HARDCORE: H-Field and Power Loss Estimation for Arbitrary Waveforms With Residual, Dilated Convolutional Neural Networks in Ferrite Cores

Wilhelm Kirchgässner, *Member, IEEE*, Nikolas Förster, Till Piepenbrock, Oliver Schweins, and Oliver Wallscheid, *Senior Member, IEEE*

***Abstract*—The MagNet challenge 2023 called upon competitors to develop data-driven models for the material-specific, waveform-agnostic estimation of steady-state power losses in toroidal ferrite cores. The following HARDCORE (H-field and power loss estimation for arbitrary waveforms with residual, dilated convolutional neural networks in ferrite cores) approach shows that a residual convolutional neural network with physics-informed extensions can serve this task efficiently when trained on observational data beforehand. One key element is an intermediate model layer, which first reconstructs the BH curve, and then, estimates the power losses based on the curve's area rendering the proposed topology physically interpretable. In addition, emphasis was placed on expert-based feature engineering and information-rich inputs in order to enable a lean model architecture. A model is trained from scratch for each material, while the topology remains the same. A Pareto-style tradeoff between model size and estimation accuracy is demonstrated, which yields an optimum at as low as 906 parameters and down to below 8% for the average 95th percentile of the relative power loss error across diverse materials. This contribution has won the first place in the performance category of the MagNet challenge 2023, which further highlights the effectiveness of the proposed model.**

***Index Terms*—Convolutional neural network (CNN), machine learning (ML), magnetics.**

I. Introduction

VIRTUALLY all power electronic devices depend to some extent on magnetic circuits—be it in consumer hardware, such as mobile applications, or industrial power sources. The ever-increasing miniaturization of electronic components and chips on system boards renders thermal and electromagnetic aspects an emerging challenge. A high precision in manufacturing and layout during the design phase becomes more and more sought-after, but the mathematical modeling foundation behind power magnetics lacks rigorousness. The common practice is, thus, often marked by overdimensioning and large material margins for meeting the requirements reliably.

A. Related Work

Driven by this design issue, the power loss modeling of soft-magnetic materials has been the subject of increasingly active research. The basic Steinmetz equation [1], [2] describes hysteresis and eddy current losses with sinusoidal B-fields for a fixed temperature, and its publication ranges back to 1894. This approach has been extended, only using three parameters, for nonsinusoidal waveforms with subloops by the modified Steinmetz equation [3], the generalized Steinmetz equation (GSE) [4], and the improved generalized Steinmetz equation (iGSE) [5] from 1996 to 2002.

The improved-improved generalized Steinmetz equation (i2GSE) [6], [7] published in 2012 considers the magnetic relaxation effect, for increased accuracy in case of zero-voltage periods. Therefore, new material parameters are introduced.

The latest low-dimensional, physics-derived core loss model is the Stenglein method [8] introduced in 2021. This approach uses a quasi-static hysteresis energy loss at low frequencies together with a frequency and rate-dependent term. Just a few measurements are necessary to implement this approach claiming an accuracy of 15% mean-squared error (MSE), which can be considered the best-in-class accuracy for very low-dimensional, physics-driven fitting models. However, previous studies of the Stenglein model only covers a rather small dataset while the application to a comprehensive dataset addressing various materials and operation conditions has not yet been considered.

Hence, as part of a preinvestigation for this contribution, the six parameters of the Stenglein equation have been fitted to the given MagNet dataset (cf., next section), together with a standard temperature correction factor approach (additional three parameters) according to [9]. Note that fitting these parameters on real measurement data is more challenging than using ideal waveforms as the second derivative in the Stenglein equation contains a lot of noise, which needs to be filtered accordingly. The found results are summarized in Table I. Here, it can be noted that the Stenglein model's performance is similar compared to the iGSE approach from [10] reporting 95th quantile relative

Received 18 April 2024; revised 8 August 2024 and 16 September 2024; accepted 15 October 2024. Date of publication 30 October 2024; date of current version 18 December 2024. Recommended for publication by Associate Editor M. Shen. *(Corresponding author: Wilhelm Kirchgässner.)*

Wilhelm Kirchgässner, Nikolas Förster, Till Piepenbrock, and Oliver Schweins are with the Department of Power Electronics and Electrical Drives, Paderborn University, 33095 Paderborn, Germany (e-mail: kirchgaessner@lea.upb.de; foerster@lea.upb.de; piepenbrock@lea.upb.de).

Oliver Wallscheid is with the Chair of Interconnected Automation Systems, University of Siegen, 57076 Siegen, Germany (e-mail: oliver.wallscheid@uni-siegen.de).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2024.3488174.

Digital Object Identifier 10.1109/TPEL.2024.3488174

TABLE I
STENGLEIN METHOD RELATIVE ERRORS ON THE FIVE FINAL MAGNET CHALLENGE MATERIALS

	3C92	T37	3C95	79	ML95S
Training data	2432	7400	5357	580	2013
Test data	7651	3172	5357	7299	3738
Model parameters	9	9	9	9	9
Avg. relative error	23.8 %	15.2 %	18.2 %	23.7 %	16.8 %
95th-quantile rel. err.	55.2 %	40.1 %	52.4 %	76.9 %	44.6 %

errors in the range of 17%...47% (using a similar dataset). These observations lead to the conclusion that conventional, physics-informed fitting models reach their limits when applied to comprehensive datasets covering multiple magnet materials at various operation conditions. This motivates the need of machine learning (ML) assisted models whose flexibility and adaptability allow additional degrees of freedom in this challenging modeling task.

Recently, the publication of a comprehensive dataset with hysteresis loss measurements [11], [12], [13] has spurred ML approaches to this problem [14]. Among those, a power loss estimation based on a long short-term memory (LSTM) encoder–decoder [15] is proposed, and a transformer-based architecture is shown in [16] and [17]. These methods could improve the power loss estimation accuracy by relying on a vast amount of model coefficients and an according surge in computational demand for processing any given B sequence. Also, the considered ML-models are black-box approaches, i.e., lacking interpretability since available domain knowledge is abandoned. Another particular drawback of the LSTM and transformer architecture in those publications is the susceptibility to phase shifts of the B sequence, which is eliminated in this article. In summary, the well-known and challenging issue of accurate core losses for arbitrary waveforms remains common for both analytical as well as ML modeling.

B. MagNet Challenge 2023

In order to further foster research and collaboration in this field, the MagNet Challenge 2023 had been launched [18]. The competition hosts provided measurement datasets for 15 materials in total, where ten materials were fully disclosed including manufacturer, type, data sheets, and all target features (3C94, 78, N30, 3E6, 3F4, N87, 3C90, N49, N27, and 77). The task was to estimate the steady-state volumetric power loss p in $\mathrm{W m^{-3}}$ or, alternatively, the magnetic field strength curve H in $\mathrm{A/m}$ from frequency f, temperature T, and a magnetic flux density curve B of $M = 1024$ time steps per period (regardless the frequency). A wide range of operation points were measured covering sine, trapezoidal, and square waveforms, as well as ringing due to a high switching speed of the used semiconductors. The frequency varies from 50 to 800 kHz, the temperature from 25 °C to 90 °C, and the volumetric power losses range up to 14.66 $\mathrm{MW m^{-3}}$. An in-depth description of the dataset, the occurring waveforms, the data acquisition process, the lab setup, and the data quality control can be found in [14].

