

How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

15-10-2022 / 01-02-2023

CITATION

Chen, Minjie (2022): How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.21340998.v3>

DOI

[10.36227/techrxiv.21340998.v3](https://doi.org/10.36227/techrxiv.21340998.v3)

How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics

Haoran Li, *Student Member, IEEE*, Diego Serrano, *Member, IEEE*, Thomas Guillod, *Member, IEEE*, Shukai Wang, *Student Member, IEEE*, Evan Dogariu, *Student Member, IEEE*, Andrew Nadler, *Student Member, IEEE*, Min Luo, *Senior Member, IEEE*, Vineet Bansal, Niraj Jha, *Fellow, IEEE*, Yuxin Chen, *Senior Member, IEEE*, Charles R. Sullivan, *Fellow, IEEE* and Minjie Chen, *Senior Member, IEEE*

Abstract—This paper applies machine learning to power magnetics modeling. We first introduce an open-source database – MagNet – which hosts a large amount of experimentally measured excitation data for many materials across a variety of operating conditions, consisting of more than 500,000 data points in its current state. The processes for data acquisition and data quality control are explained. We then demonstrate a few neural network-based power magnetics modeling tools for modeling the core losses and B – H loops. Machine learning allows multiple factors that may influence the magnetic characteristics to be modeled in a unified framework, while provides insights to quantify the complexity of magnetic characteristics and reduce the size of the measurement data required to build a precise model. Neural network models are found to be effective in compressing the measurement data and predicting the material characteristics. The behaviors of a typical power magnetic material (TDK N87) across a wide range of operating conditions (e.g., temperature, waveform, dc-bias) can be well described by a small-scale neural network (204 KB) which is 2,500 times smaller than the raw measured time-series data (512 MB), paving the way for “neural networks as datasheet” to assist power magnetics design.

Index Terms—power magnetics, core loss, hysteresis loop, open-source database, machine learning, data-driven method, neural network

I. INTRODUCTION

MAGNETIC components, such as inductors and transformers, are critical in almost all power electronics systems. These magnetic components are typically the largest

This paper is a combination and extension of four previously published conference papers, “MagNet: a machine learning framework for magnetic core loss modeling” in IEEE COMPEL 2020 [1], “Transfer Learning Methods for Magnetic Core Loss Modeling” in IEEE COMPEL 2021 [2], “MagNet: an open-source database for data-driven magnetic core loss modeling” in IEEE APEC 2022 [3], and “Neural Network as Datasheet: Modeling B-H Loops of Power Magnetics with Sequence-to-Sequence LSTM Encoder-Decoder Architecture” in IEEE COMPEL 2022 [4]. This work was jointly supported by ARPA-E Differentiate Project and the Schmidt DataX Fund at Princeton University made possible through a major gift from the Schmidt Futures Foundation. (*Corresponding Author: Minjie Chen.*)

H. Li, D. Serrano, S. Wang, E. Dogariu, N. Jha, V. Bansal, and M. Chen are with the Department of Electrical and Computer Engineering and the Andlinger Center for Energy and the Environment at Princeton University, Princeton, NJ, 08540, USA (e-mail: {haoranli, ds9056, sw0123, edogariu, jha, vineetb, minjie}@princeton.edu).

T. Guillod, A. Nadler, and C. R. Sullivan are with Dartmouth College, Hanover NH, USA (e-mail: {Thomas.Paul.Henri.Guillod, Andrew.B.Nadler.TH, charles.r.sullivan}@dartmouth.edu).

Y. Chen is with the Department of Statistics and Data Science at the Wharton School at University of Pennsylvania, Philadelphia PA, USA (email: yuxinc@wharton.upenn.edu).

M. Luo is with Ekarus Engines GmbH, Essingen, Germany (e-mail: m.luo@ekarusengines.com).

in volume and have significant power loss, and therefore have an adverse impact on the system performance. While there have been major strides in the modeling and analysis of power semiconductor devices and circuit topologies, the necessary advances in the modeling and design of power magnetic components and materials are lagging [5]–[9]. Currently, power magnetics models are usually developed and tested on different private datasets [10]–[13] with either unknown or unreported data quality. Data and its models cannot be rapidly compared or cross-validated. An open-source large-scale power magnetics research platform with controlled data quality and state-of-the-art software tools is needed and serves as the basis of this research.

Modeling magnetic materials is challenging due to the complicated material excitation-response mechanisms, the numerous factors involved (i.e., temperature, dc bias, memory effects), and the fact that no fully satisfactory first-principle model is yet known. Figure 1 compares multiple measured B – H loops for N87 ferrite material as an example, where the material characteristics differ significantly under different conditions. These intertwined influence factors are quantified in [14]. They typically co-exist and change at the same time in real applications, which renders the modeling of magnetic materials extremely difficult. A widely used method to model the core loss is the Steinmetz Equation (SE) [15], [16], which is an empirical equation based on curve-fitting, employed to calculate the core loss per unit volume in magnetic materials subjected to sinusoidal magnetic flux. However, most of the magnetic components in power electronics systems often have magnetizing currents with significant harmonic components, e.g., triangular, trapezoidal, or piece-wise linear waveforms.

Many advanced methods have been developed for modeling core loss under non-sinusoidal excitations. Some of these, including the modified Steinmetz equation (MSE), improved generalized Steinmetz equation (iGSE), and improved-improved generalized Steinmetz Equation (i²GSE) [17]–[20], are listed in Table I. All these models have known accuracy limitations for specific waveform types. They usually do not have a clear pathway toward capturing the impact of temperature, dc-bias, and relaxation effects.

Recent advances in data-driven methods, especially neural networks and other machine learning techniques, have proved extremely effective in solving non-linear multi-variable classification and regression problems, such as those in computer vision and speech recognition [21]–[24]. They also have been

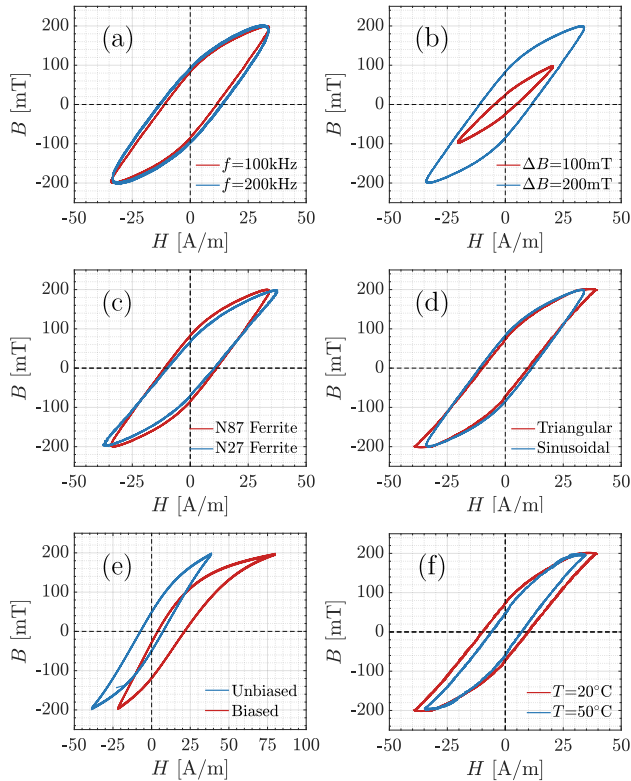


Fig. 1. Examples of measured B – H loops for ferrite materials under different conditions: (a) frequency, (b) peak flux density, (c) magnetic material, (d) shape of waveform, (e) dc bias, and (f) temperature. Except for the specified property, other properties are approximately kept the same for each measurement (TDK N87 ferrite material, 200 mT peak flux density, 100 kHz, 25 °C, and without dc bias).

TABLE I
NUMBER OF PARAMETERS USED BY CORE LOSS MODELS

Method	Core Loss (P_V)	#Param.
SE	$k f^\alpha \hat{B}^\beta$	3
iGSE	$\frac{1}{T} \int_0^T k_i \left \frac{dB}{dt} \right ^\alpha (\Delta B)^{\beta-\alpha} dt$	3
i ² GSE	$\frac{1}{T} \int_0^T k_i \left \frac{dB}{dt} \right ^\alpha (\Delta B)^{\beta-\alpha} dt + \sum_{l=1}^n Q_{rl} P_{rl}$	8
ML	Neural Network	$\gg 100$

applied to power electronics design and optimization [25]–[29]. The main strength of a neural network approach is the possibility of capturing many intertwined influence factors in a unified framework. Neural networks have been used to model magnetic hysteresis loops [30]–[38]. However, existing neural network models usually have limited data sizes, unclear data quality, and sometimes outdated network structures.

The main contributions of this paper include: (1) We presented a large-scale open-source database - *MagNet* - for power magnetics research; (2) We show an end-to-end machine learning framework which can capture the many factors that many impact the magnetic material characteristics in a unified setup; and (3) we demonstrate that a neural network can effectively store material characteristics information and has the potential to be used as an active datasheet for

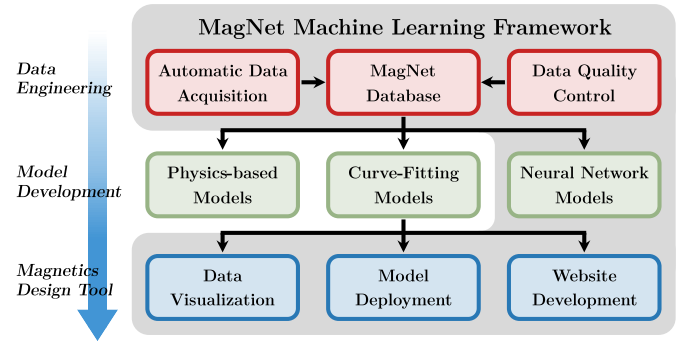


Fig. 2. Overview of the MagNet framework from data engineering, model development, to magnetics design tool. The building blocks in the shaded area are covered in this paper.

describing power magnetic materials in the design pipeline.

The key workflow of MagNet is illustrated in Fig. 2. Just as *ImageNet* advances computer vision research [24], the goal of developing MagNet is to advance research in data-driven power magnetics modeling by providing a common ground for testing and comparing different magnetic materials, modeling methods, and design optimization tools. The accuracy of equation-based models and data-driven models both rely heavily on data size and data quality.

The rest of the paper is organized as the following: Section II introduces the automatic data acquisition system of MagNet, including the hardware setup and software configurations, with extended details in Appendix A; Section III discusses the considerations on data quality evaluation and methods to improve the data quality, with extended details in Appendix B; Section IV introduces the database structure and data format in their current states; Section V presents a few example ways of applying neural networks to model power magnetics for different purposes, including scalar-to-scalar core loss prediction, sequence-to-scalar core loss prediction, sequence-to-sequence B – H loop prediction, data augmentation for increasing the model generality, transfer learning for reducing the training data size. Section VI introduces the “Neural Network as Datasheet” concept. Appendix C introduces the web-based open-source platform of MagNet.

II. DATA ACQUISITION

Large-scale and high-quality database lies the foundation of machine learning and data-driven modeling methods, which fundamentally bounds the accuracy of models. However, the behaviors of power magnetics, especially the core loss, can be impacted by many factors including frequency, flux density, dc-bias, waveform shape, and temperature, among others. The large number of degrees of freedom leads to an extremely large parameter space to sweep and measure. To capture the impact of all these factors, a fully automated data acquisition with carefully evaluated accuracy is needed. Increasing the automation level also reduces the error caused by human factors during the measurement.

The most common procedure for B – H loop (and core loss) characterization is the use of the two-winding method, also

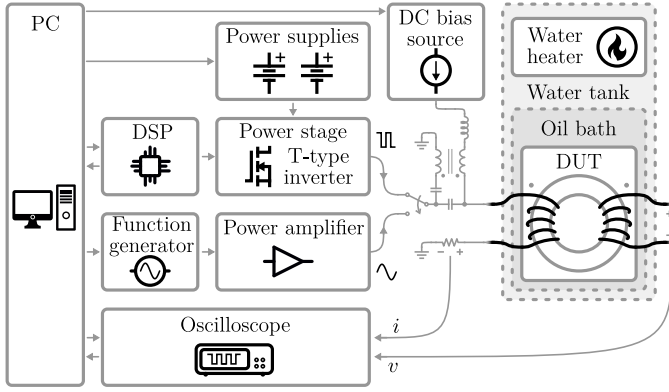


Fig. 3. Overview of the automated data acquisition system of MagNet.

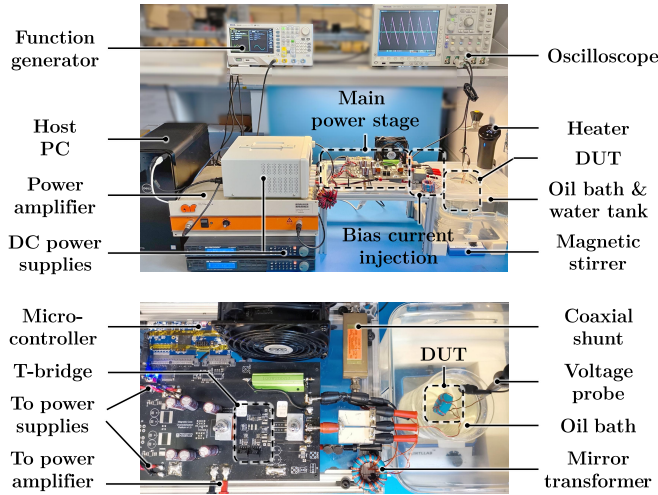


Fig. 4. Experiment setup and circuit configuration of the magnetic core loss data acquisition system of MagNet.

referred to as the voltamperometric method [17], [39], [40], where two separated windings are used. The excitation is applied to the primary, where the current is measured to obtain H . The voltage across the secondary winding is measured to obtain B as the voltage drop in the primary winding resistance or leakage inductance is not reflected in the secondary [40].

We adopted the two-winding method for magnetic characterization. Fig. 3 depicts an overview of the fully automated data acquisition system, comprising a power stage that is capable of generating different excitation waveforms, the device under test (DUT), voltage measurement, current measurement, auxiliary stage for the dc bias, and temperature control. Fig. 4 shows a picture of the experimental setup.

In this particular design, the excitation of the magnetic core is synthesized and generated by a T-type inverter in the power stage (for non-sinusoidal waves) and a power amplifier with a function generator (for sinusoidal waves). Both the secondary-side voltage waveform and the primary-side current waveform are captured and measured by an oscilloscope, where a wide-band coaxial shunt is used to enable accurate current measurement at high frequency. An optional dc bias injection circuitry is implemented to excite the magnetic core with a

non-zero bias current. An external water heater, water tank, and oil bath are implemented to provide temperature control for the measurement under different temperature conditions. A software system is programmed on the host PC to control and coordinate with the hardware system to enable fully automated equipment settings and measurements. More details about the automated power magnetics data acquisition system including the hardware configuration and implementation, the measurement equipment, and the software programming are provided in Appendix A.

This data acquisition system can automatically excite and drive the DUT with pre-programmed excitations and measure the material responses. With this system, the B - H loop and core loss can be directly measured and calculated via voltamperometric method [12], [39], [41], [42] by:

$$P_{loss} = \frac{1}{NT} \int_{t_0}^{t_0+NT} v_L(t) \cdot i_L(t) dt \quad (1)$$

$$B(t) = \frac{1}{A_e \cdot n_2} \int_{t_0}^{t_0+NT} v_L(t) dt \quad (2)$$

$$H(t) = \frac{n_1}{l_e} \cdot i_L(t) \quad (3)$$

where v_L and i_L are the measured secondary-side voltage and primary-side current, respectively. n_1 and n_2 refer to the number of turns of each winding. A_e is the effective cross-section area of the magnetic core and l_e is the effective length. NT is the total duration of the measurement. The duration is intentionally configured to ensure that the measured waveform contains an integer number of periods.

With this system, the time it takes to complete one measurement is around 1.5 seconds. The actual measurement duration is 100 μ s, and the rest time is used for control, communication, and relaxation of the material. The system can, fully autonomously, collect around 2,400 data points per hour. A complete characterization of one material usually takes a few hours, during which no human operation is needed. Please note that certain materials may require longer relaxation time between measurements to resume to the demagnetized state, leading to slower data acquisition speed. One should carefully understand and pre-calibrate the material to select the appropriate measurement time interval.

