
Transformers for Deterministic Magnetic Field Control in Particle Accelerators

Anton Lu*

Technische Universität Wien
1040 Wien, Austria
anton.lu@cern.ch

Verena Kain

CERN
1211 Geneva, Switzerland
verena.kain@cern.ch

Abstract

Accurate control of accelerator magnet fields is limited by hysteresis and other rate-dependent dynamics that are not captured by static current-to-field maps, which constrains beam quality and operational flexibility. We present a data-driven approach using transformer-based models for high-precision prediction and control of magnetic fields in the CERN Super Proton Synchrotron (SPS). The model, pre-trained on simulated sequences and fine-tuned on magnetic measurements, achieves prediction accuracy near the field sensor noise floor of 2×10^{-5} T across operational conditions, including sequences not explicitly observed during training. It has been successfully tested across all standard sequences in accelerator operation, as well as on magnetic sequences run without the usual energy-intensive magnetic precycles, significantly improving beam reproducibility, enabling automated feed-forward control of the main dipole magnets, and reducing energy consumption. Our results suggest that machine learning models could complement or replace traditional tabular approaches to accelerator magnet control, providing a practical route for feedback-free, high-precision control in accelerator operation.

1 Introduction

Particle accelerators employ magnets of various field configurations to control beam position and transverse distributions, and they require high field quality and reproducibility. Static field models derived from coil geometry and iron yokes, typically fine-tuned through magnetic measurements, are well established, and simple lookup tables are used to map excitation current $I(t)$ to magnetic field $B(t)$. However, magnetic hysteresis in the iron yokes introduces a memory effect in which the magnetization depends on the excitation history, and this phenomenon lacks a closed-form description.

In practice, hysteresis is mostly neglected in accelerator control, which constrains allowed cycle sequences and requires manual tuning whenever the cycle sequences change. The CERN Super Proton Synchrotron (SPS), which serves both as an injector to the LHC and as a fixed target machine, requires field reproducibility at the level of 1×10^{-5} T. Feedback systems with reference magnets are available in some accelerators, but they are prohibitively costly, limited in precision, and not applicable to higher-order fields. The SPS, therefore, relies on precycles to force magnets into reproducible states. In 2024, precycling accounted for about 20 % of beam time and consumed approximately 5 GWh, while still necessitating manual field corrections when sequences are changed and magnetic history is altered.

This work presents a data-driven alternative in which the magnetic field is forecast from current sequences with sequence models, and feedforward corrections are applied cycle-by-cycle. We evaluate encoder-decoder LSTMs [7], attention-augmented variants, and the Temporal Fusion Transformer

* also at CERN, Geneva, Switzerland

(TFT) [9], and we show that pretraining on simulated hysteresis improves robustness when the model is fine-tuned on magnetic measurements. Building on earlier work [11, 12], we demonstrate that transformers achieve reproducible SPS dipole fields even without the use of precycles.

2 Sequence models for hysteresis modeling

While phenomenological hysteresis models do exist, such as the operator-based Preisach [13] and ODE-based Jiles-Atherton [8] models, they are empirical and not derived from physics. Their parameters can be adjusted or fitted to data [15] to match measured data under specific excitation conditions, but are not able to reach the required accuracy below 1×10^{-4} T, nor reach generalization requirements for accelerator magnet design and control. We model the magnetic field response directly from measurement data by treating the relationship between the excitation current and the field as a time-series forecasting problem, which considers both the current and its historical values..

To address this forecasting challenge, we exploit an online field-measurement system, the B-Train [1], for the main dipoles. While not used for feedback, the B-Train provides cycle-by-cycle labels with a $\approx 5 \times 10^{-5}$ T noise floor and 2×10^{-5} T maximum precision, but can accumulate several 1×10^{-4} T of error over long horizons due to flux integration drift. The accelerator logging system [17] continuously records the current being played and the measured field at 1 kHz for downstream supervised learning.

The task is to forecast B from I as a history-dependent sequence problem by capturing temporal relations in the measured data. However, unlike traditional hysteresis modeling, the objective is not to reproduce arbitrary major or minor hysteresis loops, but rather to capture cycle-to-cycle differences within the same programmed loops, which determine reproducibility in operation. As a result, modern approaches like using neural operators [2] are not suitable for our application. We therefore adopt encoder-decoder models, in which the encoder captures past I/B context and the decoder predicts B from future current I . This approach views field modeling as multivariate time-series forecasting, accommodating long-range dependencies and high-precision constraints, and focuses on fitting the data rather than the underlying dynamics directly.

