

Physics-Informed Neural Anhysteresis Surrogate for Magneto-Elastic Vector Hysteresis in Device Simulations

K. Roppert^{1,2}, L. Domenig¹, A. Reinbacher-Köstinger¹, M. Kaltenbacher¹, and L. Daniel^{3,4}

¹Institute of Fundamentals and Theory in Electrical Engineering, 8010 Graz, Austria

²Institute for Accelerator Science and Electromagnetic Fields (TEMF), Technische Universität Darmstadt, 64289 Darmstadt, Germany

³CentraleSupélec, CNRS, Laboratoire de Génie Électrique et Électronique de Paris, Université Paris-Saclay, 91192 Gif-sur-Yvette, France

⁴CNRS, Laboratoire de Génie Électrique et Électronique de Paris, Sorbonne Université, 75252 Paris, France

In many applications, accurately capturing the magneto-mechanical coupling and dissipative effects at the material level is essential for realistic simulations. Embedding the simplified multiscale model (SMSM) inside an energy-based hysteresis framework yields high fidelity but is computationally intensive for 3-D finite element (FE) analyses. This article introduces NNSMSM, a physics-informed multi-task deep neural network that emulates the expensive SMSM operator. A hybrid Latin-hypercube (LH)/Sobol sampling strategy efficiently explores the magneto-mechanical loading space. The network is trained with a composite loss that simultaneously fits magnetization and magnetostrictive strain while enforcing reciprocity and positive definiteness of the susceptibility tensor. The traced TorchScript model is linked to the open-source FE software openCFS, replacing the SMSM inside the vector play model (VPM) hysteresis model with zero code changes. The benchmark of a permanent magnet synchronous machine (PMSM) device simulation shows a speed-up of wall clock time by a factor of 11 while preserving global accuracy of hysteresis losses.

Index Terms—Deep learning, magneto-mechanics, mechanical stress, multiscale model, surrogate modeling, vector hysteresis, vector play model (VPM).

I. INTRODUCTION

DEVICE-SCALE simulations of electrical machines and power transformers demand material models that capture both vector hysteresis and magneto-elastic coupling. A broad range of coupled magneto-mechanical effects strongly influences a machine's electromagnetic behavior [1], [2], [3]. Conventional material models found in commercial software often fail to reflect these complexities, resulting in uncertain predictions and potentially suboptimal designs. This uncertainty is particularly critical when considering thermal management, since imprecise loss predictions force the use of large safety margins, reducing overall efficiency. If engineers can reliably predict these losses, machines can be built smaller without requiring additional cooling, leading to weight reduction and improved performance.

In the preceding work [4], the simplified multiscale model (SMSM) was embedded in an energy-based vector hysteresis framework [5], [6], demonstrating that a physics-based anhysteretic description of the material significantly impacts the prediction of local field distribution and global (hysteresis) losses. However, that fidelity and physics awareness had the downside of a high computational cost. Evaluating the SMSM (and its inclusion in the hysteresis framework) inside every Gauss point and time step increased the FE simulation

time significantly, compared to classically used anhysteretic functions, like Langevin or atan.

This work addresses this challenge and introduces NNSMSM, a physics-informed multi-task deep neural network [7] that emulates the expensive SMSM operator and plugs seamlessly into the energy-based vector hysteresis framework, in our case, the vector play model (VPM) [8]. The combined model (anhysteretic NNSMSM surrogate and hysteretic VPM) is called NNSMSM+VPM in the following. Previous investigations [4] discussed the inclusion of the anhysteretic SMSM into the VPM hysteresis framework and evaluated global hysteresis losses, highlighting the substantial impact of mechanical stress states in PMSM simulations. Compared with the (full fidelity) SMSM, the surrogate (NNSMSM) accelerates material point evaluations by several orders of magnitude while preserving thermodynamic properties like symmetry and positive definiteness of the magnetic susceptibility tensor and reciprocity, required for robust convergence of a Newton scheme, used in FE simulation. The idea of replacing computationally demanding local continuum constitutive relations with machine learning approaches follows earlier works in mechanics, such as [9] and [10].

Currently, the NNSMSM is designed and trained for the 2-D case, the extension to 3-D, however, is straightforward. To showcase its practical value, the combined NNSMSM+VPM was implemented into the open source FE software openCFS [11] and a permanent magnet synchronous machine (PMSM) is simulated under a realistic press fitting stress state. The NNSMSM+VPM reproduces the full

Received 3 July 2025; revised 2 September 2025; accepted 22 September 2025. Date of publication 24 September 2025; date of current version 4 November 2025. Corresponding author: K. Roppert (e-mail: klaus.roppert@tugraz.at).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TMAG.2025.3614063>.

Digital Object Identifier 10.1109/TMAG.2025.3614063

SMSM+VPM reference to within 2% of global hysteresis losses. This proof of concept highlights how physics-informed constitutive surrogates can open the possibility to routine device-level optimization.

A. Modeling Aspects and Main Contributions

Building on the earlier work that coupled the SMSM [12], [13] with an energy-based vector hysteresis model [4], in this contribution, the dissipative part remains untouched, and the focus is exclusively on accelerating the *anhysteretic* model. The full fidelity SMSM is replaced by a physics-informed surrogate NNSMSM, while the VPM provides the irreversible part, exactly as described in [4].

In summary, this article achieves several key advancements as follows.

- 1) *Physics-Informed Surrogate*: A multi-task network that predicts the magnetization $\mathbf{M}$ and magnetostrictive strain tensor $\boldsymbol{\epsilon}^\mu$ as well as their field gradients. Point-wise speed-up over the full fidelity SMSM of around 1000x (at every Gauss point) and 11x at FE level is achieved.
- 2) *Data-Efficient Sampling*: An 8-D Latin-hypercube (LH) design for material/stress parameters combined with a low discrepancy Sobol grid in the H plane is used.
- 3) *Automated Two-Stage Hyperparameter Search*: Using Optuna [14] for finding hyperparameters that balance value and physics losses across data-rich and data-scarce regimes.
- 4) *Seamless FE Integration*: A TorchScript export of the surrogate is used in *openCFS*, the FE solver, where the reciprocity and positive-definiteness are used to keep Newton iterations from diverging.
- 5) *Device-Level Validation*: A press-fit PMSM benchmark shows $\leq 2\%$ hysteresis loss error and 11 times wall-clock reduction when using NNSMSM+VPM versus SMSM+VPM.

II. MODEL DESCRIPTION

The inclusion of the SMSM into the energy-based vector hysteresis framework and especially into the VPM approximation was derived in detail in [4] and [15], respectively. In the following, only the required notation for the embedding of the newly developed NNSMSM surrogate into the VPM is introduced.

Anhysteretic Operator: For any magnetic field and mechanical stress state combination $(\mathbf{H}, \boldsymbol{\sigma})$ the SMSM [12], [13] returns the anhysteretic magnetization and magnetostrictive strain

$$(\mathbf{M}, \boldsymbol{\epsilon}^\mu) = \text{SMSM}(\mathbf{H}, \boldsymbol{\sigma}, A_s, M_s, \lambda_s) \quad (1)$$

with the constant material parameters A_s , controlling the initial slope of the stress-free anhysteresis curve [16], λ_s the magnetostriction parameter, and M_s the saturation magnetization.

VPM incremental updates: The VPM has been previously described in the literature (with [17] introducing the vector stop model and [18] extending it to include the vector stop and vector play *hysteron*), it is briefly described in the following.

Given the previous reversible magnetic field $\mathbf{H}_{\text{rev},p}$ the current state follows from the explicit VPM update rule:

$$\mathbf{H}_{\text{rev}}^{\text{VPM}} = \mathbf{H} - \kappa \frac{\mathbf{H} - \mathbf{H}_{\text{rev},p}}{\|\mathbf{H} - \mathbf{H}_{\text{rev},p}\|}. \quad (2)$$

This update rule is derived using the following simplifying assumptions.

- 1) *Approximation 1*: The direction of the magnetization change $(\mathbf{M} - \mathbf{M}_{\text{prev}})/(|\mathbf{M} - \mathbf{M}_{\text{prev}}|)$ is identified as the direction of the irreversible magnetic field strength $\mathbf{e}_{\text{irr}} = (\mathbf{H} - \mathbf{H}_{\text{rev},p})/(\|\mathbf{H} - \mathbf{H}_{\text{rev},p}\|)$.
- 2) *Approximation 2*: Replace the current reversible magnetic field strength $\mathbf{H}_{\text{rev}}$ by the previous reversible field $\mathbf{H}_{\text{rev},p}$ to obtain $\mathbf{H}_{\text{rev}}^{\text{VPM}}$.

The connection to the SMSM (and its surrogate, the NNSMSM) is introduced after each update step (2), when the new magnetization vector, based on the reversible field component ($\mathbf{H}_{\text{rev}} = \mathbf{H}_{\text{rev}}^{\text{VPM}}$) is evaluated via the *anhysteretic* relationship

$$\mathbf{M} = \mathbf{M}(\mathbf{H}_{\text{rev}}). \quad (3)$$

This anhysteretic relationship is then represented via (1).

A. Composite Dissipative Model

A single pinning force κ yields unrealistically sharp kinks in the hysteresis loops. In reality, magnetization changes occur on a much finer scale, leading to smoother transitions observed in actual hysteresis loops. A more realistic behavior is achieved by introducing multiple pinning forces with varying magnitudes. The overall magnetization is then obtained through a weighted superposition of these individual contributions

$$\mathbf{M} = \sum_{i=1}^N \omega^{(i)} \mathbf{M}^{(i)} \quad (4)$$

where N is the number of pinning forces, and $\omega^{(i)}$ is the respective weight of the i th force. These weights typically follow a particular Rayleigh-like distribution, as described in [19].

Sections III and IV describe how SMSM is replaced by the physics-informed surrogate NNSMSM and how it is integrated in a magnetostatic hysteretic FE formulation.

III. CREATING THE MULTI-TASK DEEP-NEURAL-NETWORK SURROGATE (NNSMSM)

A. Data Generation and Standardization

The surrogate is trained on the 8-D input vector

$$\mathbf{x} = [H_x, H_y, \sigma_{xx}, \sigma_{yy}, \sigma_{xy}, A_s, M_s, \lambda_s]^T \quad (5)$$

which contains the Cartesian components of the magnetic field strength H_x and H_y , which are also used for evaluating the partial Jacobian (differential susceptibility tensor) $\chi_{\text{diff}} = \partial \mathbf{M} / \partial \mathbf{H}$ via automatic differentiation; three SMSM material parameters A_s, M_s, λ_s ; and the plane stress tensor components $[\sigma_{xx}, \sigma_{yy}, \sigma_{xy}]$.