III. DATA QUALITY CONTROL

The accuracy of a data-driven method is highly dependent on the data size and data quality. Measuring the B - H loops and core losses accurately across a wide operation range is challenging. The error distribution may be impacted by many factors including: parasitics, oscilloscope limitations, timing skew between channels, misbehavior of the micro-controllers, electrical noise and quantization noise, temperature variation, and many others. The real-time measured voltage and current signals can be decomposed as:

$$\begin{aligned} v_L(t) &= G_V(V_0 + V_{DC} + v_{AC}(t)) \\ i_L(t) &= G_I(I_0 + I_{DC} + i_{AC}(t - \theta)) \end{aligned} \quad (4)$$

where G_V represents the gain factor of the voltage measurement. V_0 is the zero-drift offset voltage introduced by the equipment. V_{DC} , and v_{AC} are the dc and ac components in the

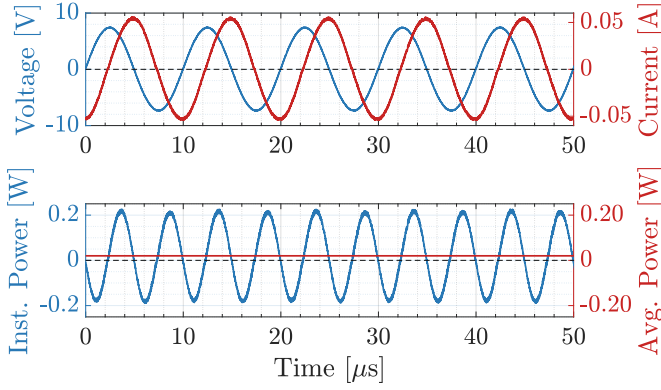


Fig. 5. Example measured voltage, current, and power waveform of TDK N87 ferrite material at 100 kHz.

measured voltage signal, respectively, in the periodic steady state. θ denotes the time skewing between the voltage and current measurement results. Similar definitions go for all the current-related variables. Based on Eq. (1), the average power loss across N cycles is:

$$\begin{aligned}
 P_{loss} &= \frac{1}{NT} \int_{t_0}^{t_0+NT} v_L(t) \cdot i_L(t) dt \\
 &= G_V G_I (V_0 I_0 + V_0 I_{DC} + V_{DC} I_0) \\
 &\quad + G_V G_I V_{DC} I_{DC} \\
 &\quad + \frac{G_V G_I}{NT} \int_{t_0}^{t_0+NT} v_{AC}(t) \cdot i_{AC}(t - \theta) dt
 \end{aligned} \tag{5}$$

Equation (5) provides useful information for the semi-quantitatively understanding of the error associated with gain and offset. Errors in G_V , G_I , V_{DC} , I_{DC} , V_0 , I_0 , v_{AC} and i_{AC} all have an impact on the core loss error. Fig. 5 illustrates an example voltage and current waveform measured with N87 ferrite material at 100 kHz, as well as the instantaneous power and average power (i.e., the power loss). The instantaneous power is much larger than the average power. A minor error in either the voltage or the current, or a phase mismatch between them, may lead to a significant percentage error in power loss.

All equipment used in the data acquisition system is evaluated and calibrated. Experiments to calibrate the oscilloscope against a digital multi-meter are conducted, where the relative error of the mean dc voltage measured by the oscilloscope is 0.25%, and the relative error of the root-mean-square (RMS) ac voltage is 0.67%. Auto-calibration of the oscilloscope is conducted before the measurement iteration starts to minimize the undesired zero-drift offset and deskew the voltage and current channels. The parasitics introduced by the power stage circuit and the cable connections are also minimized in order to further reduce the potential time skewing between the voltage and current measurement.

A model-driven method combining the physics-based virtual measurement simulation and the Monte Carlo experiments is proposed to quantify the measurement error and estimate the error distribution. Appendix B analyzes the systematic error

and statistical error of the system. The analysis helps to determine the range of measurement and maintain high data quality. The analysis also shows that the geometry variation can significantly impact the core loss. Similar phenomenons are reported in [14], where the maximum geometry-to-geometry variation of core loss density can be more than 10%, larger than the impacts of most other sources of error.

Finally, a data-driven algorithm is developed to detect and remove the anomaly outliers in the collected dataset, as they are impossible to be completely avoided in the large-scale automated data collection. The essential idea of the algorithm is to evaluate the smoothness of the measured data points within a certain range of flux density and frequency, based on the curve-fitting of the local Steinmetz Equation. For a certain data point, an expected value of core loss can be inferred based on other adjacent data points, then the discrepancy between the expected value and the measured value can be calculated. Suspicious data points that are likely to be outliers can be removed. Extended details about data quality control are provided in Appendix B.

IV. DATABASE CONSTRUCTION

The fully automated data acquisition system enables rapid measurement of B - H loop data. Fig. 6 shows the voltage and current waveforms of four examples of measured data, including sinusoidal, triangular, symmetric trapezoidal, and asymmetric trapezoidal, all measured with N87 ferrite material at 100 kHz. The sampling time step of the measured waveform sequence is set as 10 ns, and each waveform sequence contains 10,000 sampling points for a 100 μ s measurement period.

Better data documentation enables better data usage. Fig. 7 shows the data format of MagNet in its current state, which comprises three data domains: 1) information about the DUT, including the material type and the geometry parameters; 2) raw measured time-series data, including the voltage, the current, and the corresponding time stamps; and 3) post-processed data, including the frequency, the peak flux density, the dc bias, the duty ratio, the temperature, the volumetric power loss, and the single-cycle B - H loop sequences. The frequency is post-calculated from the data using the Welch's method [43], which estimates the power spectral density of the signal, and identifies the frequency with the highest power spectral density near the commanded frequency as the fundamental frequency. The flux density is calculated by the integral of the voltage signal together with the geometry parameters. The duty ratio is detected based on the zero-crossing point for each section. The single-cycle B - H loop data is produced by averaging the waveforms of different periods across the entire sequence and applying a 100-step interpolation within the averaged waveform. The single-cycle data captures the majority shape of the B - H loop with much less amount of data, but meanwhile loses resolution, especially near the switching events. [14] provides extended details about the data processing methods used to construct MagNet.

Data are open-sourced in four different formats, including ".mat", ".json", ".hdf5", and ".csv". This data structure is designed to contain sufficient information that facilitates the

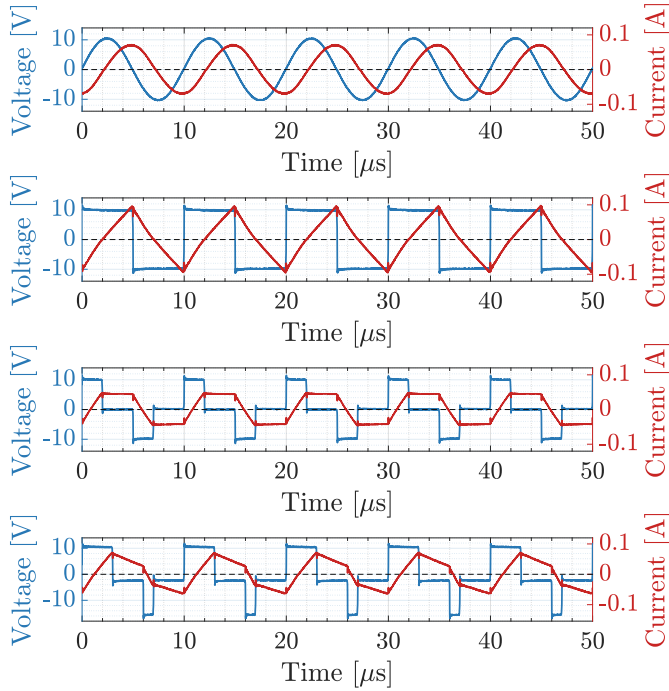


Fig. 6. Example voltage and current waveforms of sinusoidal, triangular, symmetric trapezoidal, and asymmetric trapezoidal excitations.

	Field	Value
DUT Information	Date_processing	'2022-08-10'
	Material	'N87'
	Core_Shape	'R34.0X20.5X12.5'
	Effective_Area	8.2600e-05
	Effective_Volume	6.7780e-06
	Effective_Length	0.0821
Raw Time Series Data	Primary_Turns	5
	Secondary_Turns	5
	Voltage	24773x10000 double
	Current	24773x10000 double
Post-processed Data	Time	24773x10000 double
	Sampling_Time	1.0000e-08
	Frequency	24773x1 double
	Flux_Density	24773x1 double
	Hdc	24773x1 double
	Duty_P	24773x1 double
	Duty_N	24773x1 double
	Temperature	24773x1 double
	Volumetric_Loss	24773x1 double
	B_Field	24773x100 double
H_Field	24773x100 double	

Fig. 7. Data format of the MagNet with four different types of contents.

research community to compare, verify and reproduce the core loss measurement, and trace the potential error mechanisms in the automatic data acquisition process.

Table II lists the size of the MagNet dataset in its current state. The sizes of the data for the ten materials are slightly different because of their various designated operation ranges for the parameter sweeping. Details about the range of measurement (e.g., flux density, frequency, dc bias, and temperature) are provided in Appendix. A. The total number of data points is more than 500,000 so far. Measurements for

TABLE II
NUMBER OF DATA POINTS CURRENTLY IN THE MAGNET DATASET

Material	Sine	Tri.	Trap.	Total
TDK N27	1,612	13,480	27,856	42,948
TDK N30	741	4,254	9,139	14,134
TDK N49	1,392	13,591	26,185	41,168
TDK N87	3,495	46,973	92,403	142,871
Ferroxcube 3C90	4,008	34,833	69,653	108,494
Ferroxcube 3C94	5,130	35,442	73,119	113,691
Ferroxcube 3F4	925	18,210	31,495	50,630
Ferroxcube 3E6	503	2,045	44,48	6,996
Fair-Rite 77	1,115	9,316	19,555	29,986
Fair-Rite 78	1,000	7,437	15,654	24,091
Total	19,921	185,581	369,507	575,009

other materials are in progress and the scale of MagNet dataset is expanding constantly.

Figure 8 illustrates the magnetic core loss density of N87 ferrite material as an example to visualize MagNet. The magnetic core is excited with triangular excitations with different duty ratios. Fig. 8a shows the core loss variation against the peak flux density with the frequency fixed at 200 kHz, Fig. 8b illustrates the variation against the frequency with the peak flux density approximately fixed at 120 mT, and Fig. 8c presents the variation against the duty ratios at different flux density level with the frequency fixed at 200 kHz, all of which are measured at 25°C. Fig. 8d depicts the core loss variations at different temperatures, with the duty ratio fixed at 0.5 and the frequency at 200 kHz. Each figure demonstrates a different nonlinear relationship in terms of different impact factors, and these factors typically co-exist in real applications. An extended discussion on these impacts is provided in [14]. The complexity of power magnetics characteristics motivates the use of machine learning.

An open-source webpage-based platform with a graphical user interface (GUI) – MagNet¹ – has been developed. The MagNet platform offers access to searching, visualizing, and cloning all the aforementioned measured datasets. It also provides a user-friendly interface to calculate and simulate the magnetic core loss using neural network models introduced in Sec. V with the support of a PLECS simulation engine. The website, models, and datasets have been open-sourced in GitHub². More details about the MagNet platform are provided in Appendix C.

V. NEURAL NETWORK MODELS

The MagNet database can be used in many different ways. For power magnetic designs with sinusoidal, triangular, or trapezoidal excitations, one can simply plot the data and read the core loss under a particular operating condition, and use the values in the design process with or without interpolation. MagNet can also be used to develop equation-based analytical models for magnetic core loss, such as identifying the Steinmetz parameters, forming a loss map, or extracting parameters of the Jiles-Atherton model. In this paper, however,

¹Princeton MagNet website: <https://mag-net.princeton.edu/>

²MagNet GitHub repository: <https://github.com/PrincetonUniversity/magnet/>

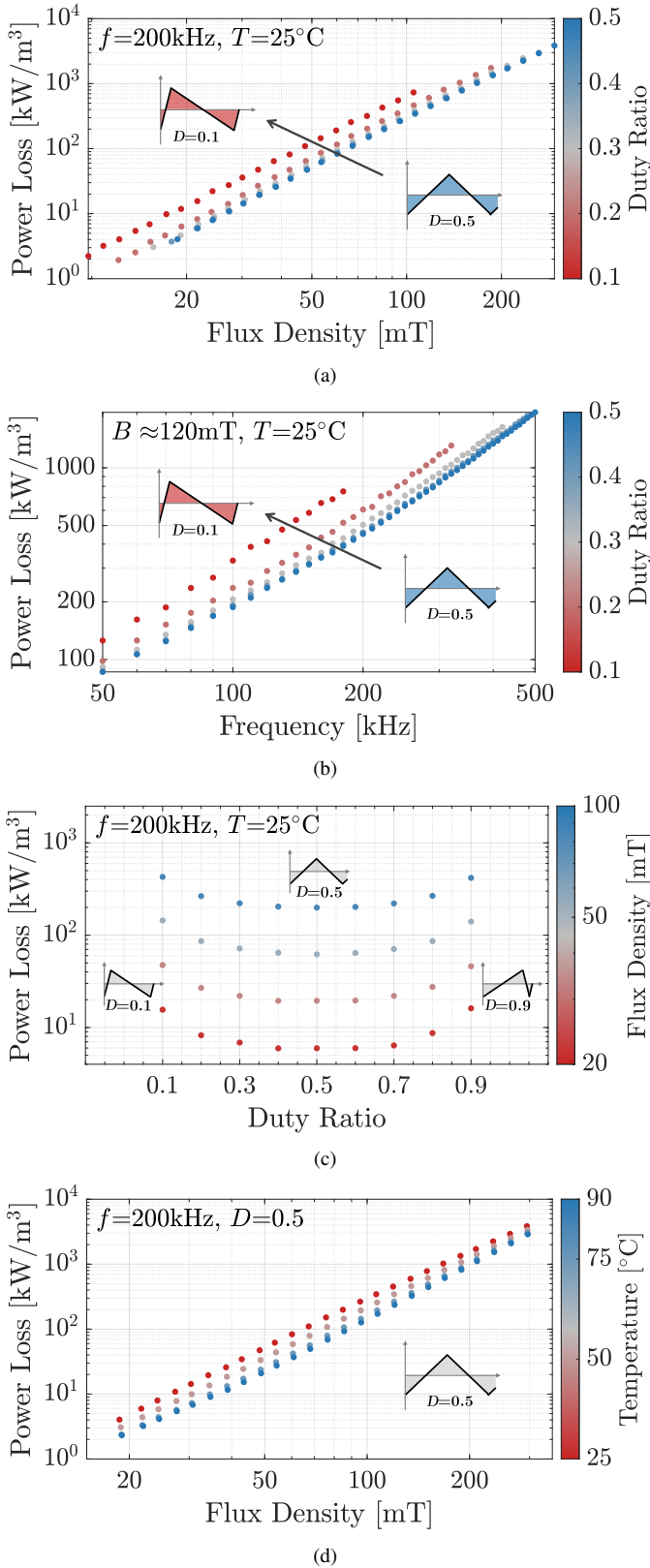


Fig. 8. Data visualization of the measured core losses under triangular excitation for N87 material: (a) core loss versus peak flux density with frequency at 200 kHz; (b) core loss versus frequency with peak flux density around 120 mT; (c) core loss versus duty ratios at different flux density level with frequency at 200 kHz; (d) core loss versus peak flux density at different temperature with frequency at 200 kHz and duty ratio at 0.5.

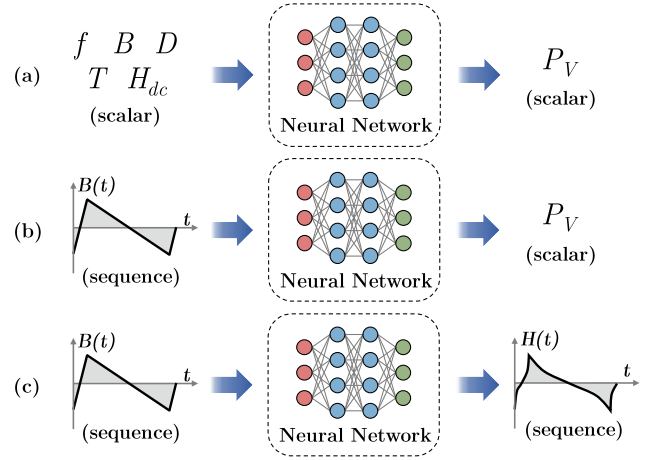


Fig. 9. Three example ways of modeling the behavior of magnetic materials with neural networks: (a) scalar-to-scalar, (b) sequence-to-scalar, and (c) sequence-to-sequence.