2.1 Architectures for field modeling

The Temporal Fusion Transformer (TFT) represents the state of the art in multivariate time series forecasting. It combines LSTM backbones with attention mechanisms, gating, and deep feature mixing. Although our field prediction task does not exploit categorical or static features and relies less on heavy feature mixing, the TFT provides a high-capacity upper bound for time-series modeling.

As a baseline model, we use an LSTM [7], but implement it as an encoder-decoder (EDLSTM), where an encoder LSTM takes I/B from the past and produces a hidden state that initializes the decoder. The decoder then infers the field response from the current. To bridge the gap between the baseline and TFT, we implement another encoder-decoder LSTM with cross-attention over the encoder and decoder on the LSTM outputs (ATTNLSTM) before the fully connected layers. Finally, we enrich an encoder-decoder LSTM with transformer blocks with self-attention in the encoder, and cross attention to the decoder on each transformer block layer (TFLSTM), similar to the traditional transformer architecture [16]. For our task, we choose to exploit LSTMs as backbones instead of MLPs due to their capacity to capture short-range dependencies, while introducing an explicit recursive time dependency without using positional encodings. To capture global patterns such as hysteresis, we resort to attention mechanisms like the TFT.

3 Model Training and Finetuning

The training dataset consists of 108 uninterrupted measurement sets, each containing at least one magnetic sequence change that excites hysteresis, and a single sequence change is reserved for validation. Signals are low-pass filtered and downsampled by a factor of 20, yielding an effective sampling interval of 50 ms, which allows us to capture transient effects while avoiding aliasing. After downsampling, the data comprises roughly 41 000 000 training steps and 300 000 validation steps recorded between 2023 and 2025. Since the SPS operates only about twenty distinct cycle programs, the validation program also appears in training. Consequently, performance is evaluated under varying machine conditions rather than through unseen hysteresis loops.

For supervised training, input samples are drawn with sliding windows, and the maximum context and prediction horizon are each 1020 steps, covering the longest operational cycle of 20.4 s. To improve robustness to variable cycle lengths, the model samples random context and horizon lengths between 180 and 1020 steps. To enable this, inputs use packed sequences with adapted attention masks during the forward pass, and padded elements are masked during loss computation.

A hyperparameter scan selected a model dimension of 256 with four attention heads for the transformer components, corresponding to about 5.5 million parameters for the TFT. The same hidden size is applied to the LSTM baselines for a fair comparison. Models are trained with AdamW [10] using a learning rate of 1×10^{-4} , batch size 64, early stopping on validation error, and regularization with 20 % dropout and weight decay of 1×10^{-4} .

To increase data diversity and improve robustness, we pre-train the TFT on simulated sequences generated by a simple Jiles–Atherton hysteresis model [8], extended to include rate-dependent effects. The simulation parameters are tuned to reproduce the general behavior of iron-yoke magnets and to match the overall magnitude of hysteresis observed in the magnetic measurements. While these simulations cannot accurately capture the dynamics of the SPS magnets, they provide a representation of the underlying hysteretic behaviour and help the model explore a large portion of the target feature space. Fine-tuning then proceeds in two stages: first unfreezing only the output head until convergence; then unfreezing the whole network with a reduced learning rate.

4 Quantitative Evaluation and Metrics

Table 1: RMSE calculated between the measured magnetic field from SPS main dipoles and the hysteresis model predictions with different evaluation methods, including autoregressive (AR), on the validation and unseen test set.

Model	validation RMSE (10^{-4} T)			test RMSE (10^{-4} T)		
	Non-cycled	Cycled	AR	Non-cycled	Cycled	AR
EDLSTM	0.36	0.42	0.42	1.48	1.56	6.01
ATTNLSTM	0.34	0.43	0.46	2.07	1.36	5.70
TFLSTM	0.27	0.38	0.39	1.09	1.27	5.28
TFT	0.42	0.52	0.97	1.53	1.31	5.81
TFT (finetuned)	0.18	0.41	0.55	0.33	0.42	0.46

Table 1 reports RMSE results showing that the TFT with transfer learning achieves the best performance across all modes, validating our pre-training approach. We use standard windowed evaluation with max target length, per-cycle prediction to assess variable sequence lengths, and cycled autoregression, where each predicted cycle feeds the next as context.

