Editors' Suggestion

Differentiable Preisach Modeling for Characterization and Optimization of Particle Accelerator Systems with Hysteresis

R. Roussel,* A. Edelen, and D. Ratner
SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA

K. Dubey, J. P. Gonzalez-Aguilera, and Y. K. Kim
University of Chicago, Chicago, Illinois 60637, USA

N. Kuklev
Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA

(Received 15 February 2022; accepted 5 April 2022; published 17 May 2022)

Future improvements in particle accelerator performance are predicated on increasingly accurate online modeling of accelerators. Hysteresis effects in magnetic, mechanical, and material components of accelerators are often neglected in online accelerator models used to inform control algorithms, even though reproducibility errors from systems exhibiting hysteresis are not negligible in high precision accelerators. In this Letter, we combine the classical Preisach model of hysteresis with machine learning techniques to efficiently create nonparametric, high-fidelity models of arbitrary systems exhibiting hysteresis. We experimentally demonstrate how these methods can be used *in situ*, where a hysteresis model of an accelerator magnet is combined with a Bayesian statistical model of the beam response, allowing characterization of magnetic hysteresis solely from beam-based measurements. Finally, we explore how using these joint hysteresis-Bayesian statistical models allows us to overcome optimization performance limitations that arise when hysteresis effects are ignored.

DOI: 10.1103/PhysRevLett.128.204801

Hysteresis is a well-known physical phenomenon where the state of a given system is dependent on its historical path through state-space. This property is evident in physical, biological, chemical, and engineering processes, including the magnetization of ferromagnetic materials [1], the activation of embryonic cells [2], the charging and discharging cycles of nickel-metal hydride batteries [3], and the driving of mechanical actuators [4]. In particular, hysteresis effects in magnetic [5], mechanical [6], and material [7] elements of particle accelerators make optimizing the performance of current accelerator facilities used for scientific discovery challenging.

State-of-the-art optimization algorithms, such as Bayesian optimization (BO) [8,9], use online computational models of objective functions to tackle optimization tasks at accelerator facilities [10–12]. However, models used in these algorithms ignore hysteresis effects entirely, limiting optimization performance due to errors caused by hysteresis (e.g., see Refs. [13,14]). These limits on optimization performance are further amplified when considering ambitious targets for future accelerators, [15,16] which are increasingly sensitive to hysteresis effects. Incorporating an accurate description of hysteresis into models used for online accelerator optimization would substantially improve the performance of current and future particle accelerators.

Nonparametric Preisach modeling [17] is a flexible approach for accurately describing systems that exhibit hysteresis behavior. Unfortunately, fitting these models to experimental measurements using numerical optimization techniques has been shown to be computationally expensive [18,19] when a large number of free parameters are present, due to the so-called "curse of dimensionality" [20].

In this Letter, we construct a *differentiable*, nonparametric Preisach model, which when used in conjunction with gradient-based optimization, significantly reduces the computational cost of model identification. We explore how our technique enables accurate online modeling of the beam response with respect to controllable accelerator parameters through the use of joint hysteresis-Bayesian statistical models. We experimentally demonstrate how this enables the characterization of hysteresis properties in magnetic beamline elements from beam-based measurements. Finally, we explore how the joint model improves optimization of a realistic beamline containing magnetic elements exhibiting hysteresis.

The Preisach model of hysteresis [21,22] is composed of a discrete set of *hysterons*, which when added together, model the output of a hysteretic system $f(t)$ for a time dependent input $u(t)$. Given a set of discrete time ordered inputs $u_i = u(t_i)$, the hysteron state is represented by the hysteron operator $\hat{\gamma}_{\alpha\beta}$ shown in Fig. 1(a), which has an