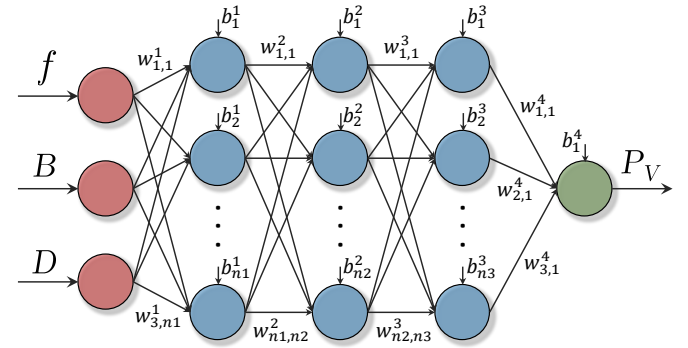


Fig. 10. Structure of an example 4-layer feed-forward neural network with three inputs (f, B, D) and one output (P_V). The structures and number of neurons in the hidden layers can be optimized. There is a trade-off between model size and model accuracy.

we demonstrate and highlight the neural network modeling method of power magnetics based on the MagNet database. As illustrated in Fig. 9, we explore three ways of modeling the behavior of magnetic materials with neural networks:

(a) **Scalar-to-scalar**: similar to the Steinmetz Equation, a neural network can be implemented as a scalar-to-scalar model to map multiple scalars, such as the frequency, peak flux density, and duty ratio, to a scalar value describing the magnetic core loss. This type of model is suitable for predicting the core loss under specific types of waveform shapes and operating conditions.

(b) **Sequence-to-scalar**: similar to the improved Generalized Steinmetz Equation (iGSE), a neural network can function as a sequence-to-scalar model, which takes the entire excitation waveform (e.g., flux density) as the input, and builds a regression mapping to the scalar value of the core loss. Compared to the scalar-to-scalar model, this type of sequence-to-scalar model is more feasible for core loss prediction under arbitrary excitation waveforms. It is no longer required to extract parameters from the waveform, reducing errors.

(c) **Sequence-to-sequence**: similar to the Jiles-Atherton

model [44], a neural network can also function as a sequence-to-sequence model to predict the magnetic responses (e.g., $B(t)$) due to an excitation waveform (e.g., $H(t)$). A sequence-to-sequence model can be potentially included in time-domain circuit simulations, such as SPICE.

A. Scalar-to-Scalar Model: Feed-forward Neural Network

Feed-forward neural networks (FNNs) are some of the simplest and most widely used artificial neural networks, and have proved to be effective in solving multi-variable nonlinear regression problems. As illustrated in Fig. 10, an FNN comprises one input layer, one output layer, and multiple hidden layers. The connections of the parameters in the feed-forward neural network can be described as:

$$z_i^j = \sigma \left(\sum_{k=1}^n (w_{k,i}^j \cdot z_k^{j-1} + b_k^j) \right) \quad (6)$$

where w is the weight between each pair of hidden neurons, and b is the bias of each hidden neuron. The subscript stands for the index of hidden neurons, and the superscript stands for the index of hidden layers. The function $\sigma(x)$ is the nonlinear activation function of the hidden neuron, which provides the network with the capability of learning non-linearity. z calculates the output value of each hidden neuron, where the subscript and the superscript share the same definitions as those of w and b .

We use a 4-layer FNN as an example to develop a scalar-to-scalar core loss model for ferrite materials under triangular excitations at a fixed temperature without dc bias. This particular network has one input layer, one output layer, and three hidden layers. The input layer takes three post-processed scalar parameters as the input variables: the fundamental frequency f , the peak flux density B , and the duty ratio D of the triangular waveform. The output layer has one parameter: the magnetic core loss density of the material P_V . Given the pre-known fact that the core loss P_V approximately changes exponentially in terms of f and B , these three variables are transformed into the logarithm values to enable a better convergence of the network. Each of the three hidden layers has multiple neurons. This model has a similar input-output configuration as the Steinmetz equation but has much more parameters available to function across a wide operating range.

The network model is synthesized and trained with PyTorch [45]. The activation function is ReLU. The loss function for the network training is selected as the mean-squared error (MSE) of the logarithm value of core loss to ensure uniform performance across the different orders of magnitude of the core loss in the operation range. The training optimizer is set as Adam [46]. An exponentially decayed learning rate strategy is implemented to yield a better model convergence, where the initial learning rate is 0.02 and the decaying rate is 50% per 200 epochs. To start with, the dataset of N87 ferrite material in the MagNet database is selected, and the dataset is randomly split into two parts as 80% and 20%. The first part is further split into five subsets to conduct a K -fold ($K = 5$) training and cross-validation of the candidate networks, while the second part is kept aside untouched as the test set.

TABLE III
PERFORMANCE OF A FEW DIFFERENT FNN MODELS WITH DIFFERENT SIZES FOR N87 MATERIAL WITH TRIANGULAR-WAVE EXCITATION

Neuron Search Range	Neurons in Hidden Layers	Total Number of Parameters	Avg. Relative Error on Test Set	Max. Relative Error on Test Set
[1, 4]	(2,1,3)	21	19.76%	68.59%
[4, 8]	(5,8,4)	109	9.81%	47.72%
[8, 16]	(15,15,9)	454	5.33%	21.36%
[16, 32]	(29,27,23)	1594	1.81%	9.47%
[32, 64]	(44,57,47)	5515	1.77%	8.62%

To determine the number of neurons in each hidden layer, we need to consider the trade-off between the network size and network performance. A small network may have limited learning capability and cannot provide good predictions, while a large network is more prone to overfitting, and also require a larger amount of training data. To analyze the relationship between the network size and prediction performance, we first set multiple boundaries for the number of neurons in each layer. A hyperparameter optimization tool – Optuna [47] – is selected to automatically search for the optimal number of neurons in each layer within each range. Table III lists the search range for the number of neurons in each layer, the local optimal number of neurons, and the average and maximum relative error of the prediction results on the test set for five FNNs with different scales. The total number of parameters in each neural network is also listed. As expected, prediction performance boosts as the size of the neural network increases.

Specifically, three neural networks with different numbers of hidden layers neurons are selected for the case study, including NN(2,1,3), NN(5,8,4), and NN(44,57,47), noted as a small-scale network, a medium-scale network, and a large-scale network, respectively. Fig. 11 compares the predicted core loss curves of the three different neural network models for N87 material under triangular excitation with different duty ratios. The results demonstrate that a small neural network is only capable of capturing part of the magnetic core loss characteristics (e.g., the slope of core loss curves), but unable to distinguish the impact of duty ratios. As the scale of the network increases, the model starts to capture the nonlinear impact of duty ratio, and eventually achieves a very close match with the measured core loss curves, given a large number of hidden neurons. This validates the effectiveness of feed-forward neural network modeling on magnetic core loss under certain shapes of excitation waveforms.

This FNN-based model can be easily extended with more input variables to capture other impact factors, such as the temperature and the dc bias, while special attention must be paid when considering multiple variables with different orders of magnitude, in which case the normalization operation is necessary in order to uniform the data distribution, as well as avoid undesired numerical problems during the training.

One of the limitations of this FNN-based model is that the prediction relies on the scalar representation of waveform shapes. Different types of waveform shapes may require a different number of scalars to be described. In the example above, the shape of the triangular wave is abstracted by the duty ratio.

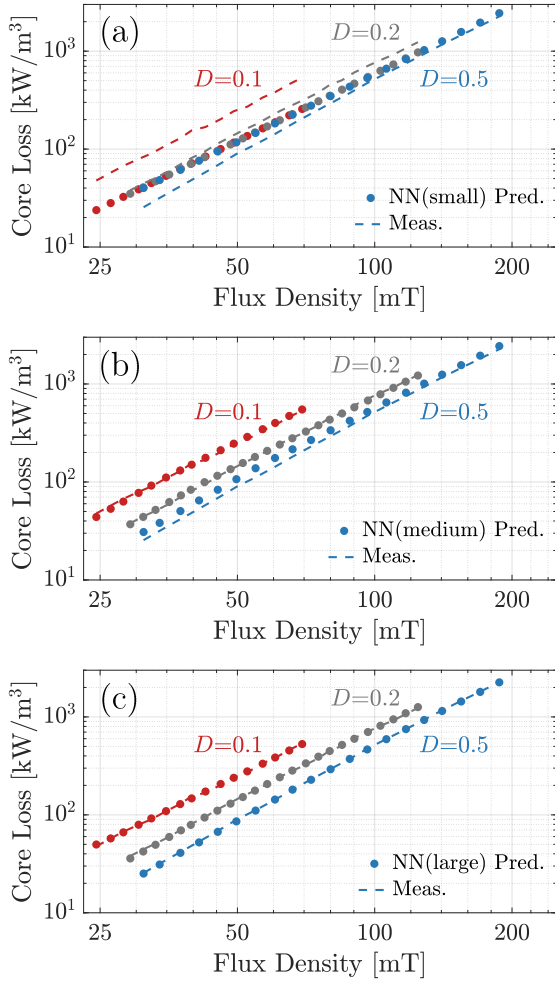


Fig. 11. Prediction results of three neural network core loss models for the N87 material at 300 kHz: (a) small scale; (b) medium scale; (c) large scale. The prediction accuracy increases as the number of hidden neurons scales up.

Other piece-wise waveforms such as the trapezoidal wave require numerous duty ratios to be fully described. However, there are also various types of nonlinear waveforms that cannot be described simply with scalars, which limits the applicability of scalar-to-scalar model, and motivates sequence-to-scalar and sequence-to-sequence models.

B. Sequence-to-Scalar Model: LSTM Network

It is important to consider the waveform information when modeling power magnetics with non-sinusoidal excitations. The long short-term memory (LSTM) network [48], [49] is one of the most commonly used neural networks for regression problems with sequential input. An LSTM has feedback connections with the capability of memorizing information across the sequences of data. LSTM networks are well-suited for classifying and making predictions based on time series data, especially if there are sophisticated correlations (e.g., memory effects) in the time domain. Modeling the unclear sequential causality relationships between $B(t)$, $H(t)$, and core losses is LSTM models' forte.

Figure 12a shows the basic structure of a standard LSTM cell. The fundamental mechanism of an LSTM cell that distinguishes it from other types of recurrent neural networks is the implementation of the input gate i_t , the forget gate f_t , and the output gate o_t . With these gates, the cell is able to regulate the information flow and selectively memorize the important information across a specific time interval within the sequence rather than across the entire sequence. Mathematically, the operation of the LSTM cell at time t can be described as:

$$\begin{aligned}
 f_t &= \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \\
 i_t &= \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \\
 g_t &= \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}) \\
 o_t &= \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \\
 c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
 h_t &= o_t \odot \tanh(c_t)
 \end{aligned} \tag{7}$$

where c_t and h_t refer to the cell states and the hidden states, respectively. These states are the recurrent variables that will be fed back to the LSTM cell and thus provide the memorizing capability. The function $\sigma(x)$ is the Sigmoid function that operates as the activation function to provide the nonlinear learning capability. As in an FNN, W and b are the weights and biases, and the subscript refers to the source and target variables that they are applied to. The operator \odot stands for the Hadamard product, which performs an element-wise product for all the elements of two matrices.

To start, based on the inputs x_t and the previous hidden states h_{t-1} , the forget gate f_t determines to what extent the cell states impact the calculation at the current time step. Then similarly, the input gate i_t and the cell gate g_t jointly determine how the cell states are updated. Finally, the output gate o_t regulates and updates the hidden states h_t , which are typically considered as the output of an LSTM cell.

In this particular example, we develop an LSTM-based sequence-to-scalar model with time sequences of excitation flux density $B(t)$ as the input (full waveform with multiple cycles, fixed length, rather than the single-cycle data), and the volumetric power loss as the output. An example structure of the model is demonstrated in Fig. 12b. The input layer of the LSTM takes the entire flux density waveform as a sequence input. The output of LSTM is aggregated and loaded with an FNN to perform the core loss regression. In this example design, the LSTM has 32 cell states and 32 hidden states, while the FNN comprises three hidden layers. The output of the FNN is the volumetric magnetic core loss. This example model contains 5,569 parameters in total.

To validate the effectiveness of this LSTM-based core loss model, the network shown in Fig. 12 is synthesized and trained with PyTorch. A merged dataset that contains all three types of waveforms (sinusoidal, triangular, and trapezoidal) for N87 material is now selected, instead of the single-shape waveform in the scalar-to-scalar cases. Due to the fact that different types of waveform shapes have different degrees of freedom (e.g., the amplitude, the frequency, and the duty ratio), the numbers of original data points for the sinusoidal wave, the triangular wave, and the trapezoidal wave vary significantly, as listed in Table II. Such an unbalanced dataset may potentially impact

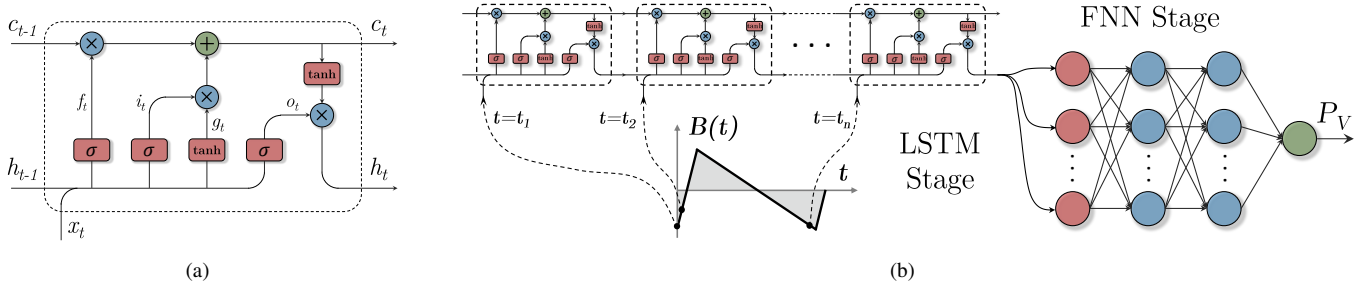


Fig. 12. The network structure of the long short-term memory (LSTM) network for sequence-to-scalar model: (a) the basic structure of a standard LSTM cell, which contains the input gate, the forget gate, and the output gate; (b) the structure of an example LSTM-based sequence-to-scalar magnetic core loss model, which takes the time sequence as inputs.

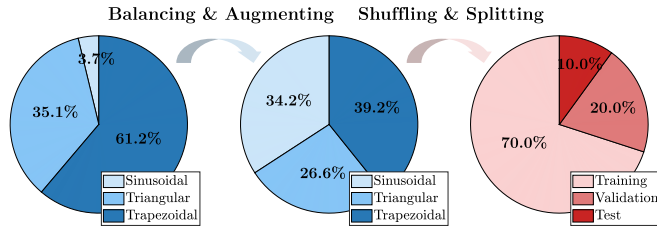


Fig. 13. The data preparation process of sequence-based network training. The original dataset is augmented and balanced by assigning a random phase shift, then shuffled and randomly split into the training set, the validation set, and the test set.

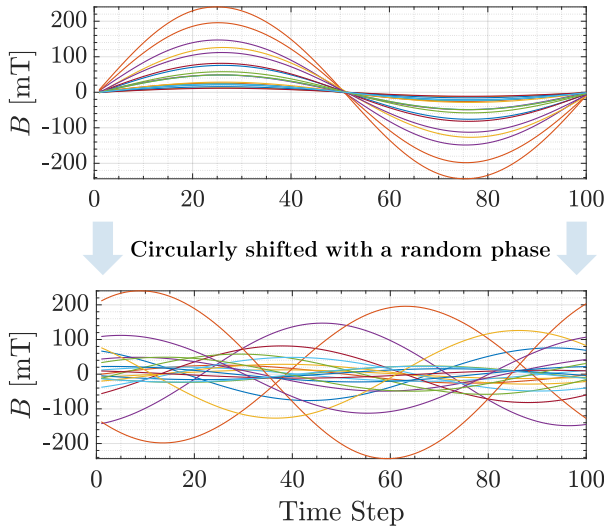


Fig. 14. An example illustration of the data augmentation for sinusoidal waveforms. Each waveform sequence is circularly shifted with a random phase. Theoretically, the magnetic core loss does not change in the steady state regardless of the starting phase.

the performance of the neural network model. Data augmentation techniques can be used to increase the scale of the original dataset and balance the data distribution. In this work, the augmentation technique consists of circularly shifting and adding noise, where each of the waveforms is circularly shifted with a random phase [50] and then superposed with white noise. Theoretically, the magnetic core loss stays constant in the steady state regardless of the starting phase. Hence, this

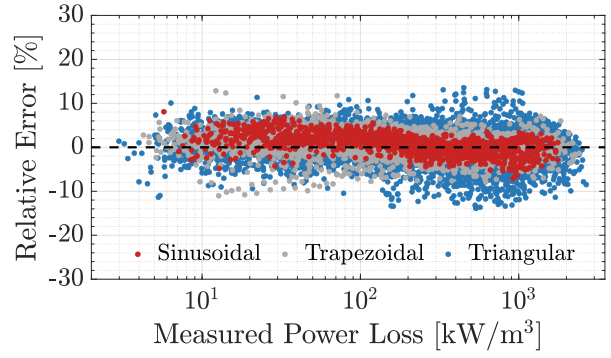


Fig. 15. The distribution of the testing relative error for the LSTM-based core loss model on the N87 ferrite material.