TABLE I
LH RANGES

Parameter	Min	Max
A_s	$1 \cdot 10^{-3} \text{ J/m}^3$	$5 \cdot 10^{-2} \text{ J/m}^3$
M_s	1.4 MA/m	1.8 MA/m
λ_s	$1 \cdot 10^{-6}$	$5 \cdot 10^{-5}$
$\sigma_{xx}, \sigma_{yy}, \sigma_{xy}$	-120 MPa	$+120 \text{ MPa}$

The network outputs five macroscopic quantities (magnetization vector and magnetostriction strain tensor)

$$\mathbf{y} = [M_x, M_y, \varepsilon_{xx}^\mu, \varepsilon_{yy}^\mu, \varepsilon_{xy}^\mu]^\top. \quad (6)$$

All features are standardized according to

$$x_i^{\text{std}} = \frac{x_i - \mu_i}{\sigma_i}, \quad y_i^{\text{std}} = \frac{y_i - \mu_i}{\sigma_i} \quad (7)$$

with μ_i as the mean value and σ_i as the standard deviation of quantity x_i or y_i computed on the training split only.

B. Sampling Strategy

Two datasets were generated in exactly the same way but at two very different resolutions so that the influence of data scarcity could be studied systematically.

Step 1: Choose Material and Stress Parameters: All six “slow” variables [A_s , M_s , λ_s , σ_{xx} , σ_{yy} , σ_{xy}] are drawn from the LH design summarized in Table I. A single LH draw is called a *material instance* in the remainder of the article.

Step 2: Probe the Magnetic Loading Space: For every material instance, the SMSM is evaluated on a 2-D Sobol grid spanning the H plane. The reason for using Sobol sampling and not just random draws is the (deliberately chosen) low number of 10 H values between 0 and the maximum H value of 500 kA/m. This approach spreads the samples more uniformly (low discrepancy) throughout the interval and guarantees that no important regions in the H plane are missed. We generate N_ϕ equi-distributed directions and N_H magnitudes. To avoid a bias toward saturation, the magnitudes are sampled in three concentric bands as follows.

- 1) 50% of samples in $[0, 300] \text{ A/m}$.
- 2) 30% of samples in $[300, 5000] \text{ A/m}$.
- 3) 20% of samples in $[5000, 500\,000] \text{ A/m}$.

The distribution of the points is based on the change of magnetization with respect to magnetic field amplitude, which is highest in the low field region, drops in the “knee” region of the BH curve and reaches vacuum permeability (constant value) in full saturation, which is the reason why only 20% of samples are placed into the highest H band. The reason for that is to counteract saturation biasing, which means if we were to sample H uniformly, there would be more data points in the saturation region and the important linear and knee region of the typical BH curve would be underrepresented. To prevent that, the three bands were introduced. The optimal sampling, of course, requires knowledge of the output prior to the model’s evaluation, which is the reason why the bands are roughly tuned for electric steel sheet material. The directions are rotated randomly for every material instance to suppress

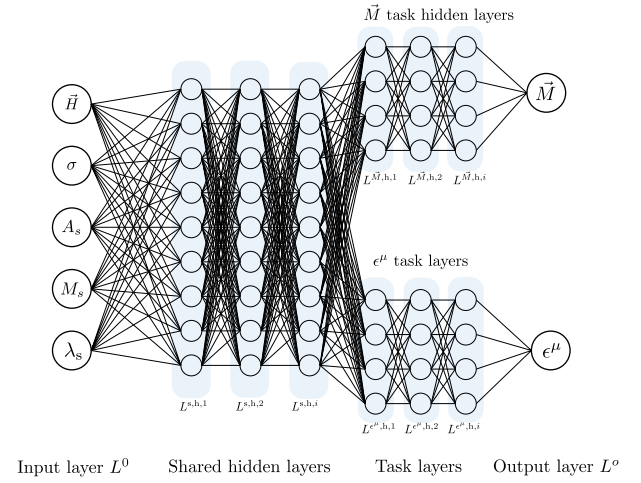


Fig. 1. Architecture of the multi-task neural network.

artificial anisotropy. During the evaluation (testing) process, special care was taken to spot signs of overfitting in certain H value regions, which could have been introduced via the banded H sampling approach but none were observed.

Terminology: Throughout the article the term *sample* denotes one 8-D input vector $\mathbf{x}$ and its 5-D target $\mathbf{y}$. The terminology *feature* is used for an individual component of $\mathbf{x}$, and *dataset* or *split* for a collection of samples such as the training set.

Large Versus Small Dataset: With $(N_\phi, N_H) = (120, 10)$ and 63 material instances the procedure yields 75 600 samples (input–output pairs). This collection is referred to as the *large dataset*. Using $(N_\phi, N_H) = (72, 10)$ and just 28 material instances produces a 2016 sample *small dataset*. Both sets are sampled independently of each other and are partitioned 80%/10%/10% into training, validation, and test subsets.

The two-sized design allows us to quantify how much each loss term in Section II-A behaves under data abundance (large set) and under severe data scarcity (small set). All ablation studies in Section III-H and FEM evaluations in Section IV, therefore, report results for both resolutions.

C. Network Architecture

The NNSMSM surrogate is implemented as a fully connected, multi-task feed-forward network. A schematic depiction can be seen in Fig. 1. The single shared backbone $f_\theta: \mathbb{R}^8 \rightarrow \mathbb{R}^{256}$ receives the 8-D input vector and is subsequently split into two task-specific branches (*heads*). The backbone uses six identical layers, each consisting of a linear projection with 256 neurons, followed by LayerNorm and a GELU (Gaussian error linear unit) activation. Layer normalization stabilizes the activations across the entire mini batch, which proved more robust than batch normalization for our batch size of 1024, while the GELU nonlinearity improves the gradient flow in low field regimes [20].

From the shared backbone, the latent features are split into two disjoint branches: the *magnetization head* f_θ^{mag} and the *strain head* f_θ^{str} . The magnetization head contains three hidden layers of 96 neurons each (Linear-LayerNorm-GELU)

and outputs a 2-neuron linear layer that predicts (M_x, M_y) . The strain head is shallower (result of the hyperparameter study described in Section III-H), containing just one layer of 128 neurons followed by a 3-neuron linear projection.

This branched design offers two advantages over a single large network. The first one is that the backbone learns a common representation of the material state, while each head can adapt individually. Furthermore, the separation prevents gradient interference between the heterogeneous output scales, which seemed to accelerate convergence compared with a monolithic architecture. In early prototyping, a monolithic variant was pursued which consistently underperformed the later developed multi-task architecture. It showed slower convergence, elevated RMSE, and it was finally replaced with the multi-task architecture. Although the multi-task network has more hyperparameters, the training, including Optuna hyperparameter optimization, was significantly faster than for the large monolithic network. These prototyping results are not included, as a fair comparison would require a full hyperparameter retuning and ablation on the monolithic baseline, which is out of scope. Since the branched model converged robustly and outperformed the monolithic variant early on, we focused our resources there.

Exact partial derivatives of the magnetization with respect to the standardized field components, $\partial \mathbf{M} / \partial \mathbf{H}$, are obtained via PyTorch's reverse mode automatic differentiation. The gradients are subsequently rescaled by the feature standard deviations to obtain the physical susceptibility tensor used in the Sobolev, consistency, and positive definiteness penalties, defined in the preceding section. Because all operations are analytic and differentiable, the resulting partial Jacobian¹ is exact up to machine precision and introduces a negligible overhead.

D. Composite Loss

The total loss $\mathcal{L}$ consists of four parts

$$\mathcal{L} = \mathcal{L}_{\text{val}} + w_{\text{Sob}} \mathcal{L}_{\text{Sob}} + w_{\text{cons}} \mathcal{L}_{\text{cons}} + w_{\text{pd}} \mathcal{L}_{\text{pd}} \quad (8)$$

where all losses are computed in *standardized* space (7) unless stated otherwise. The weights w_{Sob} , w_{cons} , and w_{pd} were obtained by the Optuna [14] hyperparameter search described in Section III-H.

Value Loss: This is the classical loss term that tries to minimize the difference between the output of the neural network and the features, given as

$$\mathcal{L}_{\text{val}} = \|\mathbf{M}_{\text{pred}}^{\text{std}} - \mathbf{M}_{\text{ref}}^{\text{std}}\|_2^2 + \|\boldsymbol{\epsilon}_{\text{pred}}^{\mu, \text{std}} - \boldsymbol{\epsilon}_{\text{ref}}^{\mu, \text{std}}\|_2^2 \quad (9)$$

where $(\cdot)_{\text{pred}}$ represents the predicted value and $(\cdot)_{\text{ref}}$ is the reference (ground truth) from the physical full fidelity SMSM operator.

Sobolev Loss (Diagonal Differential Susceptibility): This loss term acts as an additional loss term penalizing the difference between the gradients of the model and the sampled gradients. In other words, the network must fit the slope of the MH plane in two orthogonal directions (in 2-D). This aims to

reduce overfitting and a smoother interpolation between sparse field magnitudes

$$\mathcal{L}_{\text{Sob}} = \sum_{i \in \{x, y\}} \|\chi_{\text{diff}, ii, \text{pred}}^{\text{phys}} - \chi_{\text{diff}, ii, \text{ref}}^{\text{phys}}\|_2^2 \quad (10)$$

with

$$\chi_{\text{diff}, ij, \text{pred}}^{\text{phys}} = \frac{\partial M_i^{\text{std}}}{\partial H_j^{\text{std}}} \frac{\sigma_{M_i}}{\sigma_{H_j}} \quad (11)$$

where σ_{M_i} and σ_{H_i} are the standard deviations of the i th-component of the magnetization and magnetic field strength, respectively. The weight w_{Sob} is held at zero for the first 30 epochs [waiting for the first major drop in value loss (9)] and linearly ramped over the next 20 epochs, allowing the network to capture the coarse landscape before it is asked to match derivatives.

Positive-Definite Susceptibility: Ensuring $\det \boldsymbol{\chi} \geq 0$ guarantees that the magnetic energy density remains convex, which is a fundamental requirement of thermodynamics. The additional loss term eliminates spurious “negative differential permeability” spikes observed in early experiments and avoids convergence issues when using Newton solvers. For a 2×2 symmetric tensor positive definiteness is equivalent to $\chi_{xx} \geq 0$ and $\det \boldsymbol{\chi} \geq 0$. Violations are penalized via

$$\mathcal{L}_{\text{pd}} = \text{ReLU}(-\chi_{xx}^{\text{phys}}) + \text{ReLU}(-\det \boldsymbol{\chi}^{\text{phys}}). \quad (12)$$

Because the ReLU acts only on positive arguments, the term is zero once both inequalities are satisfied.

