

Bayesian optimization of the beam injection process into a storage ringChenran Xu (徐晨冉)^{*}, Tobias Boltz[†], Akira Mochihashi, Andrea Santamaria Garcia,
Marcel Schuh, and Anke-Susanne Müller*Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany*

(Received 17 November 2022; accepted 15 February 2023; published 3 March 2023)

We have evaluated the data-efficient Bayesian optimization method for the specific task of injection tuning in a circular accelerator. In this paper, we describe the implementation of this method at the Karlsruhe Research Accelerator with up to nine tuning parameters, including the determination of the associated hyperparameters. We show that the Bayesian optimization method outperforms manual tuning and the commonly used Nelder-Mead optimization algorithm in both simulation and experiment. The algorithm was also successfully used to ease the commissioning phase after the installation of new injection magnets and is regularly used during accelerator operations. We demonstrate that the introduction of context variables that include intrabunch scattering effects, such as the Touschek effect, further improves the control and robustness of the injection process.

DOI: [10.1103/PhysRevAccelBeams.26.034601](https://doi.org/10.1103/PhysRevAccelBeams.26.034601)**I. INTRODUCTION**

Particle accelerators contribute to major discoveries in particle physics and are used to generate synchrotron radiation for photon science applications [1]. They consist of a large and varied number of components, from magnets to steer, focus, or quickly redirect the particle beam, to radio frequency (rf) sources to accelerate it. This leads to a large number of possible parameter combinations for tuning, which is a complex and time-consuming task. In addition, the beam dynamics are influenced by many nonlinearly correlated parameters and are subject to physical phenomena such as the Touschek effect [2]. Although advanced control systems are employed, the fine-tuning of beam properties is and will be a challenge, especially for future large-scale and compact accelerators. While intervention is required for several accelerator operation tasks, manual tuning may not find the global optimum, especially in the shortest possible time and does not adapt well to short-term and long-term drifts of the accelerator condition due to the large number of correlated parameters that influence it.

In such a case, computer algorithms can be introduced to assist the operator and eventually automate the tuning process [3]. Metaheuristics such as the evolutionary

algorithm [4] and particle swarm optimization [5] need a large number of evaluation steps to converge and are not suited for online tuning. Gradient-based methods like robust conjugate direction search [6,7] are successfully used for online accelerator tuning, but they are prone to get stuck in local optima and observation noise. Another promising local optimization method is the model-independent extremum seeking (ES) algorithm. Despite also being sensitive to local optima, ES can be used as a feedback control to track the optimum settings with respect to drifts of the accelerator components. It has been successfully applied to tune time-varying accelerator systems with a large number of parameters [8,9]. As a method to globally optimize an unknown function with expensive evaluations, Bayesian optimization (BO) is shown to perform well among other approaches [10]. With observed data, BO builds a surrogate model of the unknown function using a Gaussian process (GP) [11] and uses an acquisition function to guide the search efficiently. Recently, BO has been successfully applied at the LCLS [12,13] and the SwissFEL [14] for free-electron laser performance tuning, motivating the usage of Bayesian optimization for other accelerator tuning tasks [15,16].

The process of beam injection can be approximately viewed as a black-box function optimization problem, which makes BO a well-suited method for the injection tuning. To investigate this specific optimization problem, we use the accelerator test facility Karlsruhe Research Accelerator (KARA), which is a 110-m storage ring and part of the KIT Light Source. It produces synchrotron radiation for photon science users and accelerator physics experiments at a top energy of 2.5 GeV. The electron storage ring is filled at 500 MeV by a chain of

^{*}chenran.xu@kit.edu[†]Present address: SLAC, Menlo Park, USA

preaccelerators with a repetition rate of 1 Hz. The injection tuning is performed manually and initial studies have shown that BO is well suited to optimize this task [17]. Furthermore, it is known that the injection condition may vary from one day to another and operators need to retune the injection elements. This might result from a change in the environmental conditions such as the ambient temperature or the quality of the vacuum, variables that are either noncontrollable or can only be changed slowly. In this article, besides showing the implementation of BO at a storage ring with several examples and use cases, we also take into account the effect of the environment by generalizing the standard BO method to contextual Bayesian optimization (CBO) [18] to optimize the injection efficiency under the given context.

II. BAYESIAN OPTIMIZATION

Bayesian optimization (BO) is a method designed to optimize a black-box function f in a sample-efficient way. The BO algorithm used in this paper is shown in pseudocode in Algorithm 1 [19]. It makes use of a statistical surrogate model of the objective function f , generally built with a Gaussian process (GP), based on the belief that the objective function f is drawn from some prior probability distribution $p(A)$. After initialization and observations $f(x)$, the posterior distribution $p[A|f(x)]$ is built according to the Bayes' theorem

$$p[A|f(x)] \propto p[f(x)|A]p(A). \quad (1)$$

The posterior distribution is further used to build an acquisition function $\alpha(x)$, which determines the next point for evaluation.

Algorithm 1 Bayesian optimization algorithm.

-
- 1: Define the prior for the GP.
 - 2: Observe objective function f on n_0 initial points to get initial dataset D_0 .
 - 3: **for** $t = 1, 2, \dots$ **do**
 - 4: Build a GP using available data D_{t-1} .
 - 5: Find the next evaluation point
 $x_t = \arg \max_x \alpha(x|D_{t-1})$.
 - 6: Observe y_t at point x_t .
 - 7: Augment the dataset $D_t = D_{t-1} \cup (x_t, y_t)$.
 - 8: **end for**
-

A. Gaussian process

A Gaussian process (GP) is a generalization of the multivariate Gaussian distribution and is characterized by its mean and covariance functions $f(x) \sim \text{GP}[\mu(x), k(x, x')]$. In this study, we set the prior mean function to be $\mu(x) = 0$, which is commonly used when the shape of the objective function is unknown. The covariance function $k(\cdot, \cdot)$, also known as kernel, measures the similarity between data points. Based on the assumption that f is continuous, the

data points x that are close to each other are expected to have similar output values to $f(x)$. In this study, we use the widely used radial basis function (RBF) as the kernel function

$$k_{\text{RBF}}(x, x') = \exp \left[-\frac{1}{2} \sum_{i=1}^d \left(\frac{x_i - x'_i}{l_i} \right)^2 \right], \quad (2)$$

where l_i represents the length scale of the i th input dimension. It roughly corresponds to the distance along one input axis, at which the two data points become uncorrelated and the function values can change significantly.

In order to emulate the noise present in the real observed signal, we explicitly incorporate stochastic noise by adding Gaussian distributed noise to the covariance function as diagonal terms. Thus, the kernel becomes

$$k(x_i, x_j) = \sigma^2 k_{\text{RBF}}(x_i, x_j) + \sigma_{\text{noise}}^2 \delta_{ij}. \quad (3)$$

The signal variance σ^2 , the noise variance σ_{noise}^2 , and the length scales l_i are referred to as the hyperparameters, which determine the behavior of the GP. Since no dedicated optimization data are stored in the database, several measurements are performed to estimate the hyperparameter settings.