The encoder–decoder LSTM provides a strong baseline, indicating that much of the field behavior can already be captured through short-term temporal dependencies. Given that the baseline EDLSTM with a model dimension of 256 already contains about 2.5 million parameters, a substantial part of the field patterns is likely memorized directly in the model weights. In fact, even smaller LSTMs with only a few hundred parameters are able to reproduce the field response within 4×10^{-4} T, indicating that a significant fraction of the behavior can be captured with low-dimensional representations.

Adding attention layers in the ATTNLSTM and TFLSTM further improves performance by enabling the model to capture long-range dependencies. The fine-tuned TFT combines these advantages by adding fully connected layers for gating and variable selection, yielding the lowest RMSE across most modes. On the unseen test set, all models perform worse than on the validation set, yet the fine-tuned TFT remains below 1×10^{-4} T in all modes, and demonstrates model robustness against error accumulation in autoregressive operation. In practice, we often see even lower errors than the aggregate RMSE, depending on the rate of change of the magnetic cycle, since errors tend to be higher during fast ramps. The achieved accuracy approaches the magnetic measurement precision of 2×10^{-5} T and is sufficient to control the main dipoles for high energy operation, where the tolerance is 1×10^{-4} T.

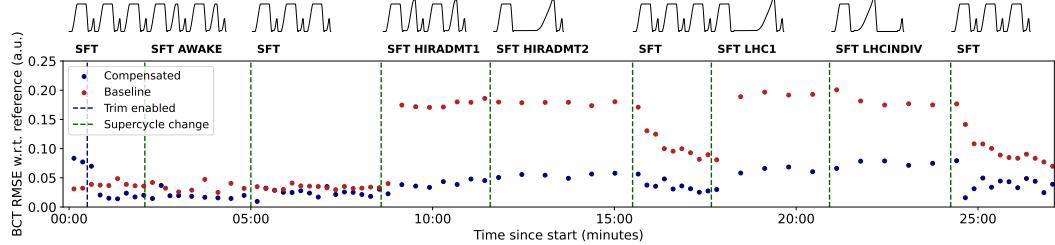


Figure 1: Operational sequence changes, with and without field compensation. The metrics show extracted beam quality via RMSE between normalized measured and reference beam parameters (lower RMSE indicates better reproducibility) for the SFT cycle. In this case, the cycle is always pre-cycled to represent current operational conditions. The magnetic sequences are shown at the top.

5 Active Field Compensation in the SPS

A key achievement of this work was demonstrating that such a data-driven model can be used in practice for active field compensation of the main dipoles in the SPS. Using the model predictions, we can perform feedforward correction and keep the magnetic field reproducible with respect to a chosen reference. In practice, the reference is set to the model’s first field prediction, and subsequent cycles are corrected to match this reference using the model’s prediction of the field.

The model is fully integrated into CERNs accelerator control system [14, 4, 3] hierarchy, and the control loop can send ΔB corrections for every cycle to be played in the SPS, with a 2500 ms forewarning, and the control system propagates a corresponding ΔI to the power converters. The inference is done on a single NVIDIA A100 GPU, and runs inference between 100-200 ms per cycle for the biggest TFT model.

At the time of writing, the compensation is applied to the SPS proton Fixed Target physics cycle (SFT) at the extraction energy, which is most sensitive to magnetic field changes due to keeping the accelerator on resonance (required for stable extraction) for roughly 5 s to achieve uniform particle extractions, occupying 50 % of SPS machine time in 2024. In fact, the SFT cycle is the only cycle in the SPS requiring hysteresis compensation in normal operation. Errors as small as 1×10^{-4} T already introduce significant perturbations during beam extraction in this cycle, while changes in the magnetic sequence can introduce deviations up to 4×10^{-4} T field error during normal operation. Figure 1 shows the reproducibility of the beam parameters at extraction on the SFT cycle with and without field compensation. Reproducibility is quantified as the RMSE between the measured beam current transformer (BCT) signal and a fixed reference. With field compensation, the magnetic field is kept stable with respect to the reference below an error tolerance of 5×10^{-5} T, keeping the extracted beam parameters reproducible.

In addition to this short 30-minute test shown in Figure 1, we have successfully performed operational tests over 36 hours, demonstrating stable autonomous compensation and improved beam quality over extended periods while being fully transparent to machine operators on shift. By successfully applying field compensation on this cycle, we demonstrate that our approach significantly improves existing operation by reducing the need for manual adjustments when switching magnetic sequences, and enabling more dynamic beam operation.

