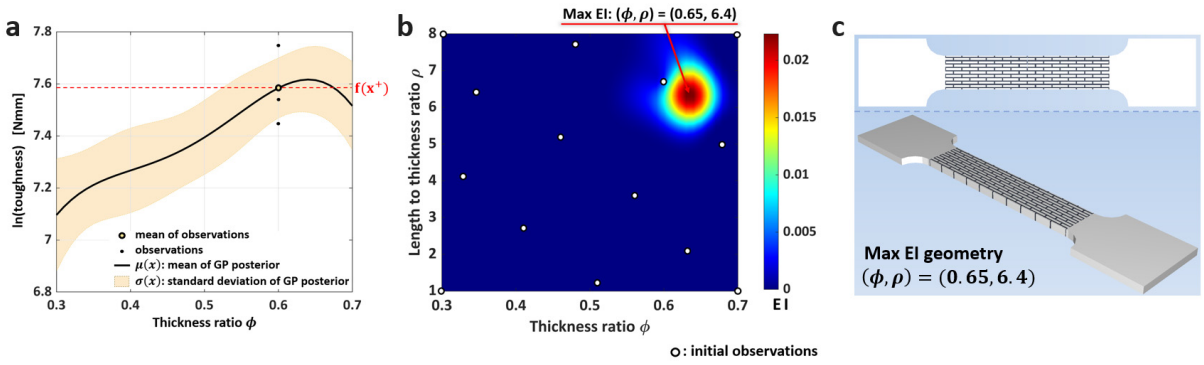



**Figure 7. Characterization of the next data collection point by the expected improvement (EI) algorithm**

(a) 2D representation of the GPR at a length-to-thickness ratio (ρ) of 6.7. The solid line indicates the regression mean, and the shaded region represents the standard deviation of the regression.

(b) Heatmap of expected improvement. The white dots indicate the positions of the initial training data. The highest value of expected improvement is observed at (ρ, φ) = (0.65, 6.4), and this point was selected as the next data collection point.

(c) 3D modeling of the tensile specimen with the maximum expected improvement geometry.