First, the noise σ_{noise} is determined by fixing the accelerator settings and measuring the fluctuation of the objective function, i.e., the injection efficiency, as defined in Eq. (7). By the assumption that the noise mostly comes from the statistical fluctuation between different injection shots, σ_{noise} can be extracted as the standard deviation of a Gaussian fit.

Second, the length scales l_i and the signal variance σ^2 are estimated from one-dimensional parameter scans. Figure 1 shows the scan result of the septum voltage, which corresponds to its magnetic field strength. We extract the GP hyperparameters with the maximum likelihood fit.

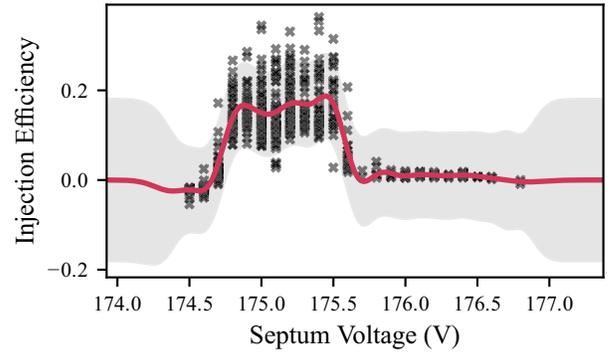


FIG. 1. Example of a parameter scan measurement of the septum magnet voltage, where all other accelerator tuning parameters are kept constant during the scan. The GP hyperparameters $\{l_i, \sigma^2\}$ are fitted to the measured data points (crosses) via a log-likelihood fit. The solid line depicts the GP posterior mean and the shaded region shows the 95% confidence level.

B. Acquisition function

With a GP model inferring the posterior distribution of the objective function, an acquisition function α is built to determine the next sample point to evaluate so that the number of required physical observations is effectively reduced. Concretely, the objective f is sampled at $\arg \max_x \alpha(x|D)$ at each step, where D is the observed dataset. We use two different acquisition functions in this study: the upper confidence bound (UCB) and the expected improvement (EI). UCB explicitly controls the exploration-exploitation trade-off with a parameter κ

$$\alpha_{\text{UCB}}(x) = \mu(x) + \kappa\sigma(x), \quad (4)$$

where $\mu(x)$ and $\sigma(x)$ are the GP posterior mean and standard deviation. For high κ values, the contribution of the variance term becomes large and points with high uncertainty are sampled, which leads to more exploration. On the contrary, small κ values lead to more exploitation of observed peaks, as they have a higher posterior mean.

An empirical choice of the parameter is $\kappa = 2$, which corresponds to the usual 95% confidence bound for Gaussian distributed values. Nevertheless, κ can also increase along with the evaluation steps to ensure that BO converges to the global optimum [19,20].

The expected improvement (EI) calculates the expected value of the improvement of a proposed point x over the best-observed value f_{best} [21]

$$\begin{aligned} \alpha_{\text{EI}}(x) &= \mathbb{E}[\max(\mu(x) - (f_{\text{best}} + \xi), 0)] \\ &= (\mu(x) - (f_{\text{best}} + \xi))\Phi(Z) + \sigma(x)\phi(Z), \end{aligned} \quad (5)$$

with the parameter Z describing the normalized improvement

$$Z = \frac{\mu(x) - (f_{\text{best}} + \xi)}{\sigma(x)}, \quad (6)$$

where ϕ is the probability distribution function and Φ is the cumulative distribution function of the standard normal distribution. The exploration-exploitation trade-off in EI is determined by a positive parameter ξ . In general, higher ξ values lead to more exploration. In this study, we use the recommended value of $\xi = 0.01$ according to Ref. [22], which emphasizes exploitation.

III. IMPLEMENTATION AT A STORAGE RING

In this section, we describe the implementation of the BO method to improve the injection efficiency into a storage ring, namely the storage ring of the KIT Light Source, the Karlsruhe Research Accelerator (KARA).

The electron storage ring is filled at 500 MeV with a repetition rate of 1 Hz by a chain of preaccelerators, consisting of an electron gun, a microtron, and a booster synchrotron. Further details can be found in Ref. [23].

The storage ring employs a three-kicker injection scheme [24], as shown in Fig. 2, to create a closed orbit bump. When the kicker magnets are powered, they steer the closed orbit away from the nominal trajectory and near the septum magnet. Kicker magnets are fast pulsed (in the order of μs) with a relatively low field strength, so they are combined with septa magnets that provide a stronger field. The septum magnet deflects the beam into the aperture of the storage ring and provides a separation space between the circulating and injected beam. The separation barrier (septum) is usually as thin as possible with minimum field leakage. The injection bump orbit at KARA spans over one quadrant of the storage ring. There are several sextupoles in the injection bump which add nonlinearities to the beam dynamics of the injection process. The stray field of the septum magnet and the potential energy mismatch between the booster and the storage ring also affect the settings required for a closed bump orbit. Moreover, due to the lack of nondestructive beam diagnostics in the transfer line except for one current monitor, the tuning is partly blind and relies mostly on the operators' experience.

In order to obtain a good injection rate, operators have to adjust multiple parameters like the magnet currents and radio frequency (rf) parameters. The manual tuning often ends up at a local optimum with a lower injection rate, which eventually reduces the beamtime availability. The tuning task after a shutdown period can be especially difficult since previous accelerator settings can no longer be used and it costs valuable beamtime to recover the expected injection performance. We show that the BO method outperforms the simplex methods and achieves a sample-efficient optimization.

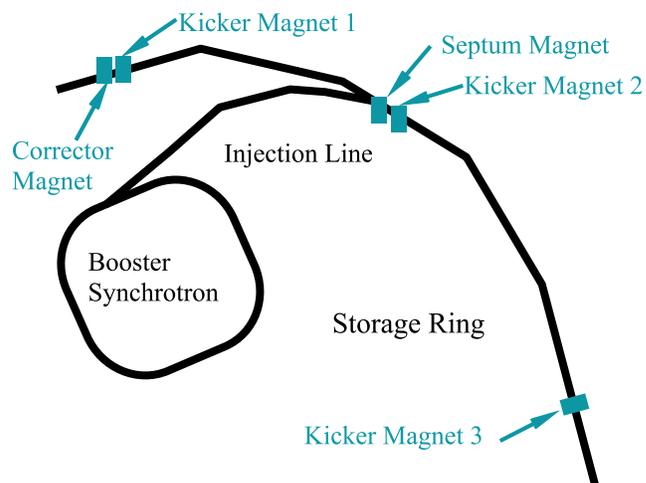


FIG. 2. In the injection scheme, electron bunches are injected from a synchrotron via an injection line into the storage ring. The relative positions of the magnets and their location are illustrated as they are placed at the storage ring KARA at KIT. See Ref. [23] for further details. The tuning parameters of the magnets and rf cavities, not shown in this figure, are used as optimization parameters.

