Embedding physical symmetries into machine-learned reduced plasma physics models via data augmentation

M. C. McGrae-Menge [1,*] J. R. Pierce [1] F. Fiuza [2] and E. P. Alves [1,3,†]

[1]*Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA*

[2]*GoLP/Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal*

[3]*Mani L. Bhaumik Institute for Theoretical Physics, University of California at Los Angeles, Los Angeles, California 90095, USA*

(Received 6 June 2025; accepted 23 January 2026; published 23 February 2026)

Machine learning is offering powerful tools for the development and discovery of reduced models of nonlinear, multiscale plasma dynamics from the data of first-principles kinetic simulations. However, ensuring the physical consistency of such models requires embedding fundamental symmetries of plasma dynamics. In this work, we explore a symmetry-embedding strategy based on data augmentation, where symmetry-preserving transformations (e.g., Lorentz and Galilean boosts) are applied to simulation data. Using both sparse regression and neural networks, we show that models trained on symmetry-augmented data more accurately infer the plasma fluid equations and pressure tensor closures from fully kinetic particle-in-cell simulations of magnetic reconnection. We show that this approach suppresses spurious inertial-frame-dependent correlations between dynamical variables, improves data efficiency, and significantly outperforms models trained without symmetry-augmented data, as well as commonly used theoretical pressure closure models. Our results establish symmetry-based data augmentation as a broadly applicable method for incorporating physical structure into machine-learned reduced plasma models.

DOI: 10.1103/9fq1-vpqz

I. INTRODUCTION

Plasmas are hot, ionized gasses that exhibit complex nonlinear behavior across a broad range of spatial and temporal scales. Notable examples include inertial and magnetically confined fusion plasmas [1,2], where the coupling between microturbulence and transport scales plays a central role in plasma confinement, and astrophysical plasmas where dissipation of large-scale magnetic fields is regulated by small kinetic-scale plasma processes that further mediate the acceleration of energetic particles [3,4]. A comprehensive understanding of such multiscale plasma phenomena remains a major scientific challenge, requiring improved theoretical frameworks and modeling tools to make progress. First-principles approaches based on plasma kinetic theory and simulations offer accurate descriptions of plasma dynamics but are computationally prohibitive for capturing the full range of scales in many realistic systems of interest. In contrast, fluid models are efficient at capturing large-scale plasma behavior but often miss important effects associated with small-scale kinetic processes. To effectively model such multiscale plasma dynamics, it is important to develop reduced plasma models that capture the key coupling between small-scale kinetic processes and large-scale plasma fluid behavior, without explicitly resolving the small scales.

Traditionally, reduced plasma models have been developed using analytic theory, which are typically valid in linear plasma dynamics regimes [5] or in well-defined asymptotic limits [6]. However, the assumptions that underpin these models often break down in many scenarios of interest, where the physics is nonlinear and multiscale. Current state-of-the-art kinetic plasma simulations are beginning to be able to capture sufficiently large dynamical ranges to study nonlinear multiscale plasma dynamics from first principles [7,8]. The data from these simulations are creating opportunities to guide the development of reduced plasma descriptions in these regimes.

Machine learning (ML) techniques are emerging as promising tools to develop reduced plasma models from the data of first-principles simulations. Deep neural networks (DNNs), in particular, have been used to learn pressure and heat flux closures for fluid models of magnetic reconnection (MR) [9], to capture kinetic effects like Landau damping in fluid plasma models [10–12] as well as in Vlasov-Poisson models using a truncated Fourier-Hermite basis [13], and to efficiently simulate the dynamics of one-dimensional collisional plasmas [14]. Indeed, DNNs have been shown to excel at approximating the complex nonlinear dependencies associated with plasma dynamics, but the inherent complexity of these models hinders their interpretability and often limits their ability to generalize to conditions outside the training data. These limitations have motivated growing interest in interpretable machine learning techniques, particularly symbolic [15–18] and sparse [19–22] regression methods. These techniques produce compact mathematical expressions—often in the form of partial differential equa-

*Contact author: madoxmm@physics.ucla.edu

†Contact author: epalves@physics.ucla.edu

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

tions (PDEs)—that offer insight into the underlying physics, though their more restricted functional forms may trade-off accuracy for complexity in highly nonlinear, complex systems. In plasma physics, sparse regression (SR) has been used to recover the fundamental hierarchy of reduced plasma models [from Vlasov to single-fluid magnetohydrodynamics (MHD)] from fully kinetic plasma simulation data [23], to infer fluid moment closures for magnetic reconnection [24] and electrostatic plasma phenomena [25], and to infer reduced models of magnetized plasma turbulence [26]; other closely related applications of sparse regression include the development of low-dimensional reduced-order models for fluid plasma dynamics [27,28], recovering the MHD equations from 3D turbulence data [29], stellarator design optimization [30,31], development of plasma profile evolution models near divertors [32], and characterization of plasma propulsion dynamics [33].

An essential, yet underdeveloped, aspect of ML-based reduced plasma modeling is the incorporation of fundamental physical symmetries. These include, for example, inertial-frame invariance (to Lorentz transformations), rotational and translational invariance, and energy conservation. Embedding these symmetries is crucial to ensure the physical consistency and generalization capabilities of the inferred models [34]. Here, we focus on frame invariance. The Vlasov equation, which forms the fundamental mathematical framework for collisionless plasma dynamics, is Lorentz invariant [35–37], and therefore any reduced plasma description derived from Vlasov should respect this symmetry.

In general, there are two dominant approaches for embedding symmetries into ML models: (1) hard constraints, which guarantee symmetry preservation, and (2) weak constraints, which encourage, but do not guarantee, symmetry preservation. The first strategy, hard constraints, can be most directly applied in ML by selecting model forms that explicitly satisfy the intended symmetry. For example, by exclusively using symmetry-respecting model terms in SR, the inferred reduced models will preserve the corresponding symmetries, such as rotational and translational invariance [38]. In the context of plasma physics, low-dimensional proper orthogonal decomposition-Galerkin models for magnetohydrodynamics have been shown to embed global energy conservation by restricting model term nonlinearities to quadratic [28]. A more involved method of embedding hard constraints is by building the symmetry into the ML architecture, known as equivariant ML, which has emerged as a powerful paradigm in scientific ML applications [39]. It builds on early work embedding discrete rotational symmetry into convolutional neural networks (NNs) [40]. Recent developments in equivariant ML are beginning to tackle more complex space-time symmetries like Lorentz invariance, for example, Lorentz-equivariant architectures for applications in high-energy physics [41], and even Lorentz-equivariant neural PDE solvers for fluid and electrodynamic systems [42]. While these hard-constrained methods enforce the targeted symmetries, they often come with increased implementation complexity and reduced flexibility in model form.

The second strategy, weak constraints, can be embedded into ML models through data augmentation or loss-based regularization. In loss-based regularization, an additional term in the objective function penalizes deviations from symmetry [43,44]. However, this requires a careful balance between loss coefficients. Data augmentation, on the other hand, incorporates the symmetry by applying symmetry-preserving transformations to the training data, thereby guiding the learned model to respect those transformations [45]. This method has been widely used in computer vision [46,47], where invariance to translations or rotations is critical. While weak constraints have received less attention in recent ML for physics applications due to the increasing popularity of hard constraint approaches, recent works have demonstrated data augmentation's effectiveness in domains such as particle physics for its ease of embedding complex symmetries. For example, Lorentz-boosted data augmentation has been shown to improve model performance in jet tagging tasks [48].

In this work, we revisit data augmentation as a method for embedding physical symmetries in data-driven plasma models inferred from fully kinetic particle-in-cell (PIC) simulations. We specifically focus on the effects of embedding frame-invariance symmetry, according to Lorentz (and Galilean) transformations, in the data-driven inference of reduced plasma models of collisionless magnetic reconnection. In Sec. II A, we demonstrate the recovery of the two-fluid equations for collisionless magnetic reconnection via sparse regression and show that symmetry-augmented data (1) yield more accurate coefficients, (2) suppress spurious symmetry-violating terms, and (3) reduce the amount of expensive laboratory-frame PIC data required. Building on these results, we apply this methodology in Sec. II B to the problem of inferring an accurate closure model for the electron fluid pressure tensor in collisionless magnetic reconnection. Specifically, we apply Galilean-boosted data augmentation (the low-velocity limit of Lorentz boosts) to embed Galilean frame invariance in the closure model and show that it not only outperforms the model obtained from laboratory-frame-only data, but also exceeds the accuracy of commonly used analytical closure models. We then further show that symmetry embedding via data augmentation is not limited to SR. Neural networks trained with Galilean-augmented data also yield improved pressure closures, underscoring the broader applicability of this approach. Finally, we conclude in Sec. III by discussing the implications of symmetry embedding for the future development of reduced plasma models capable of faithfully capturing multiscale dynamics.

II. RESULTS

We examine the effectiveness of data augmentation to embed physical symmetries into machine-learned reduced plasma models from the data of first-principles kinetic simulations. As a case study, we focus on the plasma physics problem of developing reduced fluid plasma descriptions of collisionless magnetic reconnection, which we introduce and motivate below.

A. Recovering fluid plasma models of magnetic reconnection from kinetic simulation data

Magnetic reconnection [49–53] is a fundamental plasma physics process where a change in the magnetic topology

converts energy originally stored in the magnetic field to plasma heat, bulk flow kinetic energy, and energetic particle production; it is ubiquitous to both laboratory and astrophysical plasmas. For example, in magnetically confined fusion plasmas magnetic reconnection is responsible for disruptions and sawtooth crashes, while in solar physics MR is responsible for solar flares [54]. MR is a quintessential multiscale process as there exists a hierarchy of interconnected length scales and timescales. Changes in the magnetic topology typically occur within a diffusion region at the ion skin depth scale, which is much larger than the electron skin depth and much smaller than the system size [55]; a plethora of coherent structures and waves interconnects these disparate scales. Reduced plasma models help solve the multiscale problem by integrating out the finest scale details leading to computational speedup. However, reduced plasma models can also serve a vital role for improving theoretical understanding. Despite significant progress over the last decades, there are still important theoretical challenges in MR, including the origin of the fast reconnection rate [55,56] and the dominant mechanisms of anomalous resistivity [57]; the discovery, implementation, and examination of reduced plasma models of MR are important to shed light into these open questions.

We generate high-fidelity data of the dynamics of collisionless, magnetic reconnection using the fully kinetic PIC simulation code *OSIRIS* [58,59]. Specifically, we simulate collisionless, low plasma beta (0.1), nonrelativistic magnetic reconnection in a pair plasma ($m_e = m_p = 1$). To isolate the fundamental plasma process of magnetic reconnection, we initialize a doubly periodic force-free Harris current sheet [60] in the (x, y) plane; details about the physical and numerical parameters of the simulation are described in Appendix A. While the spatial grid is in two dimensions, the PIC method evolves particle velocities in all three (x, y, z) directions. At each time step over all spatial grid points, we output fluid moment diagnostics, which are integrals of the particle distribution function with respect to powers of the phase space momentum $\boldsymbol{p}$. In general, the kth order moment is defined as [37]

$$\mathcal{M}^{\alpha_1\cdots\alpha_k} \equiv \int \frac{d\boldsymbol{p}}{p^0} p^{\alpha_1}\cdots p^{\alpha_k} f, \tag{1}$$

where $f(\boldsymbol{x}, \boldsymbol{p}, t)$ is the particle distribution function. We also output electromagnetic field diagnostics and can construct the electromagnetic field strength tensor $F^{\mu\nu}$. Since the PIC method keeps track of the fluid moment and electromagnetic fields at different space-time locations during the loop, we postprocess our data by half-time-stepping all particles and spatially centering the electromagnetic field data to ensure all quantities are known at the same space-time locations. The PIC simulation data used in this paper can be found on Zenodo [61].

1. Lorentz data augmentation

Our goal is to examine the effects of embedding Lorentz symmetry into reduced fluid models inferred through sparse regression. We accomplish this by augmenting our MR dataset using analytic Lorentz transformations (also known as Lorentz boosts). This is the same underlying principle as embedding rotational/translational symmetry by augmenting the data with rotations/translations, with the only difference that the data augmentation operator is Lorentz boosting. For a Lorentz boost velocity $\boldsymbol{\beta} = (\beta_x, \beta_y, \beta_z)$ $[c]$ (with corresponding Lorentz factor $\gamma_{\boldsymbol{\beta}} \equiv \frac{1}{\sqrt{1-(\frac{\boldsymbol{\beta}}{c})^2}}$) relative to the simulation, laboratory frame, we are able to construct the fluid moments as observed by the reference frame moving at velocity $\boldsymbol{\beta}$. We denote quantities observed in the Lorentz-boosted reference frame with primed variables,

$$\mathcal{M}^{\alpha'\beta'\cdots\xi'} = \Lambda^{\alpha'}_{\mu}\Lambda^{\beta'}_{\nu}\cdots\Lambda^{\xi'}_{\rho}\mathcal{M}^{\mu\nu\cdots\rho}, \tag{2}$$

where Λ^{μ}_{ν} is the Lorentz transformation tensor for the given boost $\boldsymbol{\beta}$. We construct the electromagnetic fields as seen by the Lorentz moving observer as

$$F^{\alpha'\beta'} = \Lambda^{\alpha'}_{\mu}\Lambda^{\beta'}_{\nu}F^{\mu\nu}. \tag{3}$$

We also analytically boost temporal and spatial derivatives of the fluid moments and electromagnetic fields in the laboratory frame using analytic Lorentz transformations,

$$\partial_{\alpha'} = \Lambda^{\mu}_{\alpha'}\partial_{\mu}. \tag{4}$$

These Lorentz transformations can all be applied to calculate the physical quantities and their space/time derivatives under any boost from a single laboratory-frame simulation. In our numerical experiments, we will apply random Lorentz boost transformations of the data along the two principle axes (x, y), forward and backward, for varying Lorentz factors $\gamma_{\boldsymbol{\beta}}$.