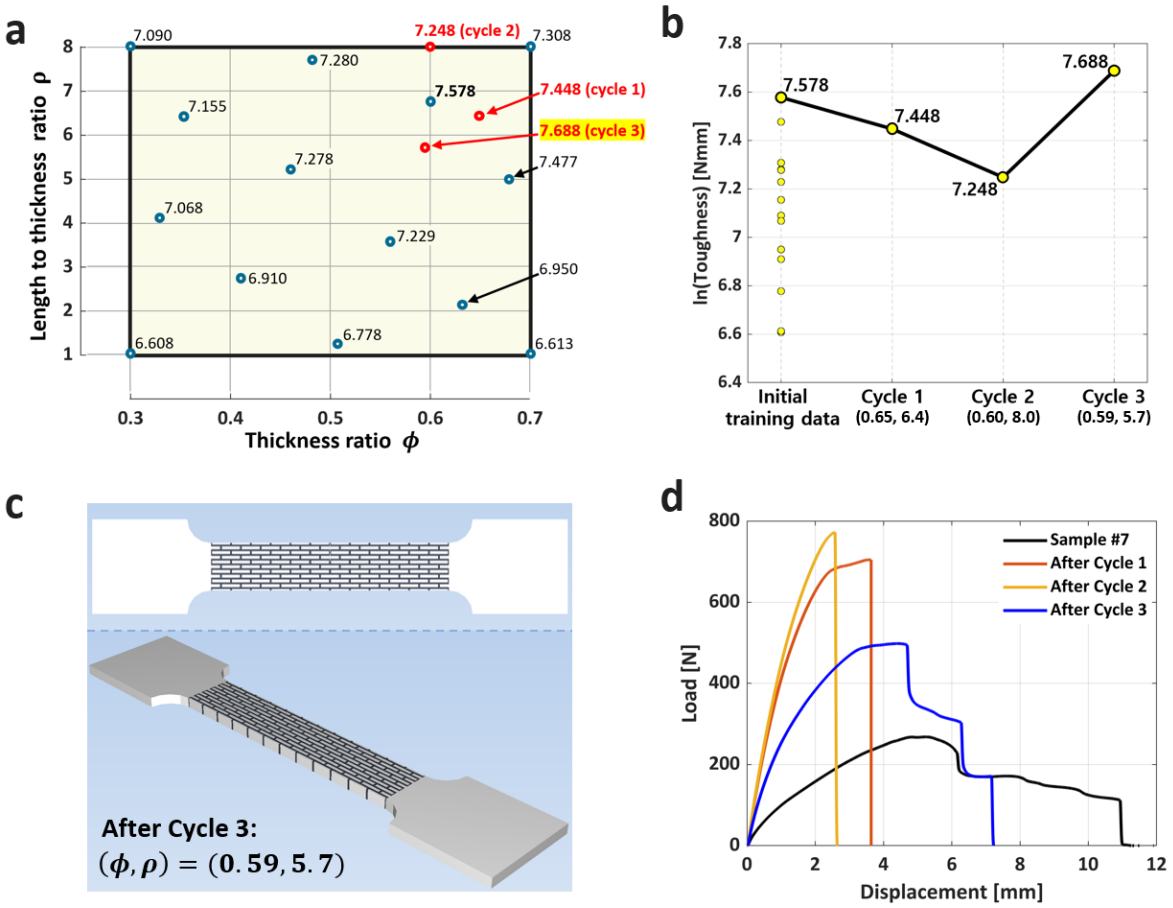



**Figure 8. Result of staggered platelet pattern optimization**

(a) All experimental data points obtained in this research are represented in the design space. The blue dots indicate the 14 initial training data points, and the red dots indicate the new data points added after the optimization cycles. The ln(toughness) values for each data point are indicated.

(b) Improvement in ln(toughness) throughout the optimization process.

(c) Higher toughness geometry obtained after the third optimization cycle.

(d) Load–displacement curves of the best sample from the 14 initial training sets and the samples after each iteration cycle.

| Optimization cycle | φ | ρ | Toughness [Nmm] | | | Average Toughness [Nmm] | ln(Toughness) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Mean μ | Variance $\sigma^2$ |
| Initial training data | 0.30 | 8.0 | 961.0137 | 1220.432 | 1473.141 | 1218.196 | 7.090 | 0.03055 |
| | 0.33 | 4.1 | 836.685 | 1245.004 | 1554.684 | 1212.124 | 7.068 | 0.06569 |
| | 0.35 | 6.4 | 1160.487 | 1096.702 | 1648.188 | 1301.792 | 7.155 | 0.03247 |
| | 0.30 | 1.0 | 626.112 | 612.195 | 1060.324 | 766.2103 | 6.608 | 0.06441 |
| | 0.51 | 1.2 | 642.4007 | 909.3786 | 1160.111 | 903.9634 | 6.778 | 0.05883 |
| | 0.41 | 2.7 | 739.9603 | 1131.371 | 1201.738 | 1024.356 | 6.910 | 0.04656 |
| | 0.48 | 7.7 | 1386.025 | 1273.72 | 1730.274 | 1463.34 | 7.280 | 0.01669 |
| | 0.56 | 3.6 | 1236.68 | 1356.495 | 1561.792 | 1384.989 | 7.229 | 0.00921 |
| | 0.46 | 5.2 | 1260.769 | 1460.451 | 1651.26 | 1457.493 | 7.278 | 0.01217 |
| | 0.63 | 2.1 | 940.8218 | 983.7896 | 1225.102 | 1049.904 | 6.950 | 0.01331 |
| | 0.70 | 1.0 | 656.2151 | 777.4508 | 810.5611 | 748.0757 | 6.613 | 0.00834 |
| | 0.68 | 5.0 | 1735.132 | 1856.982 | 1711.88 | 1767.998 | 7.477 | 0.00127 |
| | 0.60 | 6.7 | 2314.348 | 1883.52 | 1715.85 | 1971.239 | **7.578** | 0.01563 |
| | 0.70 | 8.0 | 1490.625 | 1524.859 | 1463.129 | 1492.871 | 7.308 | 0.00029 |
| **Cycle 1** | 0.65 | 6.4 | 1520.839 | 1870.063 | 1777.152 | 1722.685 | **7.448** | 0.00773 |
| **Cycle 2** | 0.60 | 8.0 | 1688.333 | 1285.465 | 1280.228 | 1418.009 | **7.248** | 0.01677 |
| **Cycle 3** | 0.59 | 5.7 | 2413.807 | 2014.249 | 2138.972 | 2189.009 | **7.688** | 0.00566 |