We selected nine tuning parameters for the BO based on their influence on the injection process and from prior knowledge of experienced operators. The parameters are the main radio frequency (rf), the injection septum magnet strength, the strength of one horizontal corrector magnet, and the strength and timing of the three kicker magnets. The strength and timing of the three kickers define the shape of the injection bump orbit. The strength of the septum magnet changes the angle of the beam injected into the storage ring. The corrector magnet chosen here is located directly before the first kicker magnet and controls the angle of the incoming beam to the injection bump. Finally, the rf defines the orbit length and by changing it slightly, the closed-orbit condition is modified. The relative positions of the magnets relevant to the injection optimization process are illustrated in Fig. 2.

Since only the storage ring parameters are considered, the objective function is also chosen accordingly to decouple the effect of the preaccelerators. The injection efficiency is defined as the net injected current into the storage ring (SR) normalized by the current measured at the booster synchrotron (S) extraction point before the beam enters the injection line to KARA

$$f_{\text{I-eff,exp}} = \frac{\Delta I_{\text{SR}} h_{\text{SR}}}{I_{\text{S}} h_{\text{S}}}, \quad (7)$$

where h_{SR} and h_{S} are the harmonic numbers of the storage ring and the booster synchrotron, respectively. The change in the storage ring current ΔI_{SR} and the extracted bunch current from the synchrotron I_{S} are read out synchronously to calculate the objective function. Whereas for the simulation study, n_i , electrons extracted from the booster synchrotron, and n_s , already stored electrons in the storage ring, are tracked. The objective function is then calculated from the electrons that are successfully injected into the storage ring $n_{i,\text{success}}$ and the stored electrons that are lost during the injection process $n_{s,\text{lost}}$

$$f_{\text{I-eff,sim}} = \frac{n_{i,\text{success}} - n_{s,\text{lost}}}{n_i}. \quad (8)$$

During a realistic beam injection process at a storage ring, the stored current is usually much higher than the injected bunch current. Thus, a detuned parameter setting causing a beam loss can result in a very large negative objective function value and distort the landscape of the GP model. Additionally, there could be outliers due to faulty readback values or occasional beam losses. In order to mitigate these effects, the lower limit of the injection efficiency is constrained to be -1 so that the GP model is more robust to the fluctuation of readback values. This should not affect the parameter region near the optima, where the stored beam is minimally disturbed and the injection efficiency is always positive.

The BO package implemented in this study uses the PYTHON interface *pyepics* [25] for the communication with the control system of the accelerators, which is based on the Experimental Physics and Industrial Control Systems (EPICS) [26]. The software package *GPy* [27] is used for building the GP model and *scipy* [28] functionalities are used for maximizing the acquisition functions.

A. Optimization results in simulation

Before deploying the algorithms at the accelerators, the BO is first tested on a simulation. The injection model is built in Accelerator Toolbox for MATLAB (AT) [29]. We used a simplified lattice without magnet errors based on an existing MATLAB model for KARA [30,31].

In the simulation, n_i injected and n_s stored electrons are generated according to the particle distribution and tracked for the first 100 turns after injection without collective effects. This proved to be sufficient, as over 95% of the particle loss happens during the first ten turns. The injection efficiency is calculated according to Eq. (8).

The BO algorithm is first tested using two tuning parameters: the septum and the corrector magnet strength. For comparison, a grid scan, shown in Fig. 3(a), is conducted with 400 observations in total. Figure 3(b) shows the GP posterior mean function after 20 evaluations.

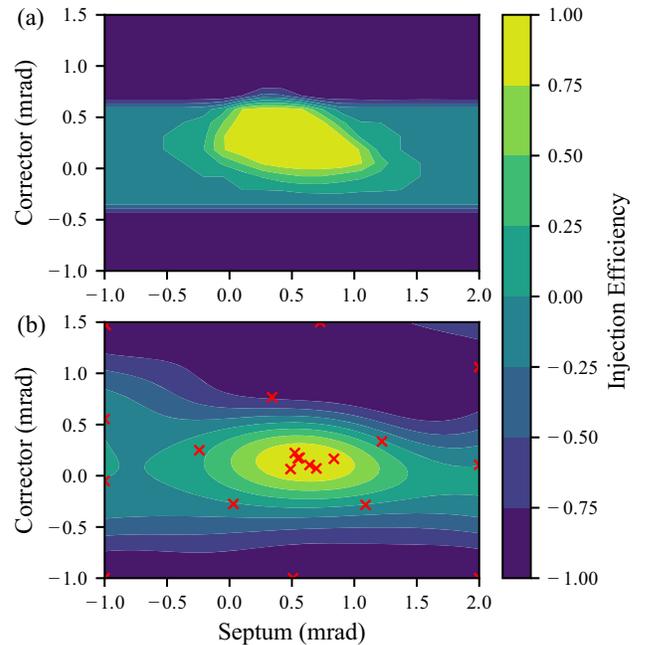


FIG. 3. Comparison of (a) a two-dimensional grid scan (grid size 400) of the injection efficiency with (b) the GP posterior mean predicted based on 20 BO evaluation steps, shown in red crosses. The tuning parameters are the strengths of the horizontal corrector magnet and the septum magnet, given in deflection angles. The BO can approximate the simple structure of parameter space in a small number of evaluation steps.

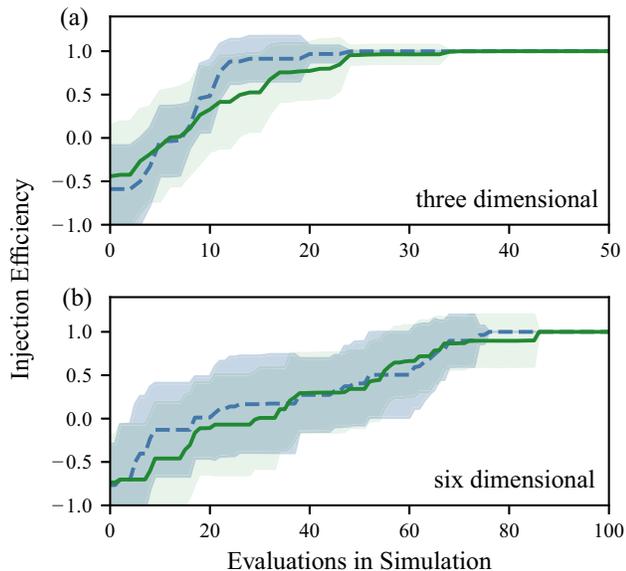


FIG. 4. Optimization results in an AT simulation model using UCB (green, solid) and EI (blue, dashed) acquisition functions for the (a) three- and (b) six-dimensional problems. The lines depict the best-evaluated injection efficiency in each run, averaged over ten independent runs. The shaded areas are the one σ spread. Within 50 and 100 steps for three- and six-dimensional problems, respectively, all the optimization runs converged to the optimum setting. The required steps for convergence increase with the input dimensions and both acquisition functions have a similar performance.

