

Physics-based deep neural networks for beam dynamics in charged particle accelerators

Andrei Ivanov^{*} and Ilya Agapov

Deutsches Elektronen Synchrotron DESY, Notkestrasse 85, 22607, Hamburg, Germany



(Received 3 April 2020; accepted 30 June 2020; published 7 July 2020)

This paper presents a novel approach for constructing neural networks which model charged particle beam dynamics. In our approach, the Taylor maps arising in the representation of dynamics are mapped onto the weights of a polynomial neural network. The resulting network approximates the dynamical system with perfect accuracy prior to training and provides a possibility to tune the network weights on additional experimental data. We propose a symplectic regularization approach for such polynomial neural networks that always restricts the trained model to Hamiltonian systems and significantly improves the training procedure. The proposed networks can be used for beam dynamics simulations or for fine-tuning of beam optics models with experimental data. The structure of the network allows for the modeling of large accelerators with a large number of magnets. We demonstrate our approach on the examples of the existing PETRA III and the planned PETRA IV storage rings at DESY.

DOI: [10.1103/PhysRevAccelBeams.23.074601](https://doi.org/10.1103/PhysRevAccelBeams.23.074601)

I. INTRODUCTION

Machine learning (ML) techniques are finding increasing usage in various aspects of particle accelerators, including fault prediction, performance optimization, and virtual diagnostics. For more information see, e.g., [1]. Many applications could benefit from having a beam optics model which on the one hand accurately represents the beam dynamics of the accelerator, and on the other hand, can be trained or adjusted on the limited amount of experimental data to serve as a model of the real machine with various imperfections. Machine learning methods and neural networks (NN) in particular hold a promise for constructing such models.

However, applying ML methods for learning the behavior of dynamical systems such as those describing linear and nonlinear dynamics in charged particle accelerators can in some cases require a prohibitively large amount of training data. There is a need for ML-based modeling approaches that reduce reliance on large data sets. A major concern is that models trained with limited observations may not generalize well beyond the training dataset and will thus exhibit poor predictive power. A significant improvement potential in the generalization ability and the training performance of ML models such as NNs lies in

incorporating physics constraints either in the NN architecture or in the training process.

NNs are often employed for system learning and control [2,3], when models are trained either with large measured or simulated datasets. By physics-inspired neural networks [4] authors generally mean either incorporating domain knowledge in the traditional NN or providing additional loss functions to ensure physically consistent predictions [5–7]. Most of the authors either incorporate numerical integrators for ordinary differential equations (ODEs) into NN or, oppositely, parametrize equations with NNs. There also exist various studies related to PNNs in the literature [8–10]. So, in [8,9] polynomial architectures that approximate differential equations are discussed. All studies related to PNNs regard the polynomial architectures as black-box models, and they do not indicate the architectures' connection to the ODEs.

We propose to incorporate the physics constraints on the single-particle beam dynamics in accelerators by introducing NN architecture which is directly derived from the Taylor maps corresponding to the accelerator components and additionally imposing symplectic constraints on the network. The approach presented in the paper is based on works [11–13], where the authors target building polynomial neural networks (PNN) that approximate the exact system of ODEs. In contrast to these works, we do not target solving exact equations and simplified physical problems, but focus on the practical application of the fine-tuning of the PNN with measured data.

In this paper, we first introduce the polynomial neural network (PNN) architecture incorporating Taylor map (TM-PNN) information. Training and symplectic regularization

^{*}andrei.ivanov@desy.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

are then discussed, followed by the application of TM-PNNs to simulating beam dynamics in storage rings. Finally, the results of fine-tuning the TM-PNN based on experimental data at the PETRA III [14] storage ring and simulations for the proposed PETRA IV [15] ring at DESY are presented.