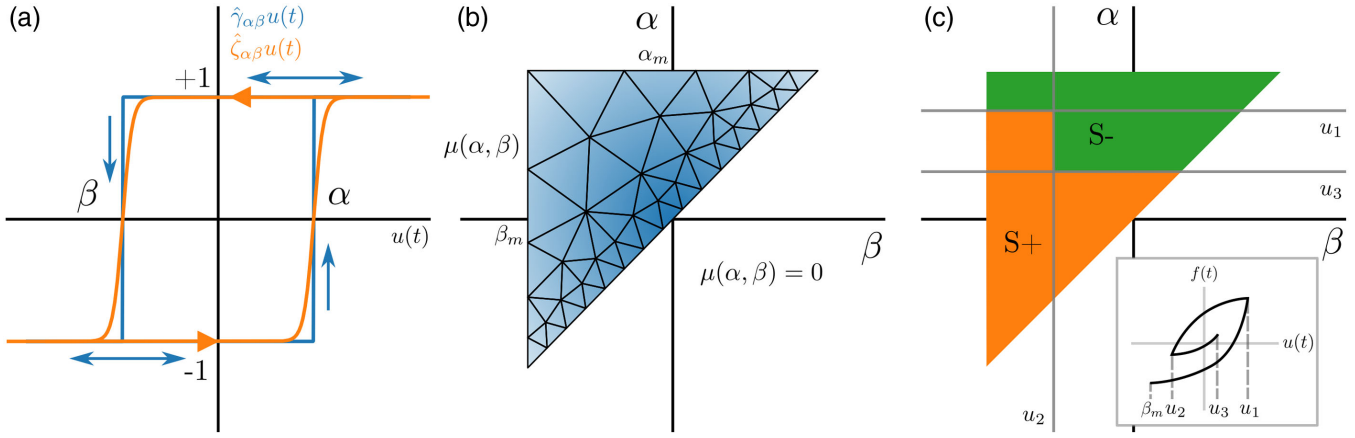

FIG. 1. Elements of the differentiable nonparametric Preisach hysteresis model. (a) Output of the hysteron operator $\hat{\gamma}_{\alpha\beta}$ and the approximate differentiable hysteron operator $\hat{\zeta}_{\alpha\beta}$ acting on the input $u(t)$. (b) Discretization of the hysteron density function on the Preisach (α-β) plane. Note that $\mu(\alpha,\beta)=0$ if $\alpha<\beta$, $\alpha>\alpha_m$, or $\beta<\beta_m$ where α_m,β_m are equal to the maximum and minimum inputs of the model respectively. (c) Positive (S^+) and negative (S^-) hysteron state subdomains after three time steps, where $u_1>u_3>u_2>\beta_m$, assuming that all hysterons are in the negative state initially. Inset: corresponding model output (not to scale).

output of ±1, where α and β describe the input required to switch a hysteron between its two possible states. The number of hysterons with values (α, β) is given by the hysteron density function $\mu(\alpha,\beta)$, plotted on the Preisach (α-β) plane [Fig. 1(b)].

The Preisach model output is represented by

$$f(t)=\hat{\Gamma}u(t)=\iint_{\alpha\geq\beta}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta \qquad (1)$$

where $\alpha\geq\beta$ results from physical conditions of the hysteron operator. This integral is evaluated through a geometric interpretation, shown in Fig. 1(c). Given the sequence of input values u_i, we can determine subregions of the Preisach plane S^+ and S^-, where hysteron operators output positive and negative states respectively. We start with the assumption that all hysterons are initially in the negative state (S^- covers the entire Preisach plane). When $u_t>u_{t-1}$, a horizontal line is swept up, flipping hysteron states from negative to positive, increasing the S^+ region. Conversely, when $u_t<u_{t-1}$, a vertical line is swept to the left, flipping states from positive to negative.

Once the regions S^+ and S^- are determined by the input $u(t)$, fitting a Preisach model to experimental data requires the determination of the hysteron density function $\mu(\alpha,\beta)$, often referred to as the *identification problem*. Approaches for solving this problem are generally divided into parametric or nonparametric methods. Parametric methods describe the hysteron density using one of several analytic functions with a small number of free parameters [19,23,24], which can be determined through numerical optimization methods given experimental data. However, this limits model flexibility, resulting in prediction errors for systems that do not match the chosen analytical function. On the other hand, nonparametric methods [18,25,26] discretize the density function using a mesh grid [Fig. 1(b)] and attempt to determine the density of hysterons at each mesh point based on experimental measurements. However, fitting a nonparametric, high-fidelity Preisach model containing thousands of mesh points is prohibitively expensive when using black box numerical optimization techniques.