TABLE IV
LSTM MODELING RESULTS FOR N87 MATERIAL AS CORE LOSS ERROR

Type of Waveform	Abs. Avg. of Relative Errors	RMS of Relative Errors	Maximum of Relative Errors
Sinusoidal	1.86%	2.33%	8.08%
Triangular	2.41%	3.19%	14.35%
Trapezoidal	1.95%	2.58%	12.85%
Overall	2.09%	2.78%	14.35%

data augmentation with random phase assignments also helps to enhance the neural network's capability of characterizing the intrinsic features of the waveforms, rather than simply memorizing the waveforms. The augmented dataset is further shuffled and randomly split into the training set, the validation set, and the test set for the LSTM model with a ratio of 70%, 20%, and 10%. Fig. 13 and 14 illustrate the process of data augmentation, balancing, and shuffling. Other techniques, such as weighting the loss function, can also help to tackle the unbalance of the dataset, yet are beyond the scope of this work.

Figure 15 shows the error distribution between the measured core loss and the predicted core loss achieved by the LSTM-based model. More prediction results are listed in Table IV. As observed, the proposed LSTM model achieves a good core loss prediction accuracy for all three types of waveforms, where the relative error approximately has an even and unbiased distribution that is close to 0%. The overall absolute average of relative error is around 2% and the maximum relative error is

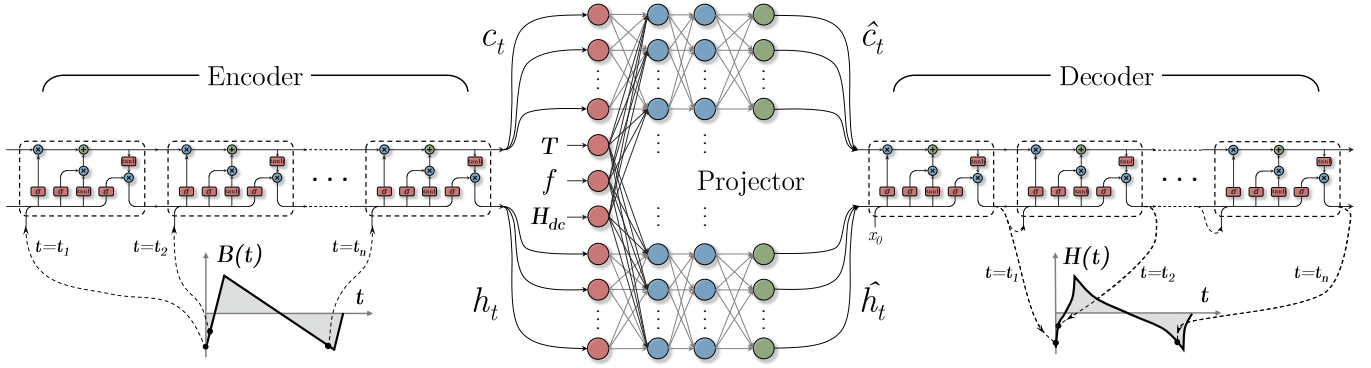


Fig. 16. Structure of an example encoder-projector-decoder network architecture that predicts the response sequence $H(t)$ based on the excitation sequence $B(t)$ and additional inputs of temperature T , frequency f , and dc bias H_{dc} .

within 15%. The LSTM model contains 5,569 parameters in total, which is close to the largest FNN mentioned in Sec. V-A. Nevertheless, this LSTM-based model is able to make core loss predictions for all three types of waveforms and beyond. The applicability of this LSTM model is not restricted by the scalar representation of waveform shapes. It can be inferred to predict the core loss for excitation waveforms that are not precisely included in the training data. This ability to generalize to arbitrary waveforms is an important objective, as with the iGSE and similar methods.

C. Sequence-to-Sequence Model: Encoder-Decoder Network

The above example proves the effectiveness of LSTM networks for solving sequence-to-scalar problems, extending from which we can further explore the concept of the sequence-to-sequence model in order to capture the magnetic material behavior more comprehensively.

The encoder-decoder network architecture [51] is one of the state-of-the-art architectures for solving sequence-to-sequence regression problems that attracts significant attention in recent years. It has proved to be successful in applications such as voice-to-voice translation and stock-price correlation, both of which take a time sequence as the input and map it to another time sequence as the output. The modeling of the magnetic B - H hysteresis behavior is faced with similar problems as the aforementioned applications. However, the modeling of the B - H loop can be more complicated to some extent, as it is also impacted by other factors besides the time sequence itself, such as the temperature and the dc bias.

Here we proposed an encoder-projector-decoder network architecture as shown in Fig. 16 for B - H loop regression. The encoder is composed of a 1-layer 32-state LSTM network. Leveraging the input gate, forget gate, and output gate mechanism of the LSTM, the input sequence $B(t)$ passes through the encoder in such order where the encoder captures and saves the characteristics within the sequence as the hidden states and cell states. The hidden states and cell states are then passed into the projector, which consists of a 3-layer FNN with 64 hidden neurons in each layer. The additional inputs, such as the temperature T , the frequency f , and the dc bias H_{dc} , are concatenated with the states and fed to the projector at the

same time. In order to avoid the potential mismatch among the scales of different input variables, a normalization process is implemented. The FNN projector modifies the state values according to the values of these additional inputs and generates the updated state values. The updated values are further loaded into the decoder to initialize its states. The decoder has the same network structure as the encoder. Based on the updated values of hidden states and cell states, the decoder generates the output sequence $H(t)$.

The difference between the encoder and the decoder is that the encoder processes the entire input sequence all at once, while the decoder generates the output sequence step-by-step, which is also known as auto-regressive inference or walk-through validation. To start with, a default value is fed to the decoder to initialize the output and generate the first time step of the output sequence. Then the generated value is fed back to the decoder to generate the next time step in iteration, until the entire output sequence is generated.

The mean-squared error (MSE) between the predicted sequence and the target sequence is selected as the loss function to update the weights and biases in the network. Other loss functions are also feasible, such as a higher-order power of error to penalize more on the data points with larger amplitude, a relative loss function to balance the error distribution across the entire amplitude range, or a phase-related loss function to minimize the phase mismatch, which is critical to the accurate prediction of the magnetic core loss. Here, we present two training examples to demonstrate the effectiveness of the proposed encoder-projector-decoder network architecture:

1) *B - H loop prediction for different excitation waveforms:* The first case study is to predict the B - H loops under different excitation waveforms. The input sequences are the flux density waveforms in sinusoidal, triangular, and trapezoidal shapes, whose amplitude varies from 10 mT to 300 mT. The additional input for the projector is the fundamental frequency of the excitation waveform in the range of 50 kHz to 500 kHz. The temperature is fixed at 25 °C and the dc bias is set to zero. Each pair of B - H sequences is circularly shifted with a random phase, as mentioned in Sec. V-B, to augment the data and minimize the impact of phase. The entire dataset contains 15,327 pairs of B - H sequences, which are split into

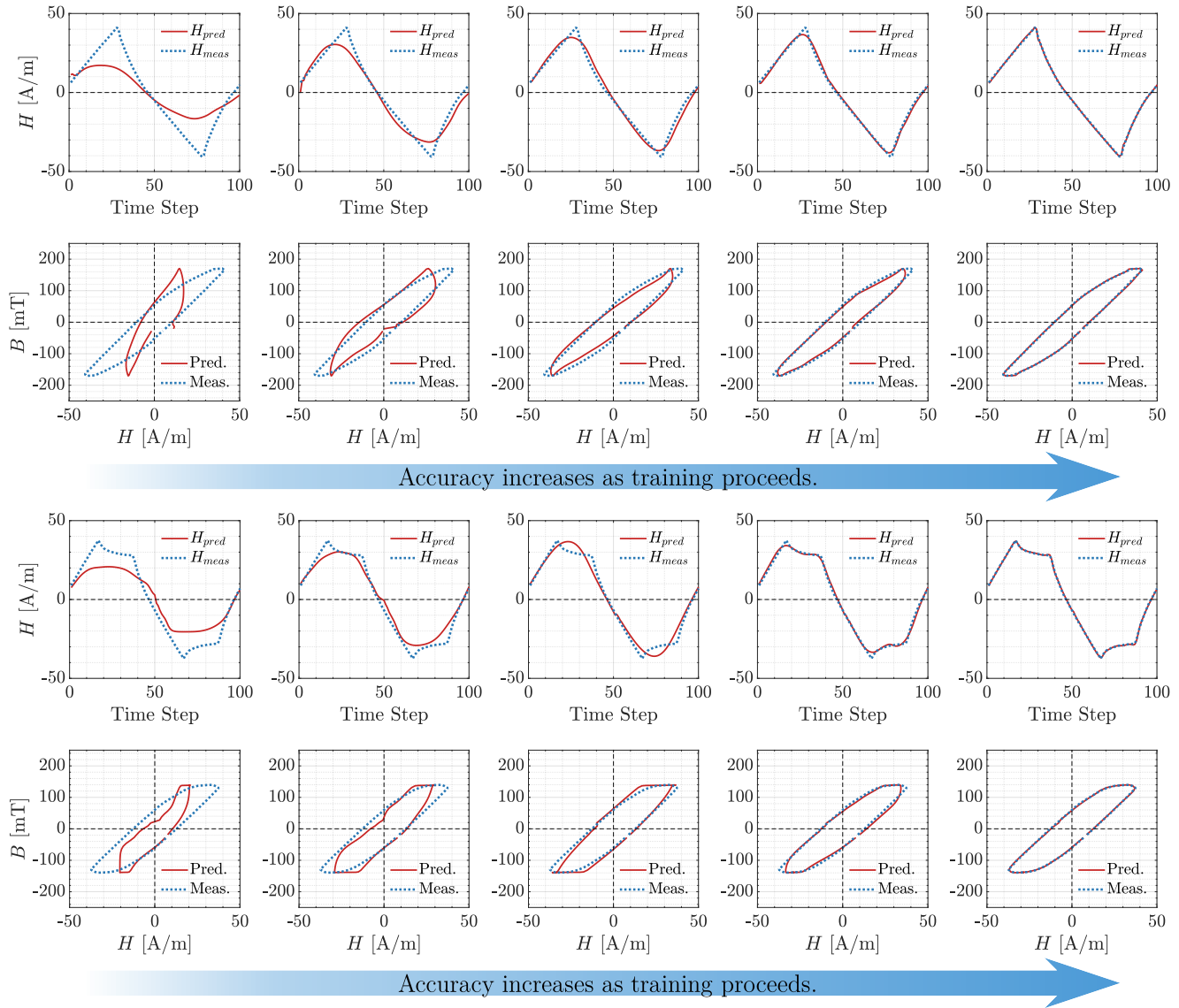


Fig. 17. Triangular and trapezoidal output sequences (N87 ferrite, 150 kHz, 25 °C) and the corresponding B - H loops predicted by the sequence-to-sequence model as it gets trained by the experimental data. The accuracy of the prediction improves with the training proceeding as the NN learns more and more data.

the training set (70%), the validation set (20%), and the test set (10%). The data is shuffled before each training process.

Figure 17 shows two examples of the comparison between the predicted sequence and the target sequence at different stages of the training. In the early stage of the training, the predicted sequence greatly deviates from the target sequence. As the training continues, the mismatch between the predicted sequence and the target sequence is gradually corrected, and eventually achieves a good match. Similar trends can be observed from the predictions of B - H loops.

To quantitatively evaluate the prediction accuracy, the rela-

tive error of sequence matching is defined as:

$$\begin{aligned} \text{Relative Err. of Sequence} &= \frac{\text{rms}(H_{pred} - H_{meas})}{\text{rms}(H_{meas})} \\ &= \frac{\sqrt{\frac{1}{n} \sum_{t=t_1}^{t_n} (H_{pred}(t) - H_{meas}(t))^2}}{\sqrt{\frac{1}{n} \sum_{t=t_1}^{t_n} (H_{meas}(t))^2}} \quad (8) \end{aligned}$$

The model is evaluated with the test set based on this definition. Fig. 18 shows the histogram of the relative error. Overall, a 3.73% average relative error of the sequence-to-sequence matching is achieved, and the maximum relative error is 18.29%. The proposed sequence-to-sequence model is capable of accurately predicting the response $H(t)$ sequence given the excitation $B(t)$ sequence across a wide frequency range and under multiple types of excitations.

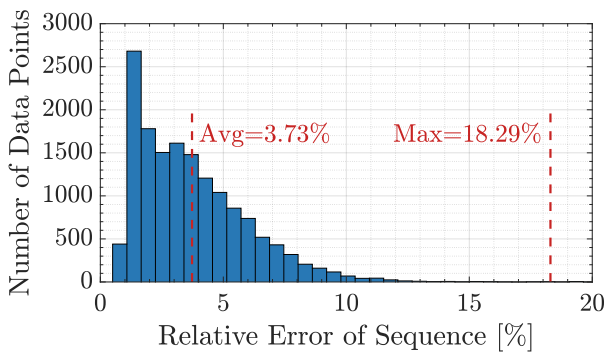


Fig. 18. The distribution histogram of the relative error of the sequence matching for N87 material under multiple types of waveforms and at 25 °C.

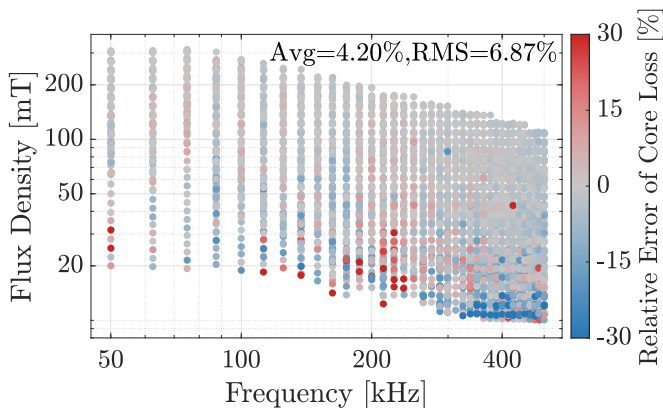


Fig. 19. The error distribution of the predicted core loss density based on the predicted B - H loops for N87 material under multiple types of waveforms and at 25 °C.

Moreover, with the predicted B - H loops, one can calculate the area of the loop as the core loss density with the integral

$$P_V = \frac{1}{T} \int_{B(0)}^{B(T)} H(t) dB(t) \quad (9)$$

to compare the predicted core loss against the measured core loss. Fig. 19 illustrates the distribution of the relative error of the predicted core loss on the f - B plane. Most of the data points are predicted with a low relative error below 15%. The high-error cases mainly concentrate in those areas with very low flux density or frequency, where the core loss itself is extremely small and close to the limit of the equipment.

It should also be noticed that the accuracy of the sequence-to-sequence regression is not necessarily consistent with that of the core loss prediction. Intuitively, a small phase mismatch in the predicted sequence may or may not lead to a large relative error in sequence-to-sequence regression, but can significantly impact the core loss prediction.

2) *B - H loop prediction under different temperatures:* The second example is to predict the B - H loops under different temperature conditions. The shape of the input flux density sequence is fixed as the sinusoidal wave, while the additional inputs of the projector now contain the temperature at different values, including 25 °C, 30 °C, 50 °C, 70 °C, and 90 °C, together with the frequency of the corresponding sequence and

zero dc bias. Phase-shift data augmentation is implemented as aforementioned. The dataset contains 4,359 pairs of B - H sequences, which are similarly split into the training set, validation set, and test set.

Figure 20 shows an example comparison between the predicted sequence and the target sequence at different stages of the training process, as well as the corresponding B - H loops. When the training begins, the predicted sequence deviates greatly from the target sequence. As the training continues, the mismatch between the predicted sequence and the target sequence is gradually reduced, and eventually, a good match is achieved. Using the same criteria defined in the above subsection, a 1.33% averaged relative error with a maximum of 8.32% of the sequence-to-sequence matching is achieved. Similarly, Fig. 21 demonstrates the histogram of the relative error. The predicted core loss is also calculated based on the predicted B - H loops, the error distribution of which is illustrated in Fig. 22 in the f - B - T space. The high-error cases concentrate in low flux density, low frequency, and high temperature areas, where the core losses are very small and a minor mismatch may lead to a large relative error.

D. Transfer Learning for Data Size Reduction

In the above examples, the large-scale database MagNet acts as a foundation that supports the training and testing of the data-driven models. Sometimes, however, it is unrealistic for designers to build a core loss measurement platform by themselves and collect a sufficiently large amount of data for model training, especially when dealing with new materials that only have a limited number of data points available, or dealing with operating conditions that are outside the ranges covered by the database or the capability of the equipment.

Transfer learning is a machine learning technique in which knowledge gained by solving one problem is applied to a similar problem. A major hypothesis behind applying transfer learning to magnetic modeling is that similar physical mechanisms govern the response of similar magnetic materials to similar excitations. As a result, one can train a generic neural network model that captures the common patterns and characteristics of magnetic materials, and further use it to support the development of the models for other new materials, excitations, temperature, or dc bias. Fig. 23 illustrates the basic principle of transfer learning. We demonstrate material-to-material and temperature-to-temperature transfer learning to elaborate the key concepts.