After the participating teams had familiarized themselves with the data of the initial ten materials and the identification task for a few months, they were provided with another five anonymized materials (3C92, T37, 3C95, 79, and ML95S) on which basis teams were scored in the challenge. The data behind those final five materials posed particular challenges stemming from the outlined training and test data split. Material 3C92 has a scarce temperature value range with only few 70 °C samples in the training dataset. Material T37, on the other hand, denotes a special material for EMI filters, unlike all other materials, which are used in power electronics inductor and transformer design. For material 79, there are very few trapezoidal waveforms available, and material ML95S has training data with high B and f values at the same time, which are not present in other materials. Other participants of the challenge considered various modeling solutions, such as low-dimensional Steinmetz extensions [10] and neural network-driven solutions including special features derived from the Landau–Lifshitz–Gilbert equation [19] or Fourier analysis [20] as well as deep black-box architectures such as generative adversarial networks [21]. While those techniques come with their individual advantages and disadvantages, none of them could reach the estimation accuracy to model complexity ratio as the subsequently introduced HARDCORE [H-field and power loss estimation for arbitrary waveforms with residual, dilated convolutional neural networks (CNNs) in ferrite cores] approach.

C. Own Contribution: The HARDCORE Approach

Compared to the general state of the art and other MagNet challenge submissions, the proposed HARDCORE approach introduces the following innovations to the field.

1) Reconstruction of the entire BH curves using a specialized CNN topology.
2) Interpretability based on an intermediate power loss estimate using the BH curve area, that is, a physics-informed ML extension leveraging domain knowledge.
3) Material-agnostic approach not relying on pre-training (from other material data), that is, a highly generalizable approach.
4) Scalability due to a Pareto optimization identifying the best compromise between power loss estimation accuracy and model complexity.
5) Best in class model accuracy for a given model size among all MagNet challenge submissions.
6) Open-source release of the HARDCORE algorithm available online.[1]
7) Ready-to-use Python package to run the HARDCORE model results available online.[2]

With a proposed model covering only 906 parameters, an unparalleled estimation performance achieving a relative power loss error of less than 8% for the average 95th percentile across diverse materials paved the way toward the MagNet challenge performance award.

[1] https://github.com/upb-lea/hardcore-magnet-challenge
[2] https://github.com/upb-lea/mag-net-hub

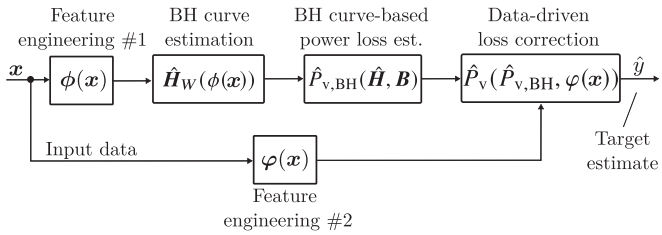


Fig. 1. Overview of the physics-inspired HARDCORE modeling toolchain.

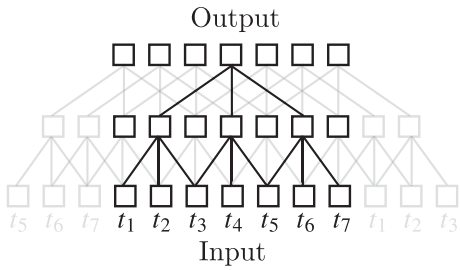


Fig. 2. Exemplary 1-D CNN with two hidden layers and circular padding in grey is shown. The second layer applies dilation.

II. Model Description

A residual CNN with physics-informed extensions is utilized for all materials. Such a CNN is trained for each material from scratch. Yet, the topology is unaltered across materials, signal waveforms, or other input data particularities.

A high-level view on the proposed residual, physics-inspired modeling toolchain is depicted in Fig. 1, which is named the HARDCORE approach. Here, $\phi(\cdot)$ represents feature engineering or a lifting function that lifts both scalar and time-series features into a richer, higher dimensional input tensor, where the input data x is fed into. Another lifting function is represented by $\varphi(\cdot)$ that utilizes the scalar feature set only. Given these features, an H estimation function approximator is denoted by $\hat{H}_{\boldsymbol{W}}(\cdot)$, which is used together with the given B sequence to calculate the area of the BH curve. This preliminary power loss estimate is ultimately corrected by a follow-up, residual correction function $\hat{P}_{\boldsymbol{v}}(\cdot)$. The details of these modeling steps are introduced in the following.

A. One-Dimensional CNNs for H-Sequence Estimation

A 1-D CNN is the fundamental building block in the HARDCORE approach. CNNs are usually employed for image processing tasks but have found their way into 1-D and 3-D problem spaces such as natural language processing and video classification ever since due to their property of utilizing local patterns [22], [23]. A 1-D CNN consists of multiple trainable kernels or filters per layer slided over the multidimensional input sequence in order to produce an activation on the following layer [22]. These activations denote the convolution (more precisely, the cross-correlation) between the learnable kernels and the previous layer's activation (or input sequence).

In this stateless architecture, circular padding ensures that subsequent activation maps are of equal size. In circular padding, samples inside a filter's view but beyond the sequence period are simply padded with samples at the other end of the sequence period, much like reading a ring buffer. Circular padding can be utilized here instead of the common zero-padding as sequences denote complete periods of the B and H curve during steady state. Moreover, a kernel does not need to read strictly adjacent samples in a sequence at each point in time, but might use a dilated view, where samples with several samples in between are used. The dilated, temporal CNN update equation for the ith filter's activation $a_i^{(l)}[k]$ at time k and layer l with the learnable coefficients $\boldsymbol{W}_i \in \mathbb{R}^{A\times\kappa}$ applied on A previous layer's filters, an uneven kernel size of $\kappa \in \{2\xi+1 : \xi \in \mathbb{N}_0\}$, and the dilation factor δ reads

$$a_i^{(l)}[k] = \sum_{\eta=0}^{A-1} \sum_{j=-(\kappa-1)/2}^{(\kappa-1)/2} \boldsymbol{W}_{i;(\eta,j)} \cdot a_\eta^{(l-1)}[k+j\delta]. \quad (1)$$

Since the task at hand does not require causality of the CNN estimates along the time domain (losses are to be estimated from entire $\boldsymbol{B}$ sequences), the sliding operation can be efficiently parallelized, and sequential processing happens merely along the CNN's depth. All 1-D CNN layers are accompanied by weight normalization [24].

A conceptual representation of a 1-D CNN for estimating a time series of an equal-length input time series is sketched in Fig. 2. In this sketch, three coefficients for each hidden layer have to be learned only.

