Bayesian Optimization of a Free-Electron Laser

J. Duris,[1] D. Kennedy,[1,2] A. Hanuka,[1] J. Shtalenkova,[1] A. Edelen,[1] P. Baxevanis,[1] A. Egger,[1] T. Cope,[1] M. McIntire,[3] S. Ermon,[3] and D. Ratner[1]

[1]*SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA*
[2]*Department of Physics, University of California, Santa Cruz, Santa Cruz, California 95064, USA*
[3]*Department of Computer Science, Stanford University, Stanford, California 94305, USA*

(Received 13 September 2019; revised manuscript received 1 February 2020; accepted 19 February 2020; published 25 March 2020)

The Linac coherent light source x-ray free-electron laser is a complex scientific apparatus which changes configurations multiple times per day, necessitating fast tuning strategies to reduce setup time for successive experiments. To this end, we employ a Bayesian approach to maximizing x-ray laser pulse energy by controlling groups of quadrupole magnets. A Gaussian process model provides probabilistic predictions for the machine response with respect to control parameters, enabling a balance of exploration and exploitation in the search for the global optimum. We show that the model parameters can be learned from archived scans, and correlations between devices can be extracted from the beam transport. The result is a sample-efficient optimization routine, combining both historical data and knowledge of accelerator physics to significantly outperform existing optimizers.

DOI: 10.1103/PhysRevLett.124.124801

Modern large-scale scientific experiments have complicated operational requirements, with performance degraded by errors in controls and dependencies on drifting or random variables. A prime example of this is the Linac Coherent Light Source (LCLS) [1], an x-ray free-electron laser (FEL) user facility that supports a wide array of scientific experiments. At LCLS, skilled human operators tune dozens of control parameters on the fly to achieve custom photon beam characteristics, and this process cuts into valuable time allocated for each user experiment. Model-independent optimizers can help automate tuning, with successful demonstrations using simplex [2,3], extremum seeking [4–6], and robust conjugate direction search [7,8]. However, these methods require a large number of expensive acquisitions and can become stuck in local optima. To improve the efficiency of optimization beyond these methods, a model of the system is necessary [9]; in this Letter, we use Bayesian optimization with Gaussian process (GP) models during live tuning of LCLS. First, we use archived data to estimate the length scales of tuning parameters in the model. Second, we show that adding physics-inspired correlations between parameters further speeds convergence. Finally, we discuss possible directions for improvement.

Bayesian optimization is a sample efficient and gradient-free approach to global optimization of black-box functions with noisy outputs [10–12]. This efficiency comes from application of Bayes's theorem to incorporate prior knowledge and previous steps to maximize the value of each new measurement. Numerical optimization of an acquisition function, incorporating the model's expectation and uncertainty, guides the selection of each new point to sample. This gives Bayesian optimization the capability of balancing exploration with exploitation, which makes it ideal for optimizing noisy or uncertain targets. Prior knowledge such as historical data, simulation, or theory can improve the efficiency of optimization and constrain the search in low signal-to-noise states. Bayesian optimization has been applied to a quickly growing number of domains; for example, resource prospecting [13,14], active policy search for reinforcement learning [15], neural network parameter tuning [16], experimental control [17], etc. Large scientific experiments are expensive to run, have noisy and uncertain inputs and outputs, and are supported by detailed physical theory, simulations, and extensive data sets; all of these qualities make complex physics experiments well-suited for control via Bayesian optimization.