Consistency: Thermodynamics also requires symmetric susceptibility tensor entries $\chi_{ik} = \chi_{ki}$, leading to the following loss term:

$$\mathcal{L}_{\text{cons}} = \|\chi_{xy, \text{pred}}^{\text{phys}} - \chi_{yx, \text{pred}}^{\text{phys}}\|_2^2 \quad (13)$$

where χ_{ik}^{phys} is the derivative of the NNSMSM's $\vec{M}_i$ output with respect to the input magnetic field vector $\vec{H}_k$ (which would be in standardized space) transformed back to physical space because the derivatives in the standardized space do not necessarily need to be symmetric due to possible slightly different scalars of the x - and y - component of the magnetization and the magnetic field vector, respectively.

E. Optimizer and Scheduler

Training employs PyTorch's [21] AdamW optimizer [22] including weight decay (coefficient is the results of the two staged hyperparameter optimization, described in the next section, and results are summarized in Tables II and III). A ReduceLROnPlateau scheduler halves the learning rate after 50 epochs without improvement, down to a floor of 10^{-5} . To reduce the high-frequency parameter noise that AdamW may introduce, we follow PyTorch's AveragedModel strategy and keep an exponential running average of the weights.

F. Hyperparameter Optimization

The hyperparameter optimization was carried out in a two-step way using OPTUNA [14]. In the first step, the parameters in Table II were optimized. Each of the 200 trials was trained for 400 epochs on 1024-sample mini-batches.

¹Partial Jacobian, since only the partial derivative of the magnetization with respect to the magnetic field strength is considered.

TABLE II
HYPERPARAMETER OPTIMIZATION STEP 1

Parameter	Optimized value
shared neurons	256
shared depth	6
magnetic head neurons	96
magnetic head depth	3
strain head neurons	128
strain head depth	1

TABLE III
HYPERPARAMETER OPTIMIZATION STEP 2

Parameter	Optimized value
Sobolev weight	0.007394
Consistency weight	0.003784
Positive definiteness weight	0.005175
Weight decay	1.034245e-05

TABLE IV
ANHYSTERETIC PARAMETERS FOR MATERIAL 1 [23]

	M_s (A/m)	λ_s (ppm)	A_s (J/m ³)
value	1.41×10^6	6.4	1.1×10^{-2}

Once the network itself was optimized, the four scalar weights for the loss term and the weight decay were tuned. Each of the 50 trials of the second optimization step was trained for 400 epochs on 1024-sample mini-batches as well. The results of the hyperparameter optimization are given in Table III.

G. Qualitative NNSMSM Validation on Two Materials

To verify that the neural surrogate generalizes to material constants and excitations not seen during training, we reproduced the rotating field for two well-documented electrical steels whose parameters are reported in the literature and lie inside the LH domain of Table I. In this qualitative comparison, only the output of the model trained on the small dataset is tested. The comparison is established such that the output of the SMSM (full fidelity physical model) is compared to the output of the trained NNSMSM surrogate model for the same excitation. Only the three anhysteretic SMSM parameters (M_s , λ_s , A_s) were changed, while the network weights remained frozen, and no additional fine-tuning was performed. In each test, the magnetic field $\mathbf{H}$ is rotated twice through 360° while its magnitude is ramped from 0 to 5kA/m. Three loading scenarios are studied as follows.

- 1) *No Mechanical Stress*: $\boldsymbol{\sigma} = \mathbf{0}$ MPa.
- 2) *Moderate Tension*: $[\sigma_{xx}, \sigma_{yy}, \sigma_{xy}] = [80, 0, 0]$ MPa.
- 3) *50% Out-of-Distribution (OOD) Tension*: $[\sigma_{xx}, \sigma_{yy}, \sigma_{xy}] = [180, 0, 0]$ MPa.

As well as two materials (both Fe-3%Si, non-oriented, and cold rolled) as follows.

Material 1: Parameters are taken from [23] and summarized in Table IV. Figs. 2 and 3 show the surrogate and reference curves for the stress-free and the 80 MPa tensile cases, respectively. In Fig. 4, the results for the 50% OOD stress values are presented.

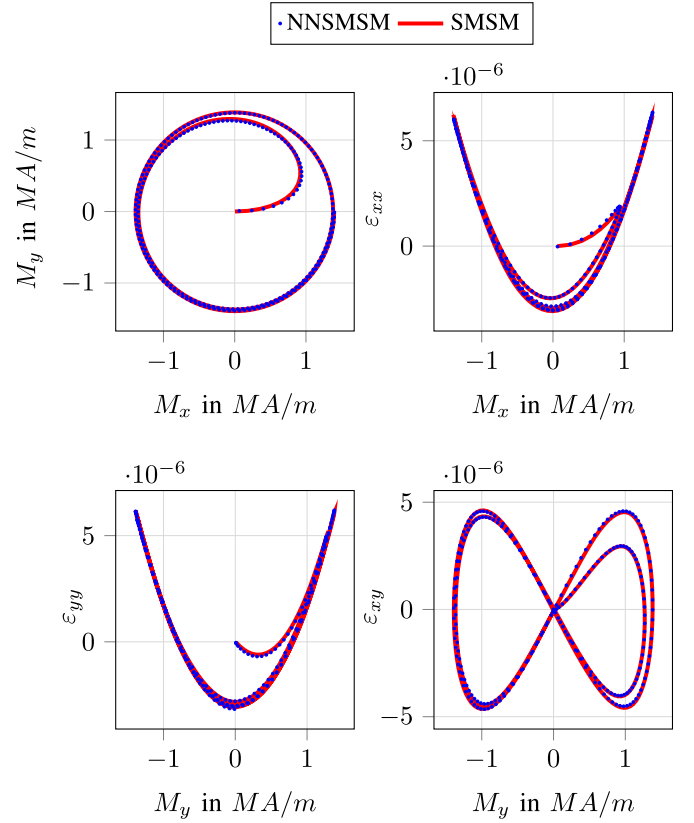


Fig. 2. Material 1, stress-free condition ($\sigma_{xx} = \sigma_{yy} = \sigma_{xy} = 0$ MPa).

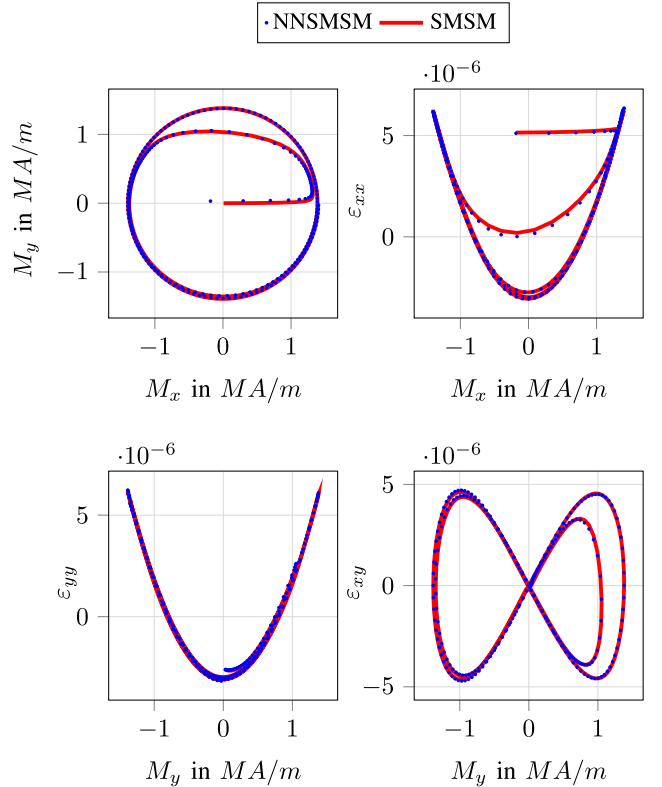


Fig. 3. Material 1 under tensile load ($\sigma_{xx} = 80$ MPa).

Material 2: Parameters are derived from [24] and the low-field susceptibility χ_0 according to [16], see Table V. The same three stress states are examined again, see Figs. 5–7.

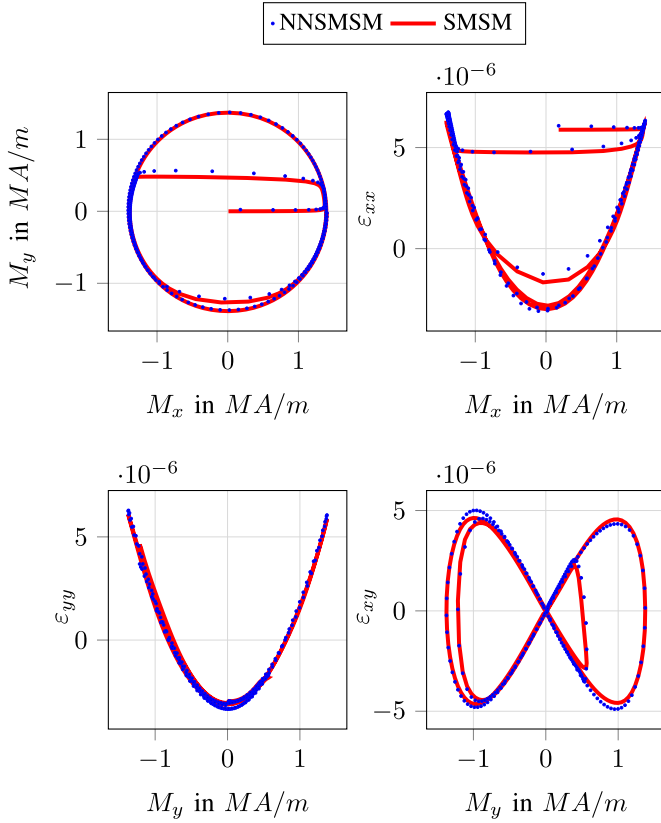
Fig. 4. Material 1 under 50% OOD tensile load ($\sigma_{xx} = 180$ MPa).

TABLE V
ANHYSTERETIC PARAMETERS FOR MATERIAL 2 [24]

	M_s (A/m)	λ_s (ppm)	A_s (J/m ³)
value	1.61×10^6	6.7	8.842×10^{-3}

The 80 MPa stress value was selected to include realistic assembly stress magnitudes reported for shrink-fit stators. Measurements and FE analyses on actual stator frame assemblies indicate circumferential compressive stresses from 10 to 20 MPa for modest material overlap [25], [26] up to 60 MPa in tighter fits or specific lab setups [27], while interface contact pressures in generic press-fit assemblies can exceed 80–130 MPa.