B. Feature Engineering

The term feature engineering encompasses all preprocessing, normalization, and derivation of additional features in an observational dataset. The input data contain the frequency f, the temperature T, the measured losses p as well as the $M = 1024$ sample points for the B and H sequences. Notably, the sample rate varies in each sequence since M is constant. Especially the creation of new features that correlate as much as possible with the target variable (here, the H curve or the scalar power loss p) is an important part of most ML frameworks [25].

1) Normalization: As is typical in neural network training, all input and target features have to be normalized beforehand. All scalar and time-series features are divided by their maximum absolute value that occurs in the material-specific dataset, with the exception of the temperature and the frequency, which will be divided by 75 °C and 150 kHz, respectively, regardless the material. These divisors will keep the temperature and the log frequency in the range of $[-2, 2]$ to support the gradient-descent-based optimization. Moreover, for an accurate H estimate, it was found to be of paramount importance to normalize each $\boldsymbol{B}$ and $\boldsymbol{H}$ curve again on a per-profile base in dependence not only on the ℓ_∞ norm of $|\boldsymbol{B}|$, but also on the maximum absolute B and H appearing in the entire material-specific dataset. The latter two values are denoted B_{lim} and H_{lim}, and can be understood as material-specific scaling constants. In particular, the per-profile

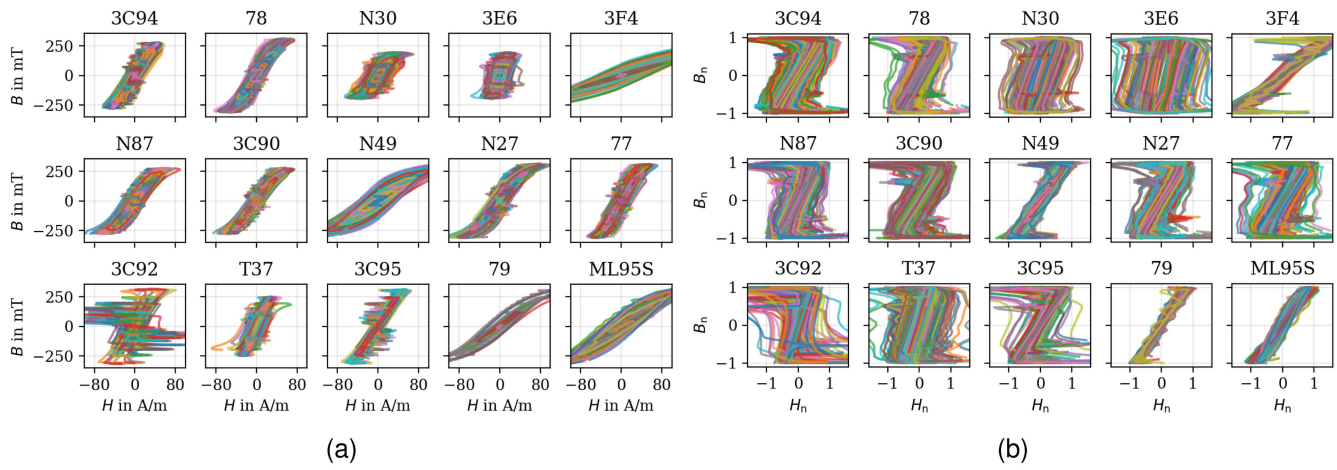


Fig. 3. Exemplary samples of the original and normalized B and H curves. (a) Before normalization. (b) After normalization.

normalized B and H curves for a certain sample read

$$\boldsymbol{B}_\mathrm{n} = \frac{\boldsymbol{B}}{\max_k |\boldsymbol{B}[k]|}, \quad \boldsymbol{H}_\mathrm{n} = \frac{\boldsymbol{H}}{H_\mathrm{lim}} \cdot \frac{B_\mathrm{lim}}{\max_k |\boldsymbol{B}[k]|} \tag{2}$$

with $H_\mathrm{lim} = \max_{i,k} |\boldsymbol{H}_i[k]|$, $B_\mathrm{lim} = \max_{i,k} |\boldsymbol{B}_i[k]|$, and i being the sample index in the entire material-specific dataset. Then, $\boldsymbol{B}_\mathrm{n}$ is added to the set of input time-series features, and $\boldsymbol{H}_\mathrm{n}$ is the target variable for the H estimation task.

The $\boldsymbol{B}_\mathrm{n}$ over $\boldsymbol{H}_\mathrm{n}$ curves are displayed in Fig. 3(b), which underlines how the BH-integral area becomes roughly unified (no large area difference between samples). In the following, all features that get in touch with the model are normalized values without any further notational indication.

2) Time-Series Features: As discussed in Section II-A, 1-D CNNs build the core of the implemented model. The inputs to the CNNs are the (per-profile) normalized magnetic flux density B_n and the corresponding first and second order derivatives ($\dot{B}_\mathrm{n}$ and $\ddot{B}_\mathrm{n}$) as time series. In a macroscopic measurement circuit context, $\dot{B}$ corresponds to the applied magnetizing voltage throughout the measurement process of the data. Accordingly, $\ddot{B}$ represents the voltage slew rate during the commutation of the switches in the test setup. Consequently, the second derivative allows to detect switching events and to characterize them according to their maximum slew rate. Fig. 4 shows, that the sinusoidal waveform (green) is generated without any fast transient switching behavior, probably with a linear signal source. The nonsinusoidal examples show typical switching behavior with different voltage slew rates during the single transitions and voltage overshoots as well as ringing. The second derivative of $\boldsymbol{B}$ informs the ML model about switching transition events and how fast changes in time are. The "tan-tan-b" feature is a double tangent of the full normalized $\boldsymbol{B}_\mathrm{n}$ sequence. It roughly follows an average $\boldsymbol{H}$ sequence and shall help to lift the input feature set into a higher dimensional nonlinear map, which might disclose relevant patterns in the data.