II. POLYNOMIAL NEURAL NETWORK INCORPORATING TAYLOR MAPS

Taylor maps are a natural and well-known approach of treating beam dynamics in particle accelerators, see e.g., [16,17]. A Taylor map \mathcal{M} is a polynomial transformation of the phase space coordinates of a particle in the form

$$\mathbf{X}(s) \equiv \mathbf{X}_1 = W_0 + W_1 \mathbf{X}_0 + W_2 \mathbf{X}_0^{[2]} + \dots + W_k \mathbf{X}_0^{[k]}. \quad (1)$$

Here $\mathbf{X}_{0,1} \in R^n$ are the n -dimensional phase space coordinates of a particle, W_i are weight matrices, and $\mathbf{X}^{[k]}$ are the k -th Kronecker powers of vector \mathbf{X} . For example, for $\mathbf{X} = (x_1, x_2)$, $\mathbf{X}^{[2]} = (x_1^2, x_1 x_2, x_2^2)$, $\mathbf{X}^{[3]} = (x_1^3, x_1^2 x_2, x_1 x_2^2, x_2^3)$ and so on. The weights of (1) can be derived from equations of particle motion that can be described in a general form with polynomial coefficients

$$\frac{d}{ds} \mathbf{X} = \mathbf{F}(s, \mathbf{X}) = \sum_{j=0}^n P_j(s) \mathbf{X}^{[j]}, \quad (2)$$

where s is an independent variable, for which in accelerator physics the path length of the reference particle is usually taken. Differentiating (1) and combining it with (2) yields an equation for weight matrices W_i

$$\frac{d}{ds} W_i = f_i(W_0, \dots, W_k, P_0, \dots, P_n), \quad i = \overline{0, k}, \quad (3)$$

where f_i are functions of matrices W_i and P_i . For instance, $f_0 = P_0 + P_1 W_0 + \dots + P_k W_0^{[k]}$. Solving (3) for W_i gives the dynamics of the system for all initial conditions by means of (1). The success of this technique for modeling beam dynamics in accelerators lies in the fact that Taylor maps for individual magnets can be readily constructed to desired order, and long sequences of magnets comprising an accelerator can be represented by concatenating the maps corresponding to individual magnets

$$\mathcal{M} = \mathcal{M}_n \circ \mathcal{M}_{n-1} \circ \dots \circ \mathcal{M}_1. \quad (4)$$

Various methods for efficiently concatenating such maps and extracting physics parameters such as betatron tunes or amplitude detuning have been developed [16]. An efficient way of constructing a neural network which could model dynamics in such a system is to represent each map in this chain by a set of polynomial neurons (layer) computing

exactly the same Taylor map, and then link these layers into a deep network. The NN has as its input the phase space coordinates and propagating the inputs through each layer is fully equivalent to applying the corresponding map.

The general architecture of the TM-PNN is shown in Fig. 1. The network has a single input and can have an arbitrary number of outputs corresponding to observables. So, to have a beam trajectory at the beam position monitor (BPM) locations as output, the drift space containing the monitor is split into two parts at the monitor location. Two layers for drifts upstream and downstream of the monitor are created, and the upstream layer additionally has beam position coordinates x and y as outputs. Generally, the network output can be an arbitrary function of the outputs of all neurons in the network. Moreover, any number of parameters of the accelerator lattice such as the strength of quadrupole magnets can be trivially introduced. To add one additional parameter, the network input vector dimension is extended by one and the parameter becomes an additional input to the neurons where it enters.

For example, the transfer matrix for horizontal motion through a focusing quadrupole of length L

$$M = \begin{pmatrix} \cos(\sqrt{k}L) & \frac{1}{\sqrt{k}} \sin(\sqrt{k}L) \\ -\sqrt{k} \sin(\sqrt{k}L) & \cos(\sqrt{k}L) \end{pmatrix}$$

can be parametrized by Taylor map and can be represented as a polynomial neuron with three-dimensional input $\mathbf{X}_0 = (x_0, x'_0, k)$ and two-dimensional output $\mathbf{X} = (x, x')$ where the weights up to the second order are

$$W_1 = \begin{pmatrix} 1 & L & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad W_2 = \begin{pmatrix} 0 & 0 & -0.5L^2 & 0 & 0 & 0 \\ 0 & 0 & -L & 0 & -0.5L^2 & 0 \end{pmatrix}. \quad (5)$$