1) *Material-to-Material Transfer Learning:* Material-to-material transfer learning is helpful if a model for a new magnetic material is needed, and only a small amount of data for this new material is available. Transfer learning can reduce the required size of the dataset needed to achieve satisfactory accuracy with a neural network model.

Figure 24 illustrates three machine learning experiments to demonstrate the principles of transfer learning: (1) selecting four materials from the MagNet database (N27, N49, 3C90, 3C94) as the existing materials, and employing a large amount of their data to train a pre-trained model similar to the FNN trained in Sec. V-A. The data points of the four materials are

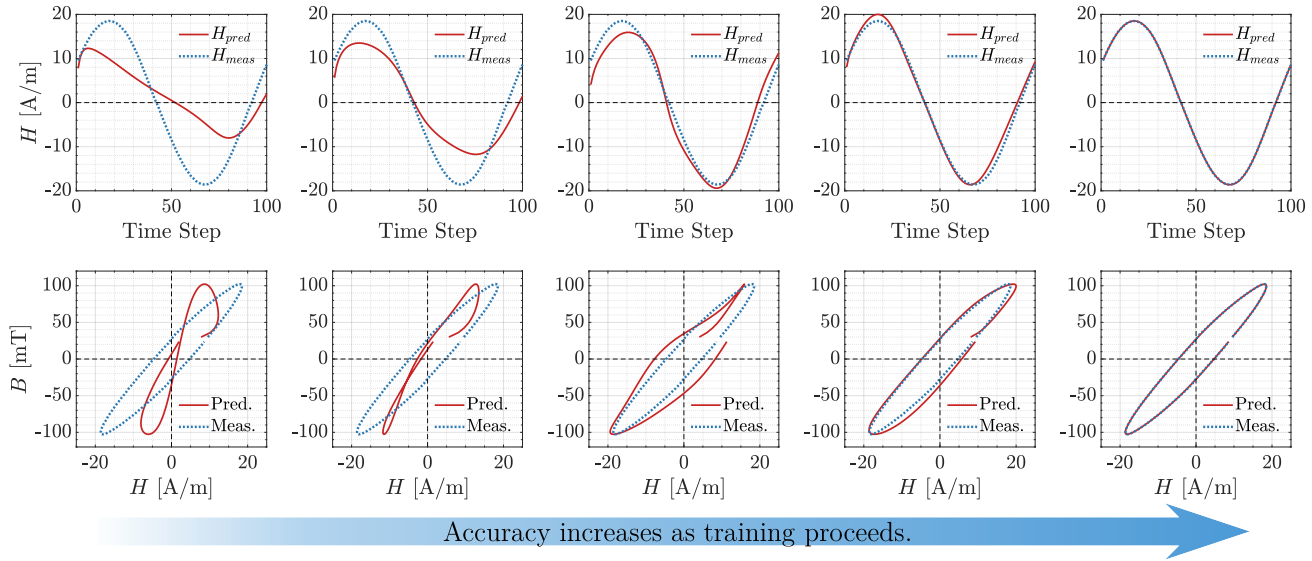


Fig. 20. Sinusoidal output sequence (N87 ferrite, 220 kHz, 50 °C) and the corresponding $B-H$ loops of the sequence-to-sequence model as it gets trained by the experimental data. The prediction accuracy improves with the training proceeding, indicating the NN is learning the patterns in the $B-H$ relationships.

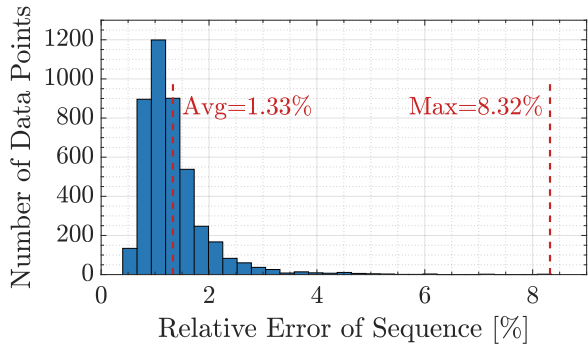


Fig. 21. The distribution histogram of the relative error of the sequence matching for N87 material under sinusoidal waveforms and at multiple temperature conditions.

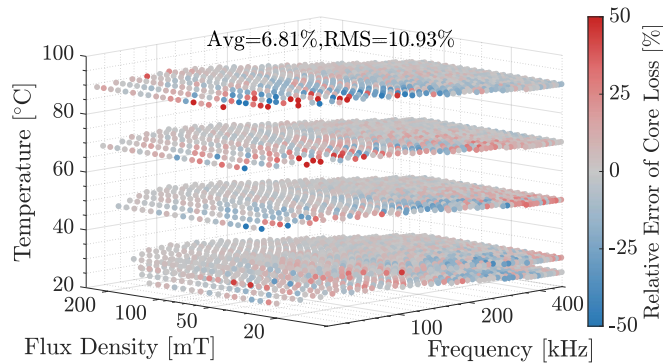


Fig. 22. The error distribution of the predicted core loss density based on the predicted $B-H$ loops for N87 material under sinusoidal waveforms and at multiple temperature conditions.

directly mixed into a larger dataset for the network training, while the material type itself is not used as the input; (2) selecting another material from the MagNet database (N87) as

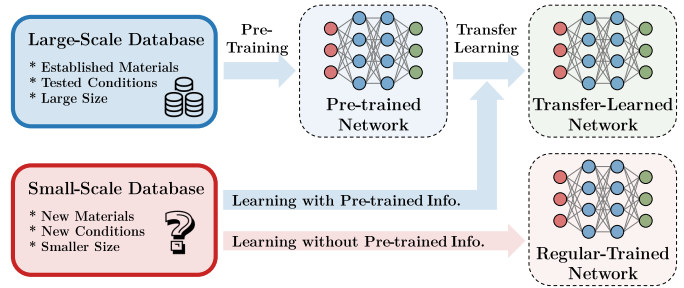


Fig. 23. Key principle of transfer learning for magnetic core loss modeling.

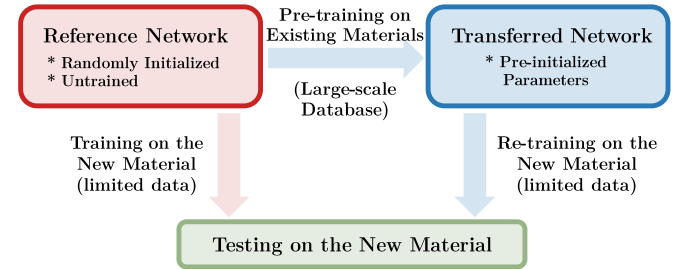


Fig. 24. Network training process of the material-to-material transfer learning.

the targeted new material that has a limited amount of core loss data, and re-training on the pre-trained model with only a small amount of data randomly selected from the database; (3) for comparison purpose, direct training of a randomly initialized neural network with the same small amount of data.

Figure 25 shows the results of material-to-material transfer learning for triangular 180 kHz excitations and three different duty ratios. The pre-trained model is trained on the large-scale data (30,705 data points in total) of the four existing materials (N27, N49, 3C90, 3C94). Fig. 25a shows the prediction results

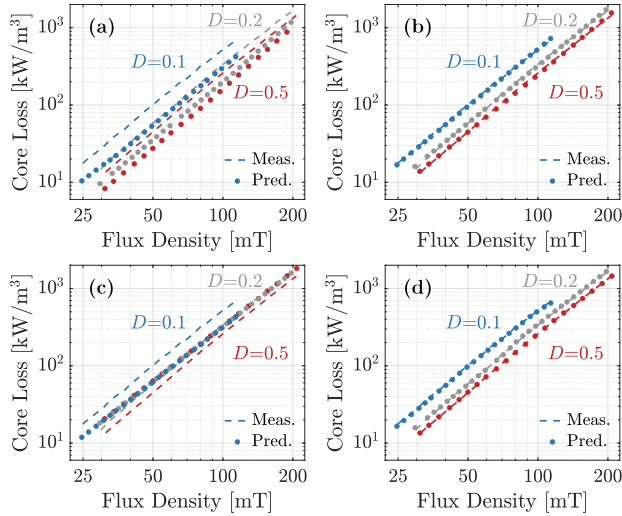


Fig. 25. Prediction results of: (a) applying a **pre-trained** model to the new material **without** re-training; (b) applying a **pre-trained** model to a new material **after** re-training with **very few** data points (100, randomly selected) from the new material; (c) applying a **randomly initialized** model trained only with **very few** data (100, randomly selected) from the new material; (d) applying a **randomly initialized** model trained with a **large amount** of data from the new material.

applying the pre-trained model directly to the new material (N87) before re-training. The pre-trained model can capture some common patterns of the magnetic core loss, such as the approximate exponential relationship between the core loss and the flux density, as well as the impact of the duty ratios, but failed to capture the details.

The pre-trained model is re-trained with 100 new data from the N87 material. Fig. 25b shows the updated prediction results. After re-training, the pre-trained model is greatly improved. The rationale behind the results is that the pre-training procedure provides a good starting point for the re-training, where the new data fine-tunes the model and greatly improves the model's accuracy. In comparison, Fig. 25c shows the results when the network is simply randomly initialized and only trained with 100 new data from the new material, without pre-training. It is observed that this model unsuccessfully captures the distribution of the magnetic core loss, and the predicted curves deviate greatly from the measured curves.

To provide a benchmark, a normal training process is also conducted, similar to the one described in Sec. V-A, where a randomly initialized neural network is trained with a large amount of data from the new material. This benchmark experiment achieves the highest prediction accuracy as expected, shown in Fig. 25d, among the four experiments aforementioned. The prediction results of the transfer-learned model, however, are almost comparably accurate as those of the benchmark model, despite the fact that only a small amount of data is available in this case. It proves the effectiveness of material-to-material transfer learning.

Figure 26 shows the overall error distribution of the normal training results and the transfer learning results, corresponding to Fig. 25b and Fig. 25c, respectively, where the duty ratio is selected to be 0.5. In the normal training case without pre-

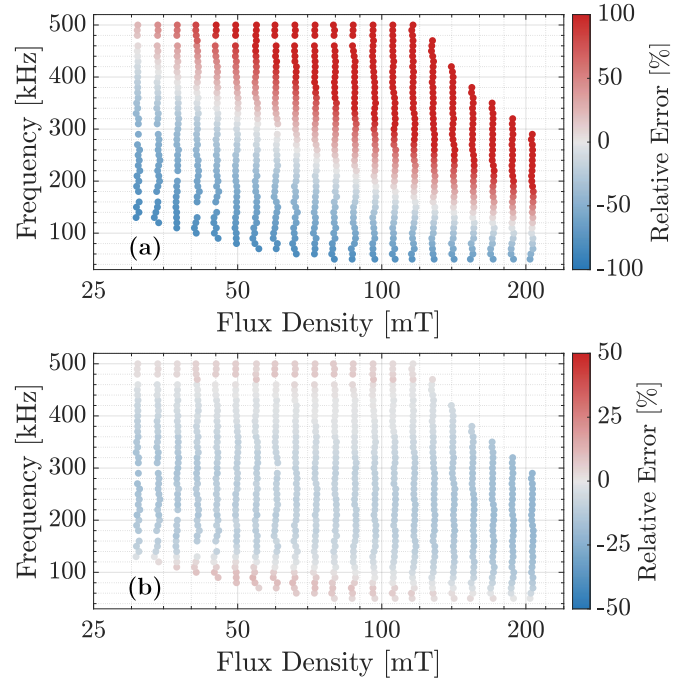


Fig. 26. Error distribution of the prediction results of: (a) applying a randomly initialized model trained only with 100 data points randomly selected from the N87 material data points (normal training); (b) applying a pre-trained model based on 4 existing materials to the N87 material data points after re-training with 100 randomly selected data points (transfer learning). The plotted points are the subset of data with a duty ratio of 0.5.

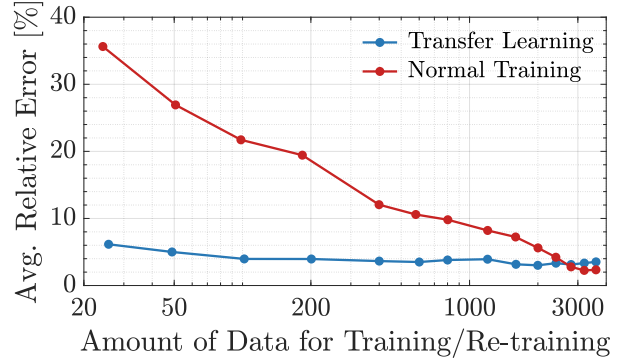


Fig. 27. Testing average relative error rates after training the normal FNN and re-training the pre-trained FNN with a varied amount of data.

training, the network performs poorly in most of the areas due to the limitation of available training data, resulting in a large average relative error of more than 50%. On the contrary, with transfer learning, the network achieves a reasonably good accuracy across the entire evaluation range, resulting in an absolute average error of 8.81%.

Furthermore, the pre-training, re-training, and testing processes are repeated many times while varying the number of data points available for the re-training step. Fig. 27 shows the testing average relative errors on the y-axis with the numbers of available data points on the x-axis ranging from 25 to 3,600. The percent errors are averaged over 10 trials to ensure consistency. The pre-trained neural networks

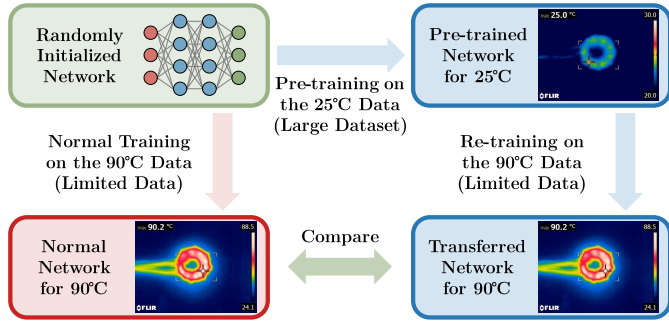


Fig. 28. Network training process of the temperature-to-temperature transfer learning. Pre-training and fine-tuning can greatly reduce the amount of data needed to model the power magnetics at a different temperature.

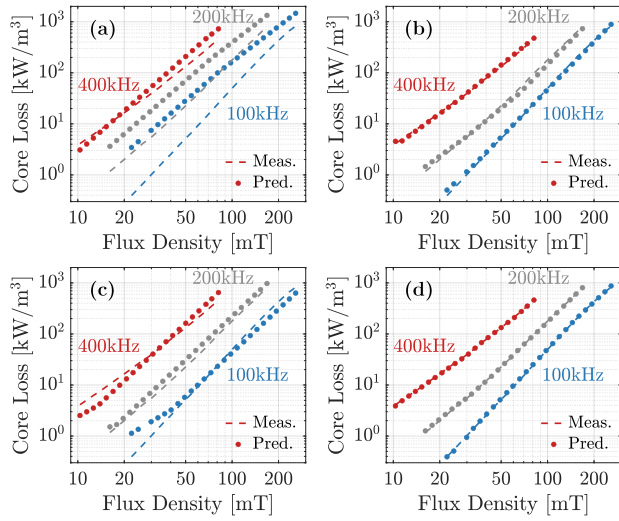


Fig. 29. Prediction results of: (a) applying a **pre-trained** 25 °C model to the 90 °C data points **without** re-training; (b) applying a **pre-trained** 25 °C model to the 90 °C data points **after** re-training with **very few** data points (10, randomly selected) from the 90 °C data points; (c) applying a **randomly initialized** model trained only with **very few** data points (10, randomly selected) from the 90 °C data points; (d) applying a **randomly initialized** model trained with a **large amount** of data (800) from the 90 °C data points.

constantly achieve good performance no matter if it is provided with 25 target data points or 3,600, whereas a normal and randomly initialized FNN requires at least 2,400 data points to consistently accomplish good and comparable performance with a similar error rate. The amount of data needed to retrain the neural network for a new material is significantly reduced by transfer learning.

2) *Temperature-to-Temperature Transfer Learning*: Temperature also greatly influences the behavior of magnetic materials. Using an established model based on the data measured under one temperature to predict the magnetic core loss under another temperature will lead to a significant mismatch. Temperature-to-temperature transfer learning method helps to build a neural network model that works for different temperature conditions, especially when the available data across different temperatures is limited.

The principles of temperature-to-temperature transfer learning are similar to that of material-to-material transfer learning.