We improve upon nonparametric modeling of hysteresis by creating *differentiable* Preisach models, which use gradient-based optimization to identify the hysteron density function at high fidelities. Differentiable modeling refers to tracking derivative information during every step of internal model calculations. This allows what is known as back propagation [27], where through the chain rule, the derivative of the model output with respect to any model parameter is analytically calculable. By combining this technique with gradient-based optimization algorithms (e.g., L-BFGS-B [28] or Adam [29]), we are able to scale nonparametric Preisach models to thousands of mesh points, while still being computationally practical for use in online modeling.

We construct a differentiable Preisach model by implementing the nonparametric version of Eq. (1) in the python library PyTorch [30]. The continuous hysteron density $\mu(\alpha,\beta)$ is replaced with a discrete one, located on a triangular mesh containing N mesh points on the Preisach plane giving $\mu_i=\mu(\alpha_i,\beta_i)$, where $i=1,\ldots,N$. We also replace the hysteron operator $\hat{\gamma}_{\alpha\beta}$ with a differentiable approximation $\hat{\zeta}_{\alpha\beta}$, enabling differentiability with respect to $u(t)$ as shown in Fig. 1(a) (see the Supplemental Material [31] for the exact form). The differentiable, nonparametric Preisach model is given by

$$f(t)=\sum_{i=1}^{N}\mu_i\hat{\zeta}_{\alpha\beta,i}u(t) \qquad (2)$$

where $\hat{\zeta}_{\alpha\beta,i}$ is the differentiable hysteron operator at the ith mesh point.

We demonstrate the effectiveness of our differentiable Preisach model by using it to analyze experimental data gathered from a quadrupole magnet at SLAC National Accelerator Laboratory. Current applied to the magnet was cycled to sample both major and minor hysteresis loops, and the integrated gradient at the magnet center was measured using a rotating coil technique [32]. Measurements were then split into training and test sets to investigate how accurately the model predicted measurement data and generalized to unknown future measurements.

Accelerator magnets pose a unique hysteresis modeling challenge, as they are designed specifically to minimize field perturbations from a polynomial function of magnet current due to hysteresis [32]. Despite this, realistic magnets contain field perturbations that cannot be tolerated in high precision applications [5]. We are interested in resolving these field perturbations, which result from nonzero hysteron densities off of the $\alpha = \beta$ line, denoted as $\bar{\mu}(\alpha, \beta)$. Resolving these small perturbations requires specialized data processing and model construction, details of which can be found in the Supplemental Material [31].

Model fitting to hysteresis perturbations observed in experiment is shown in Fig. 2 using an adaptive triangular mesh containing 7411 mesh points. We trained the model on an Intel i9-9900 K CPU at 3.6 GHz using a mean squared error loss function and the Adam algorithm with a learning rate of 0.01 over 10 k steps, which took approximately 67 s. This is roughly 2 orders of magnitude faster than a comparable analysis in previous nonparametric studies [18].

Our model captures the features of major hysteresis loops with an rms training error σ_{train} of 0.8 mT, corresponding to a percentage error ($p = 100\sigma_{\text{train}}/f_{\text{max}}$) of 0.015%. Despite only training on major hysteresis loops, our model makes accurate predictions of minor hysteresis loops and large swings in applied current with an rms error of 2.6 mT (0.051%). Our model significantly outperforms polynomial fitting of the unnormalized experimental data, which has an rms error of 12.1 mT (0.23%) over the entire dataset.

Next, we examine the case where *directly measuring hysteresis output* is impractical or impossible. For example, fields cannot be accurately characterized for magnetic elements that are already installed in accelerator beamlines. Instead, we may only observe the beam response to fields generated by these elements. To determine hysteresis characteristics in this case, we combine our hysteresis model with a Gaussian process (GP) model [9] representing beam propagation as a function of magnetic fields. We then infer hysteresis behavior from measurements of beam properties as a function of magnet current.

