Practical Complications in Applying Bayesian Optimization to Understand Tradeoffs between Substrate Fabrication Processes

Sajad Haghanifar [1] **Bolong Cheng** [2] **Michael McCourt** [2] **Paul Leu** [1]

Abstract

Designing an effective materials fabrication process demands searching the realm of possible fabrication processes to understand how desired properties are impacted by the specifications of the process. We apply active learning in the context of Bayesian optimization to propose an adaptive experimental design – our goal is to identify processes with the most desirable properties in a sample efficient fashion. In reflecting on our cross-disciplinary industry-academic collaboration between materials scientists and mathematicians, we discuss our approach to managing practical concerns and complications; these issues may be rarely addressed in theoretical literature but were necessary for our circumstances.

1. Introduction

In this work, we focus on fabricating high transmittance substrates which also have self-cleaning properties. We measure self-cleaning capacity through *superomniphobicity*, the ability to repel many liquids; superomniphobic surfaces demonstrate a static contact angle greater than 150° and low contact angle hysteresis for a variety of liquids (Pan et al., 2013; Choi et al., 2009; Tuteja et al., 2007; Kota et al., 2012). Other desirable properties include resistance to fogging (Mouterde et al., 2017; Wilke et al., 2018) and high durability against abrasion (Si et al., 2018).

Inspired by recent analysis of glasswing butterfly wings (Siddique et al., 2015), this research focuses on creating a new self-healing, durable, superomniphobic glass with *random* nanostructures as opposed to highly ordered sub-wavelength structure arrays. The glass is fabricated through a simple, scalable, two-step, maskless reactive ion etching and fluorination process, detailed in Section 2, which we demonstrate on 4 inch diameter glass wafers. This fabrication process contains a large number of decisions (e.g., regarding the fluorination process) which impact the resulting glass in often complicated and nonintuitive ways.

Traditionally, making these fabrication decisions involves using application specific expertise to explore the impact of a small number of decision parameters on a single property of fabricated glass. Because our goal in this project is to balance multiple interacting properties (to achieve superomniphobia without sacrificing ultrahigh transparency and ultralow haze), we turn to active learning, specifically Bayesian optimization (Frazier, 2018). This has the potential to adaptively search the space of fabrication strategies in a sample-efficient fashion.

Standard Bayesian optimization, though, is ill-suited for the circumstances of this materials science project; practical limitations demand adaptations, which we detail in this paper. We hope that this can serve to document some of the compromises which helped power this application of active learning in a real-world materials fabrication setting involving real-world concerns (shared lab equipment, tooling precision).

2. Fabrication Process

The nanofabrication process is performed in two steps: (a) reactive ion etching (RIE) and (b) plasma enhanced chemical vapor deposition (PECVD) and surface treatment with fluorination. This fabrication process scalably creates the nanostructures directly into the fused silica glass without the need for patterning or an external mask (Haghanifar et al., 2017; 2018). Figure 1 depicts the input and output parameters under analysis, and suggests how we will efficiently explore the fabrication parameters of this process.

The first fabrication step focuses on RIE to create sub-wavelength nanostructures in the fused silica in order to maximize the total transparency and minimize the haze at the wavelength of 550 nm. The second processing step focuses on creating re-entrant structures and a low energy surface: in this processing step, we consider the deposition of silicon dioxide (SiO_2) by PECVD on top of the sub-wavelength nanostructures in order to make the structure

[1]Laboratory for Advanced Materials at Pittsburgh, University of Pittsburgh, Pittsburgh, PA, USA [2]SigOpt, San Francisco, CA, USA. Correspondence to: Michael McCourt <mccourt@sigopt.com>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