2. Two-fluid equations

To explore the effects of including Lorentz-augmented data on the ML inference of reduced plasma physics models from kinetic simulation data, we will focus on the recovery of the well-known two-fluid plasma model equations, which we introduce below. The two-fluid equations are Lorentz invariant reduced plasma models and their inference should be improved by embedding Lorentz symmetry. The two-fluid model treats electrons and positrons as separate collisionless fluids. The fluids are coupled together by Maxwell's equations. Each species satisfies an infinite hierarchy of exact, coupled equations that describe the time evolution of the fluid moments $\partial_t\mathcal{M}^{\alpha_1\cdots\alpha_k}$. The equation for the time evolution of the kth moment is derived by taking moments of the collisionless Vlasov equation with respect to $p^{\alpha_1}\cdots p^{\alpha_k}$. Upon truncation, the two-fluid equations provide a substantial computational speedup compared with PIC as they no longer resolve dynamics of a large number of particles in momentum space as that phase-space coordinate has been integrated over. In this work, we demonstrate our reduced model discovery method on the electron moment equations up to order 2: The zeroth order continuity equation (one equation) that time evolves the timelike component of the four-current density M^{α},

$$\partial_t n_e + \partial_{x_j}[n_e\overline{v_{ej}}] = 0, \tag{5}$$

the first-order momentum equations (three equations) that time evolve the timelike components of the stress-energy tensor $\mathcal{M}^{\alpha\beta}$,

$$\partial_t[n_e\overline{p_{ek}}] + \partial_{x_j}[n_e\overline{v_{ej}p_{ek}}] = -n_e[E_k + \epsilon_{klm}\overline{v_{el}}B_m], \tag{6}$$

and the second-order energy equations (six unique equations) that time evolve the timelike components of the stress-flow tensor $M^{\alpha\beta\gamma}$,

$$\partial_t[n_e\overline{p_{ek}p_{el}}] + \partial_{x_j}[n_e\overline{v_{ej}p_{ek}p_{el}}] = -n_e[\overline{p_{ek}}E_l + \overline{p_{el}}E_k + \epsilon_{lmn}\overline{p_{ek}v_{em}}B_n + \epsilon_{kop}\overline{p_{el}v_{eo}}B_p], \tag{7}$$

where n_e is the electron density and $\overline{a} \equiv \frac{1}{n_e}\int f_e a dp_e^3$. $|q_e| = m_e = 1$ in normalized units. v_{ej} is the electron phase space velocity in the jth direction that is related to the electron phase space momentum $p_{ej} = \frac{v_{ej}}{\sqrt{1-(\frac{v_{ex}}{c})^2-(\frac{v_{ey}}{c})^2-(\frac{v_{ez}}{c})^2}}$. The one continuity equation, three momentum equations, and six energy equations comprise ten moment equations that we attempt to rediscover. In principle, we could recover moment equations higher than order 2; however, we truncate our demonstration at the level of the energy equations.

3. Embedding lorentz invariance into the sparse regression recovery of two-fluid equations

To recover the two-fluid equations from our collisionless magnetic reconnection data, we follow the algorithmic procedure as proposed in Ref. [21] PDE-FIND (the PDE generalization of SINDy [20]). PDE-FIND discovers partial differential equations directly from data by using sparse regression to fit unknown coefficients next to a prescribed basis of catalog terms to describe a target term. SR differs from traditional regression as it penalizes nonzero coefficients, resulting in a sparse approximation to the target. This leads to PDEs that capture the essence of the dynamics while still being interpretable. In this study, the regression targets are the time derivative terms from the two-fluid continuity, momentum, and energy equations as expressed in Eqs. (5)–(7). Hence, we have ten unique regression targets and perform the SR equation inference procedure in ten separate instances. For each equation inference, we provide a basis comprised of fluid moments, electromagnetic fields, their spatial derivatives, and polynomial combinations of the aforementioned terms in an attempt to describe the target dynamical term; see Appendix B for a full description of the catalog terms for each equation inference.

Importantly, to effectively handle the evaluation of PDE terms in the presence of noisy data, associated with discrete particle fluctuations (inherent to PIC simulation data), we pose the SR inference of PDEs in their integral form rather than their differential form [23,62–65]. This technique was shown to be crucial for the accurate and robust inference of plasma PDEs from the data of first-principles PIC simulations [23] and has become well established in recent years [64,65]. Thus, in this work the target time derivative terms and the library of candidate PDE terms are evaluated using finite differences on the data, followed by integrating over small compact space-time volumes $\{5\Delta x, 5\Delta y, 5\Delta t\}$ (which correspond to physical scales $\{0.5d_e, 0.5d_e, 0.33w_{pe}^{-1}\}$). In general, a suitable choice for the size of these volumes is problem dependent—they should be large enough to smooth the PIC fluctuations, yet smaller than characteristic length scales/timescales in the problem to retain information of physical variation. In our numerical experiments here, the choice of $\{5\Delta x, 5\Delta y, 5\Delta t\}$ volumes satisfies the above criteria for collisionless magnetic reconnection. We further confirmed that our results were weakly sensitive to the size of these space-time volumes, as long as they met the two above criteria. For each equation inference, we begin by performing SR on only the original laboratory-frame data, without Lorentz transformations. We use a sample of $15^3 = 3375$ randomly distributed $\{5\Delta x, 5\Delta y, 5\Delta t\}$ measurement volumes from the full MR dataset [Fig. 1(a1)]. In each of these sampled volumes, we evaluate the target terms (for example, $\partial_t n_e$, which appears in the continuity equation, and the library of catalog terms [Fig. 1(a2)]) and integrate them over each volume; see Appendix C for more data sampling details. Each of the catalog terms is attached with an unknown coefficient $\boldsymbol{\lambda}$. All measurements of a catalog term share the same unknown coefficient. SR then fits $\boldsymbol{\lambda}$ to the data in an *optimal* manner [Fig. 1(a3)] using a variation of the sequential thresholding least-squares algorithm as proposed in Ref. [20] and tenfold cross validation, and we infer a governing equation from the regression procedure [Fig. 1(a4)]. A detailed description of the SR algorithm used in this work is described in Ref. [23].

We now repeat the PDE discovery procedure using Lorentz-transformed data. Using the analytical Lorentz transformations described earlier, we boost each sampled volume into a randomly moving Lorentz reference frame as illustrated in Fig. 1(b1). Each volume is boosted in a random direction [along the two principal (x, y) axes], with a random Lorentz factor sampled uniformly in the range $1 \leqslant \gamma_{\boldsymbol{\beta}} \leqslant \gamma_{\boldsymbol{\beta}\max} = 1.5$, which corresponds approximately to $0 \leqslant |\boldsymbol{\beta}| \leqslant |\boldsymbol{\beta}_{\max}| = 0.75c$. The maximum fluid velocity in the MR simulation is $\sim 0.5c$, and we choose $\gamma_{\boldsymbol{\beta}\max}$ large enough such that some Lorentz reference frames experience velocities greater than those in the laboratory frame; we will later examine the effects of choosing different $\gamma_{\boldsymbol{\beta}\max}$. We note that although we adopt a uniform sampling of $\gamma_{\boldsymbol{\beta}} \in [1, \gamma_{\boldsymbol{\beta}\max}]$ for data augmentation, more general sampling distributions that preferentially weigh different boosts are also possible; here, however, we restrict ourselves to uniform sampling, which we empirically find to be effective. Each row of the candidate term matrix now corresponds to an observation of the measured target and catalog terms taken by a randomly moving Lorentz observer, as shown in Fig. 1(b2). While we later consider combining laboratory-frame data with arbitrarily many Lorentz-boosted samples for true data augmentation, we begin by comparing laboratory-frame-only and boosted-only datasets, with each laboratory-frame sample boosted once.

Having carried out the random Lorentz transformations of each measurement volume, we construct our framework to ensure that all Lorentz reference frames agree on the same coefficients $\boldsymbol{\lambda}'$, which embeds Lorentz invariance into SR. With the Lorentz-transformed data, we then solve the regression procedure [Fig. 1(b3)] and infer the governing PDE [Fig. 1(b4)]. We repeat this procedure for each of the ten two-fluid equations, extracting model coefficients discovered with laboratory-frame-only and boosted-only data. We note that because (1) Lorentz transformations only require simple, computationally efficient arithmetic operations

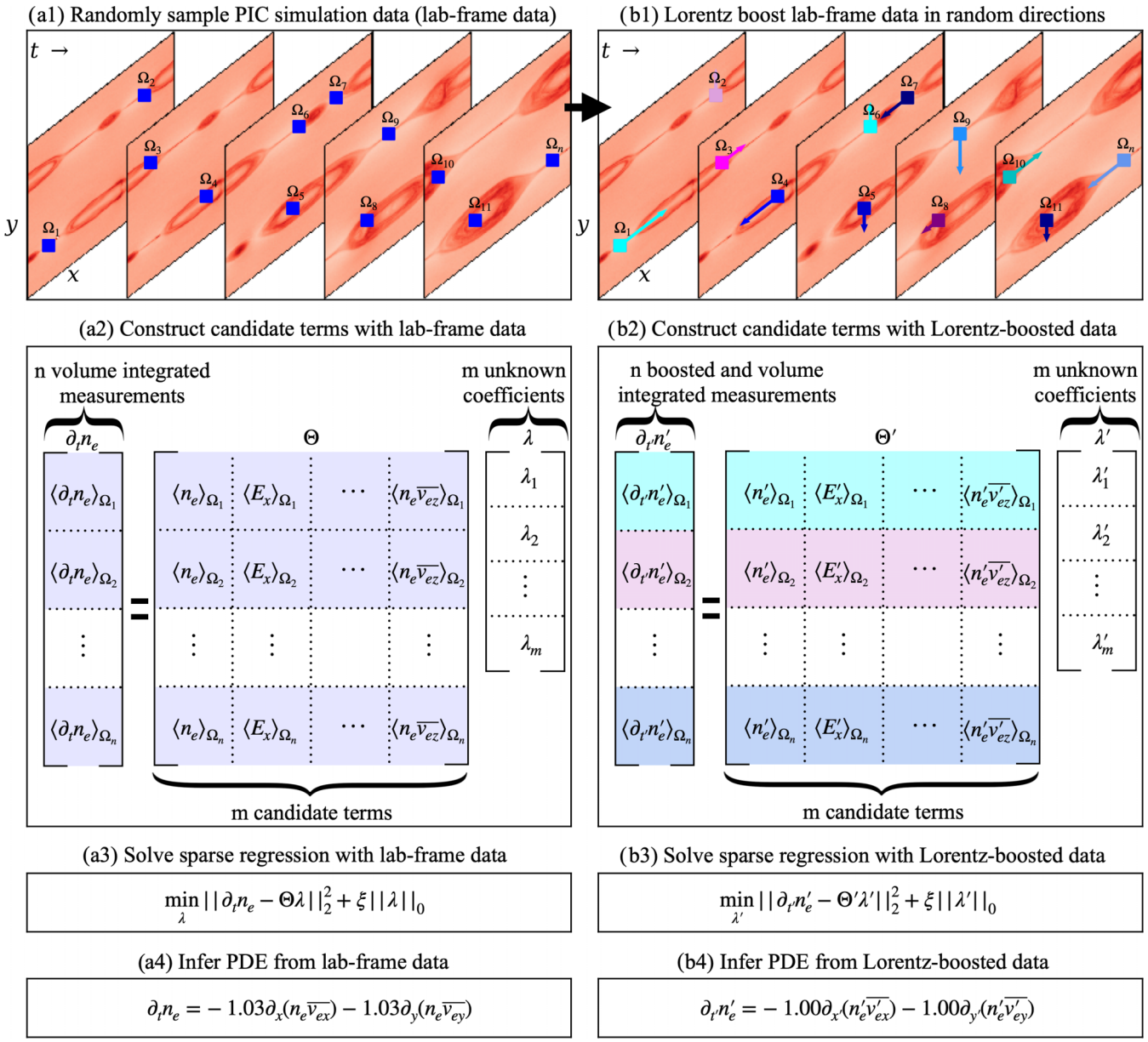


FIG. 1. Schematic of sparse regression workflow to discover reduced fluid equations with and without Lorentz-transformed data. A single PIC MR simulation performed in a stationary reference frame (which we denote from now on as the laboratory frame) provides laboratory-frame data for the electron number density n_e (shown above in color), the fluid velocities $\overline{v_{ej}}$, and the electromagnetic fields at finite temporal slices in the (x, y) plane. In step (a1), we randomly sample measurements from the laboratory-frame data, averaging over compact space-time volumes. This averaging smooths noise associated with discrete particle effects. To conduct SR with Lorentz-transformed data [panel (b1)], we use analytical transformations to relate the quantities as measured by a random Lorentz observer to the quantities as measured in the simulation, laboratory frame. Different color squares represent quantities as measured by Lorentz observers moving with different velocities. In panels (a2) and (b2), we balance m measurements of the dynamical term, $\partial_t n_e$, with measurements of catalog terms multiplied by unknown coefficients for laboratory-frame and boosted data, respectively. Crucially, in panel (b2), each Lorentz-boosted reference frame shares the same vector of unknown coefficients λ', which embeds the symmetry into the inference procedure. In panels (a3) and (b3), we solve the SR problem, and in panels (a4) and (b4) we examine the inferred PDEs that we may compare to the true continuity equation: $\partial_t n_e = -\partial_x(n_e\overline{v_{ex}}) - \partial_y(n_e\overline{v_{ey}})$. To infer higher-order moment fluid equations, like the momentum or energy equations, we follow the same algorithmic procedure but with different dynamical and catalog terms.

(2) the computational demands of machine learning methods like sparse regression scale approximately linearly with the number of training data points; Lorentz data transformations do not introduce significant, additional computational overhead. In the following, we discuss the effects of using Lorentz-transformed data on the inferred models. While the exact degree of improvement will depend on the specific plasma problem, the magnetic reconnection case studied here illustrates the main aspects of the gains achievable with Lorentz data transformations.