We further illustrate the concept by an example of a network that corresponds to a simple but practically relevant magnet arrangement, the FODO structure with sextupoles, which exhibits nonlinear behavior typical of circular accelerators. The magnetic lattice and the corresponding NN are shown in Fig. 2. This Taylor map-based polynomial neural network (TM-PNN) has 12 layers, for

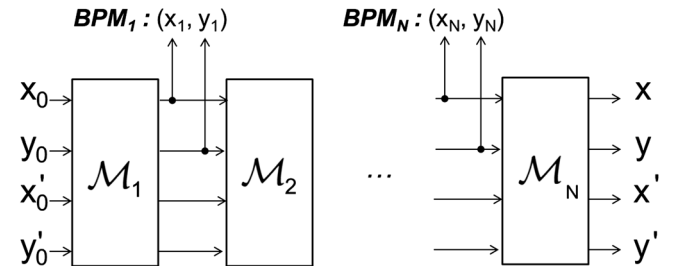


FIG. 1. Multioutput deep TM-PNN for particle accelerator.

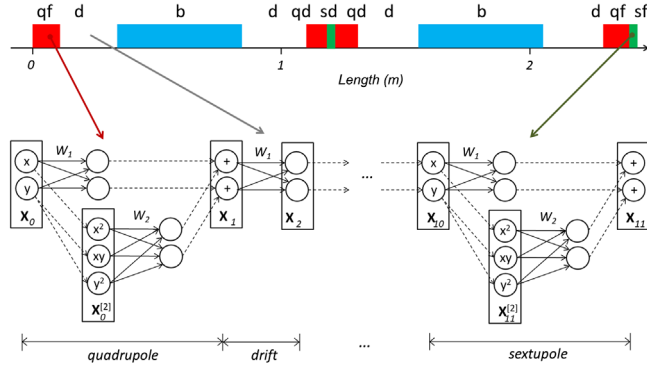


FIG. 2. Schematic representation of a FODO structure with eight magnets and a corresponding TM-PNN. The lattice consists of drifts (d), bending magnets (b), shown in blue, focusing (qf) and defocusing (qd) quadrupoles, shown in red, and focusing (sf) and defocusing (sd) sextupoles, shown in green.

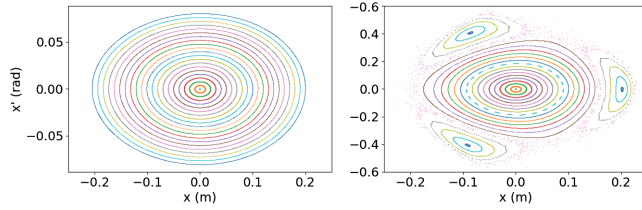


FIG. 3. Initialized from maps the TM-PNN accurately represents the parametric dependency of dynamics on magnet strength, such as the appearance of a third-integer resonance.

each component in the lattice. The resulting NN fully represents the dynamics of the lattice, including parametric dependency on quadrupole gradients. The system is non-linear and exhibits resonant behavior. Figure 3 shows the phase space portraits of this lattice computed by TM-PNN.

III. NN TRAINING AND SYMPLECTIC REGULARIZATION

A TM-PNN pre-initialized with the Taylor map coefficients extracted from the ideal magnetic lattice already provides an accurate model of the ideal accelerator and does not require training. Training may be required if it is desired to fine-tune the model based on experimental data. The training is done in the usual way by fine-tuning the network weights given a set of inputs and outputs. For training, we use the Adam [18] algorithm with gradient clipping. The examples of such training are given in the last section of this paper.

The initial network weights are derived from the maps and are guaranteed to be physically meaningful. During the network training process, we need to ensure the physical consistency of the weights, so that they are changed in such a way that the layers still correspond to Taylor maps of magnets. The Taylor maps for the systems of ordinary differential equations contain weights that vary by orders of

magnitude, and the often-used L1- or L2-norm-based regularizations of the weights during the training which try to reduce the absolute magnitude of the weights cannot be applied.