A GP is a popular choice of model for Bayesian optimization [18]. Whereas a Gaussian distribution is characterized by a mean and covariance $y \sim N(\boldsymbol{\mu}, \Sigma)$, a Gaussian *process* is determined by mean and covariance *functions*: $f(\boldsymbol{x}) \sim \mathrm{GP}[\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x})]$, where $\boldsymbol{x}$ are possible inputs to the objective. The mean $\mu(\boldsymbol{x})$ is a function encoding the expected value of the objective function. Prior to updating the model with observations, $\mu(\boldsymbol{x})$ may encode prior understanding of the objective, such as a fit to data or the most likely value of the objective given random samples on a relevant domain, or may be set to zero without loss of generality. As observations are added to the GP, the mean function begins to interpolate near measured points, and represents the most probable value of the objective given the measured data. The prediction's uncertainty $\sigma(\boldsymbol{x})$

incorporates the measured data with a covariance function or kernel $k(\boldsymbol{x},\boldsymbol{x}')$, which describes the similarity between pairs of points $\boldsymbol{x}$ and $\boldsymbol{x}'$. Parameters of the mean and covariance functions are referred to as hyperparameters and are fit to a training data set by maximizing the probability of the data given the model, which incorporates a measure of the data fit and a natural model complexity penalty that discourages over-fitting (see Sec. 5.4.1 of [18]). Whereas the mean function may change day to day, the covariance function is fundamental to the underlying physics and thus is relatively constant and more robust for online estimation of the observed data.

In contrast to parametric models such as neural networks, GPs are nonparametric models, meaning that they are constructed directly from the training instances themselves. This allows for the model's complexity to grow with the number of observations and predicts a class of functions consistent with a previously learned or physics motivated covariance structure near the current observations. Furthermore, in addition to the mean, GPs also provide predictions' uncertainties, which are needed to construct the acquisition function. For both of these reasons, GPs are popular models for Bayesian optimization.

In this Letter we apply Gaussian process optimization to the problem of maximizing the LCLS x-ray FEL pulse energy. The FEL instability is a collective effect so performance depends strongly on the current density and therefore beam size along the FEL undulator line. A series of quadrupole magnets (with positions shown in Fig. 1) transports and focuses the beam into the undulator line which is composed of 30 undulator modules of length 3.3 m, each followed by 0.6 m of space for diagnostics and corrector magnets to keep the beam on axis and quadrupole magnets to provide strong focusing and maintain a small beam size. The FEL pulse energy is therefore a function of all the quadrupole magnet strengths (field integral in kiloGauss, kG) and the input electron beam parameters. The machine is reconfigured at least twice per day, and hysteresis in magnets and motors throughout the machine necessitates retuning following each change. Consequently, LCLS operators perform tuning scans to optimize a subset (see Fig. 1) of the quadrupoles' strengths many times per day. In 2015 we found that LCLS was spending more than 500 hours/year on the single task of quadrupole optimization due to the high dimensional parameter space. Reducing this lost experimental time became an important goal for operations.

A central component of Bayesian optimization is a system model. In our case, we model the FEL pulse energy dependence on quadrupole amplitude with a Gaussian process. Given the FEL pulse energy's smooth response to quadrupole magnets when all other parameters are fixed, we choose the popular radial basis function (RBF) kernel,

$$k_{\mathrm{RBF}}(\boldsymbol{x},\boldsymbol{x}') = \sigma_f^2 \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}')^T\Sigma(\boldsymbol{x}-\boldsymbol{x}')\right] \quad (1)$$

Here, $\boldsymbol{x}$ and $\boldsymbol{x}'$ are vectors with quadrupole values as coordinates, ^{T} denotes a matrix transpose operation, σ_f^2 is the covariance function amplitude, and Σ is a diagonal matrix of inverse square length scales. The length scales set the distance over which the function changes in each dimension. The amplitude captures the variance of the target function values with respect to variations in the inputs and therefore determines the prior prediction uncertainty far from any sampled points. The amplitude and length scale hyperparameters are chosen by fits to archived data; see the Supplemental Material for more details on the choice of the kernel and its hyperparameters [19]. Learning the covariance of the function (e.g., via a GP fit) rather than the function itself (e.g., a neural network fit) provides uncertainty estimates by favoring a subset of possible functions compatible with observations and allows for robust optimization since the peak of the objective may change day to day, but the general shape remains the same.