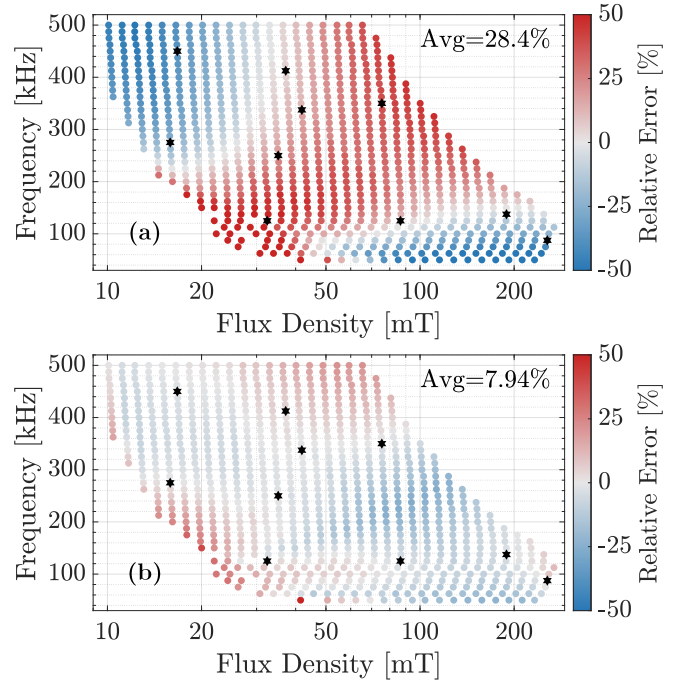


Fig. 30. Error distribution of the prediction results of: (a) applying a randomly initialized model trained only with 10 data points randomly selected from the 90 °C data points (normal training); (b) applying a pre-trained 25 °C model to the 90 °C data points after re-training with 10 data points randomly selected from the 90 °C data points (transfer learning).

Fig. 28 shows an example training process to transfer the model for 25 °C to a new one for 90 °C. In this example, the neural network is pre-trained for 500 epochs based on the core loss data of N87 ferrite material that measured at 25 °C with sinusoidal excitations. This source dataset is selected from the MagNet database, consisting of 800 data points. Then, the model is further re-trained and fine-tuned with a small number of data points that measured at 90 °C for 3,000 epochs. To make a comparison, a randomly initialized network without pre-training is also trained based on the same limited dataset with the same training settings. As a benchmark, both networks are trained with a large dataset containing 800 data points measured at 90 °C, which represents the cases where there are no limitations on the number of available data points. All the network models are again synthesized with the same structure as the NN(15,15,9) FNN mentioned in Sec. V-A.

Fig. 29 demonstrates multiple core loss curves predicted by different network models. In Fig. 29a, the model that pre-trained with the 25°C data points is directly evaluated with the 90°C data points without re-training, where the predicted core loss curves match poorly with the measured ones due to the temperature difference. After re-training with a small dataset that contains only ten 90°C data points, the model is effectively transferred, and the predicted curves match well with the measured ones, as shown in Fig. 29b. The prediction accuracy is comparable to that of the model demonstrated in Fig. 29d, which is trained with the large dataset. In comparison, the network model that is directly trained with the small dataset without pre-training only achieves a rough

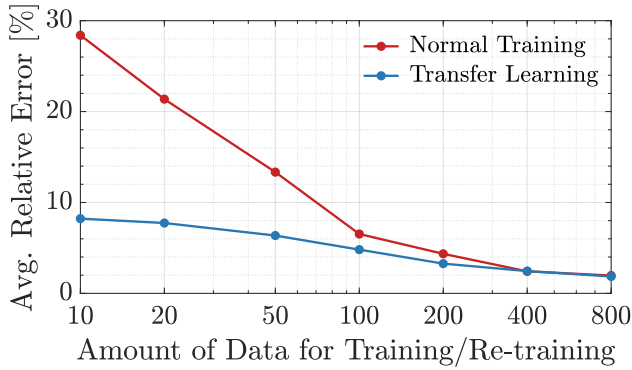


Fig. 31. Testing average relative error rates of the normal training and the transfer learning as the size of the new data increases.

prediction, as observed in Fig. 29c, where the accuracy is obviously inferior to that of the transfer learning case.

More specifically, Fig. 30 displays the overall error distribution of the normal training results and the transfer learning results, corresponding to Fig. 29b and 29c, respectively. The 10 data points included in the training and the re-training dataset are marked as black pentagrams. In the normal training case without pre-training, the network only achieves a low relative error in those areas around the 10 training data points, but performs poorly in other areas, resulting in an absolute average error of 28.4%. On the contrary, with the transfer learning process, the network not only achieves lower relative error in the areas around the 10 training data points, but also maintains a good accuracy across the entire evaluation range, resulting in an absolute average error of 7.94%.

The above temperature-to-temperature transfer learning and normal training process are repeated multiple times while sweeping the number of data points available for the training and re-training step. The number of data points in this limited dataset is selected from 10, 20, 50, 100, 200, and 400. Fig. 31 shows the testing average relative error of each case for both the normal training and the transfer learning. As observed, the amount of data needed for retraining the neural network for a new temperature is similarly reduced by transfer learning.

VI. NEURAL NETWORK AS DATASHEET

Classical datasheets for magnetic materials cannot provide sufficient information for all possible scenarios that designers may have to work with. Manufacturers sometimes provide design tools containing information that is not provided in their datasheets, where data can easily be interpolated and extracted (instead of having to read the information from graphs). However, these datasets, sometimes extensive in power loss information, still do not provide B - H loops data for a wide range of operating points and usually only work for sinusoidal waveforms. The data size increases rapidly as the number of variables increases. The capability of predicting the B - H loop under various operating conditions makes NNs a good candidate to be used as an active datasheet for power magnetics [4].

Figure 32 illustrates the concept of “neural network as datasheet”. As demonstrated in Sec. V-C, this NN would

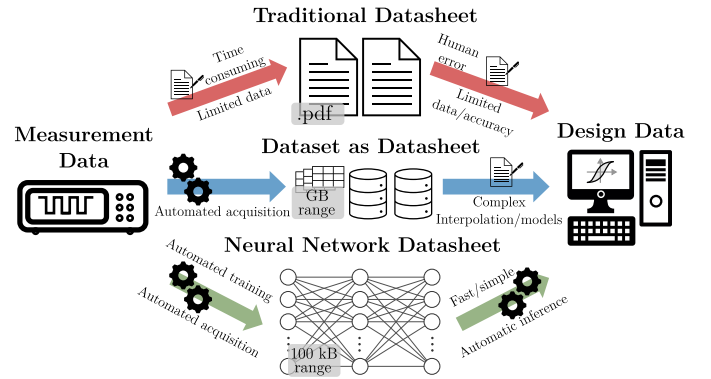


Fig. 32. Concept of neural network as datasheet. An NN is better equipped to store the shape of B - H loops for different operating conditions as compared to traditional datasheets or datasets of B - H loops that can only contain limited information. Neural networks can “compress” the information in datasets efficiently for size reduction.

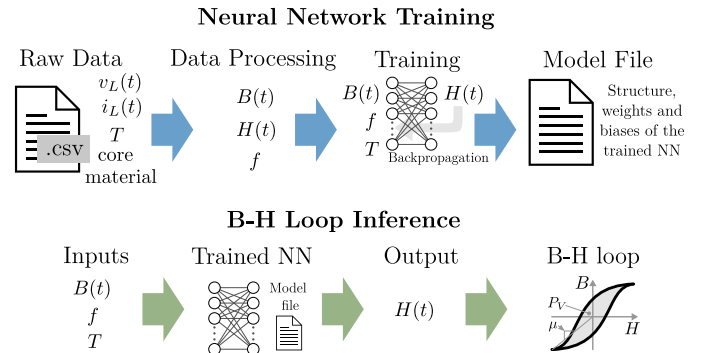


Fig. 33. Process for B - H loop modeling. From the training of the neural network to the inference of $H(t)$.

have $B(t)$ as an input and output $H(t)$ or vice versa. The input of the NN is a sequence of $B(t)$, the temperature, and frequency, and the output is $H(t)$. The user can input the desired parameters and use the trained NN to obtain the B - H loop, which can be imported into circuit or finite-element simulation models. Figure 33 describes the different steps for B - H loop inference using NNs from data acquisition, model training, to NN inference. The NN can be directly used to predict $H(t)$ for any $B(t)$ that the user defines. From the B - H loop, different magnetic properties such as the core loss or the permeability can be extracted. An example neural network has been implemented on the MagNet webpage.

Table V compares the size and accuracy of a few different data formats for an example material (TDK N87) and the corresponding neural network model. The dataset consists of sinusoidal, triangular, and trapezoidal waveforms measured at 25 °C and zero dc bias, while the NN is constructed and trained as in Sec. V-C. The NN was able to effectively compress the information contained in the dataset and reduce the data size by 2,500 times from 512 MB to 204 KB.

VII. CONCLUSIONS

This paper applies machine learning to modeling power magnetics. We first present an open-source large-scale database – *MagNet* – for data-driven magnetic components

TABLE V
COMPARISON OF DATA SIZE FOR N87 AT 25 °C (10,928 DATA POINTS).

Description	Format	Size
Raw Measurement (fixed-length)	“.csv”	5.63 GB
Raw Data (fixed-length)	“.mat”	512 MB
Processed Data (single-cycle)	“.json”	23.5 MB
Neural Network Model	“.sd”	204 KB

modeling. The data quality of MagNet is carefully evaluated and controlled to ensure model accuracy. With a large amount of data in the MagNet database, several example neural network modeling applications of the MagNet database have been explored, including the scalar-to-scalar model, the sequence-to-scalar model, the sequence-to-sequence model, as well as the transfer learning methods, which prove the effectiveness of neural networks in modeling the behavior of power magnetics. We anticipate that with constantly increasing scale, data quality, and waveform diversity, MagNet can offer unique opportunities to researchers in power electronics, power magnetics, and data science.

APPENDIX A AUTOMATED DATA ACQUISITION

We introduce more details about the design of the data acquisition system in the following aspects.

A. Excitation

With this setup, the DUT can be excited with sinusoidal, triangular, and trapezoidal waveforms.

In our design, sinusoidal waveforms are synthesized and created using a function generator (Rigol DG4102) and a power amplifier (Amplifier Research 25A250AM6). The computer sets the command for the frequency and voltage for the signal generator automatically to generate different sinusoidal excitations. Calibration for each measurement is required, as the voltage gain of the power amplifier is not constant due to the changing load under different conditions. Moreover, distorted voltage and current are obtained when the core is subjected to a large B_{ac} due to the power gain of the amplifier and low load impedance.

For piece-wise linear waveforms, such as triangular and trapezoidal excitations, a T-type inverter supplied by two voltage sources (B&K Precision XLN60026) is used, as illustrated in Fig. 34. GaN devices (GaN System GS66508B) are employed to obtain fast transitions between the three voltage levels $[V_{in}; 0; -V_{in}]$. To control the waveform shape, a micro-controller (Texas Instruments F28379D controlCARD) commanded the signals for the drivers. The micro-controller and voltage source commands are synchronized by the host PC to iterate the different duty cycles, frequencies, and amplitudes of the waveform.

To block the average voltage of the switching node of the power stage or any unwanted dc current present in the power amplifier, a blocking capacitor is placed in series with the DUT. The capacitance should be large enough to avoid the

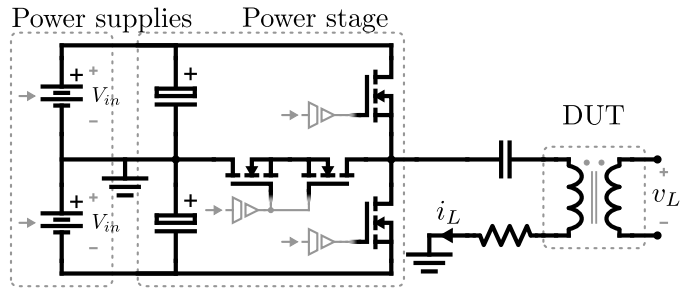


Fig. 34. Circuit schematic of the power stage for generating the excitations and measuring the magnetic component behaviors in the data acquisition system of MagNet.

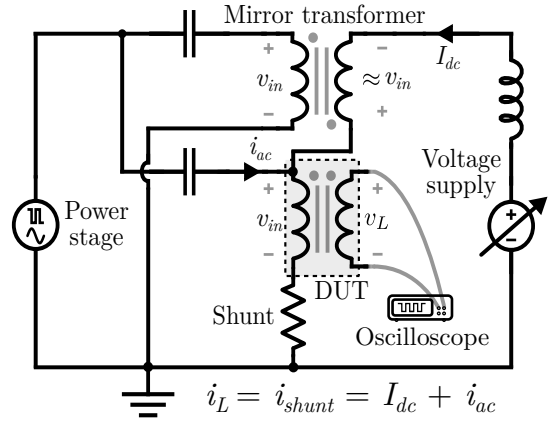


Fig. 35. Circuit schematic of the auxiliary dc-bias current injection circuitry for the measurements under dc-bias conditions.

voltage ripple distorting the excitation. For this purpose, a 100 μ F 100 V film capacitor is used.

To test the core under different dc bias conditions, an additional dc-bias current injection circuitry is included. A dc current is injected into the primary winding, after the series capacitor. This is preferred over the traditional third winding method [40], [52], [53]) as it avoids the unwanted current ripple in the dc winding. A mirror transformer and a filter inductor are added as indicated in Fig. 35, to prevent the reflected voltage of the DUT to be applied to the current source. The current source is a voltage supply (Siglent SPD3303X-E) with the current limit set by the computer automatically. Details regarding the operation and construction of the dc bias circuit can be found in [54].

Please note that a direct current is used to define the dc bias, implying that cores are tested under predefined H_{dc} rather than B_{dc} . The reason is that B_{dc} cannot be controlled properly as the initial state of magnetization (B_0) is not known. B_{dc} is not reported in this work as relating H_{dc} and B_{dc} through the initial magnetization curve might not be a reasonable approach, as described in [39].

B. Device Under Test

The DUT consists of a toroidal magnetic core, a primary winding, and a secondary winding. The primary winding is connected to the power stage and used for exciting the core,

whereas the secondary winding is open-circuited and used for inferring the magnetic flux density (B) by integrating the measured voltage across its terminals.

The DUTs considered for most of this paper are in $R34.0 \times 20.5 \times 12.5$ or similar sizes as typically used in the manufacturer's datasheet. Please note that the size and geometry of the magnetic core do affect the measured characteristics, such as the B - H loop and the core loss, primarily due to its impact on the flux density distribution within the core. The effect of geometry is beyond the scope of this paper. Discussions on the measurement results with different core sizes can be found in another publication [14].

The alternating current in the shunt and the maximum flux density set the limits for the number of turns. Given the rated voltage of the data acquisition system, the number of turns should be selected such that the available range of flux density and frequency is maximized. On the other hand, the number of turns also changes the inductance of the DUT, which is limited by the rated current of the data acquisition system. In this work, for example, the number of turns for the TDK N87 DUT is designed as 5 for both the primary and secondary windings. For the primary winding, 22 AWG Litz wire with 40 strands of 38 AWG wires optimized for 100 kHz is used. For secondary, as the current is theoretically zero, an 18 AWG round wire is used. As discussed in [55], toroidal cores without air gaps are preferred for magnetic characterization.

C. Measurement and Acquisition

The measurements for the voltage and current waveform are acquired directly with an 8-bit oscilloscope (Tektronix DPO4054). A waveform of 10,000 samples is saved for each test, with a sampling rate of 10 ns, leading to a total time for the sample of 100 μ s. Therefore, a different number of switching cycles is captured depending on the frequency of the excitation. The bandwidth of the measurement is limited to 20 MHz to avoid excessive switching noise in the triangular/trapezoidal waveforms due to the fast switching transitions.

For the voltage measurement, a low-capacitance passive probe (Tektronix P6139A) is used. For the current, a coaxial shunt (T&M Research W-5-10-1STUD) of 0.983 Ω is connected in series with the primary winding of the DUT, as it is preferred over current probes to minimize the phase mismatch [39], [40], [52]. The terminal impedance of the current measurement channel in the oscilloscope is set to 50 Ω and is accounted for in the calculation of the current [39].

D. Temperature Control

Keeping a controlled temperature for the DUT is critical since core losses are highly temperature-dependent. This is challenging as core losses during testing cause the core to heat up. To set the temperature at the desired level, the DUT is submerged in a mineral oil bath, which is inside a large water tank. The temperature of the water tank is controlled using a water heater (ANOVA AN400). The water tank is covered to ensure the oil reaches the same temperature as the water. To prevent the DUT from reaching temperatures significantly higher than the one set by the water heater, a magnetic stirrer is used (INTLLAB) to keep the oil constantly flowing.

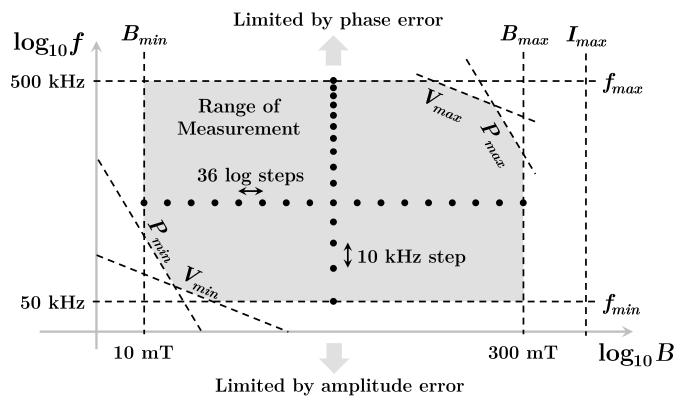


Fig. 36. Range of measurement for the flux density amplitude and the frequency.