It is however possible to introduce physical invariants of the underlying dynamical systems as regularization penalties on the weights of the TM-PNN. For Hamiltonian systems representing single-particle beam dynamics, this can be the symplecticity [19], which means that for a map $\mathcal{M}: \mathbf{X}_0 \rightarrow \mathbf{X}$

$$\left(\frac{\partial \mathbf{X}}{\partial \mathbf{X}_0} \right)^T J \frac{\partial \mathbf{X}}{\partial \mathbf{X}_0} - J = 0, \quad \forall \mathbf{X}_0, \quad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \quad (6)$$

where I is an identity matrix, and T is the matrix transpose. The symplectic property (6) for TM-PNN leads to algebraic constraints on weights $W_i = \{w_i^{jk}\}$. The symplectic penalty can be implemented for an arbitrary Taylor map as an additional term for the loss function:

$$\text{Loss} = ME(\mathbf{X}_{\text{true}}, \mathbf{X}_{\text{pred}}) + S(W_1, \dots, W_k),$$

where ME is a metric that controls accuracy on training data, and S is a symplectic regularization penalty. For example, for the second-order Taylor map for two-dimensional state vector with $W_1 = \{w_1^{ij}\}$, $W_2 = \{w_2^{mn}\}$ where i, m are row and j, n column indices in condition (6), the symplectic constraint becomes

$$\begin{aligned} w_1^{11}w_1^{22} - w_1^{12}w_1^{21} - 1 &= 0, \\ w_2^{11}w_2^{23} - w_2^{13}w_2^{21} &= 0, \\ w_2^{12}w_2^{23} - w_2^{13}w_2^{22} &= 0, \\ w_1^{11}w_2^{22} - w_1^{21}w_2^{12} + 2w_2^{22}w_1^{11} - 2w_1^{12}w_2^{21} &= 0, \\ w_1^{22}w_2^{12} - w_1^{12}w_2^{22} + 2w_1^{11}w_2^{23} - 2w_1^{21}w_2^{13} &= 0. \end{aligned} \quad (7)$$

The penalty S is the sum of squares of all left-hand-side terms in (7). Since this penalty does not depend on the inputs, the Hamiltonian structure is preserved for all new inputs which has a large impact on generalization. If the symplectic regularization is not considered during the training, the tuning of the maps leads to overfitting of the model that causes inaccurate predictions.

IV. APPLICATION TO BEAM DYNAMICS SIMULATIONS

TM-PNNs can be used for fast simulations of beam dynamics [20]. We implemented translation of beam optics models to TM-PNN using maps of up to second order for individual elements computed in OCELOT [21]. From that point on, the computations necessary for particle tracking are completely performed by standard and highly optimized NN software such as TensorFlow [22]. This makes it

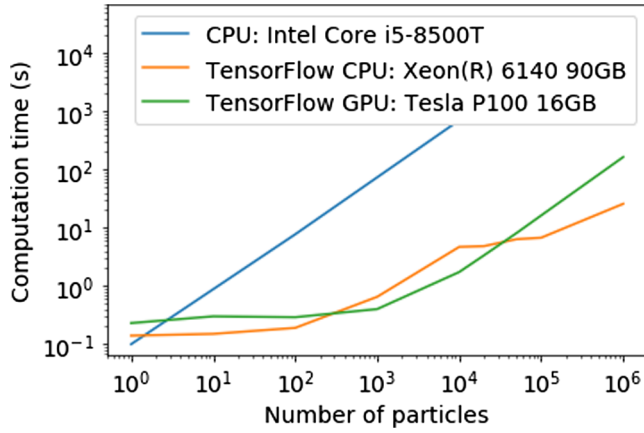


FIG. 4. Scaling of the computation time of the TM-PNN for one-turn tracking in PETRAIII lattice.

possible to have highly efficient particle tracking on various parallel architecture without any additional effort. The computations were cross-checked on the PETRA III and PETRA IV lattices. Figure 4 presents the scaling of computational speed for four-dimensional particle tracking for the PETRA III lattice. On multicore architectures, tracking of thousands of particles can be performed at the same speed as tracking of a single particle. Note that the network can deal with large lattices. So, the PETRA III lattice has 1519 elements, each of which was represented by a NN layer.