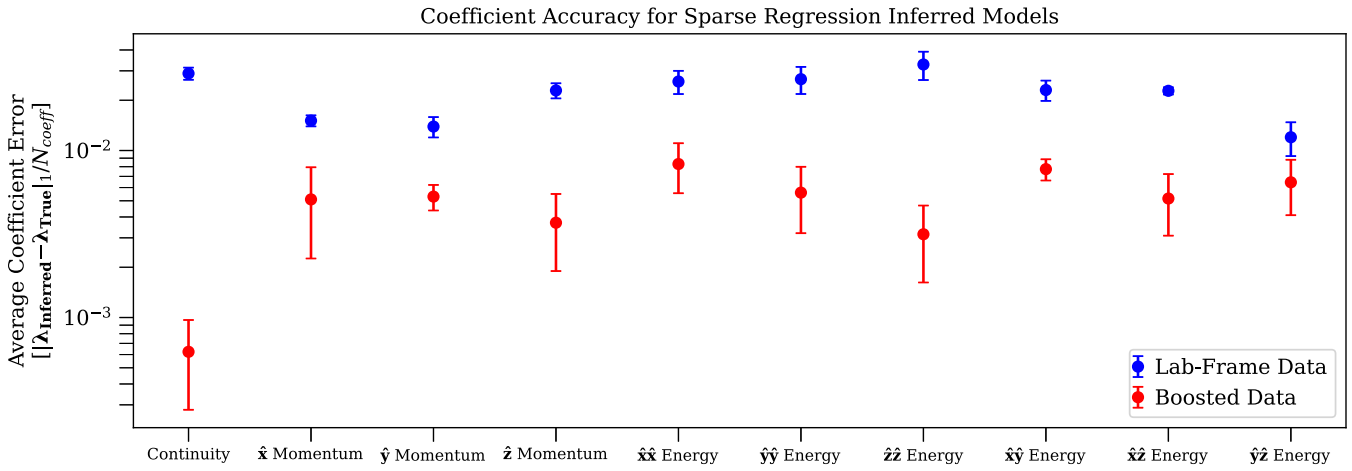


FIG. 2. Including Lorentz-boosted data in the rediscovery of the two-fluid reduced model equations greatly improves coefficient accuracy. On the vertical axis is average model coefficient error measured as the L1 norm between true and inferred coefficients normalized by the number of nontime derivative terms in the equation, while on the horizontal axis are the first ten unique two-fluid plasma equations. The blue points represent coefficient errors for SR inferred equations with only laboratory-frame data, while red points represent boosted data. The vertical error bars correspond to ±one standard deviation in average coefficient error, and the statistics come from training each model five times with different realizations of the random data sampling. For all ten equations, Lorentz boosting improves the mean accuracy of the inferred coefficients beyond one standard deviation.

4. Improved coefficient accuracy

The first and most apparent finding is that the coefficients inferred from SR are more accurate when using Lorentz-transformed data. Since the two-fluid equations can be derived analytically by taking moments of the collisionless Vlasov equation, we can compare the true two-fluid coefficients with the inferred coefficients (Fig. 2). For all ten two-fluid reduced model equations, incorporating randomly Lorentz-boosted data in the training set decreases average coefficient error beyond a standard deviation. The models discovered with laboratory-frame data have an average coefficient error of 1.34%, while for the models discovered with boosted data it is 0.15%. Relatively, the boosted models are 9 times more accurate than laboratory-frame models. See Appendix D for a table of discovered coefficients used to create Fig. 2.

We also examine how coefficient accuracy improves as a function of $\gamma_{\beta\max}$, the maximum Lorentz factor of each random Lorentz boost. Figure 3 shows coefficient accuracy as a function of $\gamma_{\beta\max}$ for the inferred continuity equation; we observe an elbow in the accuracy vs boost magnitude plot at $\gamma_{\beta\max} \sim 1.2$, corresponding to a boost magnitude $|\boldsymbol{\beta}| \sim 0.5c$, which is the maximum fluid proper velocity reached in the simulation. To reap the full benefits of Lorentz data transformations, one must Lorentz boost data into regimes outside of the existing dynamics. We show additional examples of average coefficient error decreasing with increasing $\gamma_{\beta\max}$ for the $\hat{\mathbf{z}}$ momentum and $\hat{\mathbf{z}}\hat{\mathbf{z}}$ energy equations in Appendix E.

To shed light into why coefficient accuracy improves as a function of $\gamma_{\beta\max}$, let us write down the regression loss function for the inference of the continuity equation with data that have been Lorentz transformed by $\boldsymbol{\beta} = \beta\hat{\mathbf{x}}$,

$$\mathcal{L}' \sim \partial_{t'} n'_e + \lambda'_1 \partial_{x'}[n'_e \overline{v'_{ex}}], \tag{8}$$

where λ'_1 is an unknown coefficient. From the Lorentz-boosted simulation data, we have measurements of the terms $\partial_{t'} n'_e$ and $\partial_{x'}[n'_e \overline{v'_{ex}}]$. Using the inverse Lorentz transform $(\Lambda^{-1})^{\mu}_{\nu}$, we can express $\mathcal{L}'$ in terms of the laboratory-frame moments. Crucially, we define the unknown coefficient $\lambda'_1 =$

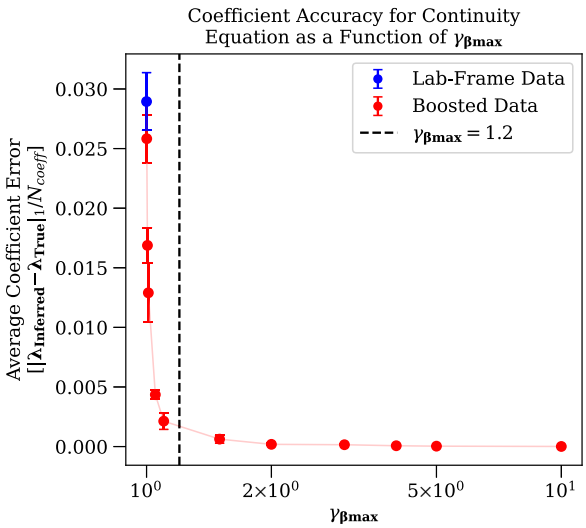


FIG. 3. Average coefficient error decreases as a function of maximum Lorentz boost ($\gamma_{\beta\max}$). The vertical axis measures the L1 difference between true and SR inferred coefficients. We randomly Lorentz transform our data in the range $1 \leqslant \gamma_\beta \leqslant \gamma_{\beta\max}$ and demonstrate that increasing $\gamma_{\beta\max}$ (embedding the symmetry more stringently) leads to more accurate coefficients for the continuity equation recovery. The vertical error bars correspond to ±one standard deviation in average coefficient error, and the statistics come from training each model five times with different realizations of the random data sampling.

λ_1 to be invariant under the inverse Lorentz transformation.

$$\mathcal{L}' \sim \gamma_{\boldsymbol{\beta}}^2\{\partial_t n_e + \lambda_1 \partial_x(n_e\overline{v_{ex}}) + \beta[\lambda_1 - 1]\left[\frac{\partial_t(n_e\overline{v_{ex}})}{c^2} - \partial_x n_e\right] - \left(\frac{\beta}{c}\right)^2[\lambda_1\partial_t n_e + \partial_x(n_e\overline{v_{ex}})]\}. \quad (9)$$

We recognize the first line of Eq. (9) to be the laboratory-frame loss function $\mathcal{L} \sim \partial_t n_e + \lambda_1\partial_x[n_e\overline{v_{ex}}]$.

$$\mathcal{L}' \sim \gamma_{\boldsymbol{\beta}}^2\{\mathcal{L} + \beta[\lambda_1 - 1]\left[\frac{\partial_t(n_e\overline{v_{ex}})}{c^2} - \partial_x n_e\right] - \left(\frac{\beta}{c}\right)^2[\lambda_1\partial_t n_e + \partial_x(n_e\overline{v_{ex}})]\}. \quad (10)$$

Equation 10 tells us that the boosted frame loss function is proportional to the laboratory-frame loss function plus additional terms tuned by the boost magnitude β. In the limit $\gamma_{\boldsymbol{\beta}} \to 1$, corresponding to $\frac{\beta}{c} \to 0$, $\mathcal{L}' \sim \mathcal{L}$. Much more interesting is the large Lorentz boost limit $\gamma_{\boldsymbol{\beta}} \to \infty$, corresponding to $\beta \to c$. In this limit, Eq. (9) becomes

$$\mathcal{L}' \sim \gamma_{\boldsymbol{\beta}}^2\left\{[1 - \lambda_1]\left[\partial_t\left(n_e - \frac{n_e\overline{v_{ex}}}{c}\right) + \partial_x(n_e c - n_e\overline{v_{ex}})\right]\right\}. \quad (11)$$

Measurements of the term $[\partial_t(n_e - \frac{n_e\overline{v_{ex}}}{c}) + \partial_x(n_e c - n_e\overline{v_{ex}})]$ from Eq. (11) are not typically equal to zero in the MR dataset; thus, to minimize the loss function in the large Lorentz boost limit the multiplicative term $[1 - \lambda_1] \to 0$, $\lambda_1 \to 1$. From theory, we know that the true continuity coefficient is indeed $\lambda_1 = 1$. In the same manner, boosting along the $\hat{\boldsymbol{y}}$ axis would fix the unknown coefficient attached to $\partial_{y'}[n_e'\overline{v_{ey}'}]$ to its true value, also 1.

For the interested reader, Appendix F examines the impact of applying a single vs multiple Lorentz boost transformations to each original laboratory-frame data point on the accuracy of the inferred models. This analysis clarifies the relative roles of (1) enriching the dataset through Lorentz-based data transformations and (2) constraining the regression to the subspace of Lorentz-equivariant models by repeatedly observing the same data in different inertial frames. In Appendix G, we further demonstrate explicitly that the use of Lorentz-transformed data not only improves the accuracy of the inferred coefficients, but also yields models that more faithfully satisfy Lorentz invariance.

5. *Elimination of spurious unphysical terms*

ML models find correlations in the data, but these correlations may not be physical. We find that Lorentz data augmentation can break unphysical correlations in ML models. For example, when discovering the $\hat{\boldsymbol{x}}\hat{\boldsymbol{y}}$ component of the energy equation with only laboratory-frame MR data, an unphysical term is consistently identified as the ninth model term by the laboratory-frame SR, even though we know from theory that the true equation has eight terms [Fig. 4(a1)]. However, upon boosting our MR data into randomly moving Lorentz reference frames, we see in Fig. 4(a2) that the regression identifies the correct eight-term model without the spurious term. Figure 4(b) reaveals how the spurious term is removed by boosting: In the laboratory frame, there exists a linear correlation between a spurious candidate term $n_e\overline{v_{ex}p_{ex}}\partial_y[n_e\overline{v_{ey}p_{ex}p_{ey}}]$ and the model residual. Hence, it is convenient for the regression procedure to attach a nonzero coefficient, -0.2298, to this term that reduces the loss function. Yet, after applying Lorentz boost transformations to the data, the once linear relation between model loss and term measurement is broken. This is because the functional form of the spurious term is not Lorentz invariant and appears different for independent Lorentz moving observers, i.e., $n_e\overline{v_{ex}p_{ex}}\partial_y[n_e\overline{v_{ey}p_{ex}p_{ey}}] \neq n_e'\overline{v_{ex}'p_{ex}'}\partial_{y'}[n_e'\overline{v_{ey}'p_{ex}'p_{ey}'}]$. This is perhaps the most important effect of Lorentz data augmentation and is crucial to enable generalizable ML models. We will return to this point in Sec. II B.

6. *Improved data efficiency*

State-of-the-art machine learning models are notoriously data hungry [66]. Generating, storing, preprocessing, transferring, and training with large amounts of data can be a significant bottleneck in the development of generalizable ML reduced models of physical systems. Here, we find that not only does Lorentz data augmentation increase the amount of physically consistent training data without having to generate more through expensive PIC simulations, but we find it improves the model accuracy significantly faster than by including more original, laboratory-frame data. We show this effect in Fig. 5, where we examine coefficient accuracy for the continuity equation as a function of varying the relative proportions of original laboratory-frame data and Lorentz-transformed data. With only $2^3 = 8$ laboratory-frame measurements, in the lower left corner of Fig. 5, we find a large average continuity coefficient error of 0.257. This is not surprising given the small amount of data used to build the model. Moving up one square, we randomly boost these 8 laboratory-frame measurements 56 times to have $4^3 = 64$ total laboratory-frame + boosted measurements and the average continuity coefficient error significantly decreases to 0.001. This same phenomenon is observed for any starting number of laboratory-frame measurements. For a fixed number of laboratory-frame samples, the coefficient error saturates after a sufficient number of augmented boosted samples. Interestingly, even with $10^3 = 1000$ laboratory-frame measurements we observe an average continuity coefficient error of 0.032, much higher than what can be achieved with only $2^3 = 8$ laboratory-frame measurements combined with Lorentz boosts. These results underscore the effectiveness of data augmentation in improving the accuracy of ML models.