One can clearly see the distortion of the non-saturated magnetization regions due to the stress-induced anisotropy. The NNSMSM curves closely superimpose on the SMSM reference. Minor deviations become visible only when zooming into the “knee region” of the BH curve, where the magnetization varies most rapidly. These comparisons show (qualitatively) that even the surrogate trained on the small 2k dataset retains excellent predictive power even for entirely new parameter sets. Furthermore, the generalization and extrapolation properties are surprisingly good, as demonstrated by evaluating the NNSMSM on a 50% OOD stress value, see Figs. 4 and 7. Regarding the performance, the evaluations for the qualitative comparisons, performed in this section, consist of 1257 evaluations of the full fidelity SMSM

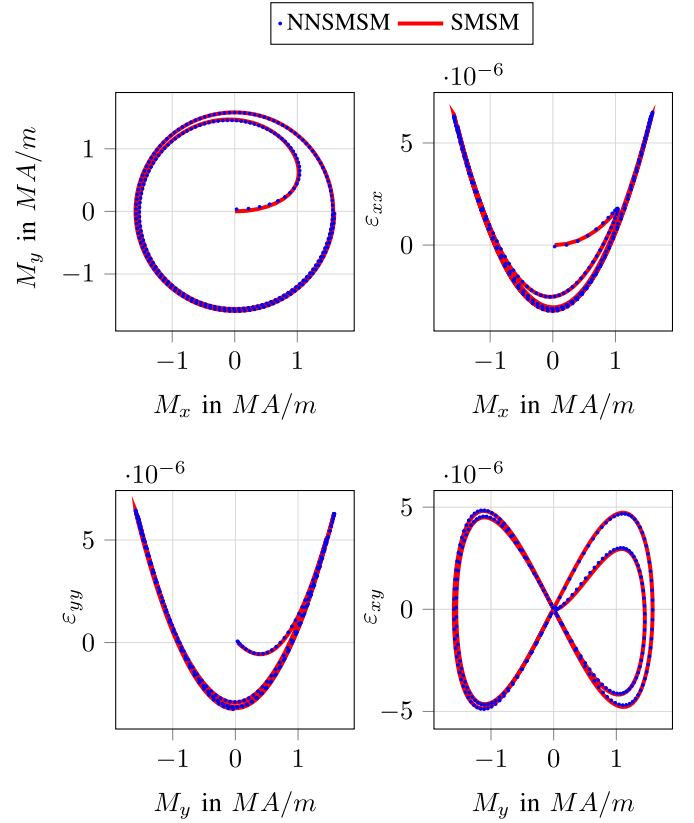
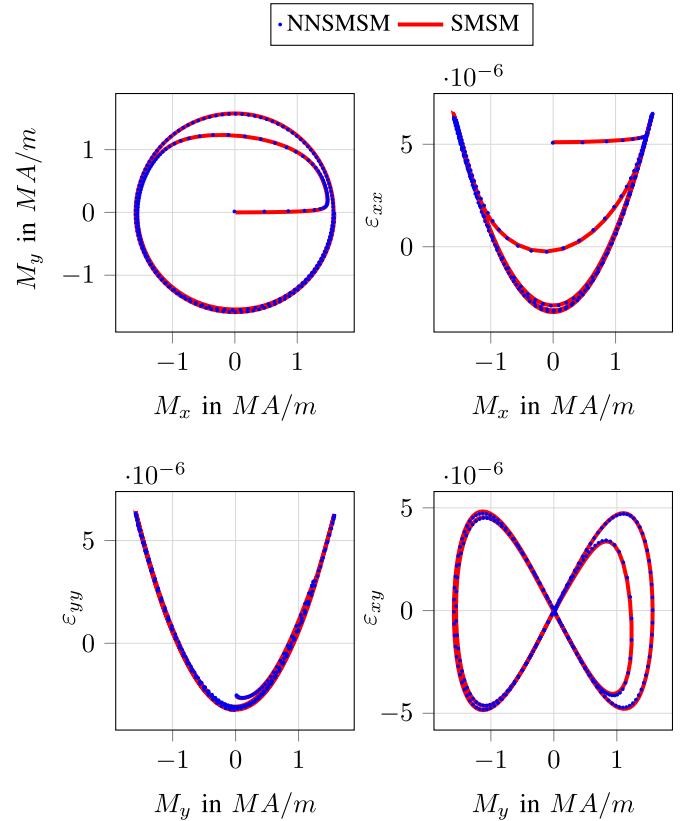
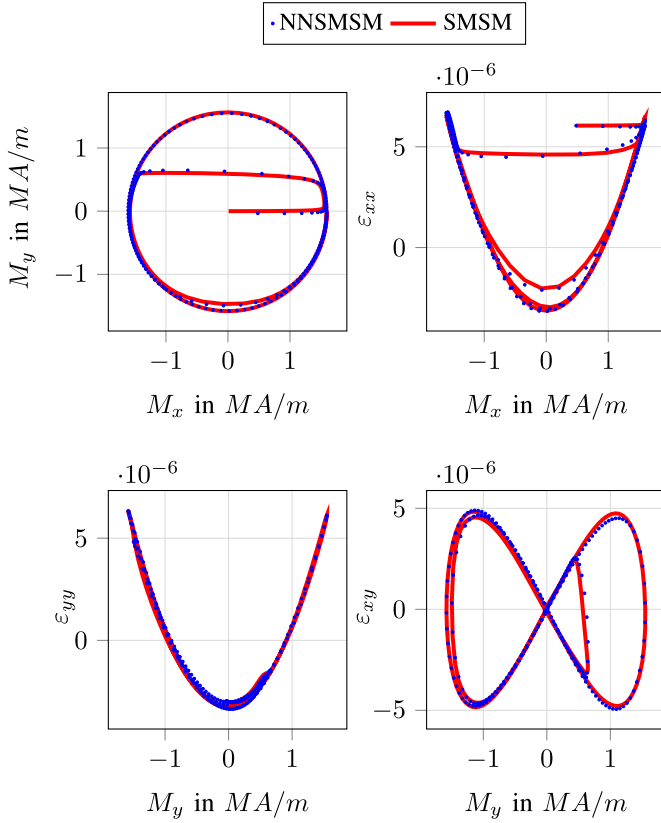


Fig. 5. Material 2, stress-free condition.

Fig. 6. Material 2 under tensile load ($\sigma_{xx} = 80$ MPa).

(25.6 s) and the NNSMSM surrogate (0.014 s), resulting in a speed-up by a factor of 1829. The excellent agreement for two

Fig. 7. Material 2 under 50% OOD tensile load ($\sigma_{xx} = 180$ MPa).

dissimilar steels, combined with the large runtime gain, confirms that NNSMSM is a reliable drop-in replacement for the analytic SMSM in engineering-scale finite element (FE) simulations, as further demonstrated in Section V.

H. Ablation Study

To evaluate the impact of every physics-informed regulariser in (8), six additional surrogates were trained that share identical architecture, optimizer, batch size, and scheduler, but with selected loss weights set to zero (see Table VI). Each network was evaluated on:

- 1) the 10% held-out test split (in-distribution) and
- 2) an OOD set created by enlarging all numerical feature ranges by 50%.

The loss weights (w_{pd} , w_{sob} , and w_{cons}) were kept identical to those obtained in the ALL_ON hyperparameter optimization (see Section III-F). This choice avoids interference of the comparison by differing optimization settings and isolates the effect of removing each physics-informed term. While it is possible that re-optimizing weights for each reduced configuration might yield slightly lower absolute errors, the qualitative impact of each constraint, e.g., increased asymmetry, is robust and does not depend on retuning.

In the following, the results of the ablation study are given for both the models trained on the large sampling set (75k samples) as well as on the smaller set (2k samples), described in Section III-B, by using the set of metrics described in the following.

TABLE VI

LOSS-WEIGHT CONFIGURATION OF THE ABLATION VARIANTS

Model tag	w_{pd}	w_{sob}	w_{cons}
ALL_ON	$\neq 0$	$\neq 0$	$\neq 0$
ALL_ZERO	0	0	0
NO_PD	0	$\neq 0$	$\neq 0$
NO_CONS	$\neq 0$	$\neq 0$	0
NO_SOB	$\neq 0$	0	$\neq 0$
NO_CONS_NO_SOB	$\neq 0$	0	0
NO_CONS_NO_PD	0	$\neq 0$	0
NO_PD_NO_SOB	0	0	$\neq 0$

Evaluation metrics: Let $\mathcal{D} = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N$ denote the standardized evaluation set. For every sample the network returns $\hat{\mathbf{y}}_n$ and the differential susceptibility $\hat{\chi}_n = \begin{bmatrix} \hat{\chi}_{xx} & \hat{\chi}_{xy} \\ \hat{\chi}_{yx} & \hat{\chi}_{yy} \end{bmatrix}_n$. Three scalar metrics are employed as follows.

- 1) *Standardized Value RMSE:*

$$\text{RMSE}_{\text{std}} = \sqrt{\frac{1}{5N} \sum_{n=1}^N \|\hat{\mathbf{y}}_n - \mathbf{y}_n\|_2^2} \quad (14)$$

averaged over the five output channels.

- 2) *Sobolev RMSE:* Relative error of the diagonal differential susceptibility

$$\text{RMSE}_{\text{Sob}} = \left[\frac{1}{2N} \sum_{n=1}^N \frac{(\hat{\chi}_{\text{diff},xx,n} - \chi_{\text{diff},xx,n})^2}{\chi_{\text{diff},xx,n}^2 + \chi_{\text{diff},yy,n}^2 + 10^{-12}} + \frac{(\hat{\chi}_{\text{diff},yy,n} - \chi_{\text{diff},yy,n})^2}{\chi_{\text{diff},xx,n}^2 + \chi_{\text{diff},yy,n}^2 + 10^{-12}} \right]^{1/2} \quad (15)$$

where the factor 10^{-12} is added in the denominator to prevent spurious blow-ups of the error due to near-zero susceptibilities, which might occur in full saturation. Due to the average value of the denominator of 10^7 , the influence of this factor is negligible.

- 3) *Consistency Error (Off-Diagonal Antisymmetry):*

$$\text{Cons} = \frac{1}{N} \sum_{n=1}^N |\hat{\chi}_{\text{diff},xy,n} - \hat{\chi}_{\text{diff},yx,n}|. \quad (16)$$

This quantifies the reciprocal-relations constraint $\chi_{\text{diff},xy} = \chi_{\text{diff},yx}$, zero means perfect symmetry.