It can be seen that the BO already converges to the maximum and is able to model the peak structure.

As a next step, the rf frequency is included as a third tuning parameter. For the optimization, the BO randomly samples 5 points to initialize the GP model. The results of the three-dimensional problem using two different acquisition functions are shown in Fig. 4(a). It can be seen that both EI and UCB are able to find the optimum in a small number of evaluation steps. Then, we include the strength of the three kicker magnets as tuning parameters and the results are shown in Fig. 4(b). Although the number of required evaluations is larger, BO can still efficiently solve the six-dimensional problem.

Nevertheless, BO does not take other physical properties into account, since it only optimizes for the objective function $f_{I\text{-eff, sim}}$ [cf. Eq. (8)]. For example, as can be seen in Fig. 5, the strength of the three kicker magnets is often not matched perfectly, which leads to a nonlocalized injection bump orbit. Although a large orbit oscillation of about 3 mm along the storage ring is visible, the electrons can be stored and no beam loss is observed.

B. Beam lifetime correction

In the simulation, the collective effects are neglected and the beam losses are only due to the detuned settings of the injection magnets, while in the actual storage ring, the

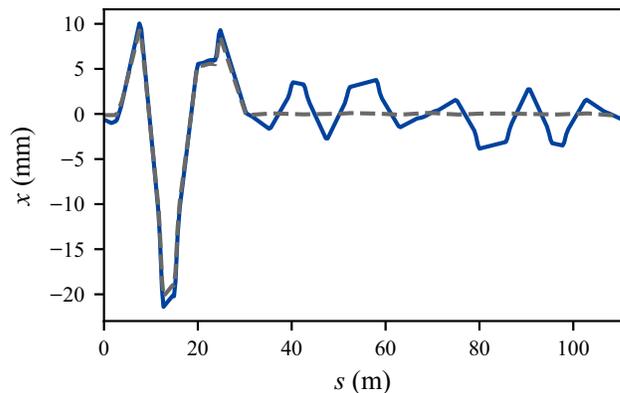


FIG. 5. Example of a nonlocalized orbit defined by the magnet strengths found by the BO algorithm (navy, solid), in comparison to the reference orbit (gray, dashed) with design values, where the injection bump orbit is fully localized. The orbit oscillation outside of the injection bump is not taken into consideration by BO, as it causes no beam loss.

electrons are lost over time due to other effects such as scattering within the bunch and current-dependent tune-shifts. In this study, we focus on the Touschek effect, which is one of the main contributions to the reduction of beam lifetime. Additionally, particles can be lost due to the scattering with the residual gas, which depends on the vacuum quality. The two effects can be combined in terms of the total lifetime τ

$$\frac{1}{\tau} = \frac{1}{\tau_T} + \frac{1}{\tau_{\text{other}}} = aI + b, \quad (9)$$

where the Touschek lifetime τ_T is dominant and inversely proportional to the current I , and other contributions τ_{other} , including beam-gas scattering, are approximately constant.

The beam lifetime decreases for higher accumulated current and leads to a decrease in the calculated injection efficiency, which is independent of the performance of the BO algorithm. As a result, the objective function will continue to decrease regardless of the sampling region, which is not taken into account in our definition of the objective. It is observed that BO often fails to optimize if this effect is not properly dealt with.

This effect can be counteracted by including the beam current as a context variable, which is discussed in Sec. III E. Alternatively, we explicitly correct this effect for normal BO by measuring the lifetime-related beam loss. For the latter, we measured the beam lifetime at the injection optics and fitted Eq. (9) to it. The injection efficiencies are then corrected accordingly in the next experiments.

C. Experimental optimization results

In the following section, we describe the results of the BO algorithms in a real-world accelerator environment.

We investigate its performance using the EI and UCB acquisition functions introduced in Sec. II B.

We use the RBF kernel [Eq. (2)] with the hyperparameters determined from the results in Sec. II A. For each test run, the machine was manually detuned to a fixed initial setting, where the beam injection was still possible with a very low injection efficiency. For the UCB acquisition function shown in Eq. (4), we chose the setting of κ according to [20], which dynamically changes the exploration-exploitation behavior

$$\kappa = \sqrt{2\nu \log(t^{d/2+2}\pi^2/(3\delta))}, \quad (10)$$

with $\nu = 1$ and $\delta = 0.1$ being free parameters. The parameter κ becomes larger with the number of evaluations t and the algorithm focuses more on the exploration.

The optimization was performed for a fixed number of steps: 50 steps for three- and six-dimensional problems and 100 steps for nine-dimensional problem. This corresponds to about 10–20 minutes for each optimization run, where most of the time was spent on setting the parameter values and evaluating the injection efficiency. The computation time of the GP model and the acquisition function is negligible. Additionally, it should be noted that we intentionally decreased the total booster synchrotron current I_S in the following experiments to reduce the radiation dose. This is expected not to influence the result, as the objective $f_{I\text{-eff,exp}}$ defined in Eq. (7) is normalized with respect to I_S .

Figure 6 shows the averaged performance over three optimization runs for each setting. For all the optimization runs, the injection efficiency clearly increased from the detuned initial setting within the allowed evaluation steps. The final achieved objective function value is, however, lower than the result obtained in simulation studies, which is due to various reasons. First, the objective function $f_{I\text{-eff,exp}}$ is normalized with respect to the current measured in booster synchrotron I_S right before extraction. The beam loss in the injection line is not accounted for, which needs to be mitigated by including additional magnets in BO. However, this is beyond the scope of this paper. Second, the current readout in the storage ring and the synchrotron booster are calibrated independently, affecting their ratio. Nevertheless, the performance of the BO can be compared and benchmarked qualitatively. It can be seen that in the three-dimensional case [Fig. 6(a)], the UCB evaluations are more noisier than the other configurations, which is mainly due to the setting of the trade-off parameter κ . However, for six- and nine-dimensional problems, the parameter space is larger and more exploration becomes indeed necessary. Therefore, the performance of both UCB and EI is quite similar in those cases [Figs. 6(b) and 6(c)]. Due to its focus on exploration, occasional unexpected beam loss is observed during the UCB runs. This can be circumvented either by introducing a safety constraint or setting κ to a lower value to focus more on exploitation.