Online optimization proceeds by first measuring the initial state of the machine and then initializing the Gaussian process model with the appropriate kernel and

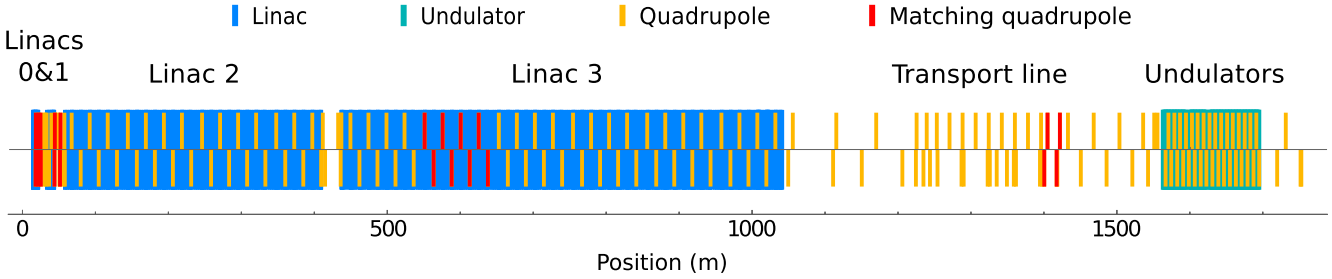
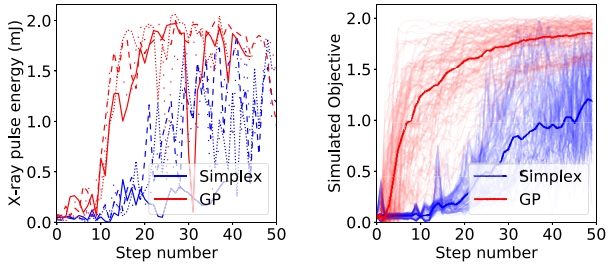


FIG. 1. LCLS beam line showing locations of linacs, quadrupole magnets, and undulator magnets. A focusing system comprised of nearly 200 quadrupoles (upper lines denote focusing whereas lower lines denote defocusing for one plane) transports the electron beam downstream (from left to right above) and several groups of matching quadrupole magnets (shown in red) are routinely used to match the beam into the focusing system. An electron beam dump and x-ray pulse energy measurement system (not shown) are positioned downstream of the undulators. Our studies focus on eight matching quads in the third linac and four in the transport line leading to the undulators.

first measured point. The GP provides a probabilistic surrogate model for the machine, and an upper confidence bound (UCB) acquisition function [23] is constructed from the GP prediction mean and standard deviation: $\mathrm{UCB}(x) = \mu(x) + \kappa\sigma(x)$, where κ is of order unity. The UCB acquisition function employs a strategy of optimism in the face of uncertainty, balancing the trade-off between exploiting known promising regions [via the prediction mean term $\mu(x)$] and exploring uncertain areas [via the prediction uncertainty $\sigma(x)$]. The point maximizing the UCB function is chosen for the next measurement, which is then acquired and added to the GP, finishing one step through the optimization process (see the Supplemental Material for details of our specific implementation [19]). The optimization continues in this way until reaching a time limit or a target performance.

There are many Gaussian process codes available [24–28], and many of these packages employ techniques to limit computation times for large data sets. In our case, we use an online GP model [29] interfaced to LCLS via the Ocelot optimization framework [3]. The online model saves frequent computations to speed subsequent predictions. In practice at LCLS, computation time per step is shorter than the acquisition time (1 to 2 seconds to set magnets and allow feedback systems to correct the trajectory and 1 second to measure FEL pulse energies for 120 shots).

Figure 2(a) shows results from live optimization of the FEL pulse energy simultaneously on 12 quadrupoles. In this example, Bayesian optimization is approximately four times faster than LCLS's standard Nelder-Mead simplex algorithm, and reaches a higher optimum. The different lines for each algorithm correspond to scans with identical starting conditions.