V. DATA-DRIVEN TUNING OF THE TM-PNN

Due to alignment errors, field imperfections, and other factors, ideal beam optics models of accelerators differ to various extents from the parameters of real machines. Adjustment of optics parameters in operation is normally performed. Usual methods of optics measurement and correction [23] exploit linear or weakly nonlinear dependency of observable parameters such as beam orbits on free model parameters such as magnet strengths and alignments. The problem of optics measurement is formulated as a least-squares minimization of free parameters given a set of observations, and then a correction is done using some of these parameters as controls to minimize the discrepancy between theoretical and observed models.

One of the general shortcomings of these methods is that the number of free parameters such as offsets and strength errors of magnets is larger than the number of observables, which is ultimately limited by the number of free parameters in the phase space transformation between adjacent beam position monitors. For example, for linear coupled motion the number of such parameters is the dimension of the group of 4D symplectic matrices which is $n(2n + 1) = 10$ so that the number of possible observable parameters in the beam optics is $10N_{\text{BPM}}$. So, for the PETRA III ring at DESY the number of magnets is $N_{\text{mag}} = 1519$ and

$N_{\text{BPM}} = 246$. With 6 free parameters for geometrical misplacement of the components, optics determination becomes strictly speaking ill-posed. Employed least-square fit usually has a different number of parameters than intrinsically present in the model, and additional assumptions or empirical knowledge is often required to successfully use such fits. In contrast, a TM-PNN with a minimal set of layers (i.e., one layer for the transfer map between two BPMs) does not have the shortcomings mentioned above.

A TM-PNN can serve as the backbone for both optics measurement and for beam control. The number of layers should be the smallest possible but include all observables. For example, maps for all magnets between BPMs can be concatenated. However, elements used for beam control should be preserved as separate layers.

The first step, model inference, is fitting all the weights in the network for a given training set of observations, such as orbit measurements. The second step, control, consists of giving the network a target function, such as the target orbit, as a new training dataset, and then using only the weights in the layers corresponding to control elements to fit the desired target. Since by construction of the network there is a unique correspondence between the weights and control parameters, the appropriate correction can be directly inferred. The TensorFlow engine can be used for both training of the network with available observations, and for finding the control parameters. This concept is demonstrated below on several examples.

A. Simulated data for PETRAIV

First, we used a simplified version of the multibend Achromat lattice of PETRA IV (see Fig. 5) [15] to model single-pass trajectory correction with a TM-PNN. Single-pass trajectory correction is an important step in machine startup when beam accumulation is not yet achieved. Random magnet misalignments were generated, and single-pass beam trajectories simulated. In the following we give an example of trajectory correction on the example of a single cell. The cell has 36 magnets and a total of 166 logical elements (free spaces, markers, etc.). For simplicity, we consider only beam motion in horizontal plane $x - x'$. The cell has 11 beam position monitors (BPMs) and 10 orbit correctors.

The orbit correction is formulated as a nonlinear optimization problem

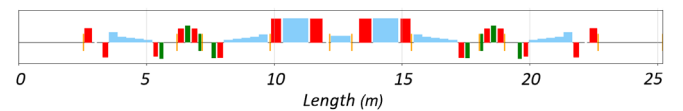


FIG. 5. The PETRA IV cell (a seven-bend achromat of ESRF-EBS type) that consists of bending magnets, shown in blue, focusing and defocusing quadrupoles, shown in red, and focusing (sf) and defocusing (sd) sextupoles, shown in green.