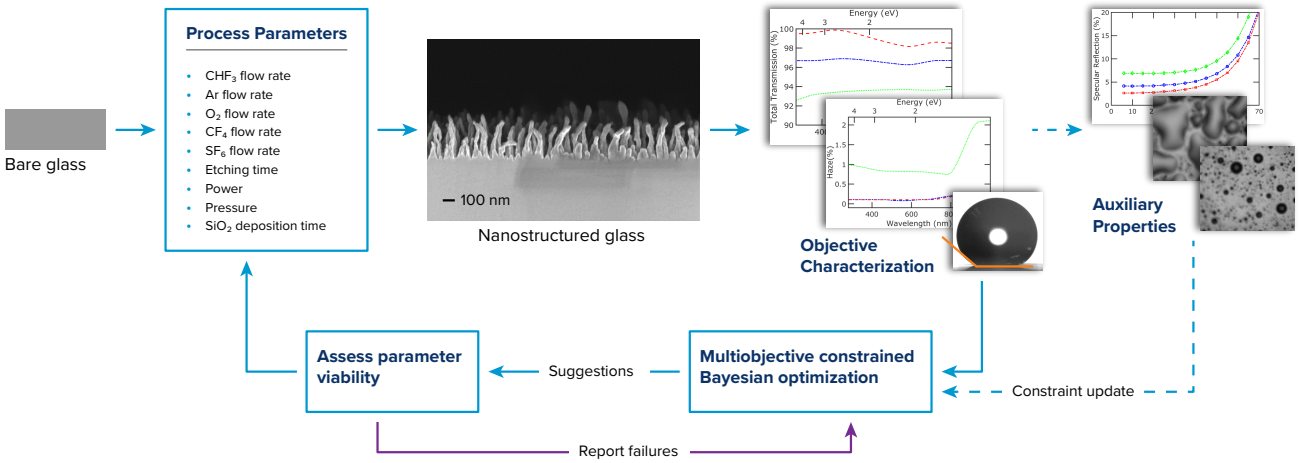

Figure 1. Schematic of experimental fabrication and Bayesian learning optimization process for nanostructured glass. The objectives, defined in Section 2 are transmission, haze and oil contact angle. The auxiliary properties, which did not drive the optimization but were monitored to determine if stronger constraints were possible, were reflection, fogging proclivity, and durability.

re-entrant followed by fluorination.

Numerous decisions made within these two steps impact the properties of the resulting glass; before an efficient search of the space of fabrication parameters could take place, we needed to agree on the appropriate parameters to search, and an appropriate domain for those parameters. Figure 1 lists the parameters which we choose to vary during our search – other parameters, such as the fluorination time, were fixed.

Because we simultaneously want high performance photon management and wetting properties, we phrase our goal in terms of a multiobjective optimization problem with solution $\mathbf{x}^*$,

$$\mathbf{x}^* \text{ satisfies } \begin{cases} \mathbf{x}^* = \arg\max_{\mathbf{x}\in\Omega} \theta_o(\mathbf{x}), \\ \mathbf{x}^* = \arg\max_{\mathbf{x}\in\Omega} T_{\text{total}}(\mathbf{x}), \text{ and} \\ \mathbf{x}^* = \arg\min_{\mathbf{x}\in\Omega} H(\mathbf{x}) \end{cases} \quad (1)$$

where Ω is the space of all possible choices of the process parameters. We denote $\mathbf{x}$ to be both the fabrication process parameters and the resulting nanostructure from using those parameters. $\theta_o(\mathbf{x})$ is the oil contact angle, T_{total} is the total transmission, and H is the haze. The total transmission and haze are optimized for wavelength $\lambda = 550$ nm, which is in the middle of the visible spectrum, and ethylene glycol was chosen as the oil. The wavelength-dependent haze $H(\lambda)$ is defined as

$$H(\lambda) = \left[\frac{T_{\text{scattered}}(\lambda)}{T_{\text{total}}(\lambda)}\right] \times 100\%, \quad (2)$$

where $T_{\text{total}}(\lambda)$ is the total transmission and $T_{\text{total}}(\lambda) = T_{\text{scattered}}(\lambda) + T_{\text{specular}}(\lambda)$, where $T_{\text{scattered}}$ is the scattered transmission and T_{specular} is the specular or direct transmission (Gao et al., 2018).

In general, there is no unique structure $\mathbf{x}^\star$ that is simultaneously optimal in all the objectives in (1). In lieu of such a point, the solution to such a multiobjective problem is often defined as the Pareto-optimal set, or Pareto-efficient frontier $\mathcal{P} \in \Omega$. Pareto optimal parameters $\mathbf{x} \in \mathcal{P}$ evince a "trade-off" between objective function values, such that no $\mathbf{x}' \in \Omega$ can yield better performance across *all* objective functions; any improvements in one metric would necessitate a loss in performance in at least one other metric. All of the points in $\mathcal{P}$ may be referred to as *efficient*, and all other points in Ω would be defined as *dominated*. A more thorough explanation of the topic can be found in multicriteria literature (Ehrgott, 2005).