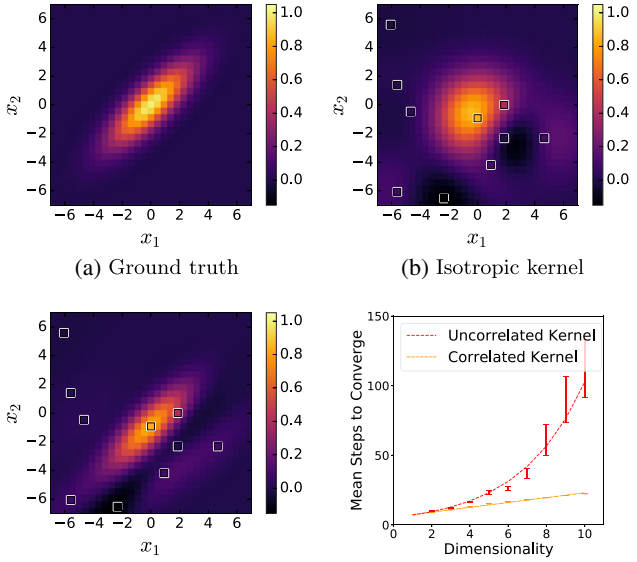


(a) Live 12 quadrupole optimization (b) Simulated 12 quadrupole optimization

FIG. 2. (a) Comparison of optimization of FEL pulse energy over 12 quadrupole magnets for Bayesian optimization (red curves) vs Nelder-Mead simplex optimization (blue). Each scan was performed four times with identical starting conditions, shown with different dashing. Each step corresponds to approximately 3 to 4 seconds. (b) Simulations using the conditions of (a). 100 individual scans for each method, with means shown by thick lines, are consistent with measurements. Note occasional drops in FEL pulse energy, e.g., at step 30 in the solid red line, are caused by momentary machine glitches. It is encouraging that the optimizer is not thrown off course by such brief glitches.

We also compare simplex and Bayesian optimization in a simulation environment. Ideally we would use physics codes such as elegant [30] and Genesis [31] to model the transport and FEL behavior, however due to mismatch between the codes and measured performance, as well as computational expense for each simulation, we instead fit a beam transport model as described below. Though the beam transport model does not capture the full complexity of the real machine, it allows us to compare the relative performance of simplex and Bayesian optimization with a simulated objective function, which we find consistent with live scans [Fig. 2(b)].

We can further improve the optimizer by leveraging our knowledge of accelerator physics to introduce correlations to the model. Correlated kernels become advantageous when a system's response to one input dimension depends strongly on one or more of the other input dimensions as is exemplified with a correlated binormal test function in Fig. 3(a). Figure 3(b) shows a GP regression on noisy samples (RMS noise is 10% of the signal peak) from a correlated ground truth [Fig. 3(a)] with an isotropic kernel, while Fig. 3(c) shows a GP regression on the *same* samples but with a kernel sharing the same correlation as the ground truth ($\rho = 0.8$). The latter model is more representative of

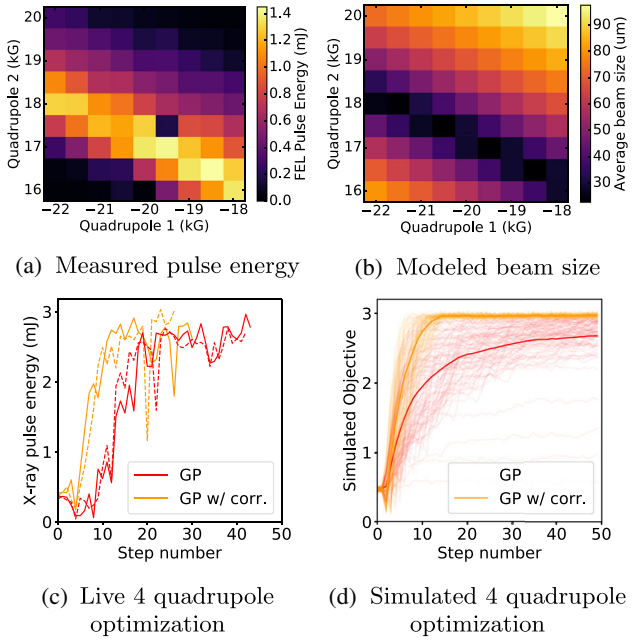


(a) Ground truth (b) Isotropic kernel (c) Correlated kernel (d) Convergence tests