B. Discovering symmetry-preserving fluid closures

1. *Sparse regression inferred closure*

Having established the benefits of Lorentz-boosted data augmentation on the recovery of known equations in Sec. II A, we now extend the application of this technique to the development of accurate, interpretable, and symmetry-preserving plasma fluid closures from our kinetic simulations of collisionless magnetic reconnection. The development of such closure models is important to enable accurate and computationally efficient fluid modeling of magnetic reconnection

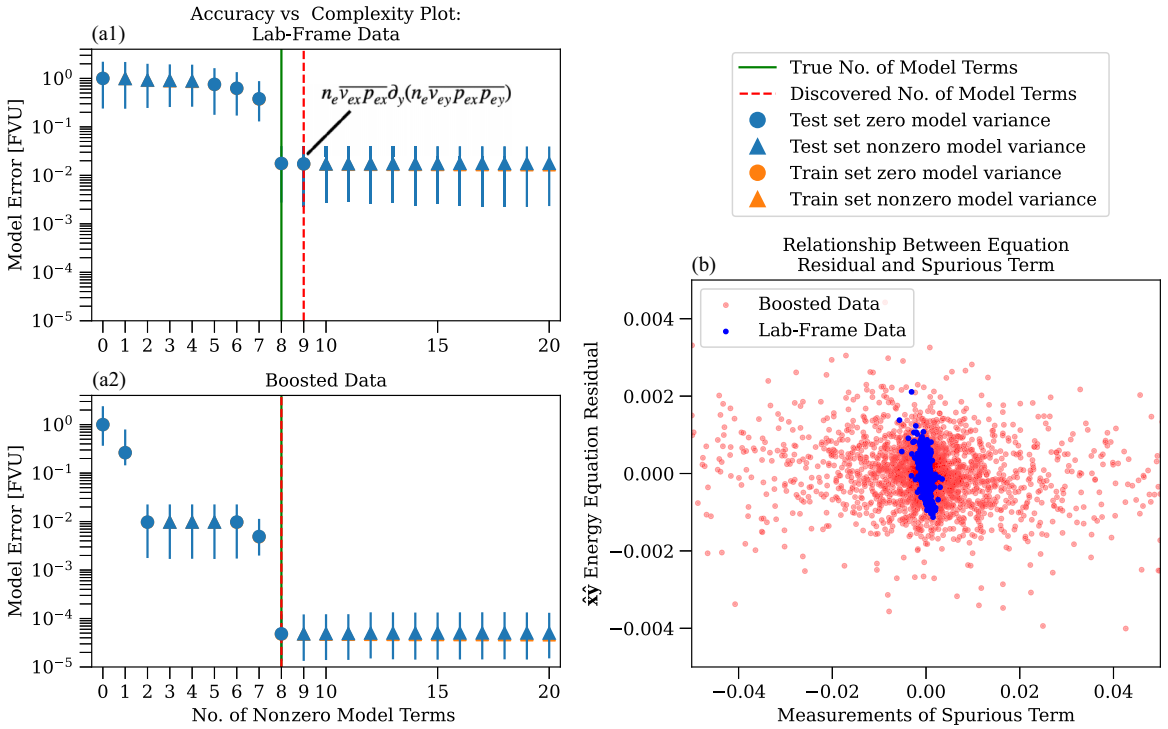


FIG. 4. Lorentz data transformation removes an unwanted, spurious model term identified by the laboratory-frame sparse regression. In panels (a1), laboratory-frame data, and (a2), Lorentz-boosted data, we display the sparse regression inferred $\hat{\mathbf{x}}\hat{\mathbf{y}}$ energy equation model error as a function of model terms. Circles represent zero model form variance for all SR K-folds, while triangles represent nonzero. See the upper right quadrant for the legend for panels (a1) and (a2). Accuracy is measured in terms of Fraction of Variance Unexplained (FVU). FVU $\equiv \frac{\mathrm{Var}(\hat{f}-f)}{\mathrm{Var}(f)}$, where f is the target variable (in this case, measurements of $\partial_t[n_e\overline{p_{ex}p_{ey}}]$) and $\hat{f}$ is the model (in this case, the SR inferred model for $\partial_t[n_e\overline{p_{ex}p_{ey}}]$). Complexity is measured in terms of the number of nonzero model terms in the equation. The true $\hat{\mathbf{x}}\hat{\mathbf{y}}$ energy equation has eight terms and is marked with a green vertical line. SR performed with laboratory-frame data identifies an incorrect nine-term model (an extra incorrect term), while with boosted data identify the correct eight-term model; these models are marked by red vertical lines. Panel (b) offers an explanation for Lorentz boosting removing the extra, spurious term; a once linear relationship between measurements of the spurious ninth term "$n_e\overline{v_{ex}p_{ex}}\partial_y[n_e\overline{v_{ey}p_{ex}p_{ey}}]$" and the $\hat{\mathbf{x}}\hat{\mathbf{y}}$ energy equation residual breaks upon Lorentz boosting. The relation becomes more nonlinear and disperse. With laboratory-frame data, the Pearson Correlation Coefficient (PCC), which measures normalized covariance of two variables, is -0.399. For boosted data, the PCC is -0.204, meaning that the linear relationship lessened upon boosting.

in multiscale plasma scenarios, such as magnetized plasma turbulence and magnetospheric physics [6,67–69].

Here, we focus on the data-driven inference of a closure model for the electron pressure tensor for collisionless magnetic reconnection. Specifically, we seek to express the electron pressure tensor $P_{ejk} \equiv \int f_e(v_{ej}-\overline{v_{ej}})(p_{ek}-\overline{p_{ek}})dp_e^3$ as a function of lower-order moments and fields to close the infinite hierarchy of two-fluid equations introduced in Sec. II A. While the pressure tensor appears in a Lorentz-invariant equation, it itself is not Lorentz invariant. However, for small Lorentz boosts (Galilean boosts) the pressure tensor remains invariant and thus we should expect our SR inferred closures to abide by this important symmetry. In the large magnetic field limit (applicable in guide field reconnection), the fluid moments and electromagnetic fields transform as follows under Galilean transformations, where β_j is a Galilean boost velocity vector and quantities as observed in the moving reference frame are denoted by primed variables:

$$n_e' = n_e, \tag{12}$$

$$\overline{v_{ej}'} = \overline{v_{ej}} - \beta_j, \tag{13}$$

$$P_{ejk}' = P_{ejk}, \tag{14}$$

$$E_j' = E_j + \epsilon_{jkl}\beta_k B_l, \tag{15}$$

$$B_j' = B_j. \tag{16}$$

Instead of discovering the pressure tensor component by component in a Cartesian basis, the more natural basis in magnetized plasmas is defined with respect to the local magnetic field. In this basis, the relevant components are the parallel ($P_{e\parallel}$) and perpendicular ($P_{e\perp}$) components of the pressure tensor.

To illustrate how data augmentation improves closure discovery, we focus on the component of the pressure tensor parallel to the local magnetic field. From Eqs. (14) and (16), we see that the electron pressure local to the magnetic field is invariant under a Galilean boost; thus, for any Galilean boost β_j, $P_{e\parallel}' = P_{e\parallel}$ and $P_{e\perp}' = P_{e\perp}$. The evolution of $P_{e\parallel}$ is more complex than the evolution of $P_{e\perp}$ in the MR regime considered in this work (medium guide field regime); thus, we focus

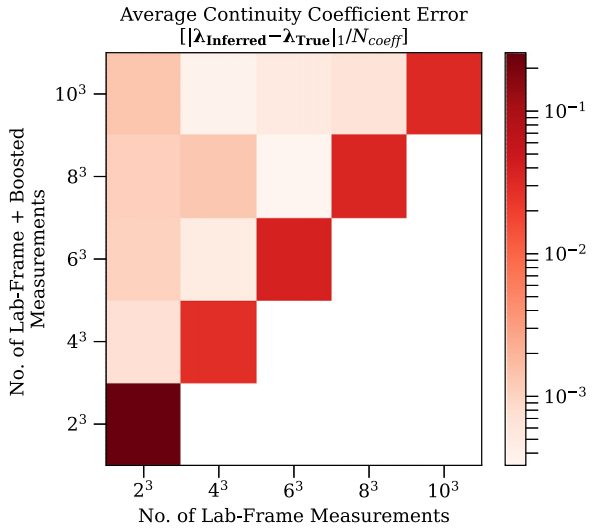


FIG. 5. Lorentz boosting improves data efficiency in reduced model discovery. The color is the average continuity coefficient error measured with the L1 norm. The horizontal axis depicts the number of randomly sampled laboratory-frame measurements from our MR simulation. The vertical axis represents this number of laboratory-frame measurements *plus* random Lorentz augmentations of those laboratory-frame measurements. The identity line corresponds to no data augmentation. For a fixed horizontal position corresponding to a fixed number of laboratory-frame data samples, moving vertically corresponds to increasing the number of Lorentz augmentations of those original, laboratory-frame samples. We see a sharp drop in average continuity coefficient error moving off of the identity line, indicating that embedding Lorentz invariance into the SR equation inference has more of an effect on coefficient accuracy than simply including more laboratory-frame data. Note that the continuity equation has two model terms in two dimensions; thus, the normalizing factor $N_{\text{coeff}} = 2$ in this case.

here in the main text on discovering a closure model for $P_{e\parallel}$; we discuss closure model discovery for $P_{e\perp}$ in Appendix H.

We sample local and instantaneous quantities at 10^3 spatial locations every $10w_{pe}^{-1}$ from $t_{\min} = 0w_{pe}^{-1}$ to $t_{\max} = 250w_{pe}^{-1}$; $t_{\max}$ also corresponds to the time at which ∼50% of the initial magnetic flux has been reconnected. In the same spirit as the two-fluid equation inference, we only sample from spatial locations where the target variable, in this case $P_{e\parallel}$, is greater than its 70th percentile value at its sampled time step. We observe how well our pressure closure generalizes to times outside of the training window, $t > t_{\max}$. We provide a library of candidate terms comprised of fluid moments and electromagnetic fields $\{n_e, \overline{v_{ex}}, \overline{v_{ey}}, \overline{v_{ez}}, B_x, B_y, B_z\}$ and their polynomial combinations; the electric field is left out as a candidate term due to its strong correlation with $\overline{\boldsymbol{v_e}} \times \boldsymbol{B}$.

Without Galilean-boost augmented data, SR infers the following seven-term electron parallel pressure model (which we refer to as the SR-Lab model):

$$P_{e\parallel\text{Lab}} = -0.069 + 0.113n_e + 0.370\overline{v_{ex}}^2 - 0.291\overline{v_{ez}}^2 - 0.078B_x^2 + 0.065B_z^2 - 0.075B_z n_e. \quad (17)$$

We readily notice that the SR-Lab closure in Eq. (17) is not Galilean invariant since the terms proportional to $\{\overline{v_{ex}}^2, \overline{v_{ez}}^2\}$ are not Galilean invariant, e.g., $\overline{v'_{ez}}^2 = (\overline{v_{ez}} - \beta_z)^2 \neq \overline{v_{ez}}^2$. The identification of these terms by SR is the result of a spurious correlation with the parallel electron pressure in the original (laboratory-frame) dataset. However, this correlation is broken when the dynamics is observed from different Galilean reference frames. Indeed, adding Galilean-augmented data in the SR closure inference removes these spurious terms as all Galilean reference frames must agree on the same closure. To embed Galilean symmetry, we double the training dataset size by including random Galilean boosts of each sampled point with $|\beta_j| \leqslant 0.4c$. With Galilean-boosted data, the following six-term model (SR-Boost) is inferred:

$$P_{e\parallel\text{Boost}} = -0.071 + 0.103n_e - 0.040B_x^2 + 0.193B_y^2 + 0.062B_z^2 - 0.062B_z n_e, \quad (18)$$

which is readily observed to be Galilean invariant. A more detailed discussion of the inferred models and their physical interpretation will be left to future work. We note that for weaker Galilean boosts, for example, $|\beta_j| \leqslant 0.1c$, the symmetry breaking terms are not removed from the closure inferred with augmented data; a sufficiently strong boost is required to break the spurious correlations between model residual and unwanted terms.

Importantly, the SR-Boost closure shows significantly better generalization to unseen times $t > t_{\max}$ compared to SR-Lab (Fig. 6). During the training window ($t < t_{\max}$), both SR-Boost and SR-Lab are found to have similar model errors (Fig. 6); the removal of the spurious terms from SR-Boost results in only a small increase of the model error with respect to SR-Lab. However, the detrimental effects of the spurious terms on the accuracy of the closure models become most noticeable for $t > t_{\max}$. Indeed, at these times we observe that the spurious correlations were not only frame dependent, but also only occurred in a narrow range of time (during the training window). Elimination of these spurious terms through Galilean symmetry embedding is crucial to enable a more robust and generalizable closure model.

In Fig. 6, we further compare the obtained SR-Lab and SR-Boost closures with commonly used analytical pressure closure models for collisionless magnetic reconnection. These include the isothermal closure (where the pressure is linearly proportional to the density), the double adiabatic Chew-Goldberger-Low (CGL) closure model [70], and Le closure [67], which interpolates between isothermal and CGL closures depending on whether the plasma is in a passing or trapped state; we note that we are extending the Le closure outside of its regime of validity, $v_{\text{the}} \gg v_A$ (v_{the} and v_A are the electron thermal and Alfvén speed, respectively), which does not hold in our low beta pair plasma where $v_{\text{the}}/v_A = 0.22$. With only the original, laboratory-frame data, the SR-Lab closure does not outperform CGL (the best performing analytical model for the simulated magnetic reconnection regime). However, including Galilean-boosted data, the SR-Boost closure outperforms all other analytical closures, having the lowest model error at late times, beyond the training time horizon $t > t_{\max} = 250w_{pe}^{-1}$ (Fig.e 6). At these late times, the analytical models greatly underestimate $P_{e\parallel}$, leading to their large errors, and the improved performance of SR-Boost over SR-Lab is primarily due to its more accurate predictions upstream of

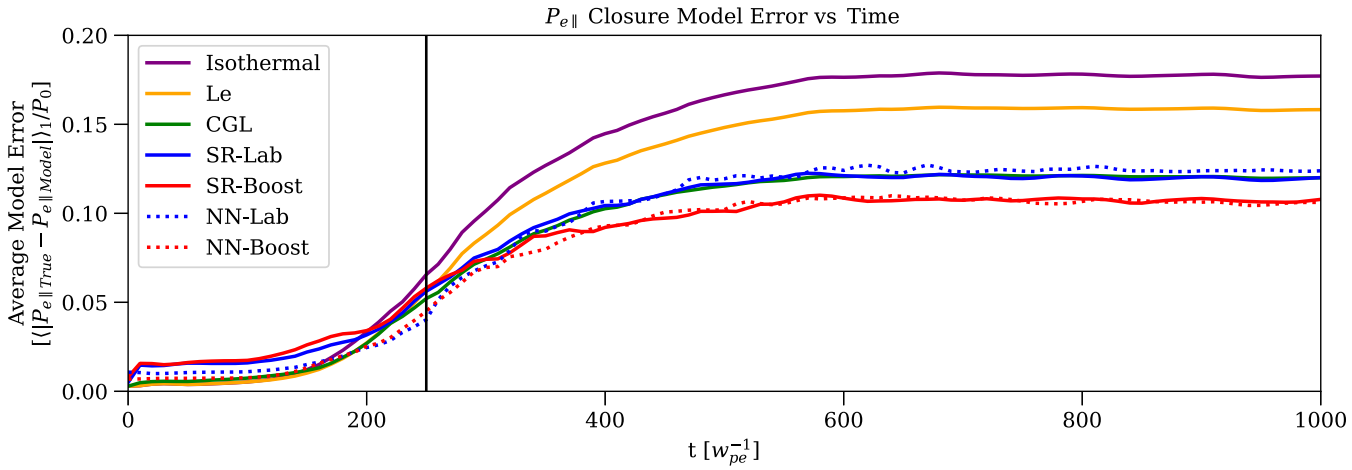


FIG. 6. Machine-learned $P_{e\parallel}$ pressure closures for collisionless magnetic reconnection—discovered with Galilean-augmented data—not only outperform models discovered without symmetry embedding, but further commonly used analytic pressure closures. The vertical axis corresponds to model error measured as the spatial mean of the absolute L1 difference between the true simulation pressure and the pressure closure model. We normalize the average pressure error to a characteristic pressure value in the simulation $P_0 = 0.1 n_0 m_e c^2$. The horizontal axis is time. The SR and NN models are trained until $t = t_{\max} = 250 w_{pe}^{-1}$ and then generalized to future times. Isothermal, Le, and CGL are commonly used pressure models with analytic functional forms. As time increases and reconnection becomes more nonlinear and complicated, all closures perform worse. At late, generalization times the machine-learned closures discovered with Galilean-augmented data have the lowest model error.

the reconnecting current sheet, outside the plasmoid regions. We present in Appendix I a more detailed discussion about the error distributions of the pressure closure models used in Fig. 6.