3) Scalar Features: Although sequence-based CNNs take up the main share of the ML model size, scalar environmental variables also have a considerable impact on $\boldsymbol{H}$ and p. While the temperature T is passed to the model unaltered (but normalized),

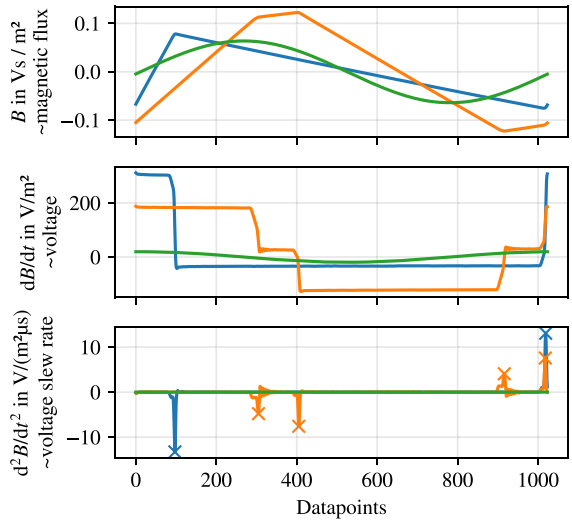


Fig. 4. Magnetic flux density examples and their first- and second-order derivatives for a sinusoidal, triangular, and one unclassified waveform with a circuit-based interpretation in terms of their proportionality to magnetic flux, voltage, and the voltage slew rate. Note, that this figure is not explicitly reporting the frequency and just shows all 1024 available data points of a given sequence, that is, the time difference per time step might be different.

the frequency is presented by its logarithm $\ln(f)$. The sample time $1/f$ is passed directly to the model. Furthermore, some B-derived scalar features are also passed to the model to feed in a priori knowledge. For example, the peak-to-peak magnetic flux ΔB as well as the mean absolute time derivative $\overline{|\dot{B}|}$ are directly fed into the network. Each waveform is automatically classified into"sine," "triangular," "trapezoidal," and "other" by template matching and consulting the form and crest factors, as well as some Fourier coefficients. The waveform classification is presented to the model by one hot encoding (OHE). A summary of all expert-driven input features is presented in Table II.

C. Residual Correction and Overall Topology

The model topology comprises multiple branches that end in the scalar power loss estimate $\hat{p}$. An overview is sketched in

TABLE II
UTILIZED INPUT FEATURES

Time series features		Scalar features	
mag. flux density	B	temperature	T
per-profile norm.	B_n	sample time	$1/f$
1st derivative	$\dot{B}_\mathrm{n}$	log-frequency	$\ln(f)$
2nd derivative	$\ddot{B}_\mathrm{n}$	peak2peak	ΔB
tan-tan-b	$\tan(0.9 \cdot \tan(B_\mathrm{n}))$	log peak2peak	$\ln(\Delta B)$
		mean abs dbdt	$\overline{\lvert\dot{B}\rvert}$
		log mean abs dbdt	$\ln(\overline{\lvert\dot{B}\rvert})$
		waveform (OHE)	

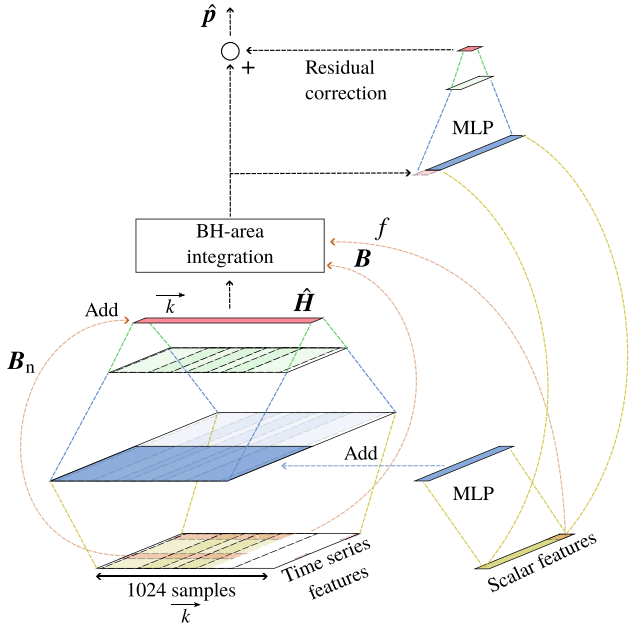


Fig. 5. Residual 1-D CNN topology is shown while applied on time-series and scalar features, which also contain engineered features from Table II. One MLP informs an intermediate 1-D CNN layer, whereas another acts as additive residual correction of the power loss estimate from the BH area integration.

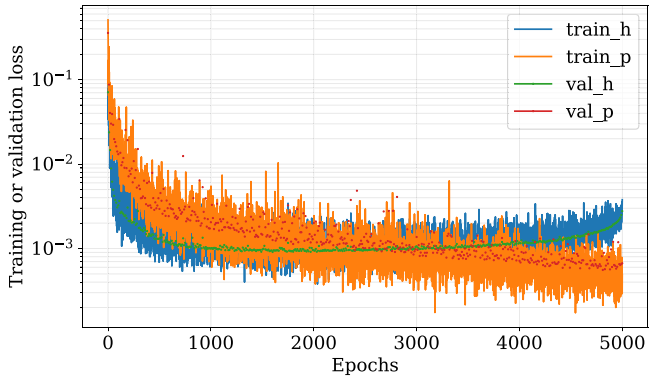


Fig. 6. Exemplary training and validation loss curve for material 3C92.

Fig. 5. Two main branches can be identified: an H-predictor and a parallel, residual p estimate correction. The H-predictor utilizes both time-series and scalar features with CNNs and multilayer perceptrons (MLPs), and estimates the full $\boldsymbol{H}$ sequence as $\hat{\boldsymbol{H}}$. This estimate is used to integrate the area under the BH curve with the trapezoidal summation formula. This sum is a preliminary power loss estimate $\hat{p}_\mathrm{pre}$, which is then corrected further by the residual correction branch. The latter accounts for the H modeling inaccuracy, and is predicted by a MLP that utilizes the scalar feature set and the preliminary power loss estimate.

The H-predictor merges time-series and scalar feature information by the broadcasted addition of its MLP output to a part of the first CNN layer output. This effectively considers the MLP-transformed scalar features as bias term to the time-series-based CNN structure.

On the merged feature set, two further 1-D CNN layers follow that end the transformation in a 1024-element sequence. The per-profile scaled $\boldsymbol{B}_\mathrm{n}$ sequence from the set of input time series is element-wise added to this newly obtained estimation (residual connection). This results in the CNN model to merely learn the difference between $\boldsymbol{H}_\mathrm{n}$ and $\boldsymbol{B}_\mathrm{n}$ [26]. Eventually, this sequence becomes the H estimation $\hat{\boldsymbol{H}}_\mathrm{n}$. Note that all such operations are still end-to-end differentiable with an automatic differentiation framework such as PyTorch [27], [28].

Since systematic modeling errors in the resulting $\hat{\boldsymbol{H}}$ cannot be ruled out, another MLP is branched off the scalar input feature set, and denotes the start of the residual error correction. This MLP inherits two hidden layers and concludes with a single output neuron and a linear activation. This branch's end, denoted by s, is added onto the preliminary log power loss estimate $\ln \hat{p}_\mathrm{pre}$ from the BH area integration

$$\ln \hat{p} = \ln \Bigg(\underbrace{\frac{1}{2} f \sum_{i=1}^{M-1} (B_i - B_{i-1})(\hat{H}_{i-1} + \hat{H}_i)}_{\hat{p}_\mathrm{pre}} \Bigg) + s. \quad (3)$$

Consequently, the residual correction branch can alter the trapezoidal summation formula result by a significant amount.