FIG. 3. (a) Ground truth of a 2D multinormal test function with unit slice widths (evaluated holding all but one coordinate fixed at a time) and correlation coefficient of 0.8. (b) GP regression with isotropic kernel on noisy samples from the ground truth. (c) GP regression with correlated kernel on identical samples as (b). (d) Bayesian optimization convergence tests on a correlated ground truth with and without kernel correlations. Each bar shows the standard error about the mean for 100 trials. The correlated GP kernel (blue linear fit) performs as well as optimization of an isotropic ground truth with an isotropic GP kernel, growing linearly with the number of dimensions. Steps to convergence with mismatched kernel grows exponentially (red exponential fit).

the system. To demonstrate the effectiveness of a correlated model for regulating the system, we perform Bayesian optimization with and without kernel correlations for various dimensional spaces with nearest-neighbor correlation coefficients of $\rho_{i,i+1} = 0.5$. Each point in Fig. 3(d) shows the average number of steps to achieve $> 90\%$ of the ground truth peak amplitude for 100 runs starting at a random position such that the starting signal-to-noise ratio is unity. The relative efficiency of the correlated kernel grows exponentially with the number of dimensions, making it attractive for high-dimensional optimization at accelerators.

FEL quadrupole optimization is an example of a highly correlated system. Strong focusing in charged particle transport relies on a series of oppositely polarized quadrupole fields [32]. Each quadrupole focuses in one transverse plane while defocusing in the orthogonal plane, and repeated application of alternate focusing and defocusing results in net focusing in both planes. Increasing the strength of one quadrupole field necessitates increasing the strength of the next quadrupole (with opposite sign) to achieve net focusing, resulting in negative correlations between nearby focusing elements. Figure 4(a) shows the average FEL pulse energy response to variation of two adjacent quadrupoles in a matching section just upstream of the FEL undulator magnets (near position 1400 m in Fig. 1). Similarly, correlations exist between all pairs of quadrupoles.



(a) Measured pulse energy (b) Modeled beam size

(c) Live 4 quadrupole optimization (d) Simulated 4 quadrupole optimization

FIG. 4. (a) Average FEL pulse energy vs two adjacent quadrupoles. The dark spot near the center of the plot was due to a momentary machine malfunction. (b) The modeled average electron beam size in the undulator vs the same quadrupoles shows the same correlations. (c) Optimization test for four quadrupoles: GP (red) vs GP with correlations (yellow). Each scan was performed twice with identical starting conditions, shown with different dashing. Each step takes approximately 3 to 4 seconds. (d) Simulations using a beam matrix model for the conditions of c. 100 individual scans for each method, with means shown by thick lines, are consistent with measurements.

To find the correlation hyperparameters, we exploit our knowledge of strong-focusing and FEL physics to calculate correlations from a beam physics transport model. The FEL pulse energy (denoted as U) is a correlation-preserving function of the transverse area of the beam, σ^2, averaged along the interaction, with $\log U \propto \langle\sigma^2\rangle^{-1/3}$ [33]. As a consequence, the beam size and pulse energy have similar correlations with respect to variations in the quadrupole magnets.

We model the average beam size in the undulator line from a linear transport model [32], with quadrupole values based on the archived settings. The result shown in Fig. 4(b) is a modeled beam size with correlations that match those of the measured FEL pulse energy [Fig. 4(a)]. Finally, we combine the correlations from the linear model with the length scales fit from archived data to give the correlated RBF kernel. In principle, we would calculate correlation hyperparameters from fits to the archive data, as was done for the length scales. However, due to sparsity in the sampled data, we find that our modeled correlations are more accurate. Conversely, in principle we could follow the same modeling procedure to calculate the length scale hyperparameters, but because our simple model does not handle details such as FEL saturation, we found empirically that the length scales from archived data are more accurate.