3. Bayesian Optimization

The substrate fabrication process described above is quite time-consuming, which necessitates an effective experimental design so as to quickly search the space Ω for input parameters which perform well for all three objective functions. Bayesian optimization is a sample-efficient iterative search framework, where the relationship between process parameters and objective function values is unknown, and function evaluations (executing the fabrication and characterizing the resulting substrate) are expensive or time consuming. Standard Bayesian optimization consists of two components: a probabilistic *surrogate model*, to model the objective function f, and an *acquisition function*, to determine which $\mathbf{x}$ parameters to next sample.

In a typical single objective Bayesian optimization setting, the objective function f is assumed to be a realization of a Gaussian process (GP) with mean function μ and a positive definite covariance kernel K, i.e., $f \sim \mathcal{GP}(\mu, K)$ (Ras-

mussen & Williams, 2005; Fasshauer & McCourt, 2015). The mean and covariance functions are often defined to have certain free parameters which are fit to the data using strategies such as maximum likelihood estimation (which was our strategy of choice in this process). In all of our modeling, we assume our GPs to have $\mu \equiv 0$ and a square-exponential K with independent length-scales in each dimension. A Tikhonov regularization parameter is fixed to be 10^{-3}, primarily to ease ill-conditioning concerns.

An acquisition function is a utility function that measures the value of sampling at different points within Ω, conditioning on the observed data. Acquisition functions balance the tradeoff between *exploitation*, suggesting input parameters near where we have observed the best results so far, and *exploration*, suggesting input parameters in regions where we have not tried out. After n different input parameters have been tested, the nth surrogate model can be created. We define the acquisition function using the surrogate model, which is then maximized to determine the next input parameter selection at which to run the fabrication process.

4. Practical Considerations for our Setting

As indicated above, the actual circumstances of Section 2 are rather different than the classical Bayesian optimization framework defined in Section 3. Of course, for this project to be successful, it was necessary that we adapt the standard approaches from active learning to be able to function effectively in this real-world setting. We list some of these driving motivations below.

- There was a limited fabrication budget (only 60 total fabrications in a 9 dimensional space) caused by the sharing of equipment between several people at the lab.
- The fabrications had to be executed in batches of 5 (5 different $\mathbf{x}$ parameter values) to accommodate the structure of the fluorination chamber.
- We wanted to be able to reject proposed fabrication as *failures* without executing them, so as to not waste lab time on something that we, as application experts, believe to be inappropriate.
- Our equipment (as with all equipment) has limited precision, i.e., parameter values which are too close may be indistinguishable.

As described in (1), the problem we consider here is actually much more costly than a single objective optimization problem—putting a high number of points on the Pareto frontier $\mathcal{P}$ is a costly proposition. After discussing the limitation of only having 60 available fabrications, we realized that, despite Bayesian optimization's sample-efficient nature, trying to fully flesh out the Pareto frontier was infeasible.

Instead, we identified an alternate strategy for finding a small number of points on $\mathcal{P}$ which represented the most interesting fabricated products. Our new glass must perform better in all metrics than standard glass; consequently, we imposed performance thresholds $\theta_o(\mathbf{x}) \geq 60^\circ$, $T_{\text{total}}(\mathbf{x}) \geq 88.5$ and $H(\mathbf{x}) \leq 1.1$. To respect these thresholds we adapt the ϵ-*constraint method* (Hwang & Masud, 1979) to produce the scalar optimization problem:

$$\max_{\mathbf{x}\in\Omega} T_{\text{total}}(\mathbf{x}), \text{ s.t. } H(\mathbf{x}) \leq \hat{H}, \theta_o(\mathbf{x}) \geq \hat{\theta}_o, \qquad (3)$$

with the analogous constrained scalar optimization problems also defined for θ_o and H. Our goal became to find as many points as possible on $\mathcal{P}$ which satisfy these constraints.