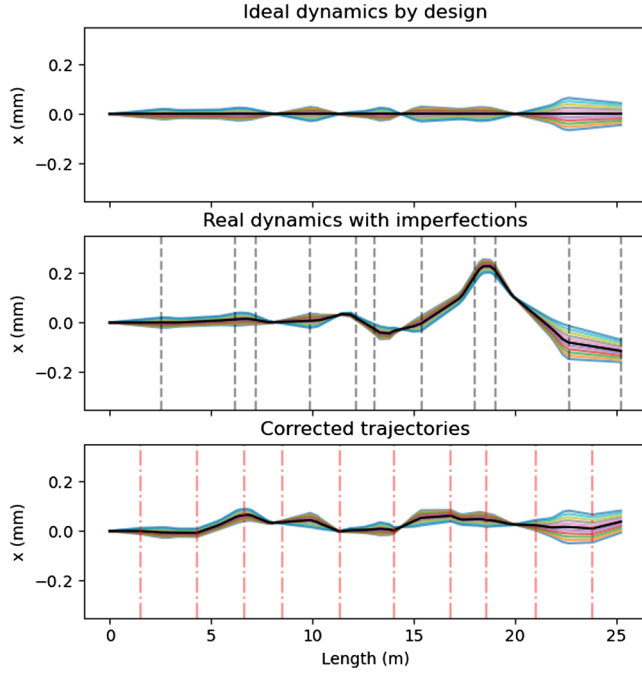


FIG. 6. Trajectory correction in a multibend achromat lattice: Trajectories in an ideal lattice (top), in a lattice with random misalignments (middle), and after correction by a TM-PNN (bottom). BPM and corrector positions are shown with black and red dashed lines respectively.

$$F = \sum_{i=0}^{10} ||x_i(c_0, c_1, \dots, c_9)|| \rightarrow 0, \quad (8)$$

where x_i is a beam coordinate measured at i th monitor, c_j is a strength of j th corrector magnet. The simulated trajectory including misalignments is fed into the TM-PNN representing the ideal lattice to recalibrate the model. The network was then used to find corrector strength resulting in zero trajectory. The correction results for one achromat cell are shown in Fig. 6. Only one trajectory is required to train the network which can then perform the orbit correction.

B. Experiments at PETRAIII

Experiments were performed at PETRA III to demonstrate applicability of TM-PNN for single-pass trajectory correction in real operation. As mentioned before, single-pass trajectory correction is an important task in commissioning and startup of accelerators where the optics model can also be subject to uncertainties. Moreover, diagnostics resolution in single-pass mode is worse than for the stored beam and the demonstration of model re-calibration thus becomes more challenging. For the measurement the beam position monitors were set up to read single-turn data and were triggered at beam injection into the ring. The first orbit corrector in the storage ring upstream from the injection point was set up such that the beam was making only one

turn. The TM-PNN model was initialized with the nominal PETRA III lattice, and the one-turn trajectories consisting of (x, y) positions at 246 BPM locations were used for training.

1. Beam threading

The first experiment demonstrated that a TM-PNN can be used for beam threading through the machine and for orbit correction. We switched off all corrector magnets which leads to the beam being able to travel through only a part of the ring. To propagate the beam further we perform orbit correction with the available (upstream from the beam loss location) correctors by the TM-PNN. For example, only 30 correctors are available before beam is lost in the first iteration, with the total amount of 210 horizontal and 194 vertical correctors in the ring. Given the measured data at each step, the TM-PNN predicts values of corrector magnets that will reduce displacement of the measured orbit. These corrector values are applied to the machine and a new measurement is acquired. After several steps, the beam is propagated along the ring. Figure 7 shows several iterations of beam threading and trajectory correction with a TM-PNN. Solid lines represent beam coordinates, dashed lines are noise downstream from the beam loss. Note that in this example model recalibration is not essential due to relatively weak nonlinearities in the machine and the primary goal was the demonstration of orbit control with the help of the TM-PNN.

2. Optics imperfection

In the second experiment we used the same beam set-up to demonstrate model recalibration on partial, noisy, and limited observations. Optics imperfections were introduced in a controlled way by detuning the strength of one

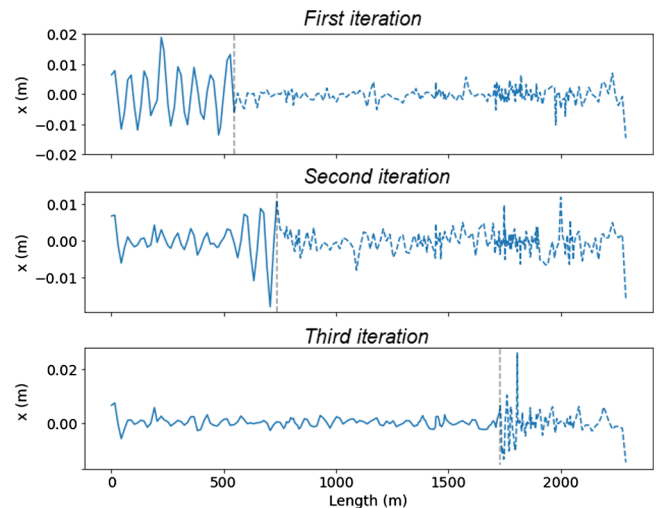


FIG. 7. Several iterations of beam threading with a TM-PNN.