To improve operational conditions, we have also applied field compensation during SFT cycles without precycles. Figure 2 shows measured magnetic loops and corresponding predictions, with conditions showing with and without precycle, and through different excitations of hysteresis. With the predictions met and downstream feedforward correction applied, the properties of the accelerated beam closely follow the chosen reference even in the absence of a precycle, suggesting that the precycle could be eliminated in the future, saving energy.

6 Limitations

The main constraints of our method lie in data coverage and label quality. While large volumes of accelerator data are available, they overrepresent a small set of standard cycle programs (often

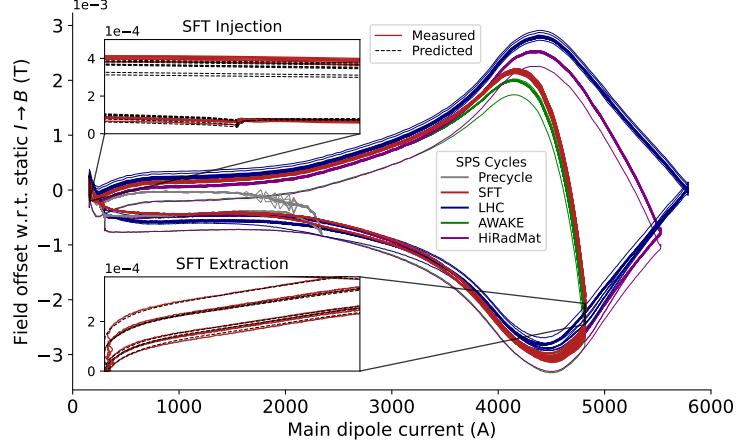


Figure 2: Hysteresis loops of magnetic cycles used for physics operation in the SPS, with cycles for accelerating beam to the LHC, the AWAKE [5] and HiRadMat [6] experiments. The insets show the prediction and ground truth in the critical regions of the Fixed Target SFT beam, with and without precycle.

with precycles). In contrast, rare conditions such as no-precycle or rapid sequence changes are underrepresented. This imbalance increases the risk of overfitting in high-capacity models like TFT, despite their superior performance when sufficient data is available. While transfer learning with the TFT improves convergence and robustness, its benefit depends on the alignment between simulated and measured sequences, and the remaining gains largely reflect higher capacity rather than interpretability. In the current SPS operation, which uses a discrete set of sequences, memorizing these patterns still enables operationally sound predictions. However, it limits the ability to generalize to unseen programs.

The method is inherently data-dependent and relies on high-precision magnetic field measurements, where the prediction accuracy must approach the precision of the data itself over long horizons. The only practical path to higher accuracy is to improve the quality of field measurements, which is highly challenging. The magnetic field is obtained by integrating the induced voltage; even a small bias of μV in the sampled signal accumulates over time, producing an integrated voltage drift. Currently, it is not possible to achieve the required precision of $1 \times 10^{-5} \text{ T}$ under pulsed conditions on the measurement bench, unlike in static measurements at a constant field. While pre-training on simulated data partially addresses the limited variety of measured cycles, enhancing data precision is currently more important than increasing data diversity. Enhancing field measurements is crucial for improving model performance and reproducibility, beyond what architectural changes alone can achieve.

7 Conclusions

This study demonstrates data-driven, feedback-free magnetic field compensation in the SPS using advanced attention-based sequence models. It demonstrates that pretraining on simulated hysteresis effectively accelerates convergence and enhances robustness when fine-tuned on actual measurements. The same method can be retrained or fine-tuned using field measurements from different magnet families. Quadrupoles exhibit an excitation pattern similar to that of dipoles, which enables the utilization of existing models as a basis for transfer learning. In contrast, sextupoles and octupoles may necessitate additional modifications to ensure optimal performance.

Operationally, this approach eliminates the need for manual corrections after magnetic sequence changes, improves reproducibility across cycles, and reduces reliance on precycles, with potential extension beyond dipoles. While further accuracy gains are primarily limited by data quality and coverage rather than model capacity, transfer learning offers a practical route to adapt these models to other magnet families such as quadrupole and sextupole circuits, and to additional accelerators.