To execute this, we adapt methods from constrained Bayesian optimization literature (Gelbart, 2015; Letham et al., 2019). After k fabrications have been conducted, Gaussian process models $s_{T,k}$, $s_{H,k}$ and $s_{\theta,k}$ are created for the transmission, haze and contact angles, respectively. These are modeled independently, though in future work we could consider a joint model. Independent modeling was preferable here because it allowed us to incorporate data from previous experiments (prior to this work) where not all of the haze, transmission and contact angle values were recorded.

Using these models, an acquisition function is defined for each component of (1); the Bayesian optimization process, as powered by this acquisition function, is presented graphically in a simplified setting in (2). This acquisition function is modified from the expected parallel improvement (Ginsbourger et al., 2008) to account for the desire for viability. Considering, at first, only the solution to (3), imposing the viability requires us to consider not only the distribution of $t \sim s_{T,k}(\mathbf{x})$ (a Gaussian distribution), but the joint distribution $t, h, z \sim s_{T,k}(\mathbf{x}), s_{H,k}(\mathbf{x}), s_{\theta,k}(\mathbf{x})$, more succinctly denoted by $t, h, z \sim s_k(\mathbf{x})$. The acquisition function (without parallel suggestions) would be defined as

$$a_{T,k}(\mathbf{x}) = \mathbb{E}_{t,h,z\sim s_k(\mathbf{x})}[(t - \tilde{t}_k)_+ I_{h<\hat{H}\cap z>\hat{\theta}_o}], \qquad (4)$$

where $\tilde{t}_k$ is the highest T_{total} value observed thus far, $(\xi)_+$ denotes $\max(\xi, 0)$, and $I_\nu = 1$ if the condition ν is satisfied and 0 otherwise (the indicator function). This is semantically equivalent to maximizing the *expected improvement* attainable for viable points; points which do not satisfy our thresholds contribute zero improvement.

To account for the desire for 5 parallel suggested parameters, we expand on the base structure of (4). This requires taking draws of our models at 5 different input parameter $\mathbf{x}_1, \ldots, \mathbf{x}_5 \in \Omega$ values. We denote this with the shorthand

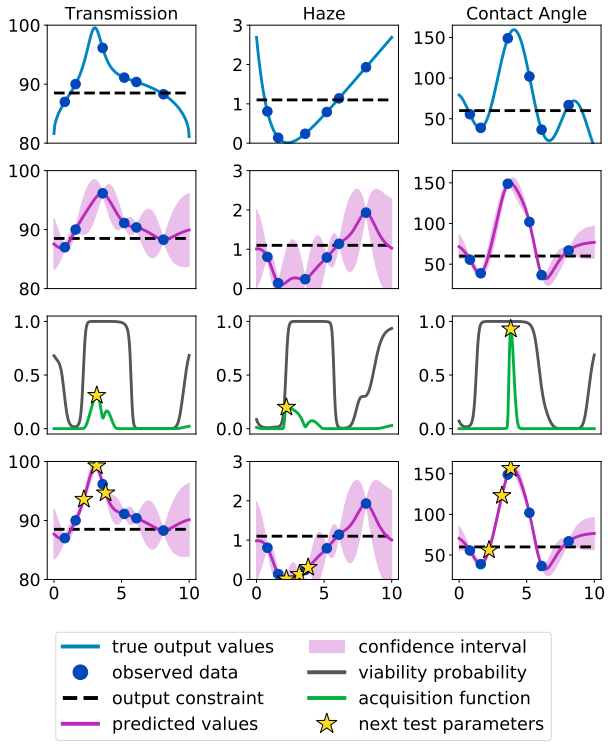

Figure 2. Sample depiction of our proposed Bayesian optimization process; each column represents one of the three output parameters under consideration. These are artificial profiles in one dimension for explanatory purposes only.
First row: The "true" output parameter to be optimized.
Second row: Statistical models built from the observed data.
Third row: The probability of an input parameter being viable (satisfying the constraints for the other two output parameters) and the associated acquisition function values along with the points which maximize the acquisition function.
Fourth row: The new observations achieved by sampling at the "next test parameters" and the new models which result from this new data.