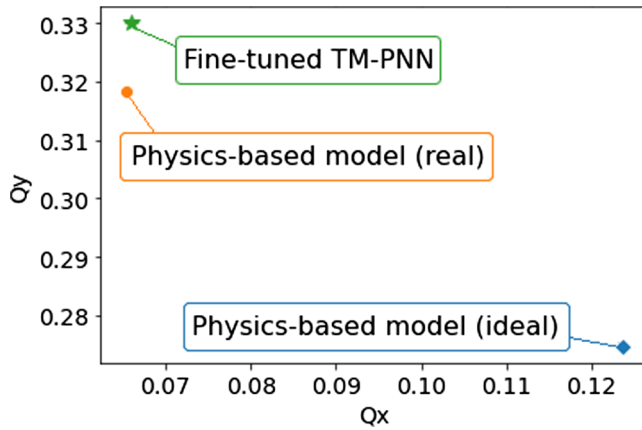


FIG. 8. Tunes in the horizontal(Q_x) and vertical (Q_y) planes.

quadrupole magnet by 20%. The network was trained on the single-pass trajectories from this setup and then used in multiturn tracking. Since the imperfection is known, we independently reproduced that optics in a traditional tracking code (OCELOT). Multiturn particle tracking was done in both codes and the tunes extracted by Fourier analysis (see Fig. 8). Note that only a single one-turn trajectory was used for model recalibration. Tune measurement based on such data with traditional methods is generally impossible.

VI. CONCLUSION

We introduced a new polynomial neural network architecture that is ideally suitable for representing beam dynamics in charged particle accelerators. The networks are already preinitialized to represent linear and nonlinear beam dynamics and can be effectively trained using the symplectic regularization technique we proposed. They provide both a backbone for fast parallel computations and a universal ML-based approach for experimental measurement and correction of charged particle beam optics. Symplectic regularization is limited to Hamiltonian systems, but the TM-PNN approach is a more general. In [12] it is shown that an arbitrary system of nonlinear ODEs can be translated to Taylor maps. So, radiation and/or space charge effects can be introduced into TM-PNN by, for example, mapping onto them the systems of equations from [24]. Also, further research on the estimation of accuracy, performance, and limitations of the proposed method as well as work on introducing it to the daily accelerator operation should be performed.

ACKNOWLEDGMENTS

The authors are grateful to Gero Kube, Joachim Keil, Gajendra Kumar Sahoo and Rainer Wanzenberg for support and active participation in measurements at the PETRAIII storage ring. The work was supported by the AMALEA Helmholtz Innovation Pool Project.

APPENDIX: SUPPLEMENTARY CODE

The implementation of the TM-PNN architecture in Keras/TensorFlow and the described examples of a FODO lattice, simulations of the PETRA IV cell, and translating of the PETRAIII lattice to the TM-PNN architecture are available under <https://github.com/andiva/TM-PNN>

<code>/tm_pnn:</code>	Implementation of the TM-PNN in Keras with TensorFlow backend With several symplectic loss functions
<code>/demos/</code>	
<code>FODO:</code>	FODO lattice with resonances
<code>PETRAIV:</code>	Orbit correction for simulated data
<code>PETRAIII:</code>	Conversion of the PETRA III with 1519 Magnets into TM-PNN