References

- [1] Simon Albright, David Giloteaux, Heiko Damerau, Alexander Huschauer, Vincenzo Di Capua, Marco Buzio, Matthias Bonora, and Gian Piero Di Giovanni. First operational results of new real-time magnetic measurement systems for accelerator control. *Proc. IPAC'23, Venice, Italy*, pages 3996–3999 pages, September 2023. ISSN 2673-5490. doi: 10.18429/JACOW-IPAC2023-THPA024. URL <https://jacow.org/ipac2023/doi/jacow-ipac2023-thpa024>.
- [2] Abhishek Chandra, Bram Daniels, Mitrofan Curti, Koen Tiels, and Elena A. Lomonova. Magnetic Hysteresis Modeling With Neural Operators. *IEEE Transactions on Magnetics*, 61(1):1–11, January 2025. ISSN 0018-9464, 1941-0069. doi: 10.1109/TMAG.2024.3496695. URL <https://ieeexplore.ieee.org/document/10750871/>.
- [3] Lajos Cseppentő and Mark Büttner. UCAP: A Framework for Accelerator Controls Data Processing @ CERN. *Proceedings of the 18th International Conference on Accelerator and Large Experimental Physics Control Systems, ICALEPCS2021*:6 pages, 0.454 MB, 2022. ISSN 2226-0358. doi: 10.18429/JACOW-ICALEPCS2021-MOPV039. URL <https://jacow.org/icalepcs2021/doi/JACoW-ICALEPCS2021-MOPV039.html>. Artwork Size: 6 pages, 0.454 MB ISBN: 9783954502219 Medium: PDF Publisher: JACoW Publishing, Geneva, Switzerland.
- [4] Jean-Baptiste De Martel, Roman Gorbonosov, and Nico Madysa. Machine Learning Platform: Deploying and Managing Models in the CERN Control System. *Proceedings of the 18th International Conference on Accelerator and Large Experimental Physics Control Systems, ICALEPCS2021*:6 pages, 0.478 MB, 2022. ISSN 2226-0358. doi: 10.18429/JACOW-ICALEPCS2021-MOBL03. URL <https://jacow.org/icalepcs2021/doi/JACoW-ICALEPCS2021-MOBL03.html>. Artwork Size: 6 pages, 0.478 MB ISBN: 9783954502219 Medium: PDF Publisher: JACoW Publishing, Geneva, Switzerland.
- [5] E. Gschwendtner, E. Adli, L. Amorim, R. Apsimon, R. Assmann, A.-M. Bachmann, F. Batsch, J. Bauche, V. K. Berglyd Olsen, M. Bernardini, R. Bingham, B. Biskup, T. Bohl, C. Bracco, P. N. Burrows, G. Burt, B. Buttenschon, A. Butterworth, A. Caldwell, M. Cascella, E. Chevallay, S. Cipiccia, H. Damerau, L. Deacon, P. Dirksen, S. Doevert, U. Dorda, J. Farmer, V. Fedosseev, E. Feldbaumer, R. Fiorito, R. Fonseca, F. Friebel, A. A. Gorn, O. Grulke, J. Hansen, C. Hessler, W. Hofle, J. Holloway, M. Huther, D. Jaroszynski, L. Jensen, S. Jolly, A. Joulaei, M. Kasim, F. Keeble, Y. Li, S. Liu, N. Lopes, K. V. Lotov, S. Mandry, R. Martorelli, M. Martyanov, S. Mazzoni, O. Mete, V. A. Minakov, J. Mitchell, J. Moody, P. Muggli, Z. Najmudin, P. Norreys, E. Oz, A. Pardons, K. Pepitone, A. Petrenko, G. Plyushchev, A. Pukhov, K. Rieger, H. Ruhl, F. Salveter, N. Savard, J. Schmidt, A. Seryi, E. Shaposhnikova, Z. M. Sheng, P. Sherwood, L. Silva, L. Soby, A. P. Sosedkin, R. I. Spitsyn, R. Trines, P. V. Tuev, M. Turner, V. Verzilov, J. Vieira, H. Vincke, Y. Wei, C. P. Welsch, M. Wing, G. Xia, and H. Zhang. AWAKE, The Advanced Proton Driven Plasma Wakefield Acceleration Experiment at CERN. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 829:76–82, September 2016. ISSN 01689002. doi: 10.1016/j.nima.2016.02.026. URL <http://arxiv.org/abs/1512.05498> [physics]. arXiv:1512.05498
- [6] Fiona Harden, Aymeric Bouvard, Nikolaos Charitonidis, and Yacine Kadi. HiRadMat: A Facility Beyond the Realms of Materials Testing. *Proceedings of the 10th Int. Particle Accelerator Conf., IPAC2019*:4 pages, 1.520 MB, 2019. doi: 10.18429/JACOW-IPAC2019-THPRB085. URL <http://jacow.org/ipac2019/doi/JACoW-IPAC2019-THPRB085.html>. Artwork Size: 4 pages, 1.520 MB ISBN: 9783954502080 Medium: PDF Publisher: JACoW Publishing, Geneva, Switzerland.
- [7] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, November 1997. ISSN 0899-7667, 1530-888X. doi: 10.1162/neco.1997.9.8.1735. URL <https://direct.mit.edu/neco/article/9/8/1735-1780/6109>. Publisher: MIT Press.
- [8] David C. Jiles and David L. Atherton. Theory of ferromagnetic hysteresis. *Journal of Applied Physics*, 55(6):2115–2120, March 1984. ISSN 0021-8979, 1089-7550. doi: 10.1063/1.333582. URL <https://pubs.aip.org/jap/article/55/6/2115/13220/Theory-of-ferromagnetic-hysteresis-invited>.
- [9] Bryan Lim, Sercan O. Arik, Nicolas Loeff, and Tomas Pfister. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting, September 2020. URL <http://arxiv.org/abs/1912.09363> [cs, stat]. arXiv:1912.09363 [cs, stat].
- [10] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization, January 2019. URL <http://arxiv.org/abs/1711.05101>. Number: arXiv:1711.05101 arXiv:1711.05101 [cs, math].
- [11] Anton Lu, Verena Kain, Carlo Petrone, Vincenzo Di Capua, Michael Schenk, and Carlo Zannini. First operational experience with data-driven hysteresis compensation for the main dipole magnets of the CERN SPS. pages 874–877 pages, 0.55 MB, July 2024. ISSN 2673-5490. doi: 10.18429/JACOW-IPAC2024-MOPS66. URL <https://jacow.org/ipac2024/doi/jacow-ipac2024-mops66>. Artwork Size: 874-877 pages, 0.55 MB ISBN: 9783954502479 Medium: PDF Publisher: JACoW Publishing.