It is reasonable to question the error residual branch. After all, a higher estimation accuracy of H inevitably leads to a higher estimation accuracy of p. In fact, if a high accuracy in H is the only concern, the error residual branch can be omitted. However, we found in our experiments that a given realization of the proposed approach, that is, an instantiation with a certain low amount of model parameters, can deliver a superior power loss estimate at the cost of some H accuracy when extended with the error residual branch (which features only few more parameters). In scenarios where the power loss estimate is more important than the H accuracy and a very small number of model parameters are required (such as in the MagNet challenge), this extension provides another edge compared to a correction-free architecture.

The physical interpretability of the intermediate estimate $\hat{\boldsymbol{H}}$ is a key advantage of the HARDCORE approach over other competing methods: First, it enables utilizing full H time-series simulation frameworks (e.g., time-domain FEM solvers). Second, for future designs of magnetic components with arbitrary core shapes, it becomes indispensable to accurately take into

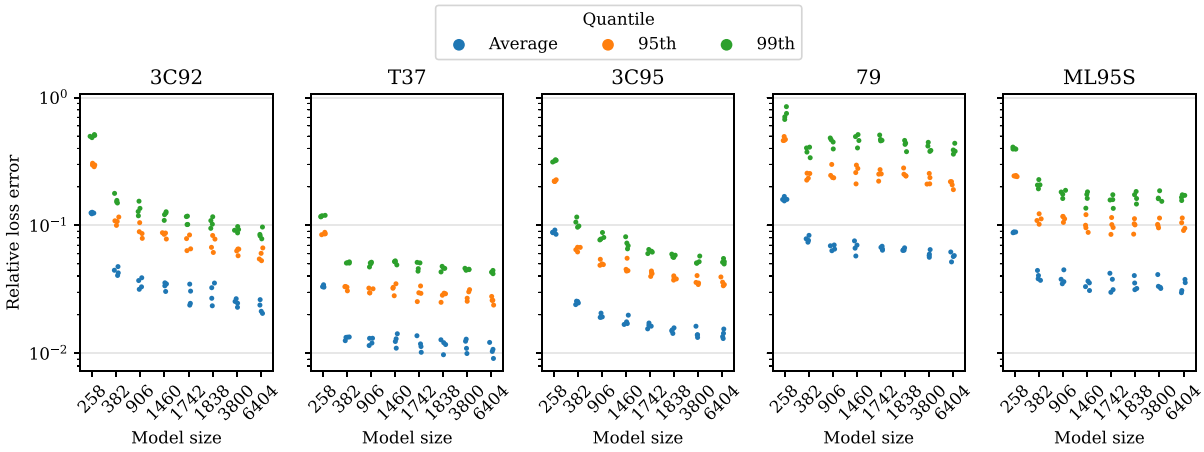


Fig. 7. Pareto front for the evaluation materials (3C92, T37, 3C95, 79, and ML95S) showing model size (amount of parameters) versus relative error of the power loss estimation. Dots represent the relative power loss error on the training dataset during a fourfold cross validation for each of four random number generator seeds.

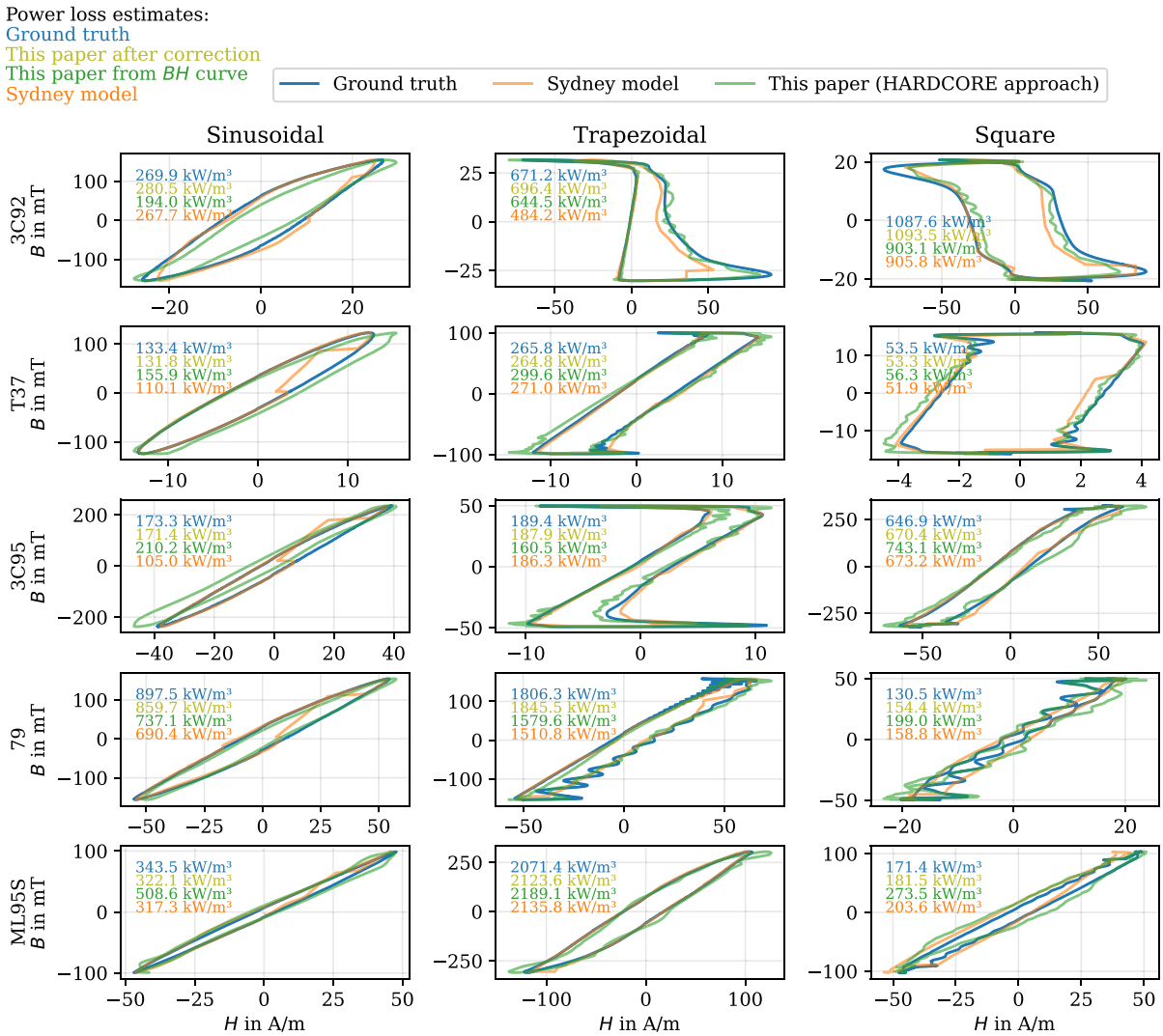


Fig. 8. Exemplary BH curves for each material and waveform are shown comparing our HARDCORE approach method with the submitted model from the University of Sydney. The H estimate is often stronger for the Sydney model but the corrected power loss estimate is better for the HARDCORE approach due to the residual correction branch.