notation

$$\mathbf{t}, \mathbf{h}, \mathbf{z} \sim s_k(\vec{\mathbf{x}}) \iff$$
$$t_1, h_1, z_1, \ldots t_5, h_5, z_5, \sim s_k(\mathbf{x}_1), \ldots, s_k(\mathbf{x}_5),$$

which allows us to write the expected parallel improvement, attenuated by viability, as

$$a_{T,k,5}(\vec{\mathbf{x}}) = \mathbb{E}_{\mathbf{t},\mathbf{h},\mathbf{z}\sim s_k(\mathbf{x})}\left[\max_{1\le i\le 5}(t_i - \tilde{t}_k)_+ I_{h_i<\hat{H}\cap z_i>\hat{\theta}_o}\right]. \quad (5)$$

Failed observations (reported prior to conducting the fabrication) were stored as the worst value observed for each metric—this allows us to avoid parameters near that region without maintaining separate models of the probability of viability. We hope to consider, in future work, developing separate or joint models of probability using these "prior beliefs". Figure 2 depicts the Bayesian optimization process in a sample problem reduced to one dimension for ease of understanding.

In practice, we estimate (5) through 4000 Monte Carlo iterations. We used the CMA-ES (Hansen et al., 2003) optimization strategy to maximize all acquisition functions. Because of the limited precision afforded from the lab equipment, the domain Ω over which the optimization takes place is

- CHF_3 flow rate: $\{0, 5, \ldots, 80\}$ sccm,
- Ar flow rate: $\{0, 5, \ldots, 100\}$ sccm,
- O_2 flow rate: $\{0, 5, \ldots, 100\}$ sccm,
- CF_4 flow rate: $\{0, 5, \ldots, 80\}$ sccm,
- SF_6 flow rate: $\{0, 5, \ldots, 80\}$ sccm,
- Etching time: $\{0, 60, \ldots, 5400\}$ seconds,
- Power: $\{20, 30, \ldots, 300\}$ watts,
- Pressure: $\{50, 100, \ldots, 250\}$ mtorr,
- SiO_2 deposition time: $\{8, 10, \ldots 500\}$ seconds.

The CMA-ES algorithm is modified to only propose points on this discrete space. The evolutionary population is 25, with 100 full iterations and 10 uniform random restarts.

5. Results

During 64 fabrications (4 of which were immediately reported as failures and thus did not count against the original budget) driven by Bayesian optimization (which, itself, was seeded with 79 fabrications executed prior to the start of this project) we produced 5 points on $\mathcal{P}$. They are presented in Table 1 and depicted in Figure 3.

Table 1. The Pareto efficient points found during the fabrication search. The point in bold is the point for which the fabrication displayed in Figure 4 was conducted.

Transmission (%)	Haze (%)	Oil Contact Angle (°)
97.01	**0.01**	**155**
95.90	0.03	157
95.65	0.02	156
94.54	1.30	158
94.36	0.60	158

We conducted scanning electron microscopy (SEM) images of the sub-wavelength, re-entrant structures. The randomness in the height and spacing provide for broadband and omnidirectional antireflection, like the glasswing butterfly wings (Siddique et al., 2015). By depositing the SiO_2, the surface area at the top of the pillars increase which provide the re-entrant structures required for omniphobicity. One such microscope image is presented in Figure 1 with the label "Nanostructured glass".

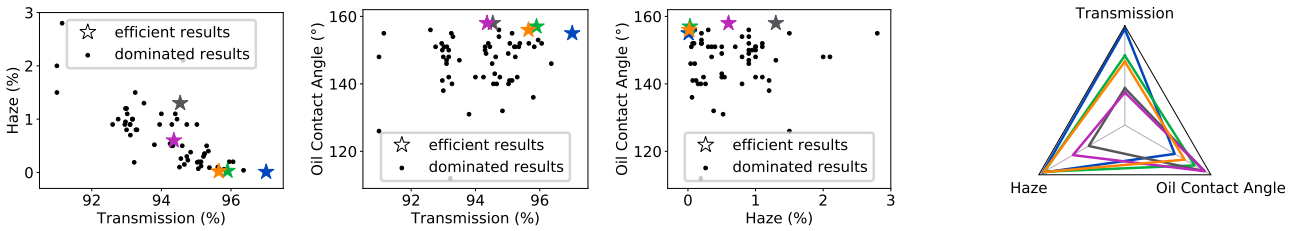

Figure 3. Depictions of the experimental design driven by our Bayesian optimization methodology. *left*: Three 2D feasible region plots of the three objectives under consideration. *right*: Radar plot of the 5 viable efficient outcomes identified during the parameter search (plot qualitatively exaggerated to account for the different scales of the three objectives). These points are listed explicitly in Table 1.