- 4) *Positive-Definiteness Violation Rate in %:*

$$\text{PD} = 100 \frac{1}{N} \sum_{n=1}^N \mathbf{1}[\det \hat{\chi}_{\text{diff},n} < 0 \text{ or } \hat{\chi}_{\text{diff},xx,n} < 0] \quad (17)$$

where $\mathbf{1}[\cdot]$ is the indicator function.

Tables VII and VIII lists all metrics for the in-distribution and OOD evaluations of the large sample set as well as Tables IX and X for the small set. Fig. 8 visualizes the spatial error for the eight ablation study variants of the large sample set for $\vec{H}$ magnitudes in $[-100, 100]$ kA/m. The faint, periodic “spokes” in those heatmaps are the imprint of the angular sampling used to build the training set with N_ϕ equally spaced directions. Even with random rotations per material instance, the distribution over φ remains discrete, so the network seems

TABLE VII

ABLATION STUDY FOR NNSMSM'S TRAINED ON LARGE SAMPLE SET
(75K SAMPLES): IN-DISTRIBUTION METRICS (10% HOLD-OUT SET)

Variant	RMSE_std	Sob.	Cons.	PD [%]
ALL_ON	0.0148	1.093×10^5	0.0314	4.35
NO_CONS	0.0138	1.093×10^5	0.0567	4.88
NO_PD	0.0153	1.085×10^5	0.0368	5.13
NO_SOB	0.0148	1.098×10^5	0.0330	3.35
NO_CON_NO_PD	0.0137	1.095×10^5	0.0504	4.23
NO_CON_NO_SOB	0.0173	1.090×10^5	0.0616	3.34
NO_PD_NO_SOB	0.0141	1.090×10^5	0.0544	4.90
ALL_ZERO	0.0144	1.093×10^5	0.0476	4.42

TABLE VIII

ABLATION STUDY FOR NNSMSM'S TRAINED ON LARGE SAMPLE SET
(75K SAMPLES): OOD METRICS (50% ENLARGED SAMPLE RANGE)

Variant	RMSE_std	Sob.	Cons.	PD in %
ALL_ON	0.0444	6.848×10^5	0.0614	8.02
NO_CONS	0.0443	6.863×10^5	0.0711	10.30
NO_PD	0.0444	6.833×10^5	0.0663	9.40
NO_SOB	0.0444	6.854×10^5	0.0648	6.52
NO_CON_NO_PD	0.0444	6.858×10^5	0.0680	7.20
NO_CON_NO_SOB	0.0449	6.849×10^5	0.0692	9.78
NO_PD_NO_SOB	0.0445	6.858×10^5	0.0676	7.99
ALL_ZERO	0.0442	6.853×10^5	0.0645	7.09

TABLE IX

ABLATION STUDY FOR NNSMSM'S TRAINED ON SMALL SAMPLE SET
(2K SAMPLES): IN-DISTRIBUTION METRICS (10% HOLD-OUT SET)

Variant	RMSE_std	Sob.	Cons.	PD [%]
ALL_ON	0.0213	8.65×10^4	0.028	4.95
NO_CONS	0.0262	8.70×10^4	0.151	5.99
NO_PD	0.0378	8.67×10^4	0.038	5.73
NO_SOB	0.0241	8.59×10^4	0.033	7.20
NO_CON_NO_PD	0.0235	8.59×10^4	0.035	5.12
NO_CON_NO_SOB	0.0251	8.73×10^4	0.027	5.82
NO_PD_NO_SOB	0.0248	8.63×10^4	0.030	3.39
ALL_ZERO	0.0215	8.79×10^4	0.029	5.82

TABLE X

ABLATION STUDY FOR NNSMSM'S TRAINED ON SMALL SAMPLE SET
(2K SAMPLES): OOD METRICS (50% ENLARGED SAMPLE RANGE)

Variant	RMSE_std	Sob.	Cons.	PD [%]
ALL_ON	0.2096	6.10×10^4	0.109	15.2
NO_CONS	0.2318	6.28×10^4	0.113	17.0
NO_PD	0.3197	5.90×10^4	0.113	17.2
NO_SOB	0.2049	6.02×10^4	0.109	17.5
NO_CON_NO_PD	0.3115	6.12×10^4	0.108	15.0
NO_CON_NO_SOB	0.3429	5.84×10^4	0.111	13.2
NO_PD_NO_SOB	0.2363	6.03×10^4	0.113	13.2
ALL_ZERO	0.2220	5.82×10^4	0.106	17.3

to learn a weak angular prior that shows up as small, high-frequency radial artifacts. A secondary contribution may come from the SMSM reference itself, which evaluates the free energy on a fixed spherical tessellation. In contrast, the broader lobes at low to mid magnetic field strengths arise due to a directional mismatch in the learned Jacobian. This can be quantified by evaluating a directional susceptibility error

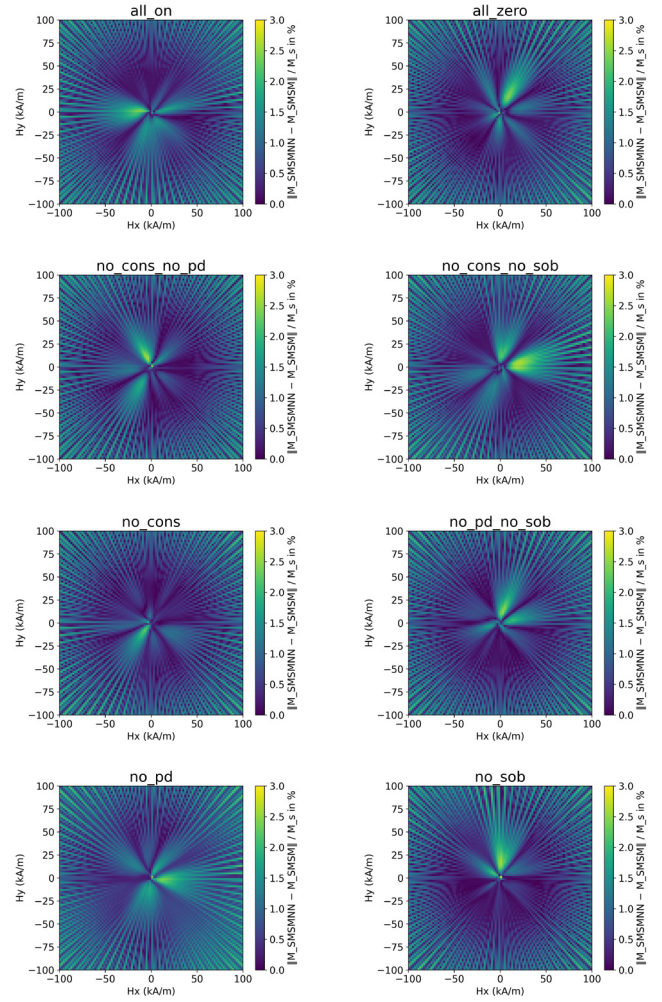


Fig. 8. Error heat-maps in the (H_x, H_y) plane for the eight model variants.

measure

$$E_{\chi}(\|\vec{H}\|_2, \varphi) := |(\hat{\chi}(\|\vec{H}\|_2, \varphi) - \chi(\|\vec{H}\|_2, \varphi)) \cdot \vec{e}_{\varphi}| \quad (18)$$

where $\hat{\chi}$ and χ are the predicted and reference magnetic susceptibility tensors evaluated at a constant magnitude of the magnetic field strength $\|\vec{H}\|_2 = \text{const.}$ and direction $\varphi \in [0, 2\pi]$, displayed in Fig. 9. Across ablation variants, the angular profiles mirror the heatmap features, confirming that the low- to mid-field lobes originate from small but systematic directional Jacobian errors, even though their absolute values remain small. Finally, it should be noted that the apparent smoothness of the ALL_ZERO variant in Fig. 9 compared with partial ablations must be interpreted with care. Since all ablations were trained with the same hyperparameters optimized for the ALL_ON case, those settings are not guaranteed to be well balanced once a constraint is removed. This explains why ALL_ZERO can look deceptively better, despite lacking the physics constraints that are essential for robustness under scarcity and extrapolation.

Tables VII–X reveal that with abundant data, the value loss already defines the landscape and auxiliary physics terms serve mainly as subtle regularisers.

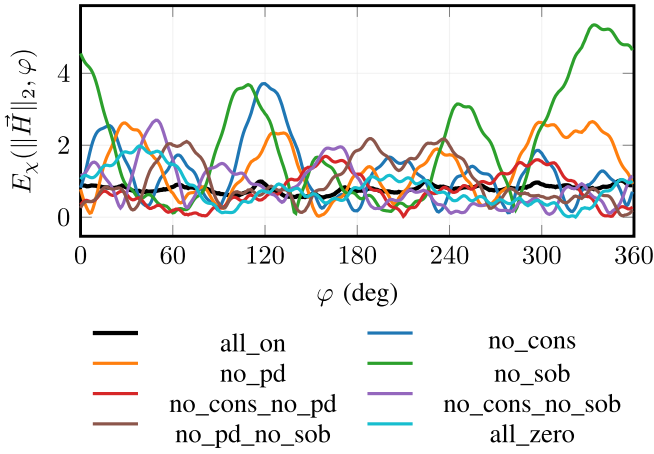


Fig. 9. Directional susceptibility error $E_\chi(\|\vec{H}\|_2, \varphi)$, exemplary for $\|\vec{H}\|_2 = 10$ kA/m, providing an explanation for the low- to mid-field lobes in Fig. 8.

In the small sample set, the observations are different. NO_PD inflates RMSE_{std} by 70%, NO_CONS multiplies the symmetry error by five, and NO_SOB drastically increases the PD-violation rate. This means that under data scarcity, the constraints replace the missing prior knowledge, and each loss term is responsible for keeping a different failure mode bounded (PD spikes, antisymmetry, and noisy susceptibility).

Concluding, it can be said that all three physics-informed terms are necessary when the training set is small or when significant extrapolation is expected. For large datasets, the lighter NO_SOB variant offers a good compromise between hyperparameter tuning complexity and local accuracy, see Table VII.

IV. INCLUDING NNSMSM+VPM IN AN FE FORMULATION

Including the trained NNSMSM models in the VPM vector hysteresis model, described in Section II, enables us to use the fast neural network surrogate model as the anhysteresis curve generator in a vector hysteresis model, which can be used in macroscopic FE simulations, depicted in Fig. 10.