We note that at early times ($t < t_{\max}$), corresponding to the training phase, the analytical closure models outperform both SR-based approaches. This indicates that the current SR library lacks the expressiveness needed to accurately capture the pressure closure across both early and late time regimes. One possible remedy is to enrich the SR candidate library by incorporating more complex nonlinear terms. Alternatively, neural networks (being universal function approximators [71]) offer a flexible framework to explore more expressive functional forms. Motivated by this, we turn to neural network-based closures in the next section.

2. *Neural network inferred closure*

NNs, such as multilayer perceptrons, are able to model much more sophisticated functional forms than those in the simple library of catalog terms used in SR to infer the $P_{e\parallel}$ closure. This comes at the cost of reduced interpretability of the inferred model. Nevertheless, NNs can be used as tools to investigate the existence of more accurate closure models than those found by SR, by leveraging their ability to approximate more complex functional forms.

We therefore design NNs to approximate $P_{e\parallel}$ based on local and instantaneous values of lower-order fluid moments and fields $\{n_e, \overline{v_{ex}}, \overline{v_{ey}}, \overline{v_{ez}}, E_x, E_y, E_z, B_x, B_y, B_z\}$ as inputs. Two NN models are trained using laboratory-frame-only data (which we refer to as NN-Lab) and Galilean-augmented data (NN-Boost). Training data are sampled and augmented using the procedure previously applied to SR. As before, the NNs are trained up to $t < t_{\max} = 250 w_{pe}^{-1}$ and tested on their ability to generalize to later times; we explore the impact of varying $t_{\max}$ in Appendix J. The NN architectures are composed of three hidden layers with 25 nodes per layer. They are trained in PyTorch [72] for 2000 epochs with an AdamW optimizer, learning rate of 0.001, and batch size of 10. These training parameters consistently lead to well-converged models.

The accuracy of the NN closure models (as measured by their mean absolute error) is shown in Fig. 6. For $t < t_{\max}$, both NN models are observed to perform similarly to the analytical closure models and outperform the SR pressure closure models. The improved NN model error at early times suggests the existence of other functional forms that better approximate the pressure closure model. These results can motivate a reexamination of the library of candidate terms used in SR to capture such improved behaviors in an interpretable framework.

For $t > t_{\max}$, beyond the training horizon, we see a notable improvement in the performance of NN-Boost over NN-Lab (Fig. 6). These results reveal that NN-Boost generalizes better than NN-Lab, underscoring the benefits of symmetry embedding via data augmentation in improving the generalization capabilities of ML models other than SR. However, despite their greater flexibility, the NN models do not outperform the SR-Boost closure in this regime. This indicates that the quadratic polynomial basis used in SR is already sufficient to capture the essential behavior of $P_{e\parallel}$ at late times.

These numerical experiments illustrate the importance of embedding fundamental physical symmetries in machine-learned-based fluid closure models from first-principles kinetic simulations. However, there are several important considerations that are worth discussing before applying these techniques to study magnetic reconnection or other plasma phenomena more broadly. First, it is important to recognize that despite embedding fundamental physical symmetries, the validity of ML fluid closure models may be limited to the physical regimes in which they are trained. For example, ML

closures trained on kinetic simulations of subrelativistic magnetic reconnection may not generalize to relativistic magnetic reconnection regimes, where fundamentally unique behavior can arise, such as relativistic plasma effects and highly efficient nonthermal particle acceleration. These effects may require fundamental structural changes to the closure model or to the reduced plasma framework itself (such as going from a pure plasma fluid framework to a hybrid kinetic-fluid plasma framework). Therefore, it is important to identify the range of physical regimes of interest at the onset and define an appropriate reduced theoretical plasma framework accordingly. The kinetic simulation data used to train the ML closure models of such a framework should then span the physical regimes of interest—with the aid of symmetry-guided data augmentation.

Second, the ML pressure closure models explored in this work could potentially be more accurate by allowing for nonlocal dependencies (in space and in time) on low-order plasma moments and fields. Indeed, the inclusion of kinetic effects in fluid plasma models has been shown to require closure models that are inherently nonlocal [5,73]. Such nonlocal relations can be captured using convolution neural networks [11,74–77], for example. In addition, the development of ML closure models at a higher-order moment of the distribution function, such as the heat flux tensor for example, could also be fruitful. An ML closure model for the heat flux tensor (or higher-order moment) may turn out to be simpler than that for the pressure tensor for a given problem, and may ultimately even facilitate gaining insights into the underlying physics.

Finally, to properly assess the performance of ML closures, it is important that these be implemented into a fluid (or hybrid) simulation framework, enabling self-consistent calculations with this model and verifying whether key observables of interest (such as the reconnection rate and energy partitioning, for example) are well captured. There has been recent progress in this direction [10–14,76,78,79], but extending these approaches to magnetic reconnection and to higher-dimensional plasma phenomena remains a valuable direction for future work.

III. DISCUSSION

In this work, we showed that data augmentation is an effective and flexible method to embed physical symmetries into machine-learned reduced plasma models trained on data from first-principle plasma simulations. This approach works across different types of machine learning models, physical problems, and physical symmetries. As a proof of concept, we showed that by Lorentz transforming PIC simulation data of collisionless magnetic reconnection we were able to embed Lorentz invariance into the SR inference of the relativistic two-fluid equations. Discovering reduced plasma models that abide by fundamental physical symmetries such as Lorentz invariance is crucial for their physical consistency and generalization performance. The act of embedding Lorentz invariance into the inference of the two-fluid equations resulted in more accurate coefficients and the elimination of unphysical, spurious model terms. We further showed that with a limited amount of original, simulation data combined with Lorentz-boosted data augmentation, we could discover more accurate models than with orders of magnitude more original simulation data. This has important implications for reducing the volume of data produced by state-of-the-art simulations, which is an increasingly pressing issue in the era of exascale computing [80].

We further demonstrated the importance of embedding symmetries into the inference of pressure closures for collisionless magnetic reconnection in the low Lorentz boost limit (Galilean boosts). We obtained a parallel pressure closure model with SR that was not only Galilean invariant, but furthermore outperformed commonly used analytical closure models. To show the methodological generality of symmetry embedding with data augmentation, we trained NN models to predict the parallel pressure with and without Galilean boosts. The symmetry-preserving models outperformed their original data counterparts. We envision further improvements to the development of ML fluid moment closures for magnetic reconnection by extending the library of SR terms or NN architectures to account for nonlocal effects [77].

There are several promising opportunities for extending this work, particularly in enabling reduced plasma models that generalize to different plasma conditions beyond the training data and thus accelerate the computational exploration of unexplored regimes. One direction is the development of symmetry-preserving closures trained on ensembles of kinetic simulations with varied initial conditions, which could enable reduced models capable of interpolating across parameter space and accelerating systematic scans of reconnection dynamics. Another avenue is to train on data with varying scale separations (e.g., ion-electron mass ratio) to explore whether and how ML reduced models can capture asymptotic scaling and remain valid beyond the regimes directly represented in training. Beyond fluid moment closures, the symmetry-embedding techniques explored here can also be applied to the development of subgrid models for anomalous resistivity and viscosity, which are important ingredients for theoretical models of magnetic reconnection at very large scales [57,81,82]. Future work on these fronts may benefit from implicit sparse regression methods [83] and greedy singular value decomposition algorithms [38,84] to efficiently capture nonlinearities without significant library modification.

We note that rather than weakly embedding Galilean and Lorentz symmetries through data augmentation, we could have hard constrained the symmetries by selecting inputs that were invariant by construction. In our example, this would correspond to formulating a library of catalog terms comprised of Lorentz tensors. This is certainly a reasonable approach, however demands a careful method to construct the library of catalog terms that guarantees invariance. We aim to explore this approach in future work. Additionally, discovering fluid closures in their Lorentz tensorial form for complicated relativistic plasma systems is an important yet largely unexplored direction for future work.

In summary, we have shown that symmetry-guided data augmentation provides a powerful and versatile approach for embedding physical symmetries into machine-learned reduced models of plasma dynamics constructed from the data of first-principles kinetic simulations. This approach is agnostic to the learning framework, integrating seamlessly with techniques that range from sparse regression to deep neural networks. Our results help lay the foundation for the

development of more accurate, physically consistent, and data-efficient reduced plasma models, which are key to enabling computationally efficient multiscale simulations of complex plasma systems.

ACKNOWLEDGMENTS

The authors would like to thank M. Almanza, N. Barbour, D. Carvalho, W. Dorland, G. Guttormsen, N. Loureiro, G. Morales, I. Pusztai, A. Vanthieghem, and A. Velberg for helpful discussions. The authors would also like to thank the reviewers for their helpful comments. The authors acknowledge the OSIRIS Consortium, consisting of UCLA, University of Michigan, and IST (Portugal) for the use of the OSIRIS 4.0 framework. This work was supported by the National Science Foundation Grants No. PHY-2108087 and No. PHY-2108089. F.F. acknowledges support from the European Research Council (ERC-2021-CoG Grant XPACE No. 101045172). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility using NERSC award FES-ERCAPm1157.

DATA AVAILABILITY

The data that support the findings of this article are openly available [61].

APPENDIX A: SIMULATION DETAILS

Here, we detail the physical and numerical parameters of our particle-in-cell simulation of collisionless magnetic reconnection. We construct a force-free current sheet with the following magnetic field profiles:

$$B_x(y) = B_{\text{reconn}}\left[\tanh\left(\frac{y+\frac{L_y}{4}}{\lambda}\right) - \tanh\left(\frac{y-\frac{L_y}{4}}{\lambda}\right) - 1\right], \tag{A1}$$

$$B_z(y) = \sqrt{B_{\text{reconn}}^2 + B_{\text{guide}}^2 - B_x(y)}. \tag{A2}$$

This results in two current sheets spatially separated in $\hat{\boldsymbol{y}}$ with out-of-the-plane, $\hat{\boldsymbol{z}}$ currents in opposite directions. It is periodic in both (x, y). These magnetic field profiles satisfy the force-free equilibrium condition $\boldsymbol{J} \times \boldsymbol{B} = 0$. There is no pressure gradient in the initialization. To perturb the system in an asymmetric manner, we initialize

$$B_y(x) = \delta B\left[\sin\left(\frac{2\pi x}{L_x}\right) + \frac{1}{4}\exp\left(-\left(\frac{x-\frac{L_x}{4}}{\left(\frac{L_x}{8}\right)}\right)^2\right)\right]. \tag{A3}$$

In our system, $B_{\text{reconn}} = 0.5$, $B_{\text{guide}} = 0.2$, $\delta B = 0.01$, $L_x = 48d_e$, $L_y = 64d_e$, and $\lambda = 1d_e$. This is the medium guide field regime as $\frac{B_{\text{guide}}}{B_{\text{reconn}}} = 0.4$. From the magnetic fields, we calculate the self-consistent current density. We partition the initial density and momentum equally between electrons and positrons ($n_e = n_p = n_0 = 1$, $\boldsymbol{v_e} = -\frac{\boldsymbol{J}}{2}$, $\boldsymbol{v_p} = \frac{\boldsymbol{J}}{2}$). We choose $\beta_e = \beta_p = \beta_0 = 0.1$, where $\beta_s \equiv p_s/(B^2/8\pi)$, which sets the electron and positron thermal velocities $v_{\text{th}} = \sqrt{\frac{\beta_0}{8}} \approx 0.112c$. For Sec. II A, the numerical PIC parameters used were 1600 particles per cell, linear particle shapes, grid spacing $\Delta x = \Delta y = 0.1d_e$, and time step $\Delta t = 0.0667\omega_{pe}^{-1}$. The simulation is nonrelativistic as the initial ratio of thermal to magnetic energy $\sim\frac{v_{\text{th}}^2}{B_{\text{reconn}}^2} \approx 0.05$. We end the simulation at $t_{\max} = 10^3\omega_{pe}^{-1}$ when the reconnection rate has saturated. The data used for the pressure closure model discovery in Sec. II B were generated using identical physical parameters, but used 625 particles per cell and quadratic particle shapes.

APPENDIX B: CANDIDATE TERMS FOR TWO-FLUID EQUATION REDISCOVERY

Here, we provide details for the sparse regression inference of the two-fluid moment equations from magnetic reconnection data. For each equation inference, we provided sparse regression with the possible catalog terms to describe the target (dynamical) variable. The choice of these catalog terms is guided by the physical intuition of the problem; in this case, we know the true equations and thus build catalog term libraries that include the true model terms, as well as other reasonable model terms that may throw the sparse regression algorithm off from the true equation. The catalog terms are the building block catalog terms listed below and polynomial combinations of the building block terms up to order 2.