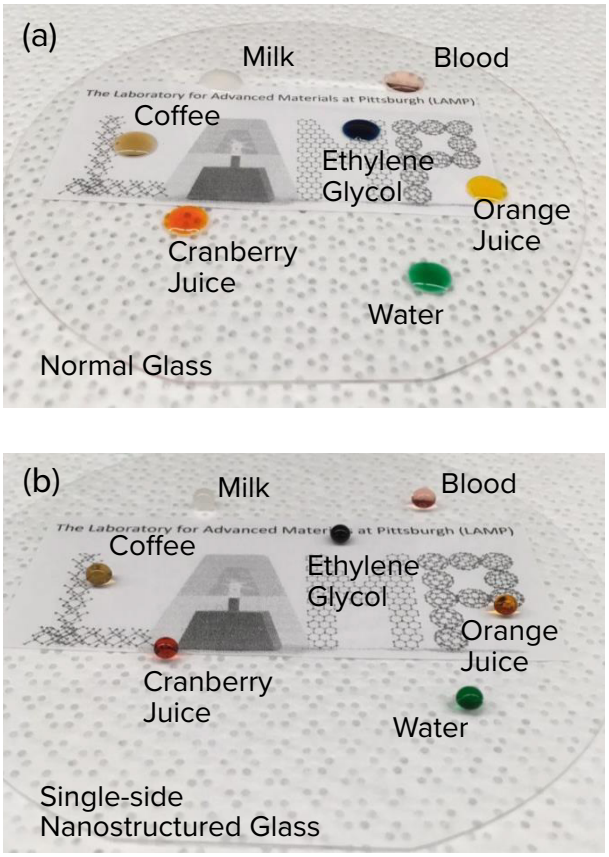

Figure 4. We placed 5 μL of various liquids on both standard glass and our newly fabricated glass (ethylene glycol was the oil for which we defined superomniphobicity). *top*: Standard glass. *bottom*: Glass fabricated using the parameters associated with the results in bold in Table 1, also in blue in Figure 3.

To investigate the omniphobic property, we deposited 5 μl drops of different liquids with different surface tensions, from water (72.8 mN/m) to ethylene glycol (47.7 mN/m), on both bare (Figure 4(a)) and nanostructured (Figure 4(b)) substrates. The bare fused silica has $42.9 \pm 1.1°$ and $18.7 \pm 0.7°$ contact angle for water and oil. By creating re-entrant structure on the bare fused silica, the water and oil contact angles increase significantly to $162.1 \pm 2.0°$ and $155.2 \pm 2.2°$, respectively. This improvement over the bare glass is substantial, and comes with improvements in transmission and haze as well.

The total transmission of double-side nanostructured glass at 550 nm is 99.5%. In both single-side and double-side nanostructured glass, the haze value reduces to less than 0.1% across a broadband range of wavelength. After analyzing specular reflection, the reflection values are always less than glass even for a high incidence angle of $70°$, which reveals the high omnidirectional, antireflective performance of our fabricated glass.

6. Conclusions

We feel that, by joining forces on this research, we were able to produce this new fabrication strategy much more efficiently than previous research would have suggested. Considerable changes were required from a standard Bayesian optimization strategy, but this work will help power future materials science developments. By attending this workshop, we hope to share our experiences and also learn from others on how active learning (and other statistical learning topics) are adapted to practical circumstances with unavoidable limitations.

In future work, we hope to adapt this active learning approach to exploring other properties of this fabricated glass which may require a more flexible feedback cycle than what is suggested in Figure 1. In particular, we envision wanting to modify the search process based on user's preferences as solicited through questions asked before providing next points (a process briefly outlined in (Dewancker et al., 2017), but requiring more development).

References

Choi, W., Tuteja, A., Chhatre, S., Mabry, J. M., Cohen, R. E., and McKinley, G. H. Fabrics with Tunable Oleophobicity. *Advanced Materials*, 21(21):2190–2195, June 2009.