The overall characteristics of GPs, defined as $g(\boldsymbol{x}) \sim \mathcal{GP}[m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})]$ with a mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})$, are governed by a set of hyperparameters $\boldsymbol{\theta}$, which describe our prior knowledge of the model's smoothness, amplitude, and noise. GP models predict the distribution of function values at a location $\boldsymbol{x}$ to be $p(g|\mathcal{D}, \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}[\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x})]$, where $\mathcal{D} = \{X, \boldsymbol{y}\}$ is the set of training samples and $\mu(\boldsymbol{x})$, $\sigma^2(\boldsymbol{x})$ are the posterior mean and uncertainty (see the Supplemental Material [31] for details). We infer hyperparameters for a GP model from

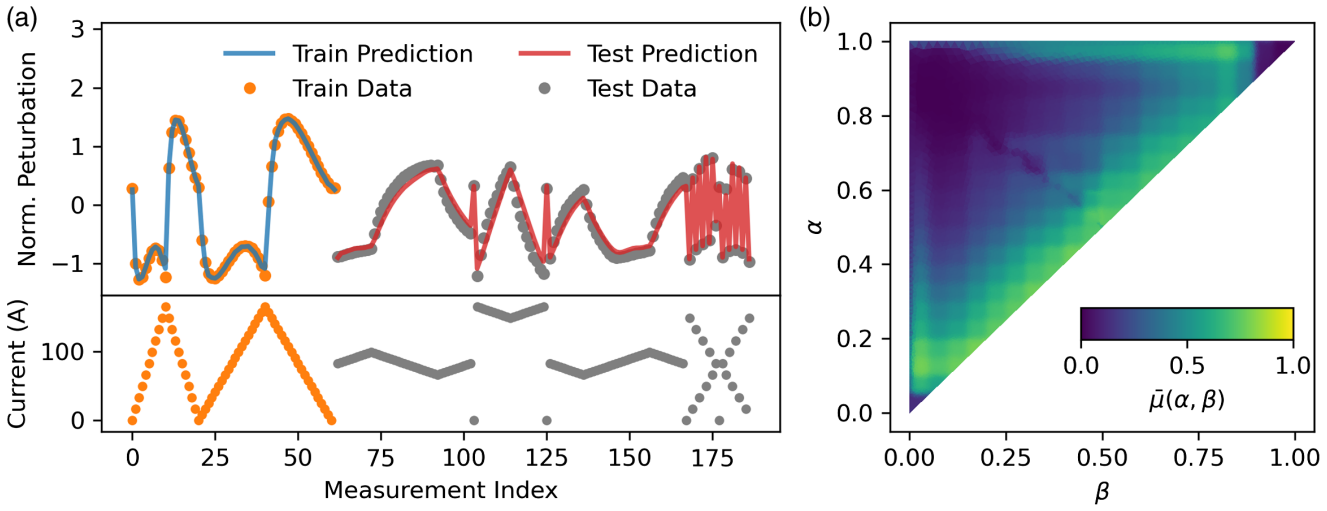

FIG. 2. Nonparametric modeling of hysteresis perturbations using direct measurements of a SLAC quadrupole magnet. (a) Normalized hysteresis error training data, test data and model predictions. Sequence of applied currents during measurements $u(t)$ is normalized during training to the unit domain, and measured field errors are transformed such that the training data has zero mean and a standard deviation of 1. (b) Normalized hysteron density representing hysteresis perturbations $\bar{\mu}_i$ on the normalized Preisach plane after model training.

training data by maximizing the marginal likelihood given by

$$p(\boldsymbol{y}|X) = \int p(\boldsymbol{y}|X,\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{3}$$

with respect to the hyperparameters $\boldsymbol{\theta}$, resulting in a model that balances the trade-off between accuracy and complexity.

We combine the hysteresis and GP models into a single joint model by treating the hysteresis output as the GP input and training both models simultaneously. The joint hysteresis-GP model is given by

$$p(g|\mathcal{D},\boldsymbol{t},\boldsymbol{\phi},\boldsymbol{\theta}) = \mathcal{N}\Big(\mu[f(\boldsymbol{t})],\sigma^2[f(\boldsymbol{t})]\Big) \tag{4}$$

where $\boldsymbol{\phi}$ represents hysteresis model parameters. The joint set of parameters $\boldsymbol{\Phi} = \{\boldsymbol{\phi},\boldsymbol{\theta}\}$ is then determined by maximizing the marginal likelihood using Eq. (3) with respect to the new set of parameters $\boldsymbol{\Phi}$.