Equation	Target variable	Building block catalog terms
Continuity	$\partial_t n_e$	$1, n_e, n_e\overline{v_{ex}}, n_e\overline{v_{ey}}, n_e\overline{v_{ez}}, \partial_x[n_e\overline{v_{ex}}]$, $\partial_y[n_e\overline{v_{ey}}], E_x, E_y, E_z, B_x, B_y, B_z$
$\hat{\boldsymbol{x}}$ momentum	$\partial_t[n_e\overline{p_{ex}}]$	$1, n_e, n_e\overline{v_{ex}}, n_e\overline{v_{ey}}, n_e\overline{v_{ez}}, n_e\overline{v_{ex}p_{ex}}$, $n_e\overline{v_{ex}p_{ey}}, n_e\overline{v_{ex}p_{ez}}, \partial_x[n_e\overline{v_{ex}p_{ex}}]$, $\partial_y[n_e\overline{v_{ex}p_{ey}}], E_x, E_y, E_z, B_x, B_y, B_z$
$\hat{\boldsymbol{y}}$ momentum	$\partial_t[n_e\overline{p_{ey}}]$	$1, n_e, n_e\overline{v_{ex}}, n_e\overline{v_{ey}}, n_e\overline{v_{ez}}, n_e\overline{v_{ey}p_{ex}}$, $n_e\overline{v_{ey}p_{ey}}, n_e\overline{v_{ey}p_{ez}}, \partial_x[n_e\overline{v_{ey}p_{ex}}]$, $\partial_y[n_e\overline{v_{ey}p_{ey}}], E_x, E_y, E_z, B_x, B_y, B_z$
$\hat{\boldsymbol{z}}$ momentum	$\partial_t[n_e\overline{p_{ez}}]$	$1, n_e, n_e\overline{v_{ex}}, n_e\overline{v_{ey}}, n_e\overline{v_{ez}}, n_e\overline{v_{ez}p_{ex}}$, $n_e\overline{v_{ez}p_{ey}}, n_e\overline{v_{ez}p_{ez}}, \partial_x[n_e\overline{v_{ez}p_{ex}}]$, $\partial_y[n_e\overline{v_{ez}p_{ey}}], E_x, E_y, E_z, B_x, B_y, B_z$
$\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}$ energy	$\partial_t[n_e\overline{p_{ex}p_{ex}}]$	$1, n_e, n_e\overline{p_{ex}}, n_e\overline{p_{ey}}, n_e\overline{p_{ez}}, n_e\overline{v_{ex}p_{ex}}$, $n_e\overline{v_{ex}p_{ey}}, n_e\overline{v_{ex}p_{ez}}, n_e\overline{v_{ex}p_{ex}p_{ex}}$, $n_e\overline{v_{ey}p_{ex}p_{ex}}, n_e\overline{v_{ez}p_{ex}p_{ex}}$, $\partial_x[n_e\overline{v_{ex}p_{ex}p_{ex}}], \partial_y[n_e\overline{v_{ey}p_{ex}p_{ex}}]$, $E_x, E_y, E_z, B_x, B_y, B_z$
$\hat{\boldsymbol{y}}\hat{\boldsymbol{y}}$ energy	$\partial_t[n_e\overline{p_{ey}p_{ey}}]$	$1, n_e, n_e\overline{p_{ex}}, n_e\overline{p_{ey}}, n_e\overline{p_{ez}}, n_e\overline{v_{ey}p_{ex}}$, $n_e\overline{v_{ey}p_{ey}}, n_e\overline{v_{ey}p_{ez}}, n_e\overline{v_{ex}p_{ey}p_{ey}}$, $n_e\overline{v_{ey}p_{ey}p_{ey}}, n_e\overline{v_{ez}p_{ey}p_{ey}}, \partial_x[n_e\overline{v_{ex}p_{ey}p_{ey}}]$, $\partial_y[n_e\overline{v_{ey}p_{ey}p_{ey}}], E_x, E_y, E_z, B_x, B_y, B_z$
$\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}$ energy	$\partial_t[n_e\overline{p_{ez}p_{ez}}]$	$1, n_e, n_e\overline{p_{ex}}, n_e\overline{p_{ey}}, n_e\overline{p_{ez}}, n_e\overline{v_{ez}p_{ex}}$, $n_e\overline{v_{ez}p_{ey}}, n_e\overline{v_{ez}p_{ez}}, n_e\overline{v_{ex}p_{ez}p_{ez}}$, $n_e\overline{v_{ey}p_{ez}p_{ez}}, n_e\overline{v_{ez}p_{ez}p_{ez}}, \partial_x[n_e\overline{v_{ex}p_{ez}p_{ez}}]$, $\partial_y[n_e\overline{v_{ey}p_{ez}p_{ez}}], E_x, E_y, E_z, B_x, B_y, B_z$
$\hat{\boldsymbol{x}}\hat{\boldsymbol{y}}$ energy	$\partial_t[n_e\overline{p_{ex}p_{ey}}]$	$1, n_e, n_e\overline{p_{ex}}, n_e\overline{p_{ey}}, n_e\overline{p_{ez}}, n_e\overline{v_{ex}p_{ex}}$, $n_e\overline{v_{ey}p_{ey}}, n_e\overline{v_{ez}p_{ez}}, n_e\overline{v_{ex}p_{ey}}, n_e\overline{v_{ex}p_{ez}}$, $n_e\overline{v_{ey}p_{ez}}, n_e\overline{v_{ex}p_{ex}p_{ey}}, n_e\overline{v_{ey}p_{ex}p_{ey}}$, $n_e\overline{v_{ez}p_{ex}p_{ey}}, \partial_x[n_e\overline{v_{ex}p_{ex}p_{ey}}]$, $\partial_y[n_e\overline{v_{ey}p_{ex}p_{ey}}], E_x, E_y, E_z, B_x, B_y, B_z$

$\hat{x}\hat{z}$ energy	$\partial_t[n_e\overline{p_{ex}p_{ez}}]$	$1, n_e, n_e\overline{p_{ex}}, n_e\overline{p_{ey}}, n_e\overline{p_{ez}}, n_e\overline{v_{ex}p_{ex}}, n_e\overline{v_{ey}p_{ey}}, n_e\overline{v_{ez}p_{ez}}, n_e\overline{v_{ex}p_{ey}}, n_e\overline{v_{ex}p_{ez}}, n_e\overline{v_{ey}p_{ez}}, n_e\overline{v_{ex}p_{ex}p_{ez}}, n_e\overline{v_{ey}p_{ex}p_{ez}}, n_e\overline{v_{ez}p_{ex}p_{ez}}, \partial_x[n_e\overline{v_{ex}p_{ex}p_{ez}}], \partial_y[n_e\overline{v_{ey}p_{ex}p_{ez}}], E_x, E_y, E_z, B_x, B_y, B_z$
$\hat{y}\hat{z}$ energy	$\partial_t[n_e\overline{p_{ey}p_{ez}}]$	$1, n_e, n_e\overline{p_{ex}}, n_e\overline{p_{ey}}, n_e\overline{p_{ez}}, n_e\overline{v_{ex}p_{ex}}, n_e\overline{v_{ey}p_{ey}}, n_e\overline{v_{ez}p_{ez}}, n_e\overline{v_{ex}p_{ey}}, n_e\overline{v_{ex}p_{ez}}, n_e\overline{v_{ey}p_{ez}}, n_e\overline{v_{ex}p_{ey}p_{ez}}, n_e\overline{v_{ey}p_{ey}p_{ez}}, n_e\overline{v_{ez}p_{ey}p_{ez}}, \partial_x[n_e\overline{v_{ex}p_{ey}p_{ez}}], \partial_y[n_e\overline{v_{ey}p_{ey}p_{ez}}], E_x, E_y, E_z, B_x, B_y, B_z$

APPENDIX C: DATA SAMPLING DETAILS

Here, we detail the data sampling procedure to extract measurements for the sparse regression recovery of the two-fluid equations. First, to remove some of the data overhead we sample our data from windows (clusters) of neighboring time steps. Beginning at time step 3500 (which corresponds to a physical time $t = 233.33w_{pe}^{-1}$), we sample our measurements from a 30 time step ($2w_{pe}^{-1}$) window. We then skip 800 time steps ($53.33w_{pe}^{-1}$) until the next 30 time step window. The process is repeated until time step 6020 ($401.33w_{pe}^{-1}$). This results in four time windows of consecutive time steps where we can sample measurements for the sparse regression equation inference. These time windows fall during the nonlinear stage of MR when complex, nonlinear, and chaotic dynamics are occurring. To ensure that the measured points reside in interesting regions in space-time where the target variables are dynamically balanced by the interplay of model terms, for each time window we calculate the 70th percentile value of the absolute target variables and only select measurements above that threshold. The results are weakly sensitive to the percentile value above 50%. However, the sampling threshold may vary across problems, so we recommend scanning this parameter for each dataset and evaluating which models generalize best. A total of $15^3 = 3375$ measurements are randomly selected—satisfying the threshold condition—from the time windows. Selecting a large amount of measurements from the simulations, as done here, helps reduce the variance of model coefficient error as observed in Fig. 2. We then integrate these measurements over small space-time cubes of size $\{5\Delta x, 5\Delta y, 5\Delta t\}$ (the integral formulation) and discover equations in their integral form. The data sampling scheme is different for closure discovery and described in the text body (Sec. II B).

APPENDIX D: TABLE OF INFERRED COEFFICIENTS

Here, we provide the two-fluid model equations inferred by sparse regression with and without Lorentz-boosted data transformations in Sec. II A. The Lorentz boosts are conducted with a maximum Lorentz boost of $\gamma_{\beta_{\max}} = 1.5$. The error bars correspond to ±one standard deviation, and the statistics come from training each model five times with different realizations of the random data sampling. We compare the laboratory-frame and Boosted Data Equations with the analytically derived true equations.

Continuity equation	
True equation	$\partial_t n_e = -\partial_x[n_e\overline{v_{ex}}] - \partial_y[n_e\overline{v_{ey}}]$
Laboratory-frame data inferred equation	$\partial_t n_e = -(1.0262 \pm 0.0028)\partial_x[n_e\overline{v_{ex}}]$ $-(1.0317 \pm 0.0022)\partial_y[n_e\overline{v_{ey}}]$
Boosted data inferred equation	$\partial_t n_e = -(1.0006 \pm 0.0007)\partial_x[n_e\overline{v_{ex}}]$ $-(1.0005 \pm 0.0001)\partial_y[n_e\overline{v_{ey}}]$

$\hat{x}$ Momentum equation	
True equation	$\partial_t[n_e\overline{p_{ex}}] = -\partial_x[n_e\overline{v_{ex}p_{ex}}]$ $-\partial_y[n_e\overline{v_{ex}p_{ey}}] - n_eE_x - n_e\overline{v_{ey}}B_z$ $+n_e\overline{v_{ez}}B_y$
Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ex}}] =$ $-(1.0189 \pm 0.0008)\partial_x[n_e\overline{v_{ex}p_{ex}}]$ $-(1.0311 \pm 0.0014)\partial_y[n_e\overline{v_{ex}p_{ey}}] -$ $(1.0071 \pm 0.0016)n_eE_x - (1.0077 \pm 0.0018)n_e\overline{v_{ey}}B_z$ $+(1.0107 \pm 0.0016)n_e\overline{v_{ez}}B_y$
Boosted data inferred equation	$\partial_t[n_e\overline{p_{ex}}] =$ $-(1.0002 \pm 0.0005)\partial_x[n_e\overline{v_{ex}p_{ex}}]$ $-(1.0016 \pm 0.0003)\partial_y[n_e\overline{v_{ex}p_{ey}}] -$ $(0.9918 \pm 0.0051)n_eE_x - (0.9917 \pm 0.0050)n_e\overline{v_{ey}}B_z$ $+(0.9931 \pm 0.0040)n_e\overline{v_{ez}}B_y$

$\hat{y}$ Momentum equation	
True equation	$\partial_t[n_e\overline{p_{ey}}] = -\partial_x[n_e\overline{v_{ey}p_{ex}}]$ $-\partial_y[n_e\overline{v_{ey}p_{ey}}] - n_eE_y - n_e\overline{v_{ez}}B_x$ $+n_e\overline{v_{ex}}B_z$
Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ey}}] =$ $-(1.0211 \pm 0.0022)\partial_x[n_e\overline{v_{ey}p_{ex}}]$ $-(1.0245 \pm 0.0017)\partial_y[n_e\overline{v_{ey}p_{ey}}] -$ $(1.0079 \pm 0.0023)n_eE_y - (1.0081 \pm 0.0023)n_e\overline{v_{ez}}B_x$ $+(1.0080 \pm 0.0021)n_e\overline{v_{ex}}B_z$
Boosted data inferred equation	$\partial_t[n_e\overline{p_{ey}}] =$ $-(1.0012 \pm 0.0006)\partial_x[n_e\overline{v_{ey}p_{ex}}]$ $-(1.0010 \pm 0.0002)\partial_y[n_e\overline{v_{ey}p_{ey}}] -$ $(0.9916 \pm 0.0018)n_eE_y - (0.9925 \pm 0.0011)n_e\overline{v_{ez}}B_x$ $+(0.9916 \pm 0.0018)n_e\overline{v_{ex}}B_z$

$\hat{z}$ Momentum equation	
True equation	$\partial_t[n_e\overline{p_{ez}}] = -\partial_x[n_e\overline{v_{ez}p_{ex}}]$ $-\partial_y[n_e\overline{v_{ez}p_{ey}}] - n_eE_z - n_e\overline{v_{ex}}B_y$ $+n_e\overline{v_{ey}}B_x$
Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ez}}] =$ $-(1.0256 \pm 0.0042)\partial_x[n_e\overline{v_{ez}p_{ex}}]$ $-(1.0322 \pm 0.0026)\partial_y[n_e\overline{v_{ez}p_{ey}}] -$ $(1.0207 \pm 0.0030)n_eE_z - (1.0187 \pm 0.0019)n_e\overline{v_{ex}}B_y$ $+(1.0173 \pm 0.0029)n_e\overline{v_{ey}}B_x$