In the following, the magnetic FE formulation of Fig. 10 will be introduced very briefly based on the previous works [28], [29], [30] with a special focus on discussing convergence-critical properties of the material model.

A. Hysteretic Magneto-Static Φ Formulation

For the FE simulation, the magnetic scalar potential formulation is used. In the nonlinear and hysteretic case, the starting point is to minimize the magnetic co-energy w_c with respect to the magnetic scalar potential Φ over the whole computational domain Ω

$$\arg \min_{\Phi \in H^1(\Omega)} \mathcal{F}_\Phi = \arg \min_{\Phi \in H^1(\Omega)} \int_{\Omega} w_c(\mathbf{H}_s - \Phi) dV. \quad (19)$$

The source magnetic field strength $\mathbf{H}_s$ fulfills Ampere's law

$$\nabla \times \mathbf{H}_s = \mathbf{J}_s \quad (20)$$

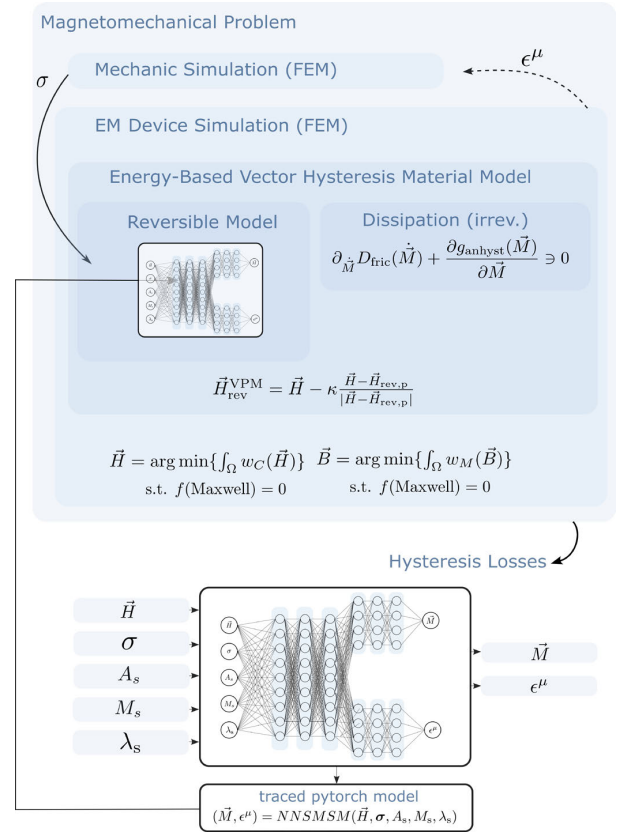


Fig. 10. Magneto-mechanical NNSMSM vector hysteresis material model in a magneto-mechanical FE simulation.

and can be computed by several methods [31]. In this work, it is solved by the 2-D penalized magnetostatic $\mathbf{H}$ formulation [32]: For a given source current density $\mathbf{J}_s$ and appropriate values for the penalty parameter ρ find $\mathbf{H}_s \in H(\text{curl}, \Omega)$ such that

$$\int_{\Omega} \mu \mathbf{H}_s \cdot \mathbf{H}'_s + \rho \nabla \times \mathbf{H}_s \cdot \nabla \times \mathbf{H}'_s dV = \int_{\Omega} \rho \mathbf{J}_s \cdot \nabla \times \mathbf{H}'_s dV \quad \forall \mathbf{H}' \in H(\text{curl}, \Omega). \quad (21)$$

The nonlinear partial differential equation equivalent to the minimization problem (19) is obtained by taking the variation with respect to the magnetic scalar potential Φ yielding the stationary points: for a given source magnetic field strength $\mathbf{H}_s$ find $\Phi \in V := \{\Phi \in H^1 \mid \Phi = \Phi_e \text{ on } \Gamma_e\}$ such that

$$\begin{aligned} \delta \mathcal{F}_\Phi[\Phi, \Phi'] &= \int_{\Omega} \frac{\partial w_c(\mathbf{H}_s - \nabla \Phi)}{\partial \mathbf{H}} \cdot \nabla \Phi' dV \\ &= \int_{\Omega} \mathbf{B}(\mathbf{H}_s - \nabla \Phi) \cdot \nabla \Phi' dV = 0 \\ \forall \Phi' \in V' &:= \{\Phi' \in H^1 \mid \Phi' = 0 \text{ on } \Gamma_e\}. \end{aligned} \quad (22)$$

The weak form (22) is then solved using the FE method in combination with the iterative Newton-Raphson (NR) method. This yields the following problem to solve in each NR iteration n :

$$\mathbf{J}^n \delta \Phi = \delta \mathcal{F}_\Phi^n \quad (23)$$

$$\Phi^{n+1} = \Phi^n + \eta \delta \Phi \quad (24)$$

with η being an appropriate value for the line search parameter. In the case of a nodal discretization in terms of FE shape

functions N , the Jacobian matrix J is computed via

$$J_{ij}^n = \int_{\Omega} \frac{\partial \mathbf{B}^n}{\partial \mathbf{H}} \cdot \nabla N_i \cdot \nabla N_j dV \quad (25)$$

where the differential magnetic permeability

$$\boldsymbol{\mu}_{\text{diff}} = \mathbf{I} + \boldsymbol{\chi}_{\text{diff}} = \partial \mathbf{B} / \partial \mathbf{H}^n \quad (26)$$

with $\mathbf{I}$ as the unit tensor, is obtained by quasi-Newton methods, e.g., the Broyden or the Broyden–Fletcher–Goldfarb–Shanno method [33]. The NNSMSM enters the formulation at exactly this location via the constitutive law

$$\begin{aligned} \mathbf{B} &= \mathbf{B}(\mathbf{H}, \boldsymbol{\sigma}) = \mu_0(\mathbf{H} + \mathbf{M}(\mathbf{H}, \boldsymbol{\sigma})) \\ &= \mu_0(\mathbf{H} + \text{NNSMSM}(\mathbf{H}, \boldsymbol{\sigma})). \end{aligned} \quad (27)$$

Since the mechanical stress state is constant during the solution of the magnetic problem, it does not introduce any additional complexity and can be considered as a constant parameter, only used in the evaluation of the NNSMSM. The right-hand side of (23) is

$$\delta \mathcal{F}_{\Phi,i}^n = \int_{\Omega} \mathbf{B}^n \cdot \nabla N_i dV. \quad (28)$$

To guarantee the convergence of the NR method, the Jacobian matrix has to be positive definite and symmetric in every iteration. This can be ensured by the following conditions:

$$\frac{\partial \mathbf{B}^n}{\partial \mathbf{H}} = \frac{\partial \mathbf{B}^{n,\top}}{\partial \mathbf{H}}, \quad \mu_1 |\boldsymbol{\xi}|^2 \leq \boldsymbol{\xi}^\top \frac{\partial \mathbf{B}^n}{\partial \mathbf{H}} \boldsymbol{\xi} \leq \mu_2 |\boldsymbol{\xi}|^2 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^2 \quad (29)$$

and $\mu_1, \mu_2 > 0$. It is realized by applying an eigenvalue decomposition to $\partial \mathbf{B} / \partial \mathbf{H}^n$ for every quadrature point and check if they are in the interval $[\mu_1 \mu_2]$. This represents an additional safety measure for guaranteed convergence, on top of the PD and reciprocity constraint in the training of the NNSMSM (17), which is enforcing exactly the same requirement. The line search parameter η is determined by minimizing (19) along the NR direction $\delta \Phi$ resulting in

$$\begin{aligned} \frac{\partial \mathcal{F}^{n+1}}{\partial \eta} &= \left(\frac{\partial \mathcal{F}^{n+1}}{\partial \Phi^{n+1}} \right) \frac{\partial \Phi^{n+1}}{\partial \eta} \\ &= \delta \mathcal{F}(\Phi^n + \eta \delta \Phi)^\top \delta \Phi \stackrel{!}{=} 0. \end{aligned} \quad (30)$$

This represents a 1-D root-finding problem, which can be solved, e.g., by Brent's method [34].

V. APPLICATION EXAMPLE: SIMULATION OF A PMSM

To evaluate the performance and accuracy of the different NNSMSM models (as described in Section III-H), all models were embedded in a macroscopic FE formulation described in the previous section. These simulations were carried out in the open-source FE software *openCFS* [11]. The simulated device is based on the cross section of a six-pole, 36-slot PMSM, using a 2-D triangular mesh, as shown in Fig. 11. Inside the permanent magnet region, a remanent flux density of $B_r = 0.6$ T is prescribed. The stator and rotor are composed of non-oriented Fe-3%Si electrical steel, with the anhysteretic parameters for the SMSM provided in Table V. The air region

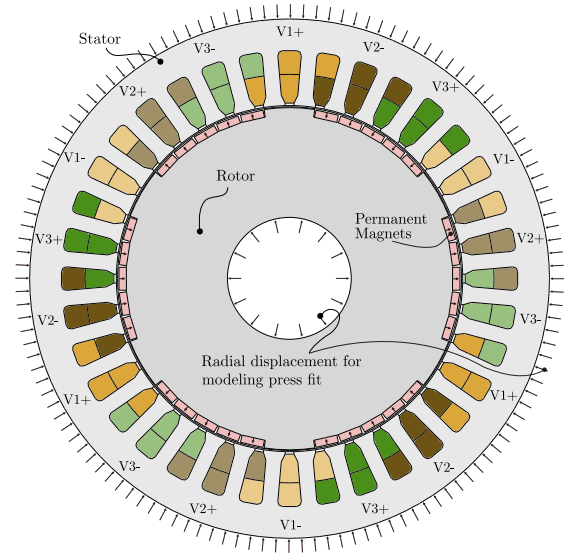


Fig. 11. Geometry of the simplified PMSM with $V_{1,2,3}$ as the three stator windings.

TABLE XI
 κ AND ω VALUES

$\kappa^{(i)}$ (A/m)	$\omega^{(i)}$
0	0.25
46.7	0.25
93.3	0.25
140.0	0.25

and the stranded coils are treated as linear with μ_0 and zero conductivity so that eddy currents are absent.