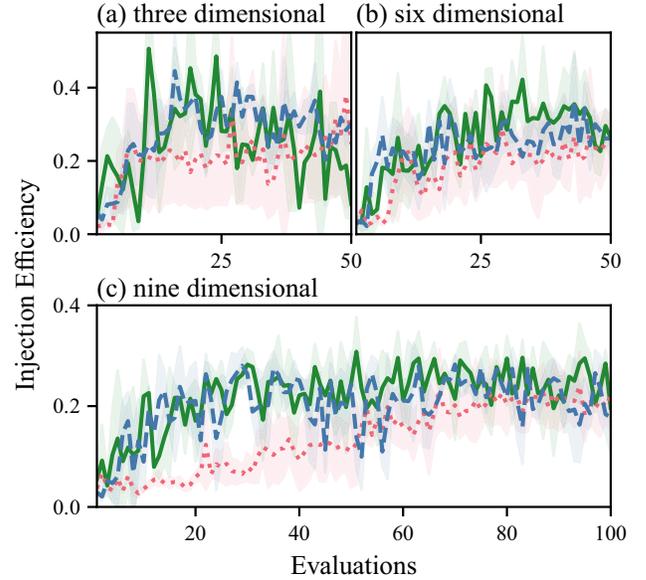


FIG. 6. Experimental optimization result of the injection efficiency using BO with two different acquisition functions and Nelder-Mead algorithm (red, dotted). UCB (green, solid) and EI (blue, dashed) have similar results and are both able to optimize the injection efficiency within a reasonable number of evaluations. For a nine-dimensional problem, NM is clearly slower than BO.

For comparison, we perform the same optimization procedure using the Nelder-Mead algorithm [32]. Nelder-Mead is a widely used numerical optimization method and is considered a standard benchmark in the accelerator community [7]. The algorithm keeps track of $d + 1$ evaluation points and uses these to build a simplex. The Nelder-Mead method searches for the optimum via geometric modifications of the stored simplex and is proven to converge relatively fast. The results of the Nelder-Mead algorithm are also plotted in Fig. 6. For the three- and six-dimensional cases, Nelder-Mead achieves a similar performance to BO, whereas for the nine-dimensional problem, the optimization speed of Nelder-Mead is clearly slower. The fact that Nelder-Mead scales not as well as BO is expected, as the local search becomes less applicable for higher dimensional spaces. The primary problem for BO to scale to higher dimensions is the computation time of the GP model and maximizing the acquisition function. However, these are still noncritical for $d < 20$ [33]. It is worth mentioning that we used randomly selected parameter settings to initialize the GP model, which often leads to a poor performance in the first few steps, as can be seen in Fig. 6 three-dimensional case. Alternatively, one can mitigate that by using the initial steps of other algorithms, such as Nelder-Mead, or some historical data to initialize the GP model. This could avoid the unstable phase of BO in the beginning and subsequently reach quicker convergence.

The distances Δx between two consecutive sampled points of BO and Nelder-Mead for the nine-dimensional problem are shown in Fig. 7. It is visible that Nelder-Mead searches only locally. Since Nelder-Mead cannot handle the noisy signal well, it easily gets stuck and continues to contract, which eventually converges to a nonoptimal setting. To prevent Nelder-Mead from breaking down and force exploration of the parameter space, it is automatically restarted when the step becomes too small. On the contrary, BO is more robust against the observation noise by explicitly modeling it as a hyperparameter. It can be seen that BO generally samples at a larger distance and does not become trapped in local optima.

In the experiments, we found that the timing of the kicker magnets is consistent with the previous set of values. As visible in Fig. 6(c), adding the timing parameters only slows down the optimization and does not improve the objective function further. Thus, we run the BO always with six parameters in the following sections.

As mentioned in Sec. III A, the settings found by the BO approach sometimes result in a not closed injection bump orbit. We also observed this effect during the experiments on KARA despite the fact that the orbit oscillation is partly mitigated by the radiation damping. Since the trajectory of the electron beam is not centered through the quadrupole and sextupole magnets, the resulting orbit is often different from the one usually obtained via manual tuning and sometimes a slight tune shift is visible. Although this effect does not affect the beam injection process for KARA, it might be critical for other accelerators with more stringent constraints on orbit and betatron tunes.

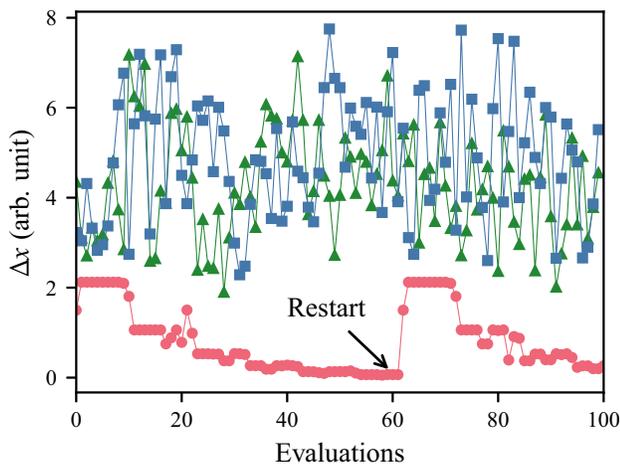


FIG. 7. Convergence behavior of BO with UCB (green, triangle) and EI (blue, square) acquisition functions and Nelder-Mead (red, circle) on the nine-dimensional problem showing the distance Δx between two consecutive sampled points in the scaled parameter space. Due to the noisy evaluation of the objective function, Nelder-Mead optimizes mostly locally and often needs to be restarted, whereas BO optimizes globally and does not require a restart.

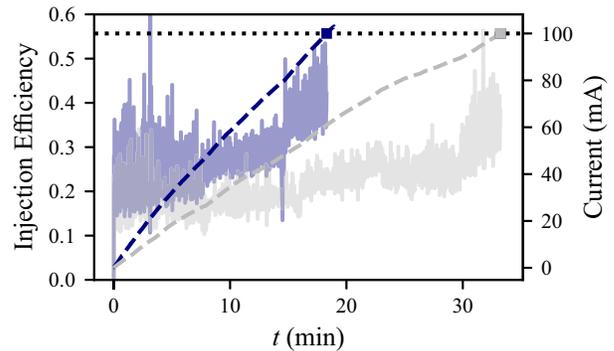


FIG. 8. Beam injection up to 100 mA using BO result (navy) compared to the manual tuning result (grey) obtained from an operation day. The dashed lines are the accumulated current of BO and manual tuning result. BO was roughly twice as fast as the manual tuning, greatly reducing the required beam injection time.