Boosted data inferred equation	$\partial_t[n_e\overline{p_{ez}}] =$ $-(1.0007 \pm 0.0005)\partial_x[n_e\overline{v_{ez}p_{ex}}]$ $-(1.0003 \pm 0.0002)\partial_y[n_e\overline{v_{ez}p_{ey}}] -$ $(0.9942 \pm 0.0029)n_eE_z - (0.9942 \pm 0.0029)n_e\overline{v_{ex}}B_y$ $+(0.9942 \pm 0.0029)n_e\overline{v_{ey}}B_x$

$\hat{x}\hat{x}$ Energy equation

True equation	$\partial_t[n_e\overline{p_{ex}p_{ex}}] = -\partial_x[n_e\overline{v_{ex}p_{ex}p_{ex}}]$ $-\partial_y[n_e\overline{v_{ey}p_{ex}p_{ex}}] - 2n_e\overline{p_{ex}}E_x -$ $2n_e\overline{v_{ex}p_{ey}}B_z + 2n_e\overline{v_{ex}p_{ez}}B_y$
Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ex}p_{ex}}] =$ $-(1.0219 \pm 0.0020)\partial_x[n_e\overline{v_{ex}p_{ex}p_{ex}}]$ $-(1.0360 \pm 0.0021)\partial_y[n_e\overline{v_{ey}p_{ex}p_{ex}}] -$ $(2.0230 \pm 0.0086)n_e\overline{p_{ex}}E_x -$ $(2.0230 \pm 0.0051)n_e\overline{v_{ex}p_{ey}}B_z$ $+(2.0255 \pm 0.0049)n_e\overline{v_{ex}p_{ez}}B_y$
Boosted data inferred equation	$\partial_t[n_e\overline{p_{ex}p_{ex}}] =$ $-(1.0010 \pm 0.0004)\partial_x[n_e\overline{v_{ex}p_{ex}p_{ex}}]$ $-(1.0095 \pm 0.0016)\partial_y[n_e\overline{v_{ey}p_{ex}p_{ex}}] -$ $(1.9881 \pm 0.0057)n_e\overline{p_{ex}}E_x -$ $(1.9871 \pm 0.0052)n_e\overline{v_{ex}p_{ey}}B_z$ $+(1.9959 \pm 0.0074)n_e\overline{v_{ex}p_{ez}}B_y$

$\hat{y}\hat{y}$ Energy equation

True equation	$\partial_t[n_e\overline{p_{ey}p_{ey}}] = -\partial_x[n_e\overline{v_{ex}p_{ey}p_{ey}}]$ $-\partial_y[n_e\overline{v_{ey}p_{ey}p_{ey}}] - 2n_e\overline{p_{ey}}E_y -$ $2n_e\overline{v_{ey}p_{ez}}B_x + 2n_e\overline{v_{ey}p_{ex}}B_z$
Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ey}p_{ey}}] =$ $-(1.0278 \pm 0.0036)\partial_x[n_e\overline{v_{ex}p_{ey}p_{ey}}]$ $-(1.0299 \pm 0.0023)\partial_y[n_e\overline{v_{ey}p_{ey}p_{ey}}] -$ $(2.0263 \pm 0.0093)n_e\overline{p_{ey}}E_y -$ $(2.0250 \pm 0.0069)n_e\overline{v_{ey}p_{ez}}B_x$ $+(2.0247 \pm 0.0059)n_e\overline{v_{ey}p_{ex}}B_z$
Boosted data inferred equation	$\partial_t[n_e\overline{p_{ey}p_{ey}}] =$ $-(1.0062 \pm 0.0027)\partial_x[n_e\overline{v_{ex}p_{ey}p_{ey}}]$ $-(1.0016 \pm 0.0004)\partial_y[n_e\overline{v_{ey}p_{ey}p_{ey}}] -$ $(1.9965 \pm 0.0087)n_e\overline{p_{ey}}E_y -$ $(1.9979 \pm 0.0062)n_e\overline{v_{ey}p_{ez}}B_x$ $+(1.9971 \pm 0.0084)n_e\overline{v_{ey}p_{ex}}B_z$

$\hat{z}\hat{z}$ Energy equation

True equation	$\partial_t[n_e\overline{p_{ez}p_{ez}}] = -\partial_x[n_e\overline{v_{ex}p_{ez}p_{ez}}]$ $-\partial_y[n_e\overline{v_{ey}p_{ez}p_{ez}}] - 2n_e\overline{p_{ez}}E_z -$ $2n_e\overline{v_{ez}p_{ex}}B_y + 2n_e\overline{v_{ez}p_{ey}}B_x$
Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ez}p_{ez}}] =$ $-(1.0287 \pm 0.0048)\partial_x[n_e\overline{v_{ex}p_{ez}p_{ez}}]$ $-(1.0361 \pm 0.0054)\partial_y[n_e\overline{v_{ey}p_{ez}p_{ez}}] -$ $(2.0440 \pm 0.0073)n_e\overline{p_{ez}}E_z -$ $(2.0290 \pm 0.0075)n_e\overline{v_{ez}p_{ex}}B_y$ $+(2.0259 \pm 0.0082)n_e\overline{v_{ez}p_{ey}}B_x$
Boosted data inferred equation	$\partial_t[n_e\overline{p_{ez}p_{ez}}] =$ $-(1.0006 \pm 0.0004)\partial_x[n_e\overline{v_{ex}p_{ez}p_{ez}}]$ $-(1.0004 \pm 0.0003)\partial_y[n_e\overline{v_{ey}p_{ez}p_{ez}}] -$ $(1.9965 \pm 0.0040)n_e\overline{p_{ez}}E_z -$ $(1.9963 \pm 0.0042)n_e\overline{v_{ez}p_{ex}}B_y$ $+(1.9961 \pm 0.0041)n_e\overline{v_{ez}p_{ey}}B_x$

$\hat{x}\hat{y}$ Energy equation

True equation	$\partial_t[n_e\overline{p_{ex}p_{ey}}] = -\partial_x[n_e\overline{v_{ex}p_{ex}p_{ey}}]$ $-\partial_y[n_e\overline{v_{ey}p_{ex}p_{ey}}] - n_e\overline{p_{ex}}E_y -$ $n_e\overline{p_{ey}}E_x - n_e\overline{v_{ex}p_{ez}}B_x - n_e\overline{v_{ey}p_{ey}}B_z$ $+n_e\overline{v_{ey}p_{ez}}B_y + n_e\overline{v_{ex}p_{ex}}B_z$
Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ex}p_{ey}}] =$ $-(1.0324 \pm 0.0033)\partial_x[n_e\overline{v_{ex}p_{ex}p_{ey}}]$ $-(1.0418 \pm 0.0023)\partial_y[n_e\overline{v_{ey}p_{ex}p_{ey}}] -$ $(1.0184 \pm 0.0049)n_e\overline{p_{ex}}E_y$ $-(1.0171 \pm 0.0038)n_e\overline{p_{ey}}E_x -$ $(1.0193 \pm 0.0031)n_e\overline{v_{ex}p_{ez}}B_x$ $-(1.0173 \pm 0.0029)n_e\overline{v_{ey}p_{ey}}B_z$ $+(1.0202 \pm 0.0037)n_e\overline{v_{ey}p_{ez}}B_y$ $+(1.0177 \pm 0.0030)n_e\overline{v_{ex}p_{ex}}B_z$
Boosted data inferred equation	$\partial_t[n_e\overline{p_{ex}p_{ey}}] =$ $-(0.9999 \pm 0.0003)\partial_x[n_e\overline{v_{ex}p_{ex}p_{ey}}]$ $-(1.0009 \pm 0.0003)\partial_y[n_e\overline{v_{ey}p_{ex}p_{ey}}] -$ $(0.9882 \pm 0.0017)n_e\overline{p_{ex}}E_y -$ $(0.9913 \pm 0.0021)n_e\overline{p_{ey}}E_x - (0.9887 \pm 0.0016)n_e\overline{v_{ex}p_{ez}}B_x$ $-(0.9912 \pm 0.0021)n_e\overline{v_{ey}p_{ey}}B_z$ $+(0.9917 \pm 0.0019)n_e\overline{v_{ey}p_{ez}}B_y$ $+(0.9882 \pm 0.0017)n_e\overline{v_{ex}p_{ex}}B_z$

$\hat{x}\hat{z}$ Energy equation

True equation	$\partial_t[n_e\overline{p_{ex}p_{ez}}] = -\partial_x[n_e\overline{v_{ex}p_{ex}p_{ez}}]$ $-\partial_y[n_e\overline{v_{ey}p_{ex}p_{ez}}] - n_e\overline{p_{ex}}E_z -$ $n_e\overline{p_{ez}}E_x - n_e\overline{v_{ey}p_{ez}}B_z - n_e\overline{v_{ex}p_{ex}}B_y$ $+n_e\overline{v_{ex}p_{ey}}B_x + n_e\overline{v_{ez}p_{ez}}B_y$
Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ex}p_{ez}}] =$ $-(1.0280 \pm 0.0024)\partial_x[n_e\overline{v_{ex}p_{ex}p_{ez}}]$ $-(1.0429 \pm 0.0033)\partial_y[n_e\overline{v_{ey}p_{ex}p_{ez}}] -$ $(1.0246 \pm 0.0025)n_e\overline{p_{ex}}E_z -$ $(1.0189 \pm 0.0018)n_e\overline{p_{ez}}E_x - (1.0150 \pm 0.0012)n_e\overline{v_{ey}p_{ez}}B_z$ $-(1.0181 \pm 0.0011)n_e\overline{v_{ex}p_{ex}}B_y$ $+(1.0175 \pm 0.0026)n_e\overline{v_{ex}p_{ey}}B_x$ $+(1.0172 \pm 0.0012)n_e\overline{v_{ez}p_{ez}}B_y$
Boosted data inferred equation	$\partial_t[n_e\overline{p_{ex}p_{ez}}] =$ $-(1.0004 \pm 0.0005)\partial_x[n_e\overline{v_{ex}p_{ex}p_{ez}}]$ $-(1.0015 \pm 0.0002)\partial_y[n_e\overline{v_{ey}p_{ex}p_{ez}}] -$ $(0.9952 \pm 0.0078)n_e\overline{p_{ex}}E_z -$ $(0.9960 \pm 0.0046)n_e\overline{p_{ez}}E_x - (0.9958 \pm 0.0045)n_e\overline{v_{ey}p_{ez}}B_z$ $-(0.9951 \pm 0.0078)n_e\overline{v_{ex}p_{ex}}B_y$ $+(0.9951 \pm 0.0076)n_e\overline{v_{ex}p_{ey}}B_x$ $+(0.9950 \pm 0.0048)n_e\overline{v_{ez}p_{ez}}B_y$

$\hat{y}\hat{z}$ Energy equation

True equation	$\partial_t[n_e\overline{p_{ey}p_{ez}}] = -\partial_x[n_e\overline{v_{ex}p_{ey}p_{ez}}]$ $-\partial_y[n_e\overline{v_{ey}p_{ey}p_{ez}}] - n_e\overline{p_{ey}}E_z -$ $n_e\overline{p_{ez}}E_y - n_e\overline{v_{ex}p_{ey}}B_y - n_e\overline{v_{ez}p_{ez}}B_x$ $+n_e\overline{v_{ex}p_{ez}}B_z + n_e\overline{v_{ey}p_{ey}}B_x$

Laboratory-frame data inferred equation	$\partial_t[n_e\overline{p_{ey}p_{ez}}] =$ $-(1.0252 \pm 0.0040)\partial_x[n_e\overline{v_{ex}p_{ey}p_{ez}}]$ $-(1.0327 \pm 0.0040)\partial_y[n_e\overline{v_{ey}p_{ey}p_{ez}}] -$ $(1.0064 \pm 0.0044)n_e\overline{p_{ey}}E_z -$ $(1.0099 \pm 0.0035)n_e\overline{p_{ez}}E_y - (1.0042 \pm$ $0.0027)n_e\overline{v_{ex}p_{ey}}B_y$ $-(1.0061 \pm 0.0023)n_e\overline{v_{ez}p_{ez}}B_x$ $+(1.0084 \pm 0.0027)n_e\overline{v_{ex}p_{ez}}B_z$ $+(1.0031 \pm 0.0020)n_e\overline{v_{ey}p_{ey}}B_x$
Boosted data inferred equation	$\partial_t[n_e\overline{p_{ey}p_{ez}}] =$ $-(1.0016 \pm 0.0015)\partial_x[n_e\overline{v_{ex}p_{ey}p_{ez}}]$ $-(1.0008 \pm 0.0004)\partial_y[n_e\overline{v_{ey}p_{ey}p_{ez}}] -$ $(0.9922 \pm 0.0033)n_e\overline{p_{ey}}E_z -$ $(0.9912 \pm 0.0031)n_e\overline{p_{ez}}E_y - (0.9934 \pm$ $0.0029)n_e\overline{v_{ex}p_{ey}}B_y$ $-(0.9908 \pm 0.0033)n_e\overline{v_{ez}p_{ez}}B_x$ $+(0.9911 \pm 0.0031)n_e\overline{v_{ex}p_{ez}}B_z$ $+(0.9921 \pm 0.0033)n_e\overline{v_{ey}p_{ey}}B_x$

APPENDIX E: ADDITIONAL EXAMPLES OF COEFFICIENT ERROR IMPROVING AS A FUNCTION OF MAXIMUM BOOST

Here, we show two additional average coefficient error vs $\gamma_{\beta\max}$ plots for the $\hat{z}$ component of the momentum equation and $\hat{z}\hat{z}$ component of the energy equation (Fig. 7). Again, the elbows fall roughly at $\gamma_{\beta\max} = 1.2$. For the results presented in Figs. 1, 2, 4, 5, and 9, $\gamma_{\beta\max} = 1.5$, to the right of the elbow. This demonstrates that the Lorentz boosts are large enough to reap the benefits of symmetry embedding.