We demonstrate the effectiveness of our joint hysteresis-GP (H-GP) model by fitting the beam response with respect to the current applied to a focusing magnet located in the Advanced Photon Source (APS) injector [33]. The current of the quadrupole magnet was varied using a sawtooth pattern from $-2A$ to $+2A$ while measuring the beam charge passing through a downstream current monitor. Measurements from this experiment, shown in Fig. 3, have two sources of uncertainty, one from random noise inherent in the accelerator (aleatoric uncertainty) and one due to the unknown properties of magnetic hysteresis (epistemic uncertainty). A normal GP model [Fig. 3(a)] does not take into account the existence of hysteresis; thus it interprets epistemic errors due to hysteresis as aleatoric uncertainty, overestimating uncertainties in portions of the input domain. However, the joint hysteresis-GP model [Fig. 3(b)] is able to resolve hysteresis cycles inside the data, removing epistemic uncertainties in the model prediction, thus improving model accuracy and reducing uncertainty.

The increase in accuracy from joint hysteresis-GP models has ramifications for model-based, online optimization of accelerators using BO. We examine how models with and without hysteresis taken into account affect optimization performance when optimizing a simulated accelerator that contains realistic magnetic elements which exhibit hysteresis. We simulate the task of sequentially optimizing currents applied to three identical quadrupoles using BO, in order to transform an incoming round beam with an rms beam size of $\sigma_{x,y} = 5$ mm to a final round beam size of $\sigma_{\text{target}} = 8$ mm, assuming that the hysteresis properties of the magnets have been previously determined. The objective function is given by a geometric mean of the beam size deviation $l = \sqrt{\Delta_x \Delta_y}$ where $\Delta_k = |\sigma_k - \sigma_{\text{target}}|$. A toy hysteresis model (described in the Supplemental Material [31]) with a tunable hysteresis magnitude was used to simulate realistic magnetic elements. Three beamlines with maximum fractional hysteresis errors $H_\epsilon = 0$, 0.1, and 0.4 were used in optimization trials to represent ideal, realistic, and extreme hysteresis effects respectively.

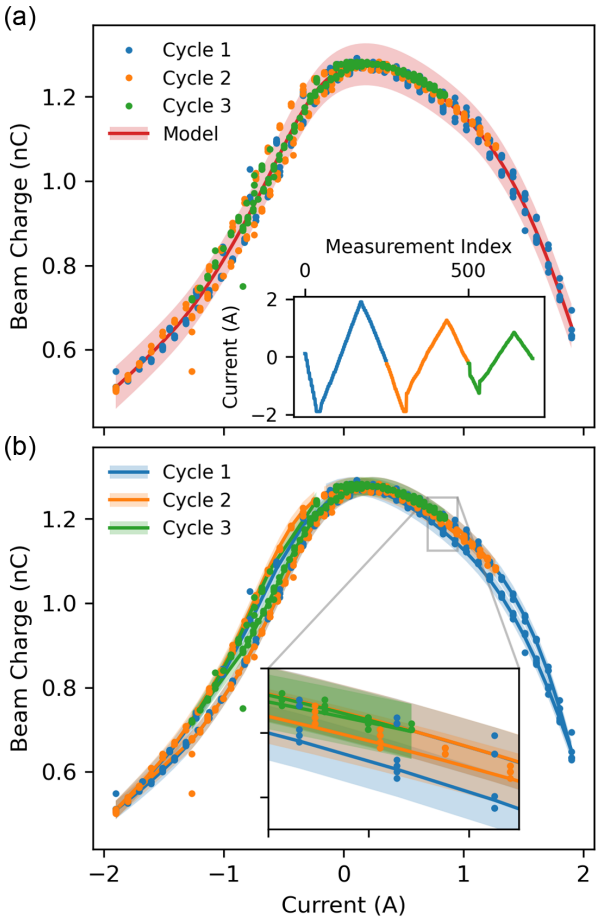

FIG. 3. Comparison between GP modeling and joint hysteresis-GP modeling of beam transmission as a function of quadrupole current at the APS injector. (a) GP model prediction with training data over three cycles (see inset). Shading denotes 2σ confidence region. (b) Hysteresis-GP model prediction, colored by cycle index.