Finally, the optimal settings found by BO are compared to the result of 1 day with dedicated manual tuning, taken a week prior to the BO experiment with a comparable accelerator condition. In this experiment, the full booster current is used in order to achieve higher injection rate. As shown in Fig. 8, BO took 18 min and manual tuning took 33 min for the injection of 100 mA current. BO clearly outperforms the manual tuning and significantly reduces the required time for beam injection.

D. Application to the commissioning of injection magnets

During a scheduled shutdown period, the power supplies of the septum and kicker magnets were exchanged. Thus, it was necessary to find new parameter settings during commissioning.

Two steps are needed to apply BO and assist in the commissioning of the new power supplies: (i) The interface with the machine is updated so that BO can control the parameters of the new magnets. (ii) Several parameter scans are conducted in order to estimate the length scales and optimization ranges for the new parameters. Once BO was reconfigured for the new machine status, it quickly found new working points for the beam injection.

The BO started from a fixed nonoptimal setting and was run multiple times using six input parameters. Figure 9 compares two of these runs with the highest and lowest achieved injection efficiency among all five runs. The evolution of the parameters during the optimization is shown within the allowed tuning range of each parameter, respectively. The rf frequency is varied very little and the set values are comparable for two runs, which is as expected. Since the rf frequency directly depends on the ring circumference, it should only change due to, for instance, the change of ambient temperature over the year. The optimized septum strength is clearly higher than the starting point. This is consistent with the expectation that

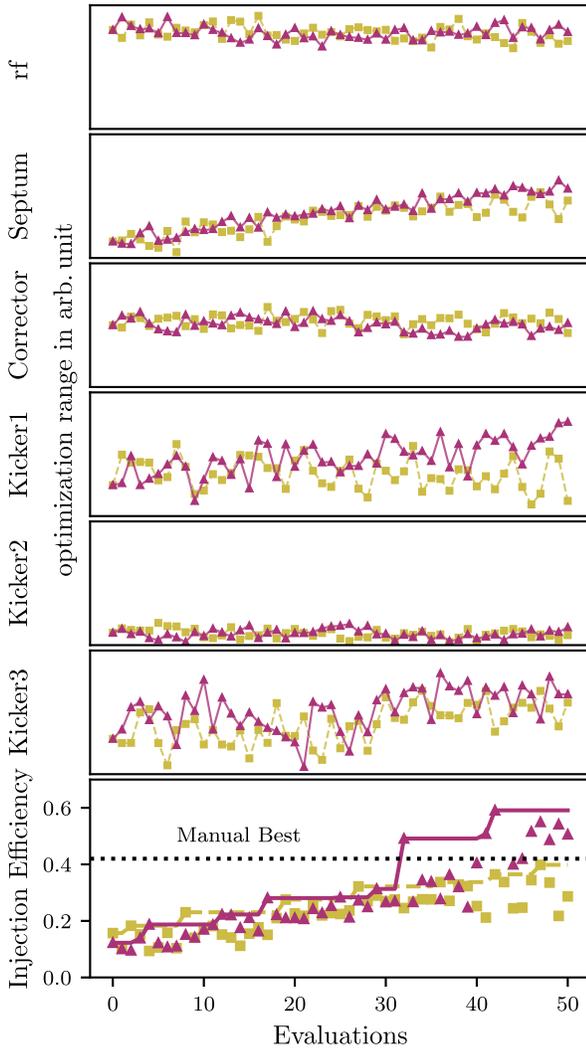


FIG. 9. Two runs with the highest (purple, solid) and lowest (yellow, dashed) optimized injection efficiency achieved are plotted out of all five BO runs with UCB acquisition using the new power supplies. The evolution of the input parameters and the resulting objective function are shown over the optimization steps, where the vertical limit is the total allowed optimization range, respectively. The bottom plot shows the measured injection functions (markers) with the cumulative best results (lines). The dotted line depicts the 90% quantile of the injection efficiency obtained by manual operation in a month prior to the shutdown period. Even the BO run with the worst performance achieved a comparable result as the manual baseline.

the septum magnet strongly influences the injection process, as it directly acts on the injected electron bunch. On the other hand, the strength of kicker 2 barely changed. Since the kicker magnet 2 is located closely after the septum magnet, it possibly implies that the BO prefers to use only the septum magnet for beam deflection. The two runs shown here have found different values for kickers 1 and 3, which lead to different bump orbit shapes and eventually different injection efficiency.

We also show the manual performance benchmark in Fig. 9, where the data are collected in a month of synchrotron light operation prior to the shutdown period with a storage ring current of up to 100 mA. This value $f_{\text{manual}} = 0.43$ is calculated as the 90% quantile of the injection efficiencies obtained by manual tuning so that the outliers due to the fluctuation of readback values are excluded. All the BO runs reached or exceeded the optimal settings found by manual tuning. This result shows that BO is robust against the change in the accelerator's condition and can indeed reduce the commissioning time after even drastic changes to the accelerator environment such as replacing elements with new model components.

E. Contextual optimization

Finally, we present the first results of including contextual parameters and extending the BO algorithm to contextual Bayesian optimization (CBO) [18].

In the above experiments, we treat the beam injection problem as stationary, which is only true as an approximation. In reality, the optimal injection condition is affected by various context variables, which are ignored by the normal BO. One example is collective effects, which become more relevant with the accumulated current. For instance, a parameter setting with a larger closed orbit is noncritical for low bunch current and could be considered good by the normal BO. However, the same setting might cause more instability and become nonoptimal for a higher current. Therefore, we include the storage ring current I as a context variable by adding the current as an additional dimension to the RBF kernel and the GP model. The acquisition function is then maximized over the parameter space with the context dimension fixed. In this case, we set the length scale of the current to be $l_I = 5$ mA by empirical testing. We do not correct the loss rate explicitly as in normal BO and let the algorithm itself recognize this effect. As before, we initialize CBO from a manually detuned

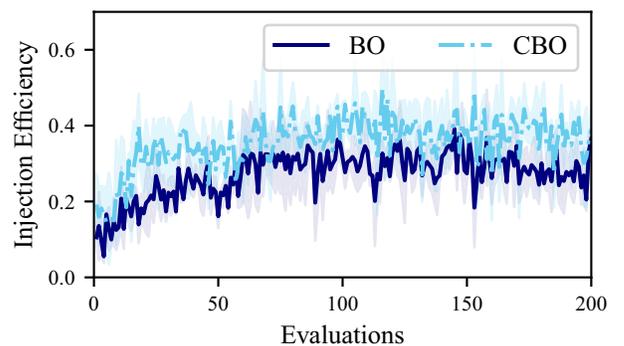


FIG. 10. Optimization results of BO (navy, solid) and contextual BO (cyan, dashdot) on the six-dimensional problem. Each line is the averaged injection efficiency from three runs. By using the stored current as a context variable, CBO optimizes faster and obtains a consistently higher injection efficiency than normal BO after about ten steps.

setting and use UCB acquisition. The number of evaluation steps is set to 200, allowing the optimizer to achieve a higher current value. The normal BO is performed with the same configurations. Figure 10 shows the results, where each line is the mean objective averaged from three runs. The CBO results are corrected afterward for the comparison with BO results. We can see that CBO can properly handle the current-dependent effect and optimizes efficiently. It also shows a consistently better performance than BO. This is partly due to the fact that the context variable provides an extra dimension for exploration. To some extent, CBO is able to track the optimal setting along the current dimension.