- [12] Anton Lu, Verena Kain, Carlo Petrone, Vincenzo di Capua, Michael Schenk, and Maurus Taupadel. Data-driven hysteresis compensation in the CERN SPS main magnets. In *Proc. IPAC'25*, IPAC'25 - 16th international particle accelerator conference, pages 1674–1677. JACoW Publishing, Geneva, Switzerland, June 2025. ISBN 978-3-95450-248-6. doi: 10.18429/JACoW-IPAC25-WEAN2. URL <https://indico.jacow.org/event/81/contributions/7155>. ISSN: 2673-5490 Number: 16 tex.paper: WEAN2 tex.venue: Taipei, Taiwan.
- [13] Franz Preisach. Über die magnetische Nachwirkung. *Zeitschrift für Physik*, 94(5-6):277–302, May 1935. ISSN 1434-6001, 1434-601X. doi: 10.1007/BF01349418. URL <http://link.springer.com/10.1007/BF01349418>.
- [14] Chris Roderick and Ronald Billen. The LSA database to drive the accelerator settings. In *Proc. ICALEPS'09*, pages 417–419. JACoW Publishing, Geneva, Switzerland. URL <https://jacow.org/icaleps2009/papers/WEP006.pdf>. tex.paper: WEP006 tex.venue: Kobe, Japan, Oct. 2009.
- [15] Ryan Roussel, Auralee Edelen, Daniel Ratner, Kabir Dubey, Juan Pablo Gonzalez-Aguilera, Young-Kee Kim, and Nikita Kuklev. Differentiable Preisach Modeling for Characterization and Optimization of Accelerator Systems with Hysteresis. *Physical Review Letters*, 128(20):204801, May 2022. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.128.204801. URL <http://arxiv.org/abs/2202.07747> [physics]. arXiv:2202.07747 [physics].
- [16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, December 2017. URL <http://arxiv.org/abs/1706.03762>. arXiv:1706.03762 [cs].
- [17] Jakub Wozniak and Chris Roderick. NXCLS - Architecture and Challenges of the Next CERN Accelerator Logging Service. *Proceedings of the 17th International Conference on Accelerator and Large Experimental Physics Control Systems*, ICALEPS2019:5 pages, 1.095 MB, 2020. ISSN 2226-0358. doi: 10.18429/JACoW-ICALEPS2019-WEPA163. URL <https://jacow.org/icaleps2019/doi/JACoW-ICALEPS2019-WEPA163.html>. Artwork Size: 5 pages, 1.095 MB ISBN: 9783954502097 Medium: PDF Publisher: JACoW Publishing, Geneva, Switzerland.