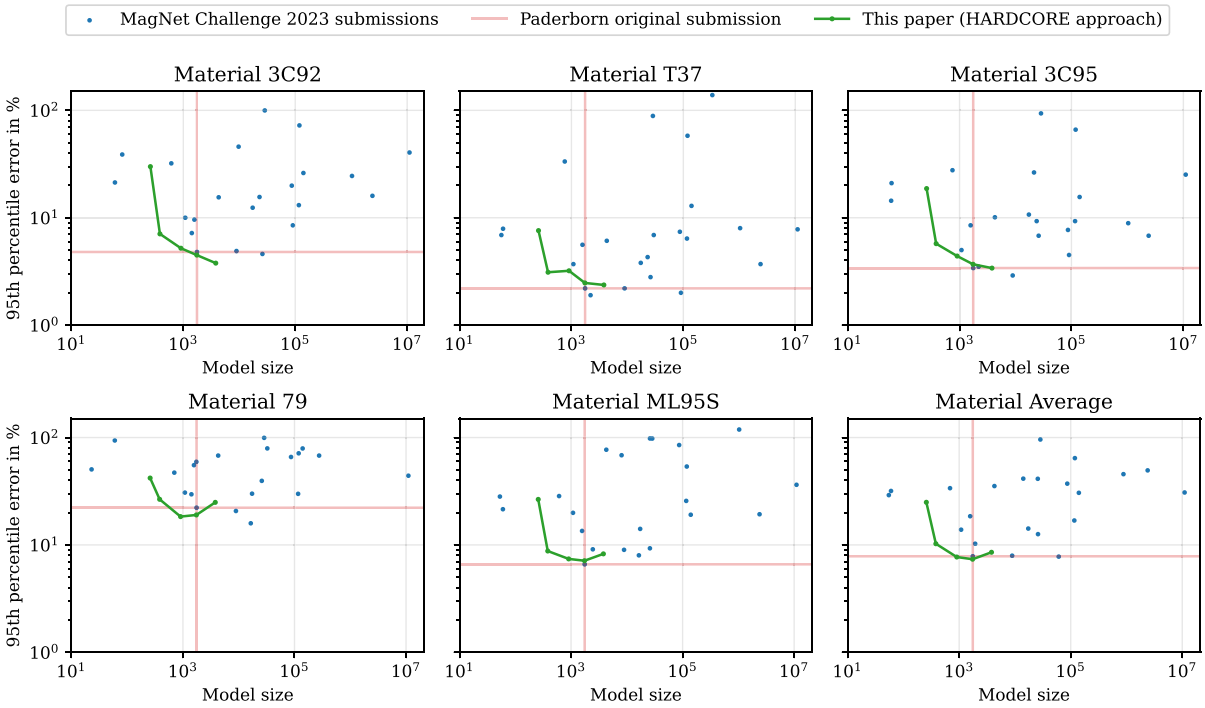


Fig. 9. Comparison of this article's HARDCORE approach with all submissions to the MagNet Challenge 2023 designated test set. The further to the bottom left, the better. The Paderborn original submission refers to [30], where the numerical integration of the BH curve was handled via the shoelace formula (now: trapezoidal rule) and where the residual power loss correction path was a multiplicative transformation (now: additive).

account also geometrical parameters of the core. With the geometry parameters, (di-)electric losses such as core eddy current losses are separately calculated from the hysteresis losses (HARDCORE approach). Also, dimensional resonance can be taken into account, which leads to an inhomogeneous magnetic flux density across the core cross-section. This effect occurs already at medium frequencies in large core designs, resulting in a drastic increase in losses and a change in inductance [29]. It is important to separate these loss components in order to interpret the results of the simulations and draw further conclusions, e.g., on the optimization of the core geometry.

D. Differences to the Originally Submitted Architecture

The original submission [30] to the challenge described a similar topology but with a different integration formula (the shoelace method instead of the trapezoidal sum). Since the trapezoidal formula was also used to determine the ground-truth power loss, additional modeling errors would be introduced by the choice of the shoelace formula. Moreover, the residual correction branch in the original architecture was constrained to alter the preliminary power loss estimate by at most 10%, which was motivated by the shoelace integration error. In this work, the residual correction is additive and incorporates not only the scalar feature set but also the preliminary power loss estimate. These changes allow for another significant decrease in the number of required model parameters, while the overall achieved model accuracy remains roughly the same compared to the original challenge submission.

E. Training Cost Functions

The training process involves two cost functions for a training dataset with size N: First, the H estimation accuracy, which is assessed with the MSE as

$$\mathcal{L}_{\text{MSE,H}} = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} (\hat{H}_{i,n} - H_{i,n})^2. \tag{4}$$

Second, the power loss estimation accuracy is to be gauged. Despite the relative error being the competition's evaluation metric, the mean squared logarithmic error (MSLE) is selected

$$\mathcal{L}_{\text{MSLE,P}} = \frac{1}{N} \sum_{n=0}^{N-1} (\ln \hat{p}_n - \ln p_n)^2 \tag{5}$$

in order to not overemphasize samples with a relatively low power loss [31]. As $\mathcal{L}_{\text{MSLE,P}}$ also depends on $\hat{\boldsymbol{H}}$ through (3), the question arises, how both cost functions are to be weighted. In this HARDCORE approach contribution, a scheduled weighting is applied with

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{MSLE,P}} + (1 - \alpha) \mathcal{L}_{\text{MSE,H}} \tag{6}$$

where $\alpha = (i_{\text{epoch}})/K_{\text{epoch}}$ with K_{epoch} being the number of training epochs, and $i_{\text{epoch}} \in \{0, 1, \ldots K_{\text{epoch}} - 1\}$ representing the current epoch index. The scheduled weighting ensures that the model focuses on $\hat{\boldsymbol{H}}$ in the beginning of the training, where more information is available. Later although, the model shall draw most of its attention to the power loss estimate, possibly at the expense of the H estimation accuracy. The loss curves during a training example for material 3C92 are depicted in Fig. 6. Note that the cost scheduling and the power loss cost function can be

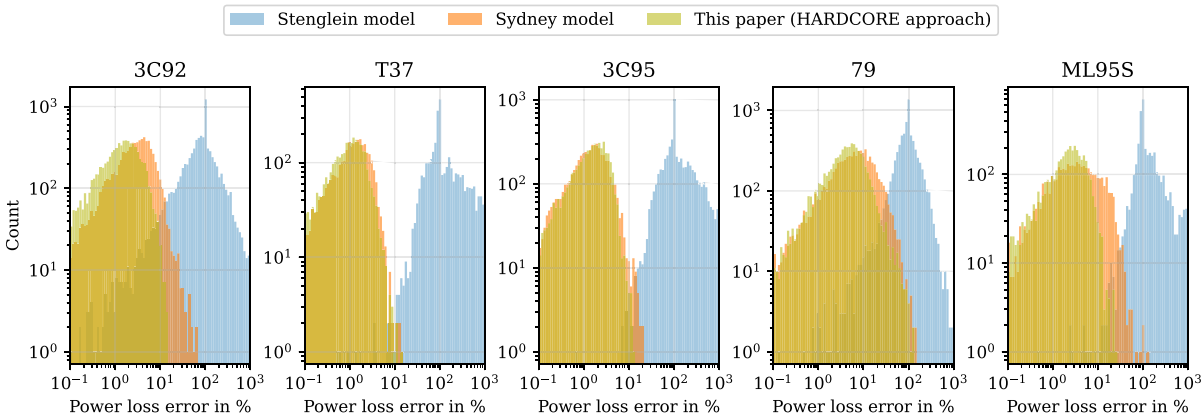


Fig. 10. Comparison of this article's HARDCORE approach with the Stenglein and Sydney model on the designated test set. The relative power loss errors are illustrated in histograms, where "count" denotes the number of samples that fall into the respective bin.

omitted altogether with the error residual branch in case a high H estimation accuracy is desired.