IV. SUMMARY AND OUTLOOK

In this article, we presented the implementation of Bayesian optimization applied to the injection task at a storage ring. The algorithms were tested in simulation and also in experiments, for which the accelerator test facility and storage ring KARA at KIT were used. The hyperparameters were determined from some simple scan measurements since no dedicated archive data with diverse parameter settings were available. We showed that BO can optimize the injection efficiency in a small number of evaluations, which makes it suitable for online optimization. We demonstrated experimentally that the BO method is robust against noisy observations, outperforming the commonly used Nelder-Mead method. We also present a real use case of BO for machine commissioning and show that BO can be quickly reconfigured to find new parameter settings after a change in the machine status. Finally, we use the accumulated current as a context variable for the optimization, which shows a better performance than explicitly correcting for the Touschek effect. The final goal of this work is to fully automate the injection tuning process, where the BO algorithm implemented in this paper serves as a basic framework. It is now used by operators routinely to assist the injection optimization task for the weekly machine start-up. Nevertheless, the current version of BO could explore excessively and lose the stored beam due to a poor setting of the kicker magnets. To mitigate that, safety constraints can be included so that the exploration is restricted to safe parameter regions. With safety constraints implemented, BO is expected to be also applicable for more use cases or even at other accelerators with more stringent machine protection requirements.

The required computation time for the presented BO with nine tuning parameters is still negligible compared to the setting of new accelerator parameters. Thus, we expect that this framework can be effortlessly extended to a few tens of tuning parameters by, for example, adding all magnets in the preaccelerators. Other context parameters, such as temperature and vacuum quality, can also be incorporated.

Optimizing only for the highest injection efficiency sometimes results in an unmatched setting of the kicker magnets where the injection bump is not localized, which changes the closed orbit and can lead to a tune shift or a lower beam lifetime. This can be mitigated by considering the beam orbit and betatron tune as additional objective functions. Generalizing the single objective BO to multi-objective optimization makes it possible to simultaneously trade off between multiple objectives, like the orbit deviation and the injection rate, and find optimal parameter settings for user-specific cases. Furthermore, the BO approach may find its application in other accelerator control tasks. It is best suited for quasistatic tuning tasks, where the optimization condition is not drastically changed within the timescale of the optimization and repeated evaluation of the same settings will result in comparable objective function values, e.g., tuning a linear accelerator or a free-electron laser (FEL). Using BO in tandem with other tuning methods will further improve its performance. For example, the Nelder-Mead can be used to provide initial samples of the GP model, and the extremum seeking algorithm can be used as a feedback system to track the optimum obtained by BO.

Whereas for solving dynamic problems for which the accelerator condition continues to change during the optimization, we expect reinforcement learning (RL) to be a promising alternative method [34–37]. We believe that combining the strength of BO and RL could automate the accelerator operation to a great extent and achieve performance beyond manual tuning.

ACKNOWLEDGMENTS

We thank E. Bründermann for reviewing and providing invaluable suggestions for improving this paper. C. Xu acknowledges the support of the DFG-funded Doctoral School “Karlsruhe School of Elementary and Astroparticle Physics: Science and Technology.”

-
- [1] E. Bründermann, H.-W. Hübers, M. Fitz, and G. Kimmitt, *Terahertz Techniques*, Springer Series in Optical Sciences Vol. 151 (Springer Verlag, Berlin, Heidelberg, 2012).
 - [2] C. Bernardini, G. F. Corazza, G. Di Giugno, G. Ghigo, J. Haissinski, P. Marin, R. Querzoli, and B. Touschek, Lifetime and Beam Size in a Storage Ring, *Phys. Rev. Lett.* **10**, 407 (1963).
 - [3] A. Edelen, C. Mayes, D. Bowring, D. Ratner, A. Adelman, R. Ischebeck, J. Snuverink, I. Agapov, R. Kammering, J. Edelen, I. Bazarov, G. Valentino, and J. Weninger, Opportunities in machine learning for particle accelerators, [arXiv:1811.03172](https://arxiv.org/abs/1811.03172).
 - [4] P. A. Vikhar, Evolutionary algorithms: A critical review and its future prospects, in *Proceedings of 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication*,