**Table 1. Result of uniaxial tensile tests**

# Supporting Information

# Designing staggered platelet composite structure with Gaussian process regression based Bayesian optimization

Kundo Park[a†], Youngsoo Kim[b†], Minki Kim[a,c], Chihyeon Song[d], Jinkyoo Park[d,*], Seunghwa Ryu[a,**]

[a] Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

[b] Department of Nature-Inspired System and Application, Korea Institute of Machinery and Materials (KIMM), Daejeon 34103, Republic of Korea

[c] Department of Integrated Systems Engineering, The Ohio State University, Columbus, OH, 43210, USA

[d] Department of Industrial & Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

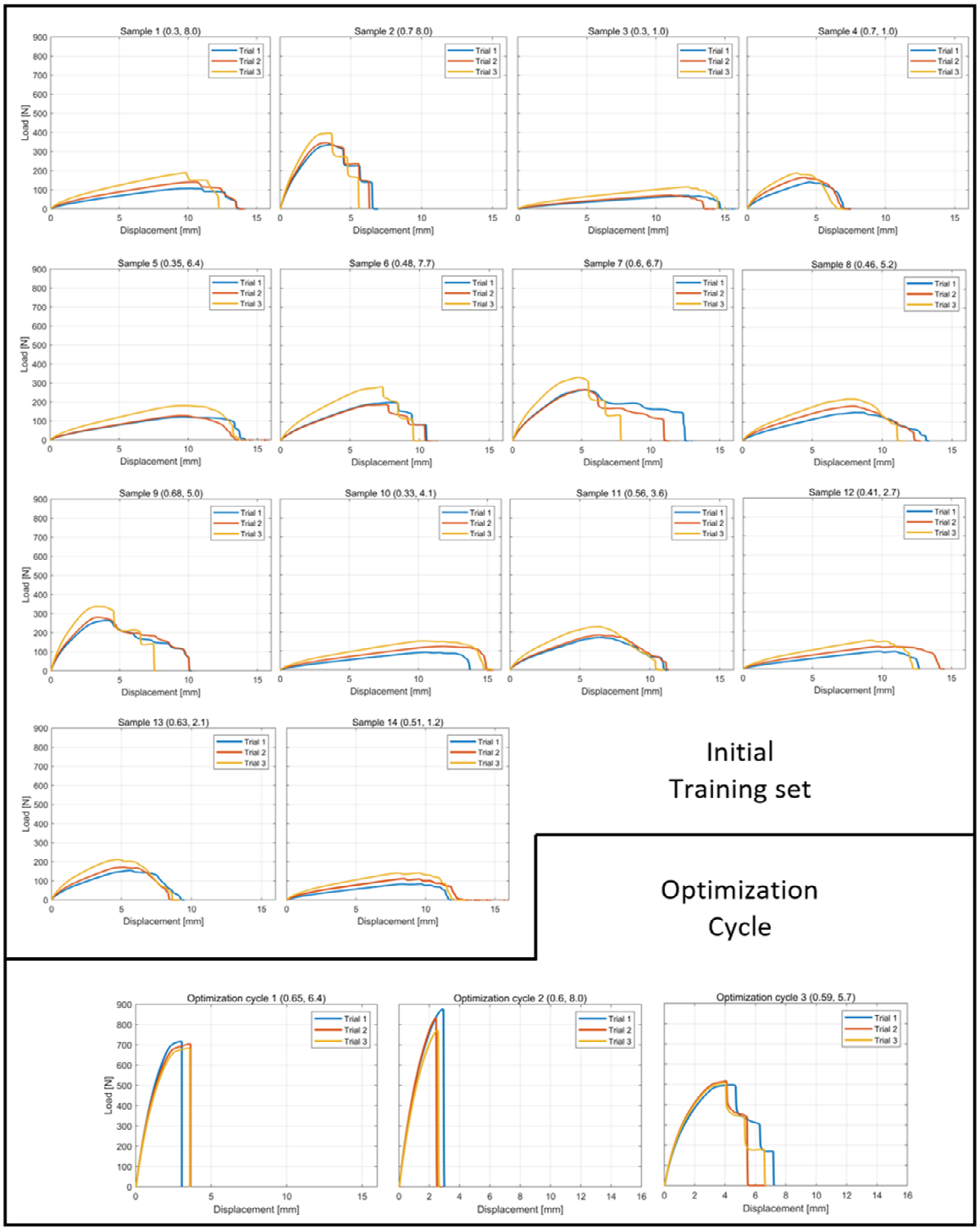
[†] These authors have contributed equally to this work.