We performed BO using the upper confidence bound acquisition function [34], first with $\beta^* = 2$ which balances exploration (sampling points in unexplored regions of input space) and exploitation (sampling points that are predicted to be at global extrema). We then repeated the experiment with $\beta^* = 0.1$, which prioritizes exploitation. Optimization results obtained over 64 trials using BO with GP and H-GP models are shown in Fig. 4.

Figure 4(a) shows that hysteresis has little effect on the performance of BO when balancing exploration and exploitation, even when extreme hysteresis errors are present. The acquisition function in this case often chooses to measure points in unexplored regions of input space and, as a result, is relatively insensitive to hysteresis errors.

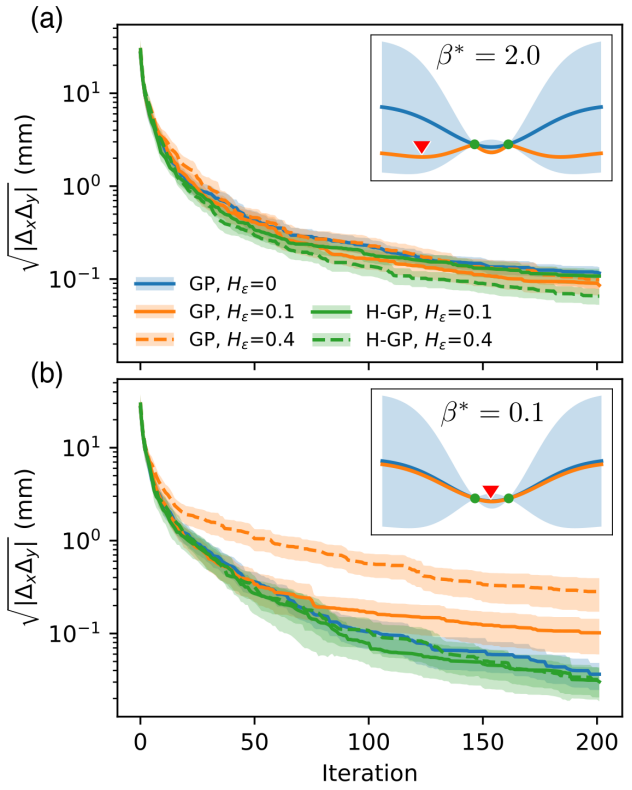

FIG. 4. BO performance for a simulated beamline optimization problem using GP and H-GP models for a varying maximum hysteresis errors H_ϵ. (a) Performance when evenly weighting exploration and exploitation ($\beta^* = 2$). (b) Strong weighing toward exploitation ($\beta^* = 0.1$). Lines denote the mean best performance of 64 optimization runs, while shading denotes standard error. Insets: cartoon showing BO observation selection (red triangle) based on an acquisition function (orange line) given a GP model mean (blue line) and 2σ uncertainty (blue shading) calculated from measurements (green points).

On the other hand, if we attempt to exploit the model as shown in Fig. 4(b), modeling errors due to hysteresis effects in normal GP models negatively impact optimization performance, depending on the magnitude of hysteresis errors. A joint hysteresis-GP model significantly improves optimization performance, matching the performance observed when optimizing an idealized beamline without hysteresis.

These results identify a clear strategy for optimizing physical systems that contain hysteresis using BO. Hysteresis effects can be neglected when coarsely searching for the global optima of a system, even when hysteresis errors are significant, since the optimization is dominated by uncertainties associated with unexplored regions in input space (especially in high dimensional input spaces where uncertainties are large). However, hysteresis effects must be taken into account when attempting to exploit extrema during optimization. In this case, a model of the hysteresis behavior can then be constructed by cycling the magnet current while directly measuring the fields inside the magnet or indirectly measuring the effect of the magnetic fields on beam properties. Once identified, the hysteresis model is combined with a GP model of the beam response to conduct BO to exploit predicted extrema. Our technique thus enables a staged approach toward optimization of systems involving hysteresis, where model and computational complexity is traded for optimization precision.