We tested FEL optimization using GPs built with and without correlations for four adjacent matching quadrupole magnets located in the transport between the accelerator and the FEL undulators. The results, presented in Fig. 4(c), show that the correlated GP model outperforms the uncorrelated GP. Furthermore, we find consistency of these results with simulations based on correlations from the beam transport model as shown in Fig. 4(d). It is interesting to note that the Bayesian optimizer maximized the FEL with a similar number of steps for the 12 quadrupole case [Fig. 2(a)], seemingly in contradiction to Fig. 3(d). The similarity is due to the fact that to leading order (assuming a monoenergetic beam and linear optics), only four quadrupole magnets are needed to match the four Twiss parameters into the undulator line. However, optimizing more quadrupoles further increases the FEL pulse energy by reducing chromatic effects which increase the electron beam emittance and suppress FEL gain. While four quadrupoles can recover a significant fraction of peak performance, in practice operators typically cycle through subsets of all of the controllable quadrupoles. As the number of control parameters increases, the GP with correlations is expected to provide an exponentially growing benefit over the standard GP.

In this Letter, we showed that online Bayesian optimization with length scales estimated from archived historical

data can tune the LCLS FEL pulse energy more efficiently than the current standard, Nelder-Mead simplex algorithm. Moreover, we showed that adding physics-informed correlations, obtained from beam optics models, further improves tuning efficiency. This latter effect was demonstrated with four control parameters, and the improvement is expected to grow dramatically as the number of dimensions increases. The flexibility of the GP enables training from archived data, simulations, and physical models, and incorporating all this information into the model allows it to improve in efficiency and adapt to new tasks—all without the need for specialized hardware such as graphics processing units (GPUs) or high performance computing clusters.

We expect Bayesian optimization will become a standard tuning method for LCLS (and later for LCLS-II) when operations resume, improving efficiency and operational ability at SLAC as well as other accelerators worldwide. Whereas the present study focused on the most time-consuming task of tuning quadrupoles, we have plans for additional applications to other accelerator and beam line tasks, e.g., controlling FEL bandwidth, optimizing undulator field strengths [34], controlling electron beam collimation and peak current, FEL focusing and alignment, etc. More expressive GP models, such as deep kernel learning [35–37] or deep GPs [38], may help represent more complicated relationships between variables and extract additional value from historical data. Furthermore, adding "safety" constraints to the acquisition function can guide exploration while ensuring operational requirements are met (for instance, avoiding transient drops in FEL pulse energy or keeping losses low) [39].

In our study, we exploited the physics of strong focusing to learn correlations between parameters while empirically learning relevant length scales. This was necessary since simulation codes that model the FEL process in the postsaturation regime are computationally expensive, limiting their usefulness in online tuning. However in other systems, simulation alone may suffice to calculate length scales and correlations [40]. Looking forward, these codes may be rewritten in a framework that supports automatic differentiation [41] to simplify and accelerate Hessian calculations. Alternatively, it may be possible to replace some simulation codes entirely with fast-to-evaluate surrogate models [42] which enable quick approximations of the system's covariance. Physics abounds with well-verified mathematical models, and incorporating additional physics knowledge into the model—either explicitly in the formulation of the kernel function as shown here or via additional training with simulation—can have a significant effect on optimization efficiency.

The authors are grateful to the SLAC Accelerator Operations group for help with live tests on LCLS, to Lauren Alsberg for a Python script to query the LCLS archive for data, and to Greg White for updated transport matrices. We would also like to thank Hugo Slepicka and Sergey Tomin for help with the details of the Ocelot-optimizer code and to Tim Maxwell, Jeremy Mock, Ahmed Osman, and Hugo Slepicka for adding helpful LCLS-specific functions to the Ocelot optimizer. We are also appreciative to Gabriel Marcus, Brendan O'Shea, and Claudio Emma for helpful discussions. This work was supported by the Department of Energy, Laboratory Directed Research and Development program at SLAC National Accelerator Laboratory, under Contract No. DE-AC02-76SF00515.