Regarding the hysteretic parameters (set of κ pinning parameters), only four pinning forces and weights were used due to a faster numerical simulation (of the FEM reference solution with SMSM+VPM), given in Table XI, which are uniformly distributed, depicting a purely isotropic dissipative behavior. In this case, the weights were chosen equally distributed and the κ parameters were obtained by fitting to uniaxial hysteresis loop measurements of a non-oriented electric steel. These dissipative parameters were used for both subsequent SMSM+VPM as well as the various NNSMSM+VPM simulations. A three-phase excitation is applied by prescribing the current density in the stator windings with a frequency of $f = 50$ Hz. The current densities in the three phases are defined as

$$\mathbf{J}_i(t) = J_i \sin(2\pi f t + \phi_i)$$

where $\phi_i \in \{0^\circ, 120^\circ, 240^\circ\}$ and $J_i = 3.468025 \cdot 10^6$ A/m². The simulation proceeds with a *pseudo* time step of $\Delta t = 0.001$ s, and it is run for 421 time steps (two full rotations of the rotor). Note that since eddy currents and rate-dependent effects are excluded, the excitation frequency does not influence the results.

The goal of this numerical experiment is to quantify the impact of choosing between the various NNSMSM versions as the anhysteretic function of the NNSMSM+VPM model at the

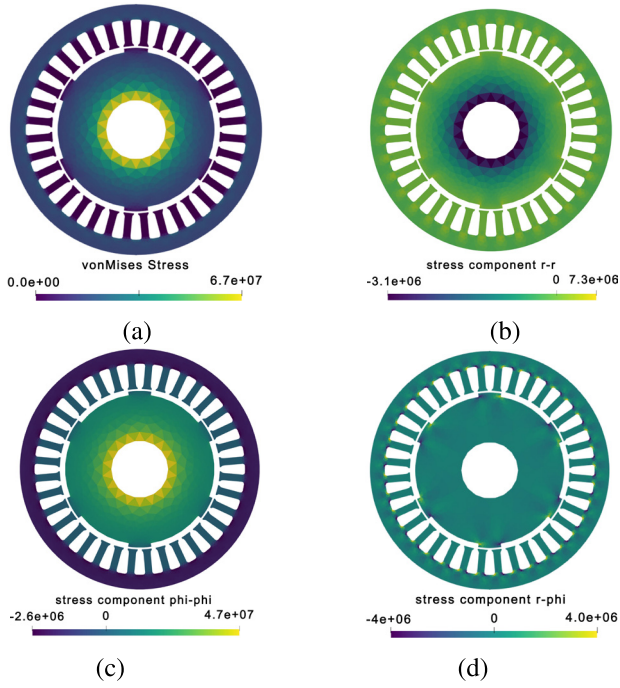


Fig. 12. Mechanical stress (in Pa) after prescribing a purely radial displacement at the rotor's inner and stator's outer edge. (a) von Mises stress. (b) σ_{rr} . (c) $\sigma_{\phi\phi}$. (d) $\sigma_{r\phi}$.

device level, extending the local material-scale comparisons presented in Section III-H. The comparison comprises the global hysteresis losses by calculating the area of the B_x - H_x and B_y - H_y loops (as described in [35]) over one cycle (one 360° rotation of the stator field), once steady-state is reached as well as the convergence metrics of the quasi Newton scheme (average iteration numbers and maximum number of iterations). The energy loss is computed element-wise and then summed over all stator and rotor elements.

For the stress state, a mechanical simulation was conducted to model a press fit of the rotor and stator sheet onto a rod or casing by prescribing a radial displacement at the rotor's inner and stator's outer edge, resulting in the stress distribution shown in Fig. 12, with a maximum von Mises stress of 67 MPa.

To ensure a fair comparison of computational performance between the SMSM+VPM (physical full fidelity material model) and the NNSMSM+VPM versions used within the FEM simulation, all nonlinear simulations were iterated until the L_2 norm of the residual reached $1 \cdot 10^{-8}$.

A. Reference Result

A two-revolution, steady state FE simulation with the full fidelity multiscale SMSM+VPM model and shrink fit stress results in the losses given in Table XII.

B. NNSMSM+VPM Surrogate FEM Results

The FE benchmarks confirm a clear distinction between the different surrogate variants and between rich and sparse training regimes. When the networks are trained on the large 75k sample set, all eight configurations reproduce the reference

TABLE XII
HYSTERESIS LOSSES FROM REFERENCE SIMULATION WITH
FULL FIDELITY MATERIAL MODEL SMSM+VPM

Variant	P_{rot} in W	P_{stat} in W	P_{tot} in W	iters avg / max
reference	1.127W	17.826W	18.953	15.5/29

TABLE XIII
ABSOLUTE HYSTERESIS LOSSES PER MECHANICAL REVOLUTION AND
NEWTON ITERATIONS FOR NNSMSM'S TRAINED ON
LARGE FEATURE SET (75K SAMPLES)

Variant	P_{rot} in W	P_{stat} in W	P_{tot} in W	iters avg / max
ALL_ON	1.080	17.604	18.683	16.0 / 31
NO_CONS	1.063	17.562	18.626	15.9 / 28
NO_PD	1.054	17.536	18.589	16.1 / 28
NO_SOB	1.105	17.591	18.696	16.1 / 27
NO_CON_NO_PD	1.062	17.596	18.658	16.2 / 32
NO_CON_NO_SOB	1.155	17.589	18.744	16.2 / 34
NO_PD_NO_SOB	1.061	17.539	18.599	16.1 / 33
ALL_ZERO	1.052	17.619	18.671	16.1 / 27

TABLE XIV
RELATIVE ERROR WITH RESPECT TO THE REFERENCE FOR NNSMSM'S
TRAINED ON LARGE FEATURE SET (75K SAMPLES)

Variant	ΔP_{rot} in %	ΔP_{stat} in %	ΔP_{tot} in %
ALL_ON	+4.26	+1.25	+1.43
NO_CONS	+5.70	+1.48	+1.73
NO_PD	+6.57	+1.63	+1.92
NO_SOB	+2.02	+1.31	+1.36
NO_CON_NO_PD	+5.79	+1.29	+1.56
NO_CON_NO_SOB	-2.40	+1.33	+1.11
NO_PD_NO_SOB	+5.94	+1.61	+1.87
ALL_ZERO	+6.70	+1.16	+1.49

hysteresis losses to <2% accuracy, see Tables XIII and XIV. The spread in total loss error stays below two percent, and the average Newton iteration count rises by no more than one iteration compared to the reference SMSM+VPM simulation. Once the data for training the models is reduced to 2k samples, the picture changes abruptly, see Tables XV and XVI. The unconstrained networks introduce errors in global hysteresis losses of up to 423% in the rotor and up to 29% in total losses (stator and rotor), a bias large enough to make the non-physically constrained networks unfeasible for FEM loss calculations. Only the ALL_ON version is able to keep the RMSE relatively low, resulting in a <2% total deviation of hysteresis losses compared to the reference result.

Regarding the speedup, all NNSMSM+VPM FEM simulations provide a reduction of simulation wall clock time of roughly a factor of 11, see Table XVII (all simulations were carried out on an Apple Mac M3 Pro). The training time for the NNSMSM amounted to around 6 h for the large feature set and around 20 min for the small feature set.

Note: Relative errors are expressed as $(P_{\text{NN}} - P_{\text{ref}})/P_{\text{ref}} \times 100\%$; negative values indicate an under-prediction by the surrogate.

TABLE XV
ABSOLUTE HYSTERESIS LOSSES PER MECHANICAL REVOLUTION AND
NEWTON ITERATIONS FOR NNSMSM'S TRAINED ON
SMALL FEATURE SET (2K SAMPLES)

Variant	P_{rot} in W	P_{stat} in W	P_{tot} in W	iters avg / max
ALL_ON	1.064	17.520	18.584	16.21 / 31
NO_CONS	1.904	17.533	19.437	16.33 / 27
NO_PD	1.854	17.536	19.389	16.10 / 28
NO_SOB	2.203	17.603	19.806	16.37 / 31
NO_CON_NO_PD	2.441	17.494	19.936	16.60 / 34
NO_PD_NO_SOB	2.742	17.481	20.222	16.04 / 31
NO_CON_NO_SOB	6.900	17.555	24.455	16.22 / 26
ALL_ZERO	5.898	17.615	23.513	16.32 / 28

TABLE XVI
RELATIVE ERROR WITH RESPECT TO THE REFERENCE FOR NNSMSM'S
TRAINED ON SMALL FEATURE SET (2K SAMPLES)

Variant	ΔP_{rot} in %	ΔP_{stat} in %	ΔP_{tot} in %
ALL_ON	-5.62	-1.71	-1.95
NO_CONS	+68.8	-1.64	+2.55
NO_PD	+64.4	-1.63	+2.30
NO_SOB	+95.4	-1.25	+4.50
NO_CON_NO_PD	+116.5	-1.86	+5.18
NO_PD_NO_SOB	+143.1	-1.93	+6.70
NO_CON_NO_SOB	+511.9	-1.52	+29.0
ALL_ZERO	+423.1	-1.18	+24.1

TABLE XVII
WALL CLOCK TIMES FOR THE DIFFERENT MODELS AND 421 TIME
STEPS (TWO FULL ROTATIONS OF THE ROTOR)

Variant	t_{WC} in min
reference SMSM+VPM	311.34
NNSMSM+VPM ALL_ON	28.45
NNSMSM+VPM NO_CONS	27.94
NNSMSM+VPM NO_PD	28.78
NNSMSM+VPM NO_SOB	28.67
NNSMSM+VPM NO_CON_NO_PD	28.67
NNSMSM+VPM NO_PD_NO_SOB	28.92
NNSMSM+VPM NO_CON_NO_SOB	28.23
NNSMSM+VPM ALL_ZERO	28.03

VI. CONCLUSION

The various simulations and comparisons in the work show that physics-guided regularization is indispensable whenever data is scarce and merely advantageous when training data is increased. With 75k samples, the shared backbone already sees enough patterns to capture the coarse anhysteretic landscape. Introducing individual penalties decreases the (local as well as global) error only slightly. After the training set is decreased to 2k points, the value loss can no longer dominate the stochastic gradient, and networks that lack either the positive definiteness hinge, the reciprocity filter, or the Sobolev slope control begin to overfit isolated directions. This seems to influence the hysteresis losses in the rotor more severely than in the stator, most likely due to the already strong bias field from the permanent magnets, shifting the operation point into the knee region of the anhysteresis curve and therefore, making it more difficult to capture the fast magnetization changes. The NNSMSM version, including all physics-informed, by contrast, keeps both local and global deviations within single-digit percentages

and maintains its runtime advantage, a cost-benefit balance that makes it a great fit for transient magneto-mechanical simulation of electrical machines.

For routine design loops, the NO_SOB variant offers a good trade-off between training cost and robustness, provided the dataset is large. If only one physics-informed loss term is kept, the parameter space of the hyperparameter search is also decreased, making the model easier to train. For high fidelity transient studies, optimization tasks with repeated extrapolation, or any workflow that must remain stable under various stress states and strong permanent magnet bias, the composite ALL_ON surrogate is the only reliable choice.