In conclusion, we have demonstrated how a differentiable Preisach model can describe arbitrary hysteretic systems using direct or indirect measurements, and can improve the model-based optimization of those systems. We collected direct and indirect measurements of systems exhibiting hysteresis and then used the procedure described here to construct a model for the hysteretic behavior. Improved identification speed of differentiable Preisach modeling enabled practical, high-fidelity regression of major and minor hysteresis loops in realistic magnetic elements. We demonstrated that our hysteresis model can be combined with GP modeling to infer hysteresis behavior without making direct measurements of the hysteresis response. Finally, we demonstrated how these joint hysteresis-GP models can be used to optimize physical systems containing hysteretic behavior, overcoming limitations faced by currently used algorithms that neglect hysteresis repeatability errors. This development enables future advanced modeling techniques of hysteresis, most notably implementing fully Bayesian hysteresis models using stochastic variational inference [35].

The authors would like to thank the SLAC Metrology group for their help gathering quadrupole data, Adi Hanuka for enriching conversations that informed the early stages of this work, and Louis Emery for early discussions on practical implementation and preliminary data. This work was supported by the U.S. Department of Energy, under DOE Contract No. DE-AC02-76SF00515 and the Office of Science, Office of Basic Energy Sciences.

*roussel@slac.stanford.edu

[1] D. C. Jiles and D. L. Atherton, Theory of ferromagnetic hysteresis, J. Magn. Magn. Mater. **61**, 48 (1986).
[2] J. R. Pomerening, E. D. Sontag, and J. E. Ferrell, Building a cell cycle oscillator: Hysteresis and bistability in the activation of Cdc2, Nat. Cell Biol. **5**, 346 (2003).
[3] D. U. Sauer, BATTERIES | ChargeDischarge curves, in *Encyclopedia of Electrochemical Power Sources*, edited by J. Garche (Elsevier, Amsterdam, 2009), pp. 443–451.
[4] M. Warnecke and M. Jouaneh, Backlash compensation in gear trains by means of open-loop modification of the input trajectory, J. Mech. Des. **125**, 620 (2003).
[5] N. J. Sammut, H. Burkhardt, S. M. White, C. Giloux, and W. Venturini-Delsolaro, Measurement and effects of the magnetic hysteresis on the LHC crossing angle and separation bumps, Report No. LHC-PROJECT-Report-1107, 2008.

[6] N. Huque, M. Abdelwhab, E. Daly, and Y. Pischalnikov, Accelerated life testing of LCLS-II cavity tuner motor, in *Proceedings of the 17th International Conference on RF Superconductivity* (JACoW, Geneva, 2015), p. THPB062.

[7] R. Willa, V. B. Geshkenbein, and G. Blatter, Probing the pinning landscape in type-II superconductors via Campbell penetration depth, Phys. Rev. **93**, 064515 (2016).

[8] J. Snoek, H. Larochelle, and R. P. Adams, Practical bayesian optimization of machine learning algorithms, in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., Red Hook, NY, 2012), pp. 2951–2959.

[9] C. E. Rasmussen and C. K. I. Williams, Bayesian Regression and Gaussian processes, in *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, 2006), Chap. 2, ISBN: 026218253X.

[10] 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).

[11] R. Roussel, J. P. Gonzalez-Aguilera, Y.-K. Kim, E. Wisniewski, W. Liu, P. Piot, J. Power, A. Hanuka, and A. Edelen, Turn-key constrained parameter space exploration for particle accelerators using Bayesian active learning, Nat. Commun. **12**, 5612 (2021).

[12] J. Kirschner, M. Mutn, N. Hiller, R. Ischebeck, and A. Krause, Adaptive and safe bayesian optimization in high dimensions via one-dimensional subspaces, arXiv:1902.03229.

[13] A. Hanuka, J. Duris, H. Shang, and Y. Sun, Demonstration of machine learning front-end optimization of the advanced photon source linac, in *Proceedings of the 12th International Particle Accelerator Conference IPAC2021* (JACoW Publishing, Geneva, Switzerland, 2021), p. 4, ISBN: 9783954502141.