E. Software System

Python-based software interface on the host PC is designed and programmed to control and coordinate with the hardware system to enable fully automated equipment settings, synchronization of the different instruments, perform the measurements, and store the acquired data.

The software has three major functions: (a) communication with the power stage (power supplies, micro-controller and function generator) to transmit the waveform properties for each test, including the frequency, voltage, and waveform shape, so that the power stage can synthesize and generate the desired excitations; (b) communication with the oscilloscope to set the configuration of signal sampling and data acquisition, perform calibration if needed, and receive the measured digitized waveforms; (c) storing the collected data points and converting them into the expected dataset format.

Specifically, the communication with the micro-controller is implemented based on the UART protocol, and the communication with equipment including the power source, the function generator, and the oscilloscope, is implemented with the support of the virtual instrument software architecture (VISA) protocol. The three aforementioned functions are executed in sequence within a multi-level iteration loop that sweeps the entire parameter space automatically. No human intervention is needed to perform a set of tests except to change the desired temperature, to change the connections from sinusoidal to piece-wise linear excitations, or to change the DUT.

F. Range of Measurement

The range of measurement is constrained by various factors, and needs to be determined carefully in order to guarantee high data quality. The proposed range of measurement for the data acquisition system in this work is illustrated in Fig. 36.

For the flux density amplitude, data is measured in the 10 mT to 300 mT range with 36 steps in the logarithm scale, leaving some distance from the saturation level provided by the material datasheet. Logarithm scales are preferred due to the exponential nature of core losses with B_{ac} .

For the frequency, data is measured from 50 kHz to 500 kHz. This range fits the operation range of the cores under

test. In order to correctly and accurately obtain the characteristics of the measured magnetic material, the frequency needs to be selected with specific frequency steps, which guarantees the measured waveform always contains complete cycles. Here, we selected a 10 kHz step of frequency given the sampling rate and the number of samples in each measurement.

Meanwhile, the ranges of flux density amplitude and frequency are also constrained by the measurement accuracy, as discussed in Sec. III and Appendix B, where data points with low amplitude or high frequency are more prone to error.

Regarding the dc bias, H_{dc} is swept in 15 A/m linear steps. For N87 ferrite, the limit is 60 A/m to leave some room for B_{ac} . To avoid running into saturation, the maximum value for B_{ac} (300 mT) is decreased with higher H_{dc} based on the maximum amplitude permeability listed on the datasheet.

Additionally, there are limitations associated with the power stage and the power amplifier. The voltage range for the tests is 1 V to 50 V for sinusoidal waveforms and 5 V to 80 V for PWM waveforms, considering the rated voltage of equipment and components, which limits the $B_{ac} \cdot f$ product of the data measured, especially for those with extreme duty cycles.

Furthermore, to avoid a significant rise in the core temperature with respect to the target temperature, measurements with extremely large estimated loss (based on iGSE) above 5000 kW/m³ are skipped. On the other hand, low loss points below the range of interest (1 kW/m³) are skipped to reduce the data collection time.

Considering the typical operating condition of the magnetic materials under test and the capability of the temperature control equipment, the temperature range for the tests is 25 °C to 90 °C. Advanced equipment such as constant temperature ovens will enable the measurement at higher temperatures above 100 °C, which is also occasionally encountered in power magnetics applications.

APPENDIX B DATA QUALITY CONTROL

We introduce more details about data quality control in the following aspects.

A. Equipment Evaluation and Calibration

The experimental setup of the MagNet data acquisition system and its calibration processes are designed following the recommendations in [39], [41], [42]. Extra attention must be paid when designing and implementing the measurement system to understand the equipment limitations.

We evaluate the error and measurement capability of the oscilloscope. The oscilloscope (Tektronix DPO4054) used in the system is calibrated against an Agilent 34401A 6½ digits multimeter to evaluate the dc and ac accuracy by measuring the same dc and ac voltage signals at the same time. Relative errors of measurements are calculated by averaging multiple testing points among the entire measurement range from 0 V to 80 V and 50 kHz to 500 kHz. The relative error of the mean dc voltage is measured as 0.25%, and the relative error of the RMS ac voltage is 0.67%, which proves the oscilloscope accuracy of V_{DC} , I_{DC} , v_{AC} , and i_{AC} (see definitions in

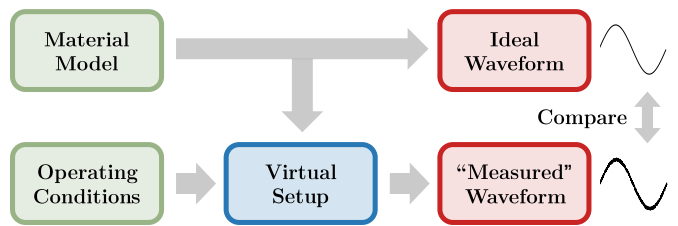


Fig. 37. Workflow of the virtual measurement simulation. The virtual measurement setup numerically simulates the impact of various sources of measurement error. The virtually measured waveform is compared against the ideal waveform to estimate the measurement accuracy.

Eq. (5)). The gain accuracy of the oscilloscope is rated as $\pm 1.5\%$ according to the equipment specifications, which quantify the error rates of G_V and G_I . Additionally, every time before the measurement iteration starts, the signal pass of the oscilloscope is reset and re-calibrated, which minimizes the undesired zero-drift offset V_0 and I_0 , and the time skewing θ between the voltage and current signals.

We then evaluate the error and measurement capability of the power stage. As mentioned in Sec. II, a wide-band coaxial shunt (T&M W-5-10-1STUD) is used for measuring the current. This current shunt has low parasitic inductance and is stable against temperature variation. A BNC connector with parasitic terminal capacitance lower than 10 pF is implemented to connect the coaxial shunt, the device under test, and the circuit board. Other parasitic capacitances on the circuit board are also minimized to our best. All these design criteria help to reduce the measurement error in I_{DC} and i_{AC} , especially minimizing the time skewing θ , which is critical to the accuracy of core loss measurement.

To highlight, the equipment calibration and measurement processes are fully automated. Human influence is minimized during the calibration and data acquisition process, which further enhanced the measurement accuracy and consistency. Repeating experiments on the same DUT consistently show a relative discrepancy lower than 3% between different trials of core loss measurements, which validates the reproducibility of the measured data.

B. Model-driven Method for Quantifying the Error

The accuracy of data-driven models is bounded by the accuracy of the data. In order to quantify the measurement error and estimate the potential error distribution of the measured results, a model-driven method combining the physics-based simulation to create virtual measurement and the different uncertainties can be assessed with Monte Carlo experiments. The error analysis also provides a baseline for setting the accuracy target for machine learning or curve-fitting methods.

Figure 37 illustrates the workflow of the virtual measurement simulation. A reference waveform is generated by the material model and passed into the virtual measurement setup. The virtual measurement setup takes various sources of measurement error into account, numerically simulates their impact on the measurement, and generates the virtually measured waveform. Comparing the virtually measured

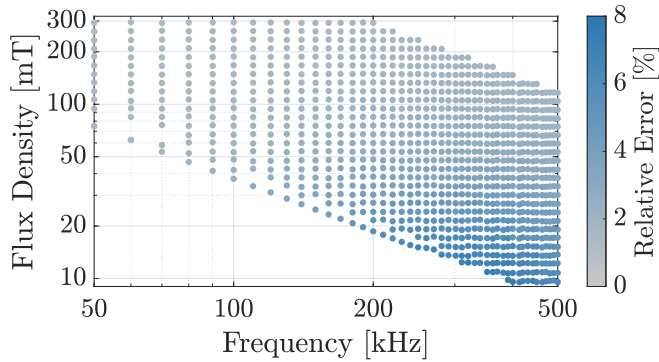


Fig. 38. Example simulation results for TDK N87 material with the virtual measurement setup and Monte Carlo experiments, where the measurement uncertainties introduced by the probe and scope are taken into consideration. Colors depict the discrepancy between the virtually measured core loss and the expected core loss.

waveform against the ideal waveform, the uncertainty of the measurement can be evaluated and estimated.

All the parameters in the virtual measurement setup are either determined according to the datasheets of equipment, components, and materials, or estimated based on the actual experimental results. Some of the most important sources of measurement error include:

- **Systematic error:** parasitics of power stage circuit, parasitics of wire and cable, parasitics of the DUT, the timing skew of passive probes (± 1.5 ns), the uncertainty of probe gains ($\pm 0.5\%$), the uncertainty of probe offsets ($\pm 0.5\%$), and the manufacturing tolerance of the core geometry (area and length, $\pm 2\%$) which affects the calculation of $B(t)$ and $H(t)$, etc.
- **Statistical error:** electrical noise due to the environment, quantization error and sampling noise due to the oscilloscope, and the undesired temperature variation of the DUT ($\pm 1.6\%$), etc.

Based on the virtual measurement setup, a series of Monte Carlo experiments is conducted, where the uncertain variables are assumed to follow the Gaussian distribution or uniform distribution with aforementioned values as their 2σ deviations. Fig. 38 demonstrates simulation results for TDK N87 material as an example, where the measurement uncertainty introduced by the probe and scope are taken into consideration and numerically simulated. Note that the result of Monte Carlo experiments for each sample point is a random distribution (the discrepancy between the virtually measured core loss and the expected core loss), and the value shown in Fig. 38 is the 95th-percentile of each distribution. As demonstrated in the figure, the majority of the samples within this examined range maintain a low measurement error rate that less than 6%. The most erroneous samples locate in the area with high frequency and low flux density, where the measurements are more prone to noise and the inaccuracy of equipment.

More specifically, the error distribution of an example point (300 kHz, 50 mT, 50% duty ratio triangular wave with zero dc bias measured at 25°C) is demonstrated in Fig. 39, where the measurement uncertainties introduced by the circuit parasitics,

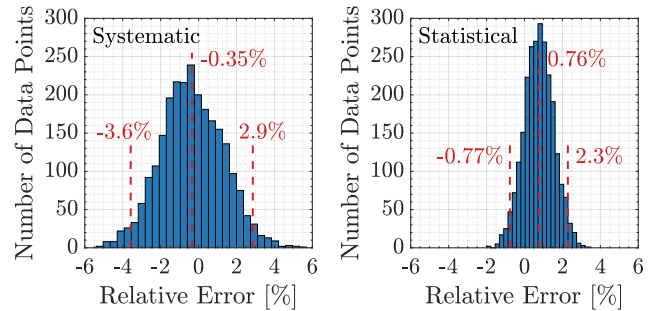


Fig. 39. Error distribution of an example point (300 kHz, 50 mT, 50% duty ratio triangular wave with zero dc bias measured at 25°C), where the measurement uncertainties introduced by the circuit parasitics, temperature variation, and geometry variations are considered. Both the systematic error and the statistical error are less than 4% for the majority of trials in the Monte Carlo experiments. The spread of systematic error is larger than that of statistical error.

temperature variation, and geometry variations are considered and simulated numerically. For the majority of trials among the Monte Carlo experiments, as shown in the histograms, the systematic error is less than 3.6% and the statistical error is less than 2.3%. Comparatively, systematic error brings a larger impact on the overall measurement accuracy than statistical error. Geometry variation contributes largely to systematic error and cannot be avoided in design, while upgrading the capability of equipment and reducing the hardware parasitics to minimize the time skewing between signals can still marginally improve the measurement accuracy. In addition, the temperature variation of DUT contributes a large portion of the statistical error, highlighting the importance of more precise temperature control.

Overall, high measurement accuracy is achieved in the main operating range of the data acquisition system, while the measurements for magnetic materials with high quality factor at high-frequency or low-amplitude ranges could be more prone to error. Based on the model-driven error analysis and the typical operating conditions provided by the material datasheet, we determine the measurement region that we are confident about the data quality, as aforementioned in Fig. 36 in Appendix A. A similar error map can be created for each material in the database for evaluating the data quality.

C. Data-driven Methods for Data Quality Control

Outliers are unavoidable for large-scale automated data collection. An algorithm was developed to detect and remove the outlier data points caused by rare anomaly operations based on smoothness analysis. As illustrated in Fig. 40, for each point in the dataset, the estimated power losses are calculated based on the Steinmetz parameters inferred from the points that are close in terms of frequency and flux density to the considered point holding constant the other variables. If the measured losses of the data point are far from this estimated value, the data point can be considered an outlier. More specifically, for a given data point, a weight reflecting the closeness is assigned to all the other data points, which is

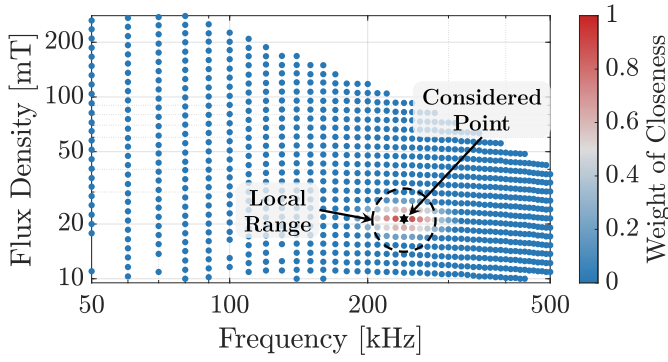


Fig. 40. Example distribution of the defined weight of closeness for a specific considered data point. The local Steinmetz will be performed within the local range that is close enough to the considered data point.

defined as:

$$w_i = \max \left\{ 1 - \frac{\sqrt{(\log f_i - \log f_0)^2 + (\log B_i - \log B_0)^2}}{w_{max}}, 0 \right\} \quad (10)$$

where (f_0, B_0) are the frequency and peak flux density of the specific data point being considered, and (f_i, B_i) refer to those of every other data point in the dataset. The square-root part quantifies the distance between two data points on the logarithm plane of f - B , and w_{max} is a parameter that can be tuned to determine the size of the neighboring area taken into consideration. Based on the definition of w_i , the closer the (f_i, B_i) is to the (f_0, B_0) , the closer w_i is to 1. Oppositely, any (f_i, B_i) that are farther from (f_0, B_0) will get a smaller w_i value and eventually 0 if the distance exceeds the w_{max} . Fig. 40 illustrates an example distribution of the weights of closeness for a considered data point, where the color of points reflects the normalized distance between any given data points and the considered data points.

Based on the weight w_i , a weighted least square regression is conducted to calculate the local Steinmetz parameters by:

$$\min_{k, \alpha, \beta} \sum_{i \neq 0} [(k f_i^\alpha B_i^\beta)^2 - P_{meas, i}^2] w_i^2 \quad (11)$$

The local Steinmetz parameters for a given data point are calculated based on the data points nearby, with which an expected core loss value can be estimated according to the Steinmetz equation. The outlier factor is defined as the relative discrepancy between the expected loss and the measured loss:

$$\text{Outlier Factor} = \frac{k f^\alpha B^\beta - P_{meas}}{P_{meas}} \times 100\% \quad (12)$$

Figure 41 shows an example of the discrepancies between the expected losses based on the Steinmetz parameters of nearby points and the measured losses for different data points. A data point with a high outlier factor is considered a low-quality measurement and is removed from the dataset.

Outlier detection is critical for data quality control. This outlier detection algorithm is just one example way of evaluating the data quality and removing abnormal data. It has its own strengths and limitations, e.g., it cannot detect systematic error and may miss unusual material characteristics.

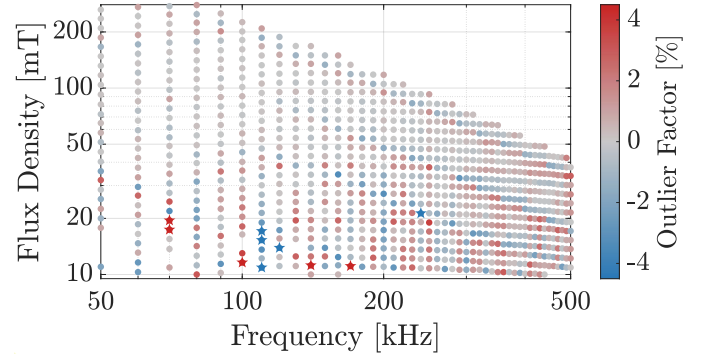


Fig. 41. Example of outlier data points in a dataset for the material N87 under sinusoidal excitation. For each point, data up to 0.1 decades far in terms of flux density and frequency are used to generate the local Steinmetz parameters. The data points discarded because the error compared to the estimation is above $\pm 4\%$ are marked as solid stars.