- ICGTSPICC, Jalgaon, Maharashtra, India* (IEEE, New York, 2016), pp. 261–265.
- [5] J. Kennedy and R. Eberhart, Particle swarm optimization, in *Proceedings of ICNN'95—International Conference on Neural Networks, Perth, WA, Australia* (1995), Vol. 4, pp. 1942–1948, [10.1109/ICNN.1995.488968](https://doi.org/10.1109/ICNN.1995.488968).
- [6] X. Huang and J. Safranek, Online optimization of storage ring nonlinear beam dynamics, *Phys. Rev. ST Accel. Beams* **18**, 084001 (2015).
- [7] X. Huang, Robust simplex algorithm for online optimization, *Phys. Rev. Accel. Beams* **21**, 104601 (2018).
- [8] A. Scheinker, X. Huang, and J. Wu, Minimization of betatron oscillations of electron beam injected into a time-varying lattice via extremum seeking, *IEEE Trans. Control Syst. Technol.* **26**, 336 (2018).
- [9] A. Scheinker, D. Bohler, S. Tomin, R. Kammering, I. Zagorodnov, H. Schlarb, M. Scholz, B. Beutner, and W. Decking, Model-independent tuning for maximizing free electron laser pulse energy, *Phys. Rev. Accel. Beams* **22**, 082802 (2019).
- [10] D. R. Jones, A taxonomy of global optimization methods based on response surfaces, *J. Global Optim.* **21**, 345 (2001).
- [11] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, Cambridge, MA, 2005).
- [12] M. McIntire, T. Cope, D. Ratner, and S. Ermon, Bayesian optimization of FEL performance at LCLS, in *Proceedings of the 7th International Particle Accelerator Conference, IPAC-2016, Busan* (JACoW, Geneva, Switzerland, 2016), pp. 2972–2975.
- [13] J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, P. Baxevanis, A. Egger, T. Cope, M. McIntire, S. Ermon, and D. Ratner, Bayesian Optimization of a Free-Electron Laser, *Phys. Rev. Lett.* **124**, 124801 (2020).
- [14] J. Kirschner, A. Adelman, N. Hiller, R. Ischebeck, A. Krause, M. Mutný, and M. Nonnenmacher, Bayesian optimisation for fast and safe parameter tuning of Swiss-FEL, in *Proceedings of 39th International Free Electron Laser Conference, Hamburg, Germany* (JACoW, Geneva, Switzerland, 2019), THP061.
- [15] A. Hanuka, J. Duris, J. Shtalenkova, D. Kennedy, A. Edelen, D. Ratner, and X. Huang, Online tuning and light source control using a physics-informed Gaussian process, [arXiv:1911.01538](https://arxiv.org/abs/1911.01538).
- [16] R. Roussel, A. Hanuka, and A. Edelen, Multiobjective Bayesian optimization for online accelerator tuning, *Phys. Rev. Accel. Beams* **24**, 062801 (2021).
- [17] C. Xu, Bayesian optimization of injection efficiency at KARA using Gaussian processes, Master's thesis, Karlsruhe Institute of Technology, 2020.
- [18] A. Krause and C. S. Ong, Contextual Gaussian process bandit optimization, in *Proceedings of the 24th International Conference on Neural Information Processing Systems, NIPS 2011* (Curran Associates Inc., Red Hook, NY, 2011).
- [19] E. Brochu, V. Cora, and N. Freitas, A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning, [arXiv:1012.2599](https://arxiv.org/abs/1012.2599).
- [20] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, Gaussian process optimization in the bandit setting: No regret and experimental design, in *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10, Haifa, Israel* (Omnipress, Madison, WI, 2010), pp. 1015–1022.
- [21] D. R. Jones, M. Schonlau, and W. J. Welch, Efficient global optimization of expensive black-box functions, *J. Global Optim.* **13**, 455 (1998).
- [22] D. J. Lizotte, Practical Bayesian optimization, Ph.D. thesis, University of Alberta, 2008.
- [23] D. Einfeld, S. Hermle, E. Huttel, R. Rossmanith, and R. Walther, The injection scheme for the ANKA storage ring, in *Proceedings of the 6th European Particle Accelerator Conference, EPAC 98, Stockholm, Sweden* (JACoW Publishing, Geneva, Switzerland, 1998), pp. 2135–2137, <https://jacow.org/e98/papers/MOP17F.pdf>.
- [24] *Proceedings of CAS-CERN Accelerator School: 5th General Accelerator Physics Course, Jyväskylä, Finland, 1992*, edited by S. Turner (CERN, Geneva, Report No. CERN-94-01, 1994), Vol. 2.
- [25] M. Newville, K. Lauer, dchabot, T. A. Caswell, M. Gibbs, A. Péteut, S. Hartman, rokvtar, R. Clarken, B. Martins, D. Allan, T. Birke, P. R. Jemian, N. Claesson, J. Adelman, J. Dwyer, H. Slepicka, G. Brandl, E. Greenberg, D. Vine, and André, *pyepics/pyepics 3.4.0* (2019).
- [26] L. R. Dalesio, M. R. Kraimer, and a. J. Kozubal, EPICS architecture, in *Proceedings of the 3rd International Conference on Accelerator and Large Experimental Physics Control Systems, ICALEPCS, Tsukuba, Japan* (KEK, Tsukuba, 1991).
- [27] GPy: A Gaussian process framework in python, <http://github.com/SheffieldML/GPy> (since 2012).
- [28] P. Virtanen and SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261 (2020).
- [29] A. Terebilo, Accelerator toolbox for MATLAB, in *Proceedings of Workshop on Performance Issues at Synchrotron Light Sources, Berkeley, CA*, Report No. SLAC-PUB-8732, 2001, <https://www.slac.stanford.edu/pubs/slacpubs/8500/slac-pub-8732.pdf>.
- [30] G. Portmann, J. Corbett, and A. Terebilo, An accelerator control middle layer using Matlab, *Conf. Proc. C* **0505161**, 4009 (2005).
- [31] S. Marsching, E. Huttel, M. Klein, and N. J. Smale, First Experience with the Matlab Middle Layer at ANKA, in *Proceedings of 13th International Conference on Accelerator and Large Experimental Physics Control Systems, ICALEPCS 2011, WTC Grenoble, France* (JACoW Publishing, Geneva, Switzerland, 2011), <https://jacow.org/icaleps2011/papers/MOPMN009.pdf>.
- [32] J. A. Nelder and R. Mead, A simplex method for function minimization, *Comput. J.* **7**, 308 (1965).
- [33] P. I. Frazier, A tutorial on Bayesian optimization, [arXiv:1807.02811](https://arxiv.org/abs/1807.02811).
- [34] J. St. John, C. Herwig, D. Kafkes, J. Mitrevski, W. A. Pellico, G. N. Perdue, A. Quintero-Parra, B. A. Schupbach, K. Seiya, N. Tran, M. Schram, J. M. Duarte, Y. Huang, and R. Keller, Real-time artificial intelligence for accelerator control: A study at the Fermilab booster, *Phys. Rev. Accel. Beams* **24**, 104601 (2021).

- [35] A. Eichler, E. Bründermann, F. Burkart, J. Kaiser, W. Kuropka, A. Santamaria Garcia, O. Stein, and C. Xu, First steps toward an autonomous accelerator, a common project between DESY and KIT, in *Proceedings of the 12th International Particle Accelerator Conference, IPAC-2021, Campinas, SP, Brazil* (JACoW, Geneva, Switzerland, 2021), pp. 2182–2185, [10.18429/JACoW-IPAC2021-TUPAB298](https://doi.org/10.18429/JACoW-IPAC2021-TUPAB298).
- [36] T. Boltz, E. Bründermann, M. Caselle, A. Kopmann, W. Mexner, A.-S. Müller, and W. Wang, Accelerating Machine Learning for Machine Physics (an AMALEA-project at KIT), in *Proceedings of the 17th International Conference on Accelerator and Large Experimental Physics Control System, ICALEPCS'19, New York, NY* (JACoW, Geneva, Switzerland, 2020), pp. 781–788, [10.18429/JACoW-ICALEPCS2019-TUCPL06](https://doi.org/10.18429/JACoW-ICALEPCS2019-TUCPL06).
- [37] W. Wang, M. Caselle, T. Boltz, E. Blomley, M. Brosi, T. Dritschler, A. Ebersoldt, A. Kopmann, A. Santamaria Garcia, P. Schreiber, E. Bründermann, M. Weber, A.-S. Müller, and Y. Fang, Accelerated deep reinforcement learning for fast feedback of beam dynamics at KARA, *IEEE Trans. Nucl. Sci.* **68**, 1794 (2021).