[1] P. Emma *et al.*, Nat. Photonics **4**, 641 (2010).
[2] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).
[3] S. Tomin, G. Geloni, I. Agapov, I. Zagorodnov, Y. Fomin, Y. Krylov, A. Valintinov, W. Colocho, T. Cope, A. Egger, and D. Ratner, *Proceedings of the 7th International Particle Accelerator Conference* (2016), http://accelconf.web.cern.ch/AccelConf/ipac2016/papers/wepoy036.pdf.
[4] A. Scheinker, X. Pang, and L. Rybarcyk, Phys. Rev. ST Accel. Beams **16**, 102803 (2013).
[5] A. Scheinker, D. Bohler, S. Tomin, R. Kammering, I. Zagorodnov, H. Schlarb, M. Scholz, B. Beutner, and W. Decking, Phys. Rev. Accel. Beams **22**, 082802 (JACoW, Busan, 2019).
[6] A. Scheinker, X. Huang, and J. Wu, IEEE Trans. Control Syst. Technol. **26**, 336 (2018).
[7] X. Huang, Phys. Rev. Accel. Beams **21**, 104601 (2018).
[8] X. Huang and J. Safranek, Phys. Rev. ST Accel. Beams **18**, 084001 (2015).
[9] R. C. Conant and W. Ross Ashby, Int. J. Syst. Sci. **1**, 89 (1970).
[10] J. Močkus, *Optimization Techniques IFIP Technical Conference, Novosibirsk* (Springer, Berlin, Heidelberg, 1975), p. 400, https://link.springer.com/chapter/10.1007%2F3-540-07165-2_55.
[11] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, Proc. IEEE **104**, 148 (2016).
[12] E. Brochu, V. M. Cora, and N. de Freitas, arXiv:1012.2599.
[13] N. Cressie, Math. Geol. **22**, 239 (1990).
[14] J. P. Delhomme, Adv. Water Resour. **1**, 251 (1978).
[15] R. Martinez-Cantin, N. de Freitas, E. Brochu, J. Castellanos, and A. Doucet, Auton. Robots **27**, 93 (2009).
[16] J. Snoek, H. Larochelle, and R. P. Adams, *Advances in Neural Information Processing Systems 25 (NIPS 2012), Harrah's Lake Tahoe, 15 US-50, Stateline, NV 89449* (Neural Information Processing Systems Foundation, 2012), p. 2951, https://papers.nips.cc/paper/4522-practical-bayesian-optimization-of-machine-learning-algorithms.pdf.
[17] M. M. Noack, K. G. Yager, M. Fukuto, G. S. Doerk, R. Li, and J. A. Sethian, Sci. Rep. **9**, 11809 (2019).
[18] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, 2006), http://www.gaussianprocess.org/gpml/chapters/.
[19] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.124801 for details of the kernel hyperparameter calculation and the acquisition function optimization, which includes Refs. [20–22].

[20] D. Duvenaud, J. Lloyd, R. Grosse, J. Tenenbaum, and G. Zoubin, *Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA* (PMLR, 2013), Vol. 28, p. 1166, http://proceedings.mlr.press/v28/duvenaud13.html.

[21] S. Sun, G. Zhang, C. Wang, W. Zeng, J. Li, and R. Grosse, *Proceedings of the 35th International Conference on Machine Learning, Vienna, Austria* (PMLR, 2018), p. 4828, http://proceedings.mlr.press/v80/sun18e.

[22] R. Akre, D. Dowell, P. Emma, J. Frisch, S. Gilevich, G. Hays, P. Hering, R. Iverson, C. Limborg-Deprey, H. Loos, A. Miahnahri, J. Schmerge, J. Turner, J. Welch, W. White, and J. Wu, Phys. Rev. ST Accel. Beams **11**, 030703 (2008).

[23] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, IEEE Trans. Inf. Theory **58**, 3250 (2012).

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011), http://www.jmlr.org/papers/v12/pedregosa11a.

[25] GPy: A Gaussian process framework in python, http://github.com/SheffieldML/GPy (2012).