APPENDIX C MAGNET WEBSITE

An open-source webpage-based platform with a graphical user interface (GUI) – MagNet – has been developed. MagNet in its current state contains over 500,000 excitation waveforms for ten ferrite materials – TDK{N27, N30, N49, N87}, Ferroxcube{3C90, 3C94, 3E6, 3F4}, and Fair-Rite{77, 78} – in the 50 kHz to 500 kHz, 10 mT to 300 mT range for the sinusoidal, triangular, and trapezoidal waveforms, collected under multiple temperatures and dc bias conditions. It is powered by Streamlit (an open-source app framework of Python for website deployment), and shared in GitHub, offering a variety of data-visualization tools with a graphical user interface for the database, magnetic core loss estimation, simulation, and prediction based on multiple models (see more details in Sec. V), as well as the access to download all the measured data points. Fig. 42 demonstrates the website architecture in its current state.

The website allows the dataset to be visualized in many ways, and enables rapid comparison of the core loss and B - H loop data of different materials. The user may specify the type of magnetic material, together with the waveform shape, frequency range, flux density range, and tolerable outlier level (described in Section B-C). The website backend searches for the requested data in the database and visualizes it in the way that the user selects. The website also provides download access to the raw data being collected from the equipment before any post-processing with the test conditions documented, and the post-processed dataset files for data-driven modeling applications.

The website embeds and deploys multiple core loss modeling algorithms, including iGSE and other data-driven methods as introduced in the following sections, for online core loss estimation. Designers can specify the expected excitation waveforms and operation conditions, based on which the website backend calculates the estimated core loss with both the iGSE equation and the neural network model.

The webpage is also connected to a circuit simulation server hosted by Plexim. The webpage feeds information to the

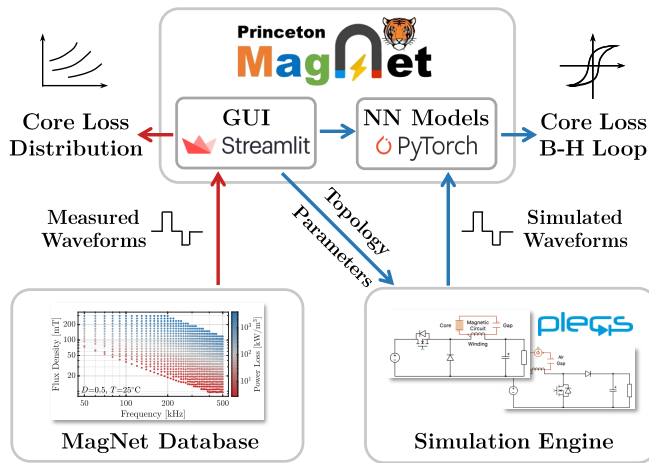


Fig. 42. The website architecture and information flow of the MagNet webpage platform, which provides users with access to download and visualize the measured data in the MagNet core loss database, as well as analyze and simulate the magnetic behaviors with the deployed neural network models and the PLECS simulation engine.

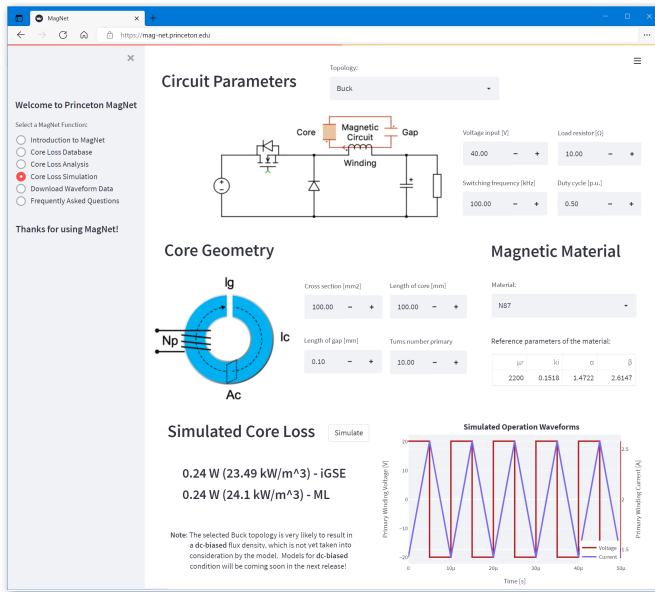


Fig. 43. An example screenshot of the core loss simulation session in the Mag-Net webpage-based magnetics database and analysis platform.

server, and the server returns inputs to the machine learning algorithms in combination with power converter operations, as shown in Fig. 43. Users can choose from a pool of common topologies (Buck, Boost, Flyback, Dual Active Bridge), specify the circuit parameters, magnetic component specifications, and operating conditions, then the simulation engine simulates and outputs the excitation waveform of the magnetic component. The MagNet server collects the waveform and predicts the core loss using the algorithms available on the web server including iGSE and neural networks.

The MagNet platform is constantly maintained and updated with new data and neural network models. Extended details are included on the website to enable trustworthy repeating measurements and cross-validation of the dataset.

ACKNOWLEDGEMENTS

This work was jointly supported by the DOE ARPA-E DIFFERENTIATE program and the Schmidt DataX Fund at Princeton University made possible through a major gift from the Schmidt Futures Foundation.

REFERENCES

- [1] H. Li, S. R. Lee, M. Luo, C. R. Sullivan, Y. Chen, and M. Chen, "MagNet: A machine learning framework for magnetic core loss modeling," in *2020 IEEE 21st Workshop on Control and Modeling for Power Electronics (COMPEL)*, 2020, pp. 1–8.
- [2] E. Dogariu, H. Li, D. Serrano López, S. Wang, M. Luo, and M. Chen, "Transfer learning methods for magnetic core loss modeling," in *2021 IEEE 22nd Workshop on Control and Modelling of Power Electronics (COMPEL)*, 2021, pp. 1–6.
- [3] H. Li, D. Serrano, T. Guillod, E. Dogariu, A. Nadler, S. Wang, M. Luo, V. Bansal, Y. Chen, C. R. Sullivan, and M. Chen, "MagNet: An open-source database for data-driven magnetic core loss modeling," in *2022 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2022, pp. 588–595.
- [4] D. Serrano, H. Li, T. Guillod, S. Wang, M. Luo, C. R. Sullivan, and M. Chen, "Neural network as datasheet: Modeling B-H loops of power magnetics with sequence-to-sequence LSTM encoder-decoder architecture," in *2022 IEEE 23rd Workshop on Control and Modeling for Power Electronics (COMPEL)*, 2022, pp. 1–8.
- [5] D. J. Perreault, J. Hu, J. M. Rivas, Y. Han, O. Leitermann, R. C. Pilawa-Podgurski, A. Sagneri, and C. R. Sullivan, "Opportunities and challenges in very high frequency power conversion," in *2009 Twenty-Fourth Annual IEEE Applied Power Electronics Conference and Exposition*, 2009, pp. 1–14.
- [6] C. R. Sullivan, B. A. Reese, A. L. F. Stein, and P. A. Kyaw, "On size and magnetics: Why small efficient power inductors are rare," in *2016 International Symposium on 3D Power Electronics Integration and Manufacturing (3D-PEIM)*, 2016, pp. 1–23.
- [7] Y. Han, G. Cheung, A. Li, C. R. Sullivan, and D. J. Perreault, "Evaluation of magnetic materials for very high frequency power applications," *IEEE Transactions on Power Electronics*, vol. 27, no. 1, pp. 425–435, 2012.
- [8] A. J. Hanson, J. A. Belk, S. Lim, C. R. Sullivan, and D. J. Perreault, "Measurements and performance factor comparisons of magnetic materials at high frequency," *IEEE Transactions on Power Electronics*, vol. 31, no. 11, pp. 7909–7925, 2016.
- [9] M. Chen, M. Araghchini, K. K. Afridi, J. H. Lang, C. R. Sullivan, and D. J. Perreault, "A systematic approach to modeling impedances and current distribution in planar magnetics," *IEEE Transactions on Power Electronics*, vol. 31, no. 1, pp. 560–580, 2016.
- [10] W. Roshen, "Ferrite core loss for power magnetic components design," *IEEE Transactions on Magnetics*, vol. 27, no. 6, pp. 4407–4415, 1991.
- [11] D. Lin, P. Zhou, W. Fu, Z. Badics, and Z. Cendes, "A dynamic core loss model for soft ferromagnetic and power ferrite materials in transient finite element analysis," *IEEE Transactions on Magnetics*, vol. 40, no. 2, pp. 1318–1321, 2004.
- [12] M. Mu, Q. Li, D. J. Gilham, F. C. Lee, and K. D. T. Ngo, "New core loss measurement method for high-frequency magnetic materials," *IEEE Transactions on Power Electronics*, vol. 29, no. 8, pp. 4374–4381, 2014.
- [13] M. Luo, D. Dujic, and J. Allmeling, "Modeling frequency independent hysteresis effects of ferrite core materials using permeance-capacitance analogy for system-level circuit simulations," *IEEE Transactions on Power Electronics*, vol. 33, no. 12, pp. 10055–10070, 2018.
- [14] D. Serrano, H. Li, S. Wang, M. Luo, T. Guillod, C. R., and M. Chen, "Quantifying the complexity of modeling power magnetic material characteristics," 2022, *IEEE Transactions on Power Electronics* [Manuscript under review].
- [15] C. Steinmetz, "On the law of hysteresis," *Proceedings of the IEEE*, vol. 72, no. 2, pp. 197–221, 1984.
- [16] J. Goodenough, "Summary of losses in magnetic materials," *IEEE Transactions on Magnetics*, vol. 38, no. 5, pp. 3398–3408, 2002.
- [17] J. Li, T. Abdallah, and C. Sullivan, "Improved calculation of core loss with nonsinusoidal waveforms," in *Conference Record of the 2001 IEEE Industry Applications Conference. 36th IAS Annual Meeting (Cat. No.01CH37248)*, vol. 4, 2001, pp. 2203–2210 vol.4.

- [18] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Improved core-loss calculation for magnetic components employed in power electronic systems," *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 964–973, 2012.
- [19] J. Reinert, A. Brockmeyer, and R. De Doncker, "Calculation of losses in ferro- and ferrimagnetic materials based on the modified Steinmetz equation," *IEEE Transactions on Industry Applications*, vol. 37, no. 4, pp. 1055–1061, 2001.
- [20] E. Stenglein and T. Dürbaum, "Core loss model for arbitrary excitations with dc bias covering a wide frequency range," *IEEE Transactions on Magnetics*, vol. 57, no. 6, pp. 1–10, 2021.
- [21] Y. LeCun, Yann Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436–444, 2015.
- [22] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, jan 2015.
- [23] B. Widrow and M. Lehr, "30 years of adaptive neural networks: perceptron, Madaline, and backpropagation," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415–1442, 1990.
- [24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [25] S. Zhao, F. Blaabjerg, and H. Wang, "An overview of artificial intelligence applications for power electronics," *IEEE Transactions on Power Electronics*, vol. 36, no. 4, pp. 4633–4658, 2021.
- [26] T. Guillod, P. Papamanolis, and J. W. Kolar, "Artificial neural network (ANN) based fast and accurate inductor modeling and design," *IEEE Open Journal of Power Electronics*, vol. 1, pp. 284–299, 2020.
- [27] E. I. Amoiralis, P. S. Georgilakis, T. D. Kefalas, M. A. Tsili, and A. G. Kladas, "Artificial intelligence combined with hybrid FEM-BE techniques for global transformer optimization," *IEEE Transactions on Magnetics*, vol. 43, no. 4, pp. 1633–1636, 2007.
- [28] C. Nussbaum, H. Pfutzner, T. Booth, N. Baumgartinger, A. Ilo, and M. Clabian, "Neural networks for the prediction of magnetic transformer core characteristics," *IEEE Transactions on Magnetics*, vol. 36, no. 1, pp. 313–329, 2000.
- [29] G. K. Miti and A. J. Moses, "Neural network-based software tool for predicting magnetic performance of strip-wound magnetic cores at medium to high frequency," *IEE Proceedings: Science, Measurement and Technology*, vol. 151, no. 3, pp. 181–187, 2004.
- [30] H. Saliyah, D. Lowther, and B. Forghani, "A neural network model of magnetic hysteresis for computational magnetics," *IEEE Transactions on Magnetics*, vol. 33, no. 5, pp. 4146–4148, 1997.
- [31] C. Serpico and C. Visone, "Magnetic hysteresis modeling via feed-forward neural networks," *IEEE Transactions on Magnetics*, vol. 34, no. 3, pp. 623–628, 1998.
- [32] E. Fallah, J. Moghani, and H. Talebi, "Modeling of magnetic hysteresis using counter propagation neural network," in *2003 IEEE International Magnetics Conference (INTERMAG)*, 2003, pp. DP–07.
- [33] H. Saliyah, D. Lowther, and B. Forghani, "Modeling magnetic materials using artificial neural networks," *IEEE Transactions on Magnetics*, vol. 34, no. 5, pp. 3056–3059, 1998.
- [34] F. Sixdenier, R. Scorretti, R. Marion, and L. Morel, "Quasistatic hysteresis modeling with feed-forward neural networks: Influence of the last but one extreme values," *Journal of Magnetism and Magnetic Materials*, vol. 320, no. 20, pp. e992–e996, 2008.
- [35] Z. Zhao, F. Liu, S. L. Ho, W. N. Fu, and W. Yan, "Modeling magnetic hysteresis under dc-biased magnetization using the neural network," *IEEE Transactions on Magnetics*, vol. 45, no. 10, pp. 3958–3961, 2009.
- [36] F. Riganti Fulginei and A. Salvini, "Neural network approach for modelling hysteretic magnetic materials under distorted excitations," *IEEE Transactions on Magnetics*, vol. 48, no. 2, pp. 307–310, 2012.
- [37] C. Grech, M. Buzio, M. Pentella, and N. Sammut, "Dynamic ferromagnetic hysteresis modelling using a Preisach-recurrent neural network model," *Materials*, vol. 13, no. 11, p. 2561, 2020.
- [38] S. Quondam Antonio, F. Riganti Fulginei, A. Laudani, A. Faba, and E. Cardelli, "An effective neural network approach to reproduce magnetic hysteresis in electrical steel under arbitrary excitation waveforms," *Journal of Magnetism and Magnetic Materials*, vol. 528, p. 167735, 2021.
- [39] E. Stenglein, B. Kohlhepp, D. Kübrich, M. Albach, and T. Dürbaum, "GaN-half-bridge for core loss measurements under rectangular ac voltage and dc bias of the magnetic flux density," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6312–6321, 2020.
- [40] V. Thottuvelil, T. Wilson, and H. Owen, "High-frequency measurement techniques for magnetic cores," *IEEE Transactions on Power Electronics*, vol. 5, no. 1, pp. 41–53, 1990.
- [41] E. Stenglein, D. Kuebrich, M. Albach, and T. Duerbaum, "Guideline for hysteresis curve measurements with arbitrary excitation: Pitfalls to avoid and practices to follow," in *PCIM Europe 2018; International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management*, 2018, pp. 1–8.
- [42] M. Luo, D. Dujic, and J. Allmeling, "Test setup for characterisation of biased magnetic hysteresis loops in power electronic applications," in *2018 International Power Electronics Conference (IPEC-Niigata 2018 -ECCE Asia)*, 2018, pp. 422–427.
- [43] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.
- [44] D. Jiles and J. Thoeleke, "Theory of ferromagnetic hysteresis: determination of model parameters from experimental hysteresis loops," *IEEE Transactions on Magnetics*, vol. 25, no. 5, pp. 3928–3930, 1989.
- [45] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "PyTorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.
- [46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [47] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.
- [48] X. Li and X. Wu, "Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 4520–4524.
- [49] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *Journal of Machine Learning Research*, vol. 3, pp. 115–143, 2002.
- [50] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [51] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, 2016.
- [52] A. Brockmeyer, "Experimental evaluation of the influence of dc-premagnetization on the properties of power electronic ferrites," in *Proceedings of Applied Power Electronics Conference. APEC '96*, vol. 1, 1996, pp. 454–460 vol.1.
- [53] C. A. Baguley, U. K. Madawala, and B. Carsten, "A new technique for measuring ferrite core loss under dc bias conditions," *IEEE Transactions on Magnetics*, vol. 44, no. 11, pp. 4127–4130, 2008.
- [54] S. Wang, D. Serrano, H. Li, A. Lin, T. Guillod, M. Luo, C. R. Sullivan, and M. Chen, "A simplified mirror transformer dc-bias injection method for power magnetics characterization," 2023, iEEE Applied Power Electronics Conference and Exposition (APEC). [Manuscript under review].
- [55] E. Stenglein, D. Kübrich, M. Albach, and T. Dürbaum, "Influence of magnetic history and accommodation on hysteresis loss for arbitrary core excitations," in *2019 21st European Conference on Power Electronics and Applications (EPE '19 ECCE Europe)*, 2019, pp. P.1–P.10.