- [1] A. Edelen, C. Mayes, C. Bowring, D. Ratner, A. Adelman, R. Ischebeck, J. Snuerink, I. Agapov, R. Kammering, J. Edelen, Y. Bazarov, G. Valentino, and J. Wenninger, Opportunities in machine learning for particle accelerators, [arXiv:1811.03172](https://arxiv.org/abs/1811.03172).
- [2] A. Nagabandi, G. Kahn, R. Fearing, and S. Levine, Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning, <https://arxiv.org/pdf/1708.02596.pdf>.
- [3] Y. Chen, Y. Shi, and B. Zhang, Optimal control via neural networks: A convex approach, <https://arxiv.org/pdf/1805.11835.pdf>.
- [4] T. Group, Physics-based deep learning, <https://github.com/thunil/Physics-Based-Deep-Learning>.
- [5] A. Karpatne, W. Watkins, J. Read, and V. Kumar, Physics-guided neural networks (pgnn): An application in lake temperature modeling, <https://arxiv.org/pdf/1710.11431.pdf>.
- [6] A. A. Sanchez-Gonzalez, V. Bapst, K. Cranmer, and P. Battaglia, Hamiltonian graph networks with ode integrators, <https://arxiv.org/pdf/1909.12790.pdf>.
- [7] S. Greydanus, M. Dzamba, and J. Yosinski, Hamiltonian neural networks, <https://arxiv.org/pdf/1906.01563.pdf>.
- [8] L. Zjavka, Differential polynomial neural network, *J. Artificial Intelligence* **4**, 89 (2011).
- [9] Y. Yang, M. Hou, and J. Luo, A novel improved extreme learning machine algorithm in solving ordinary differential equations by legendre neural network methods, *Adv. Differ. Equat.* **2018**, 469 (2018).
- [10] V. Schetin, Polynomial neural networks learnt to classify EEG signals, (1997).
- [11] A. Ivanov and S. Andrianov, Matrix lie maps and neural networks for solving differential equations, <https://arxiv.org/abs/1908.06088>.
- [12] A. Ivanov, A. Golovkina, and U. Iben, Polynomial neural networks and taylor maps for dynamical systems simulation and learning, <https://arxiv.org/abs/1912.09986>.
- [13] A. Ivanov, A. Sholokhova, S. Andrianov, and R. Konoplev-Esgenburg, Lie transform-based neural networks for

- dynamics simulation and learning, <https://arxiv.org/abs/1802.01353>.
- [14] PETRA III: A low emittance synchrotron radiation source. Technical design report, Technical Report, 2004.
- [15] C. Schroer, I. Agapov, W. Brefeld, R. Brinkmann, Y. Chae, H. Chao, M. Eriksson, J. Keil, N. G., R. Roehlsberger, O. Seeck, M. Sprung, M. Tischer, R. Wanzenberg, and E. Weckert, PETRA IV: The ultralow-emittance source project at DESY, *J. Synchrotron Radiat.* **25**, 1277 (2018).
- [16] A. Chao, Lecture notes on topics in accelerator physics, Technical report, <http://cds.cern.ch/record/747496>.
- [17] A. Dragt, Lie methods for nonlinear dynamics with applications to accelerator physics, inspirehep.net/record/955313/files/TOC28Nov2011.pdf.
- [18] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [19] V. Arnold, *Mathematical methods of classical mechanics*, Vol. 60 (Springer, New York, 1989).
- [20] A. Ivanov, S. Andrianov, N. Kulabukhova, A. Sholokhova, E. Krushinevskii, and E. Sboeva, Matrix representation of lie transform in tensorflow, *Proc. of 9th IPAC* (2018), <https://doi.org/10.18429/JACoW-IPAC2018-THPAK088>.
- [21] I. Agapov, G. Geloni, S. Tomin, and I. Zagorodnov, OCELOT: A software framework for synchrotron light source and FEL studies, *Nucl. Instrum. Methods Phys. Res., Sect. A* **768**, 151 (2014).
- [22] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.
- [23] R. Toms, M. Aiba, A. Franchi, and U. Iriso, Review of linear optics measurement and correction for charged particle accelerators, *Phys. Rev. Accel. Beams* **20**, 054801 (2017).
- [24] S. Andrianov, Analytical presentation of space charge forces, <http://accelconf.web.cern.ch/ICAP2012/papers/wep16.pdf>.