Dewancker, I., Bauer, J., and McCourt, M. Sequential preference-based optimization. In *Bayesian Deep Learn-*

ing Workshop at Neural Information Processing Systems, 2017.

Ehrgott, M. *Multicriteria optimization*, volume 491. Springer Science & Business Media, 2005.

Fasshauer, G. E. and McCourt, M. J. *Kernel-based Approximation Methods Using Matlab*. World Scientific, 2015.

Frazier, P. I. Bayesian optimization. In Gel, E. and Ntaimo, L. (eds.), *Recent Advances in Optimization and Modeling of Contemporary Problems*, pp. 255–278. INFORMS, 2018.

Gao, T., Haghanifar, S., Lindsay, M. G., Lu, P., Kayes, M. I., Pafchek, B. D., Zhou, Z., Ohodnicki, P. R., and Leu, P. W. Fundamental performance limits and haze evaluation of metal nanomesh transparent conductors. *Advanced Optical Materials*, 6(9):1700829, 2018.

Gelbart, M. A. *Constrained Bayesian optimization and applications*. PhD thesis, Harvard, 2015.

Ginsbourger, D., Le Riche, R., and Carraro, L. A Multipoints Criterion for Deterministic Parallel Global Optimization based on Gaussian Processes. Technical report, Ecole Nationale Supérieure des Mines, March 2008.

Haghanifar, S., Gao, T., Vecchis, R. T. R. D., Pafchek, B., Jacobs, T. D. B., and Leu, P. W. Ultrahigh-transparency, ultrahigh-haze nanograss glass with fluid-induced switchable haze. *Optica*, 4(12):1522–1525, December 2017.

Haghanifar, S., Vecchis, R. T. R. D., Kim, K.-J., Wuenschell, J., Sharma, S. P., Ping Lu, Ohodnicki, P., and Leu, P. W. Flexible nanograss with highest combination of transparency and haze for optoelectronic plastic substrates. *Nanotechnology*, 29(42):42LT01, 2018.

Hansen, N., Müller, S. D., and Koumoutsakos, P. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es). *Evolutionary computation*, 11(1):1–18, 2003.

Hwang, C. and Masud, A. *Multiple objective decision making, methods and applications: a state-of-the-art survey*. Lecture notes in economics and mathematical systems. Springer-Verlag, 1979.

Kota, A. K., Li, Y., Mabry, J. M., and Tuteja, A. Hierarchically Structured Superoleophobic Surfaces with Ultralow Contact Angle Hysteresis. *Advanced Materials*, 24(43): 5838–5843, 2012.

Letham, B., Karrer, B., Ottoni, G., Bakshy, E., et al. Constrained Bayesian optimization with noisy experiments. *Bayesian Analysis*, 14(2):495–519, 2019.

Mouterde, T., Lehoucq, G., Xavier, S., Checco, A., Black, C. T., Rahman, A., Midavaine, T., Clanet, C., and Qur, D. Antifogging abilities of model nanotextures. *Nature Materials*, 16(6):658–663, June 2017.

Pan, S., Kota, A. K., Mabry, J. M., and Tuteja, A. Superomniphobic Surfaces for Effective Chemical Shielding. *Journal of the American Chemical Society*, 135(2):578–581, January 2013. ISSN 0002-7863.

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

Si, Y., Dong, Z., and Jiang, L. Bioinspired Designs of Superhydrophobic and Superhydrophilic Materials. *ACS Central Science*, 4(9):1102–1112, September 2018.

Siddique, R. H., Gomard, G., and Hölscher, H. The role of random nanostructures for the omnidirectional antireflection properties of the glasswing butterfly. *Nature Communications*, 6:6909, April 2015.

Tuteja, A., Choi, W., Ma, M., Mabry, J. M., Mazzella, S. A., Rutledge, G. C., McKinley, G. H., and Cohen, R. E. Designing Superoleophobic Surfaces. *Science*, 318(5856): 1618–1622, December 2007.

Wilke, K. L., Preston, D. J., Lu, Z., and Wang, E. N. Toward Condensation-Resistant Omniphobic Surfaces. *ACS Nano*, 12(11):11013–11021, November 2018.