**Corresponding authors:**

[*] E-mail: jinkyoo.park@kaist.ac.kr

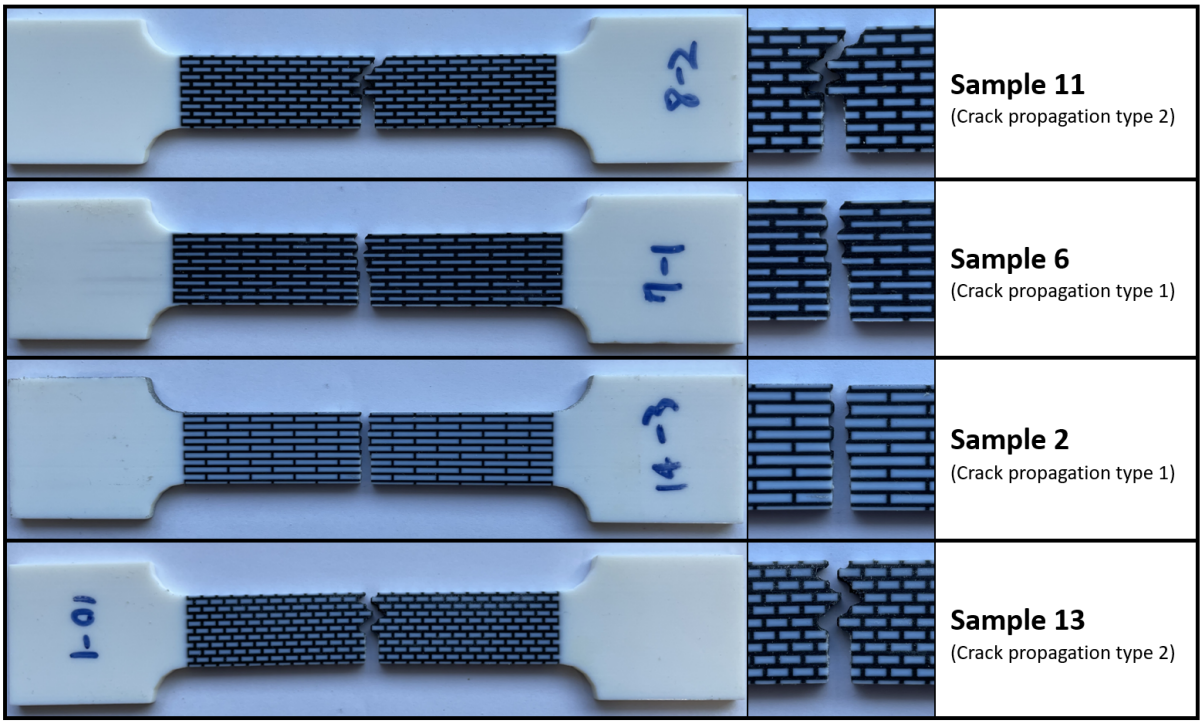
[**] E-mail: ryush@kaist.ac.kr

**Figure S1.** Load-displacement curves of all the samples tested.



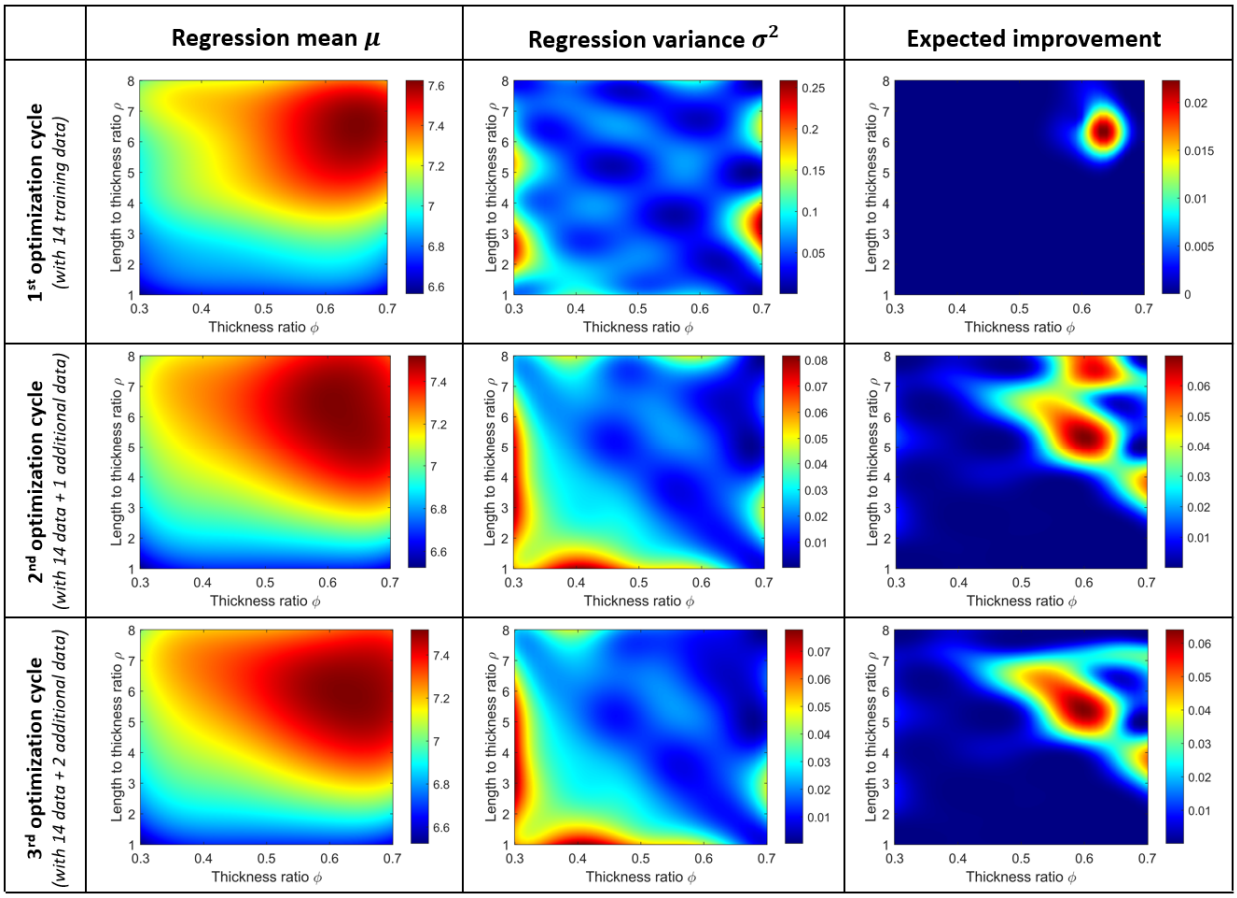

For each staggered platelet geometry, three identical samples were fabricated via additive manufacturing. The tensile test result of first, second and third sample for each staggered platelet pattern is visualized in blue, red, and yellow curve, respectively. The area under the load-displacement indicates the toughness.

**Figure S2.** The two crack propagation modes observed in the staggered platelet composite specimens.




For sample #2 and sample #6, the crack propagation is initiated at the tip of the precrack, and it propagated straight through the stiff platelets on its way. For sample #11 and #13, the crack propagation is started at the tip of the precrack and it travels only in the soft matrix, detouring around the stiff platelets in its propagating direction.

**Figure S3.** Gaussian process regression result in each of the optimization cycles.




The regression mean, regression variance, and the expected improvement in the entire design space after each of the optimization cycles are visualized as 2D heat map.