[26] A. G. d. G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrá, Z. Ghahramani, and J. Hensman, J. Mach. Learn. Res. **18**, 1 (2017), http://www.jmlr.org/papers/v18/16-537.html.

[27] C. E. Rasmussen and H. Nickisch, J. Mach. Learn. Res. **11**, 3011 (2010), http://www.jmlr.org/papers/volume11/rasmussen10a/rasmussen10a.pdf.

[28] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, PeerJ Comput. Sci. **2**, e55 (2016).

[29] M. McIntire, D. Ratner, and S. Ermon, *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence UAI16, New York* (AUAI Press, 2016), p. 517, http://www.auai.org/uai2016/proceedings/papers/269.pdf.

[30] M. Borland, Argonne National Lab. Tech., Report No. LS-287, 2000, https://doi.org/https://dx.doi.org/10.2172/761286.

[31] S. Reiche, Nucl. Instrum. Methods Phys. Res., Sect. A **429**, 243 (1999).

[32] H. Wiedemann, *Particle Accelerator Physics* (Springer, Berlin, Heidelberg, 2007), http://link.springer.com/10.1007/978-3-540-49045-6.

[33] K. Kim, Z. Huang, and R. Lindberg, *Synchrotron Radiation and Free-Electron Lasers* (Cambridge University Press, Cambridge, England, 2017), https://www.cambridge.org/core/books/synchrotron-radiation-and-freeelectron-lasers/F90638BB364B5A3EF9A78978CE15B1A5.

[34] J. Wu, X. Huang, T. Raubenheimer, and A. Scheinker, in *38th International Free Electron Laser Conference* (JACOW, Geneva, Switzerland, 2018), pp. 229–234, https://accelconf.web.cern.ch/AccelConf/fel2017/doi/JACoW-FEL2017-TUB04.html.

[35] R. Calandra, J. Peters, C. E. Rasmussen, and M. P. Deisenroth, *Proceedings of the International Joint Conference on Neural Networks, Vancouver, BC, Canada* (IEEE, 2016), p. 3338, https://ieeexplore.ieee.org/document/7727626.

[36] A. G. Wilson, Z. Hu, R. R. Salakhutdinov, and E. P. Xing, Adv. Neural Inf. Process. Syst. **29**, 2586 (2016), https://papers.nips.cc/paper/6426-stochastic-variational-deep-kernel-learning.pdf.

[37] A. Jacot, F. Gabriel, and C. Hongler, Adv. Neural Inf. Process. Syst. **31**, 8571 (2018), https://papers.nips.cc/paper/8076-neural-tangent-kernel-convergence-and-generalization-in-neural-networks.pdf.

[38] A. C. Damianou and N. D. Lawrence, *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS), Scottsdale, AZ, USA* (PMLR, 2013), Vol. 31, p. 207, http://proceedings.mlr.press/v31/damianou13a.

[39] J. Kirschner, M. Mutný, N. Hiller, R. Ischebeck, and A. Krause, *Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA* (Proceedings of Machine Learning Research, 2019), p. 3429, http://proceedings.mlr.press/v97/kirschner19a.

[40] A. Hanuka, J. Duris, J. Shtalenkova, D. Kennedy, A. Edelen, D. Ratner, and X. Huang, *Proceedings of the Machine Learning for the Physical Sciences Workshop, NeurIPS, Vancouver, BC, Canada* (Neural Information Processing Systems Foundation, 2019), https://ml4physicalsciences.github.io/files/NeurIPS_ML4PS_2019_85.pdf.

[41] J. Degrave, M. Hermans, J. Dambre, and F. wyffels, Front. Neurorobotics **13**, 6 (2019).

[42] A. Edelen, N. Neveu, C. Emma, D. Ratner, and C. Mayes, *Proceedings of the Machine Learning for the Physical Sciences Workshop, NeurIPS, Vancouver, BC, Canada* (Neural Information Processing Systems Foundation, 2019), https://ml4physicalsciences.github.io/files/NeurIPS_ML4PS_2019_90.pdf.