[14] R. Roussel and A. Hanuka, Towards hysteresis aware bayesian regression and optimization, in *Proceedings of the 12th International Particle Accelerator Conference (IPAC'21), Campinas, SP, Brazil, 2021* (JACOW Publishing, Geneva, Switzerland, 2021), pp. 2159–2162.

[15] M. Borland *et al.*, *The Upgrade of the Advanced Photon Source* (JACOW Publishing, Geneva, Switzerland, 2018), pp. 2872–2877.

[16] M. Aicheler, P. Burrows, M. Draper, T. Garvey, P. Lebrun, K. Peach, N. Phinney, H. Schmickler, D. Schulte, and N. Toge, *A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report*, CERN Yellow Reports: Monographs (CERN, Geneva, 2012).

[17] K. H. Hoffmann and G. H. Meyer, A least squares method for finding the preisach hysteresis operator from measurements, Numer. Math. **55**, 695 (1989).

[18] R. Iyer and M. Shirley, Hysteresis parameter identification with limited experimental data, IEEE Trans. Magn. **40**, 3227 (2004).

[19] H. Marouani, K. Hergli, H. Dhahri, and Y. Fouad, Implementation and identification of Preisach parameters: Comparison between genetic algorithm, particle swarm optimization, and LevenbergMarquardt algorithm, Arab. J. Sci. Eng. **44**, 6941 (2019).

[20] R. Bellman and R. E. Bellman, *Adaptive Control Processes: A Guided Tour* (Princeton University Press, Princeton, NJ, 1961).

[21] I. D. Mayergoyz and G. Friedman, Generalized Preisach model of hysteresis, IEEE Trans. Magn. **24**, 212 (1988).

[22] *The science of hysteresis*, edited by G. Bertotti and I. D. Mayergoyz, 1st ed. (Academic, Amsterdam; Boston, 2006).

[23] A. Sutor, S. J. Rupitsch, and R. Lerch, A Preisach-based hysteresis model for magnetic and ferroelectric hysteresis, Appl. Phys. A **100**, 425 (2010).

[24] K. Hergli, H. Marouani, M. Zidi, Y. Fouad, and M. Elshazly, Identification of Preisach hysteresis model parameters using genetic algorithms, J. King Saud Univ. Eng. Sci. **31**, 746 (2019).

[25] X. Tan, R. Venkataraman, and P. S. Krishnaprasad, Control of hysteresis: Theory and experimental results, in *Smart Structures and Materials 2001: Modeling, Signal Processing, and Control in Smart Structures* (SPIE, Bellingham, WA, 2001), Vol. 4326, pp. 101–112.

[26] M. Ruderman and T. Bertram, Identification of soft magnetic B-H characteristics using discrete dynamic Preisach model and single measured hysteresis loop, IEEE Trans. Magn. **48**, 1281 (2012).

[27] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Mller, Efficient BackProp, in *Neural Networks: Tricks of the Trade: Second Edition*, Lecture Notes in Computer Science, edited by G. Montavon, G. B. Orr, and K.-R. Mller (Springer, Berlin, Heidelberg, 2012), pp. 9–48.

[28] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. **16**, 1190 (1995).

[29] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[30] A. Paszke *et al.*, PyTorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d. Alch-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., Red Hook, NY, 2019), pp. 8024–8035.

[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.128.204801 for detailed algorithm, model descriptions, supplemental data and code availability.

[32] J. Tanabe, Iron dominated electromagnets: Design, fabrication, assembly and measurements, Technical Report No. SLAC-R-754, SLAC National Accelerator Lab., Menlo Park, CA, United States, 2005.

[33] Y. Sun, M. Borland, G. I. Fystro, X. Huang, and H. Shang, Recent operational experience with thermionic RF guns at the APS, in *Proceedings of the 12th International Particle Accelerator Conference (IPAC'21)* (JACoW Publishing, 2021), pp. 3959–3962.

[34] 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* (OmniPress, Madison, WI, USA, 2010), pp. 1015–1022.

[35] D. Wingate and T. Weber, Automated variational inference in probabilistic programming, arXiv:1301.1299.