Future work will explore replacing the simplified SMSM by the full multiscale model [16]. The data generation is conceptually straightforward and would allow one to include more information about the material behavior, like anisotropy and texture effects, at the cost of longer offline sampling time. A natural next step is to surrogate the entire vector hysteresis operator. That goal is significantly harder than just replacing the anhysteretic model, because hysteresis is path-dependent. The magnetic response (magnetization) at the next step depends not only on the current field but also on an internal state that stores, for example, the previous reversible field or magnetization. Capturing such vector memory would require either augmenting the surrogate with explicit state variables that are carried through the FE quasi-time steps, or using a recurrent architecture, e.g., a gated RNN or LSTM, which learns the hidden state while respecting thermodynamic constraints. Preliminary tests with vanilla RNN/LSTM models trained only on input-output relations delivered little net speed-up because, in a hysteretic FE simulation, the network sees just the current $\mathbf{H}$ and previous state and outputs the new $\mathbf{M}$. That step-by-step inference erodes the potential performance gain compared to a whole time-series of input values for the RNN/LSTM. Developing a recurrent surrogate that is both memory-efficient and computationally fast for the energy-based vector hysteresis operator in FE simulations remains a significant challenge. Recent advances in neural operator approaches for representing magnetic hysteresis [36], [37] suggest a promising direction. Finally, moving to 3-D simulations mainly requires extending the input and output vectors/tensors and adapting the loss terms accordingly, making 3-D surrogate training an attractive next step.

REFERENCES

- [1] Y. Kai, Y. Tsuchida, T. Todaka, and M. Enokizono, "Evaluation of local residual stress distribution of stator core in a rotating machine," *Electr. Eng. Jpn.*, vol. 181, no. 3, pp. 1–8, Jul. 2012.
- [2] P. Baudouin, A. Belhadj, F. Breaban, A. Deffontaine, and Y. Houbart, "Effects of laser and mechanical cutting modes on the magnetic properties of low and medium Si content nonoriented electrical steels," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 3213–3215, Sep. 2002.
- [3] R. Sundaria, A. Hemeida, A. Arkio, A. Daem, P. Sergeant, and A. Belahcen, "Effect of different cutting techniques on magnetic properties of grain oriented steel sheets and axial flux machines," in *Proc. 45th Annu. Conf. IEEE Ind. Electron. Soc.*, vol. 1, Oct. 2019, pp. 1022–1027.
- [4] K. Roppert, M. Kaltenbacher, L. Domenig, and L. Daniel, "Magneto-elastic vector hysteresis modelling for electromagnetic devices: A combination of a multiscale model with the energy-based hysteresis framework," *IEEE Trans. Magn.*, vol. 61, no. 8, Aug. 2025, Art. no. 7300414. [Online]. Available: <https://ieeexplore.ieee.org/document/11062564>

- [5] A. Sauseng, M. Kaltenbacher, and K. Roppert, "Revisiting the dry friction-like magnetic vector hysteresis model," *J. Magn. Magn. Mater.*, vol. 604, Aug. 2024, Art. no. 172285.
- [6] H. Egger, F. Engertsberger, L. Domenig, K. Roppert, and M. Kaltenbacher, "On energy consistent vector hysteresis operators," 2024, *arXiv:2410.11705*.
- [7] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, Jul. 1997, doi: [10.1023/a:1007379606734](https://doi.org/10.1023/a:1007379606734).
- [8] L. G. D. Silva, A. Abderahmane, M. Domenjoud, L. Bernard, and L. Daniel, "An extension of the vector-play model to the case of magneto-elastic loadings," *IEEE Access*, vol. 10, pp. 126674–126686, 2022.
- [9] M. Hillgärtner, K. Linka, K. P. Abdolazizi, R. C. Aydin, M. Itskov, and C. J. Cyron, "Constitutive artificial neural networks: A general anisotropic constitutive modeling framework utilizing machine learning," in *Proc. Appl. Math. Mech.*, vol. 21, 2021, Paper e202100072. [Online]. Available: <https://doi.org/10.1002/pamm.202100072>
- [10] A. Marques, A. Pereira, B. Ribeiro, and P. A. Prates, "On the identification of material constitutive model parameters using machine learning algorithms," *Key Eng. Mater.*, vol. 926, pp. 2146–2153, Jul. 2022, doi: [10.4028/p-5hf550](https://doi.org/10.4028/p-5hf550).
- [11] M. Kaltenbacher, *Numerical Simulation of Mechatronic Sensors and Actuators: Finite Elements for Computational Multiphysics*. Berlin, Germany: Springer, 2015.
- [12] L. Daniel, O. Hubert, and M. Mekik, "A simplified 3-D constitutive law for magnetomechanical behavior," *IEEE Trans. Magn.*, vol. 51, no. 3, pp. 1–4, Mar. 2015.
- [13] L. Bernard, B. J. Mailhé, N. Sadowski, N. J. Batistela, and L. Daniel, "Multiscale approaches for magneto-elasticity in device simulation," *J. Magn. Magn. Mater.*, vol. 487, Oct. 2019, Art. no. 165241.
- [14] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," 2019, *arXiv:1907.10902*.
- [15] L. G. da Silva, L. Bernard, M. Domenjoud, and L. Daniel, "A magneto-elastic vector-play model including piezomagnetic behavior," *J. Magn. Magn. Mater.*, vol. 609, Nov. 2024, Art. no. 172439.
- [16] L. Daniel, O. Hubert, N. Buiro, and R. Billardon, "Reversible magneto-elastic behavior: A multiscale approach," *J. Mech. Phys. Solids*, vol. 56, no. 3, pp. 1018–1042, Mar. 2008.
- [17] T. Matsuo and M. Shimasaki, "Isotropic vector hysteresis represented by superposition of stop hysteron models," *IEEE Trans. Magn.*, vol. 37, no. 5, pp. 3357–3361, May 2001.
- [18] T. Matsuo, "Rotational saturation properties of isotropic vector hysteresis models using vectorized stop and play hystérons," *IEEE Trans. Magn.*, vol. 44, no. 11, pp. 3185–3188, Nov. 2008.
- [19] R. Scorretti and F. Sixdenier, "An analytical formula to identify the parameters of the energy-based hysteresis model," *J. Magn. Magn. Mater.*, vol. 548, Apr. 2022, Art. no. 168748.
- [20] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.
- [21] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. NIPS*, Vancouver, BC, Canada, Dec. 2019, pp. 8024–8035. [Online]. Available: <https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>
- [22] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
- [23] L. G. da Silva, "Energy-based models for the magneto-elastic behavior of ferromagnetic materials," Ph.D. dissertation, Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, Brazil, 2024.
- [24] O. Hubert, "Multiscale magneto-elastic modeling of magnetic materials including isotropic second order stress effect," *J. Magn. Magn. Mater.*, vol. 491, Dec. 2019, Art. no. 165564, doi: [10.1016/j.jmmm.2019.165564](https://doi.org/10.1016/j.jmmm.2019.165564).
- [25] N. Simpson, R. Wrobel, J. D. Booker, and P. H. Mellor, "Multi-physics experimental investigation into stator-housing contact interface," in *Proc. 8th IET Int. Conf. Power Electron., Mach. Drives (PEMD)*, Apr. 2016, pp. 1–6.
- [26] M. Oka and M. Enokizono, "Evaluation of frame-induced compressive stress on the magnetic properties of stator cores using the excitation inner core method," *J. Electr. Eng.*, vol. 69, no. 6, pp. 477–480, Dec. 2018.
- [27] R. Raj, T. Lindberg, Z. Lin, D. Singh, K. Bergsro, and T. Thiringer, "Characterizing the effect of shrink fitting of stator housings on electric vehicle performance," *IEEE Trans. Energy Convers.*, pp. 1–11, Apr. 2025.
- [28] L. D. Domenig, K. Roppert, and M. Kaltenbacher, "Incorporation of a 3-D energy-based vector hysteresis model into the finite element method using a reduced scalar potential formulation," *IEEE Trans. Magn.*, vol. 60, no. 6, pp. 1–8, Jun. 2024.
- [29] H. Egger, F. Engertsberger, L. Domenig, K. Roppert, and M. Kaltenbacher, "On nonlinear magnetic field solvers using local quasi-Newton updates," *Comput. Math. Appl.*, vol. 183, pp. 20–31, Apr. 2025, doi: [10.1016/j.camwa.2025.01.033](https://doi.org/10.1016/j.camwa.2025.01.033).
- [30] H. Egger, F. Engertsberger, L. Domenig, K. Roppert, and M. Kaltenbacher, "On forward and inverse energy-based magnetic vector hysteresis operators," *IEEE Trans. Magn.*, vol. 61, no. 4, pp. 1–7, Apr. 2025, doi: [10.1109/TMAG.2025.3544507](https://doi.org/10.1109/TMAG.2025.3544507).
- [31] M. Kuczmann, "The Finite Element Method in Magnetics," Akadémiai Kiadó Member Wolters Kluwer Group, Budapest, Hungary, 2008.
- [32] B. Bandelier, C. Daveau, and F. Rioux-Damida, "An h -formulation for the computation of magnetostatic fields. Implementation by combining a finite element method and a boundary element method," *J. de Phys. III*, vol. 3, no. 5, pp. 995–1004, May 1993.
- [33] P. Deufhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms* (Springer Series in Computational Mathematics), vol. 35. Berlin, Germany: Springer, 2011.
- [34] R. P. Brent, *Algorithms for Minimization Without Derivatives* (Dover Books on Mathematics). Minealoe, NY, USA: Dover Publications, Mar. 2013.
- [35] G. Mörée and M. Leijon, "Iron loss models: A review of simplified models of magnetization losses in electrical machines," *J. Magn. Magn. Mater.*, vol. 609, Nov. 2024, Art. no. 172163.
- [36] Z. Guo, B. H. Nguyen, H. Hamzehbahmani, and R. V. Sabariego, "Dynamic hysteresis model of grain-oriented ferromagnetic material using neural operators," 2025, *arXiv:2504.04863*.
- [37] A. Chandra, B. Daniels, M. Curti, K. Tiels, and E. A. Lomonova, "Magnetic hysteresis modeling with neural operators," *IEEE Trans. Magn.*, vol. 61, no. 1, pp. 1–11, Jan. 2025, doi: [10.1109/TMAG.2024.3496695](https://doi.org/10.1109/TMAG.2024.3496695).