III. Hyperparameters, Pareto Front, and Results

The proposed topology features several degrees of freedom in form of hyperparameters. An important aspect is the model size, which is defined by the number of hidden layers and neurons in each layer. A simple trial-and-error investigation can provide fast insights into the performance degradation that comes with fewer model parameters.

In this work, the relative power loss error is defined as

$$\mathcal{L}_{\text{rel.err.}} = \frac{|p - \hat{p}|}{p}. \tag{7}$$

In Fig. 7, several particularly selected model topologies are illustrated against their achieved relative error versus the inherited model size. The scatter in each quantile is due to different random number generator seeds during a stratified fourfold cross-validation ($K_{\text{epochs}} = 5000$, Nesterov Adam optimizer). Topology variations are denoted by the amount of neurons in certain hidden layers. It becomes evident that there is a tradeoff between model size and accuracy, such that in case one of the two criteria can be softened, the other can be further improved.

Overall, the model performance on the different materials scales strictly with the amount of training data available. That is why the model on material 79 performs the worst (least training data available among all materials—investigation constraint by the MagNet challenge dataset), together with the fact that an entire waveform type is missing in its training dataset. The topology variation study suggests a Pareto-optimal tradeoff at 906 model parameters, which renders roughly half of the originally submitted model at 1755 parameters while the accuracy remains comparable.

Exemplary BH curves as estimated by this HARDCORE approach contribution and by a contending approach from the University of Sydney are illustrated in Fig. 8. The Sydney model carries 1084 parameters, whereas our method features 906 parameters in this plot. It becomes evident that the Sydney model's H estimate often prevails over the HARDCORE H estimate but a look at the resulting power loss estimates reveals an often more precise estimate for our method after residual correction. This makes the tradeoff between H and power loss accuracy in the HARDCORE approach more obvious.

TABLE III
Optimal Parameter-to-Accuracy Configuration for the MagNet Challenge Final Evaluation Materials

	3C92	T37	3C95	79	ML95S
Training data	2432	7400	5357	580	2013
Test data	7651	3172	5357	7299	3738
Model parameters	906	906	906	906	906
Model size in kB	42	42	42	42	42
Avg. relative error	1.8 %	1.2 %	1.8 %	6.4 %	2.8 %
95th-quantile rel. error	5.2 %	3.2 %	4.4 %	18.4 %	7.4 %
99th-quantile rel. error	8.3 %	5.1 %	6.2 %	45.1 %	11.8 %

In Table III, the model size of a corresponding PyTorch model file dumped to disk as just-in-time compilation is reported. Reported error rates are evaluated on the test set after training with a single random number generator seed. It shows effectively that any material can be modeled with the same topology at high accuracy as long as a critical training dataset size is available and all waveforms were seen during training (which is not the case for material 79; see available training data in Table III).

The accuracy of this contribution (HARDCORE approach) is compared with all other contending submissions to the MagNet Challenge 2023 in Fig. 9. All presented models were trained on all training data samples (no repetitions with different seeds), and with $K_{\text{epoch}} = 10000, \delta = 4, \kappa = 9$. The superiority of our approach becomes also evident in a comparison to the Stenglein and Sydney model in Fig. 10. There, histograms of the relative power loss error against the test set for the Stenglein, Sydney, and HARDCORE model are depicted.

IV. Conclusion

A material-agnostic CNN topology for efficient steady-state power loss estimation in ferrite cores is developed. Since the topology remains unaltered across materials and waveforms at a steadily high accuracy, the proposed model can be considered universally applicable to plenty of materials. The average 95th

quantile of the relative power loss error across the final five materials remains below 8% for a model with 906 parameters. The originally presented model has won the first place in the performance category of the MagNet challenge 2023 emphasizing its effectiveness. Considering these facts, the contributed HARDCORE approach method is proposed to become a standard way of training data-driven models for power magnetics.

It remains open how the proposed method can be transferred from idealized small toroid-based measurement data to reluctance model or FEM-based design processes of arbitrary-shaped core geometries, as further effects such as geometric resonance must be taken into account in the design phase. Moreover, it would be desirable to eliminate the tradeoff between H and power loss estimation accuracy in the future. Eventually, a systematic hyperparameter optimization might reveal even better topologies, kernel sizes, dilation factors, or optimization parameters such as the batch size or the learning rate.

ACKNOWLEDGMENT

The authors would like to thank the organizers and sponsors of the MagNet challenge 2023 for hosting a great, educational event within the power electronics community. This article would not have been possible without the provision of the MagNet dataset and the organizers' huge effort administrating the challenge.

REFERENCES

[1] C. Steinmetz, "On the law of hysteresis," *Proc. IEEE*, vol. 72, no. 2, pp. 197–221, Feb. 1984.
[2] C. P. Steinmetz, "On the law of hysteresis (part III), and the theory of ferric inductances," *Trans. Amer. Inst. Elect. Engineers*, vol. 11, pp. 570–616, 1894.
[3] M. Albach, T. Durbaum, and A. Brockmeyer, "Calculating core losses in transformers for arbitrary magnetizing currents a comparison of different approaches," in *Proc. 27th Annu. IEEE Power Electron. Spec. Conf.*, 1996, vol. 2, pp. 1463–1468.
[4] M. Lancarotte and A. de Arruda Penteado, "Estimation of core losses under sinusoidal or nonsinusoidal induction by analysis of magnetization rate," in *Proc. IEEE Int. Electric Mach. Drives Conf.*, 1999, pp. 490–492.
[5] K. Venkatachalam, C. R. Sullivan, T. Abdallah, and H. Tacca, "Accurate prediction of ferrite core loss with nonsinusoidal waveforms using only Steinmetz parameters," in *Proc. IEEE Workshop Comput. Power Electron.*, 2002, pp. 36–41.
[6] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Improved core-loss calculation for magnetic components employed in power electronic systems," *IEEE Trans. Power Electron.*, vol. 27, no. 2, pp. 964–973, Feb. 2012.
[7] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Core losses under the DC bias condition based on Steinmetz parameters," *IEEE Trans. Power Electron.*, vol. 27, no. 2, pp. 953–963, Feb. 2012.
[8] E. Stenglein and T. Dürbaum, "Core loss model for arbitrary excitations with DC bias covering a wide frequency range," *IEEE Trans. Magn.*, vol. 57, no. 6, Jun. 2021, Art. no. 6302110.
[9] S. A. Mulder, "Fit formulae for power loss in ferrites and their use in transformer design," in *Proc. Int. Power Convers. Conf.*, 1993, pp. 345–359. [Online]. Available: https://www.tib.eu/de/suchen/id/BLCP%3ACN002941661
[10] A. Arruti et al., "The composite improved generalized Steinmetz equation (ciGSE): An accurate model combining the composite waveform hypothesis with classical approaches," *IEEE Trans. Power Electron.*, vol. 39, no. 1, pp. 1162–1173, Jan. 2024.
[11] *Princeton Magnet*. Accessed: Nov. 15, 2024. [Online]. Available: https://github.com/PrincetonUniversity/Magnet
[12] H. Li et al., "Magnet: An open-source database for data-driven magnetic core loss modeling," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, 2022, pp. 588–595.
[13] D. Serrano et al., "Why MagNet: Quantifying the complexity of modeling power magnetic material characteristics," *IEEE Trans. Power Electron.*, vol. 38, no. 11, pp. 14292–14316, Nov. 2023.
[14] H. Li et al., "How MagNet: Machine learning framework for modeling power magnetic material characteristics," *IEEE Trans. Power Electron.*, vol. 38, no. 12, pp. 15829–15853, Dec. 2023.
[15] D. Serrano et al., "Neural network as datasheet: Modeling B-H loops of power magnetics with sequence-to-sequence LSTM encoder-decoder architecture," in *Proc. IEEE 23rd Workshop Control Model. Power Electron.*, 2022, pp. 1–8.
[16] H. Li, D. Serrano, S. Wang, T. Guillod, M. Luo, and M. Chen, "Predicting the B-H loops of power magnetics with transformer-based encoder-projector-decoder neural network architecture," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, 2023, pp. 1543–1550.
[17] H. Li, D. Serrano, S. Wang, T. Guillod, M. Luo, and M. Chen, "MagNet-AI: Neural network as datasheet for magnetics modeling and material recommendation," *IEEE Trans. Power Electron.*, vol. 38, no. 12, pp. 15854–15869, Dec. 2023.
[18] I. PELS-Google-Enphase-Princeton, "Magnet challenge 2023." Accessed: Nov. 15, 2024. [Online]. Available: https://github.com/minjiechen/magnetchallenge
[19] Y. Hu, J. Xu, J. Wang, and W. Xu, "Physics-inspired multimodal feature fusion cascaded networks for data-driven magnetic core loss modeling," *IEEE Trans. Power Electron.*, vol. 39, no. 9, pp. 11356–11367, Sep. 2024.
[20] Z. Li et al., "A data-driven model for power loss estimation of magnetic materials based on multi-objective optimization and transfer learning," *IEEE Open J. Power Electron.*, vol. 5, pp. 605–617, 2024.
[21] X. Shen, Y. Zuo, and W. Martinez, "Conditional generative adversarial network aided iron loss prediction for high-frequency magnetic components," *IEEE Trans. Power Electron.*, vol. 39, no. 8, pp. 9953–9964, Aug. 2024.
[22] A. Krizhevsky, I. Sulskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, vol. 60, no. 6, pp. 84–90, doi: 10.1145/3065386.
[23] J. Gu et al., "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320317304120
[24] T. Salimans and D. P. Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," 2016, *arXiv:1602.07868*.
[25] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, p. 78, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2347736.2347755
[26] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.
[27] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, vol. 32, pp. 8024–8035.
[28] A. G. Baydin et al., "Automatic differentiation in machine learning: A survey," *J. Mach. Learn. Res.*, vol. 18, pp. 1–43, 2018, *arXiv:1502.05767*.
[29] T. Piepenbrock et al., "Study on sample geometries for ferrite characterisation in the MHz range," in *Proc. Int. Exhib. Conf. Power Electron., Intell. Motion, Renewable Energy Energy Manage.*, 2024, pp. 463–471.
[30] N. Förster et al., "HARDCORE: H-field and power loss estimation for arbitrary waveforms with residual, dilated convolutional neural networks in ferrite cores," 2024, *arXiv:2401.11488* .
[31] C. Tofallis, "A better measure of relative prediction accuracy for model selection and model estimation," *J. Oper. Res. Soc.*, vol. 66, pp. 1352–1362, 2015.

Wilhelm Kirchgässner (Member, IEEE) received the B.Sc. degree in electrical engineering from the OWL University of Applied Sciences and Arts, Lemgo, Germany, in 2013, and the M.Sc. (Hons.) and doctoral (Hons.) degrees in electrical engineering from the Paderborn University, Paderborn, Germany, in 2016 and 2024, respectively.

He is currently with Beckhoff Automation, Verl, Germany, as a Data Scientist working on planar drive control and system identification. His research interests include the application of machine learning to electrical drive technologies, in particular data-driven thermal modeling of highly utilized traction drives.

Nikolas Förster received the master's degree in electrical engineering from Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany, in 2015.

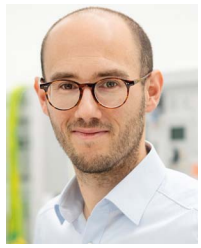
He is currently a Research Associate with the Department of Power Electronics and Electrical Drives, Paderborn University, Paderborn, Germany. His research interests include power electronics converter optimization methods and component modeling.

Till Piepenbrock received the bachelor's and master's (Hons.) degrees in electrical engineering from Paderborn University, Paderborn, Germany, in 2019 and 2022, respectively.

He is currently working as a Research Assistant with the Department of Power Electronics and Electrical Drives, Paderborn University. His research interests include high-frequency dc–dc converter design with a focus on the optimization of magnetic components.

Oliver Schweins received the bachelor's degree (Hons.) in electrical engineering in 2024 from Paderborn University, Paderborn, Germany, where he is currently working toward the master's degree in electrical engineering.

He is currently working as a Student Assistant with the Department of Power Electronics and Electrical Drives, Paderborn University.



Oliver Wallscheid (Senior Member, IEEE) received the bachelor's and master's (Hons.) degrees in industrial engineering and the doctorate (Hons.) degree in electrical engineering from Paderborn University, Paderborn, Germany, in 2010, 2012, and 2017, respectively.

He had been a Senior Research Fellow with the Department of Power Electronics and Electrical Drives and an acting Professor with the Automatic Control Department, Paderborn University. Since 2024, he has been a Full Professor with the University of Siegen, Siegen, Germany, and heading the Interconnected Automation Systems (IAS) Chair. His research interests include modeling, design, and control of electrical power systems focusing on decentralized grids, power electronics, and electrical drives.