APPENDIX F: DISENTANGLING THE ROLES OF DATA RICHNESS AND MULTIPLE SYMMETRY-RELATED DATA COPIES FOR IMPROVING COEFFICIENT ACCURACY

In this Appendix, we examine the relative contributions to coefficient improvement due to Lorentz boosting from two distinct factors: (1) data richness, interpreted as the increased variance or diversity within the dataset, and (2) symmetry enforcement achieved through repeated laboratory-data copies boosted into different Lorentz reference frames. These two factors represent natural explanations for the findings shown in Fig. 2, where coefficient accuracy improves with the inclusion of Lorentz-boosted data.

We have conducted the following computational experiments to quantify the contributions of these two effects. Starting from a set of 100 laboratory-frame measurements (original data points taken from the collisionless magnetic reconnection PIC simulation), we generated datasets in two different ways: (1) the repeated data copies method (augmentation), where each laboratory-frame measurement is boosted multiple times (either 10 or 100 times) into random Lorentz reference frames with γ_β sampled uniformly between 1 and $\gamma_{\beta\max} = \{1, 1.1, 1.5, 2, 5, 10\}$; and (2) the data transformation method, where each laboratory-frame measurement is boosted only once into a random Lorentz reference frame again with γ_β sampled uniformly between 1 and $\gamma_{\beta\max} = \{1, 1.1, 1.5, 2, 5, 10\}$. For each of these strategies, we calculated the coefficient error for the sparse regression inferred continuity equation. We also calculated the data richness, measured as the variance of the dynamical (target) term $\partial_t n_e$. We note that we repeated each of these calculations five times with different realizations of the random measurement sampling, and averaged over the results.

The results are illustrated in Fig. 8. The data richness is shown to increase with increasing $\gamma_{\beta\max}$, as expected from the broader range of boost transformations of the data.

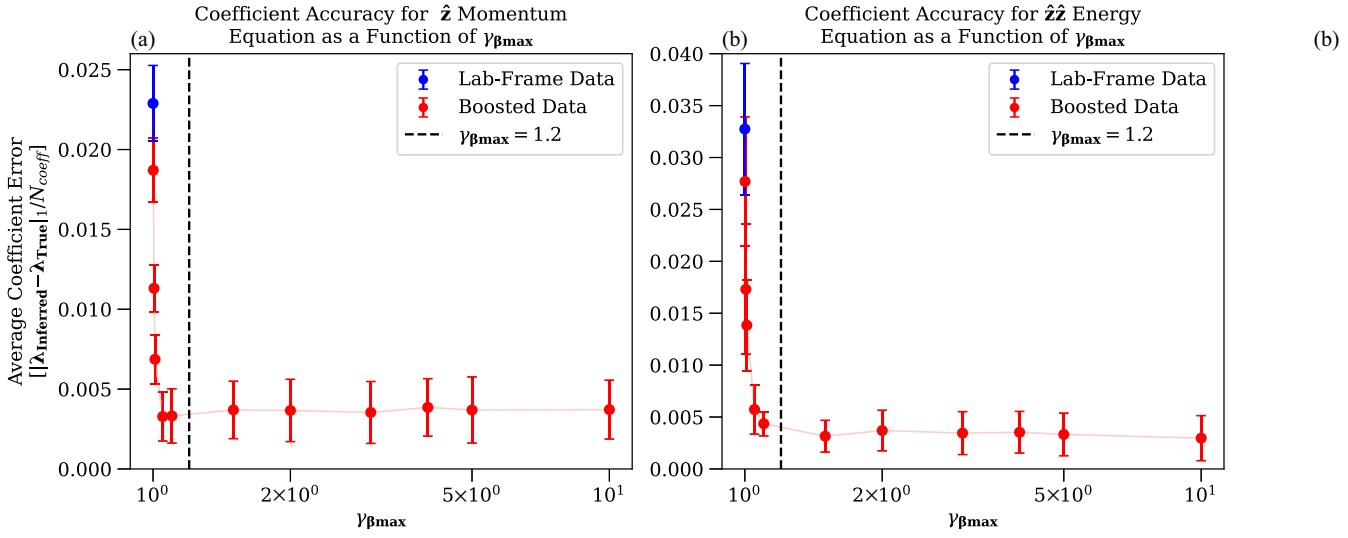


FIG. 7. Average coefficient error decreases as a function of maximum Lorentz boost ($\gamma_{\beta\max}$) for the $\hat{z}$ momentum and $\hat{z}\hat{z}$ energy equations. Panel (a) is for the $\hat{z}$ momentum equation and panel (b) is for the $\hat{z}\hat{z}$ energy equation. The vertical axis measures the L1 difference between true and SR inferred coefficients. We randomly Lorentz transform our data in the range $1 \leqslant \gamma_\beta \leqslant \gamma_{\beta\max}$ and demonstrate that increasing $\gamma_{\beta\max}$ (embedding the symmetry more stringently) leads to more accurate coefficients. The vertical error bars correspond to ±one standard deviation in average coefficient error, and the statistics come from training each model five times with different realizations of the random data sampling.

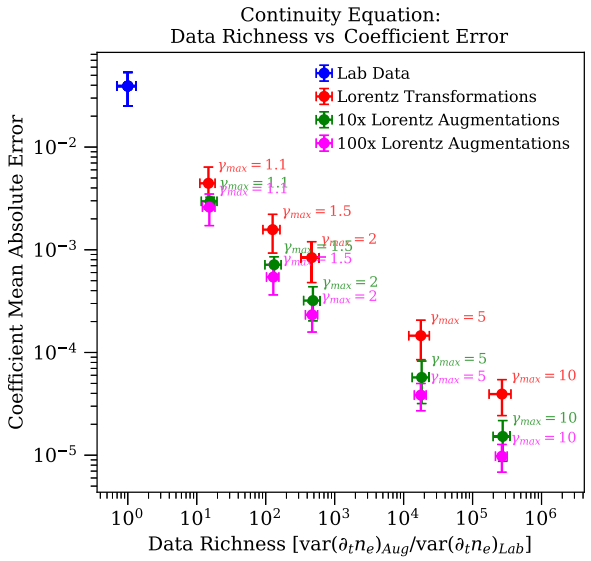


FIG. 8. Data richness vs coefficient error for the continuity equation. The vertical axis shows the mean absolute error of the sparse regression coefficients inferred for the continuity equation. On the horizontal axis, we plot a measure of data richness, defined as the variance of the target term $\partial_t n_e$, normalized by the variance computed from the original laboratory-frame data alone. The error bars correspond to ±one standard deviation, and the statistics come from training each model five times with different realizations of the random data sampling. As expected, transforming our original data into Lorentz-boosted reference frames with larger Lorentz factor $\gamma_{\beta\max}$ leads to greater data richness. Moreover, higher data richness corresponds to a substantial reduction in coefficient error for the sparse regression inferred coefficients, often by several orders of magnitude. At a fixed Lorentz boost (approximately fixed data richness), adding multiple Lorentz-boosted copies (either 10× or 100× per laboratory measurement) further decreases the coefficient error, but only by a modest factor. These findings thus suggest that the dominant factor driving improved coefficient accuracy is the increase in data variance achieved by applying more extreme Lorentz boosts.

Importantly, we find that while the data augmentation methods produce slightly lower coefficient error than the transformation method for the same $\gamma_{\beta\max}$, the main source of coefficient error reduction is due to increasing the data richness (data variance), which is controlled by $\gamma_{\beta\max}$. The coefficient accuracy at a fixed $\gamma_{\beta\max}$ appears to improve only by a factor of a few (∼2–4) when multiple symmetry-related copies of the data are added. We note that adding more data copies is computationally expensive for the sparse regression, while increasing $\gamma_{\beta\max}$ does not change the computational demand. Therefore, the data transformation method has important practical advantages since it is a more computationally efficient way of reducing the coefficient error compared to creating symmetry-related data copies with true data augmentation.

APPENDIX G: RESPECTING LORENTZ INVARIANCE

Reduced plasma models discovered with SR on Lorentz-augmented MR data are found to be more "Lorentz invariant" than models discovered on only the original, laboratory-frame simulation data. The symmetry is thus successfully embedded. Highly boosted Lorentz reference frames observe strikingly distorted dynamics relative to the simulation frame [85]; the balance of PDE terms in the boosted-frame data is different than in the laboratory frame. By including Lorentz-augmented data in the equation inference procedure, we are providing unseen, yet physically valid training data, which further restrict the solution space (coefficient manifold).

We construct a metric to measure the degree of Lorentz invariance of our inferred two-fluid equations and compare the equations discovered with Lorentz-boosted data vs with laboratory-frame data. As a metric for the degree of Lorentz invariance from our inferred equations, we add the absolute difference between the spatial derivative coefficients and 1 to the absolute difference between the electromagnetic field term coefficients and their collective mean. In Fig. 9, we show that all equations that are inferred with Lorentz-transformed data are more Lorentz invariant than when inferred solely with laboratory-frame data according to this metric.

To better understand how Lorentz boosting embeds Lorentz symmetry (and minimizes the Lorentz invariance metric described above), we return to the loss function formalism presented in Eqs. (8)–(11) and examine the regression loss function for the inference of the $\hat{z}$ momentum equation with data that have been Lorentz transformed by $\boldsymbol{\beta} = \beta\hat{\boldsymbol{x}}$.

$$\mathcal{L}' \sim \partial_t'[n_e'\overline{p_{ez}'}] + \lambda_1' n_e' E_z' + \lambda_2' n_e' \overline{v_{ex}'} B_y' + \lambda_3' n_e' \overline{v_{ey}'} B_x', \quad \text{(G1)}$$

where $\{\lambda_1', \lambda_2', \lambda_3'\}$ are unknown coefficients that the regression will try to optimize. From the Lorentz-boosted data, we have measurements of the terms $\partial_t'[n_e'\overline{p_{ez}'}]$, $n_e' E_z'$, etc. Using the inverse Lorentz transform $(\Lambda^{-1})^{\mu}_{\nu}$, we can express $\mathcal{L}'$ in terms of the laboratory-frame moments. Crucially, we define the unknown coefficients to be invariant under the inverse Lorentz transformation. Focusing solely on the Lorentz force terms in Eq. (G1),

$$\mathcal{L}' \sim \lambda_1 \gamma_\beta^2 n_e \left[1 - \frac{\beta \overline{v_{ex}}}{c^2}\right][E_z + \beta B_y] + \lambda_2 \gamma_\beta^2 n_e [\overline{v_{ex}} - \beta]\left[B_y + \frac{\beta E_z}{c^2}\right] + \lambda_3 n_e \overline{v_{ey}} B_x. \quad \text{(G2)}$$

Group like terms proportional to $n_e E_z$.

$$\mathcal{L}' \sim \gamma_\beta^2 n_e E_z \left\{\lambda_1\left[1 - \frac{\beta\overline{v_{ex}}}{c^2}\right] + \lambda_2 \frac{\beta}{c^2}[\overline{v_{ex}} - \beta]\right\}. \quad \text{(G3)}$$

Take the limit $\gamma_\beta \to \infty$, corresponding to $\beta \to c$.

$$\mathcal{L}' \sim \gamma_\beta^2 n_e E_z [\lambda_1 - \lambda_2]\left[1 - \frac{\overline{v_{ex}}}{c}\right]. \quad \text{(G4)}$$

Measurements of the term $[1 - \frac{\overline{v_{ex}}}{c}] n_e E_z$ from Eq. (G4) are not typically equal to zero in the MR dataset; thus, to minimize the loss function in the large Lorentz boost limit the multiplicative term $[\lambda_1 - \lambda_2] \to 0$, $\lambda_1 \to \lambda_2$. Lorentz boosting forces the terms proportional to electromagnetic fields to converge to the same value. For the inferred equation to be Lorentz invariant, the coefficients of terms proportional to electromagnetic field indeed must be identical in magnitude.

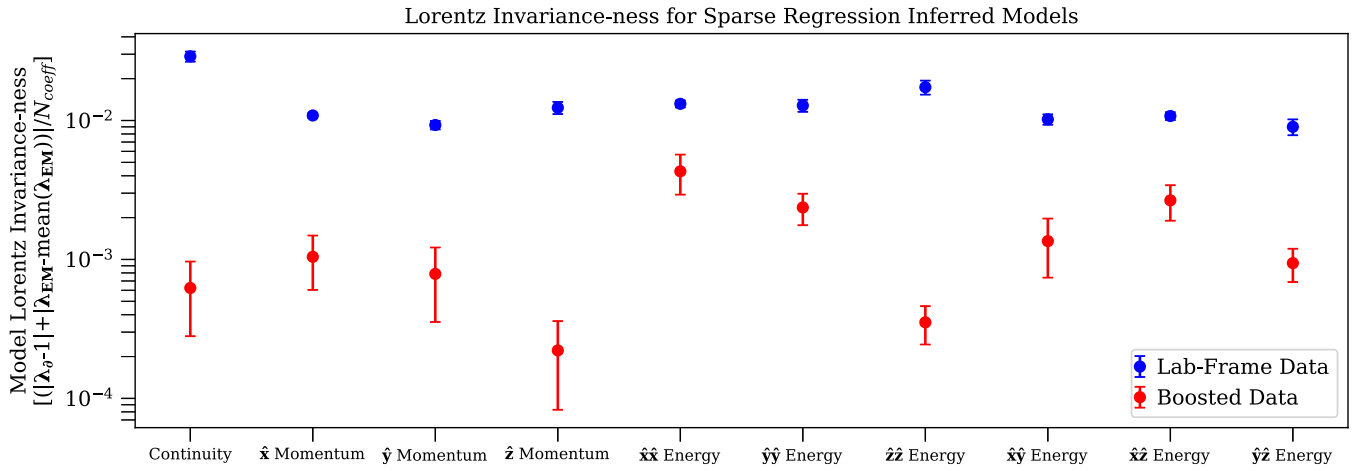


FIG. 9. Demonstrating that sparse regression inferred models with Lorentz-transformed data respect Lorentz invariance symmetry better than models discovered with only laboratory-frame data. On the vertical axis, we have constructed a metric to describe how well Lorentz invariance is satisfied by each sparse regression inferred model. A model that is perfectly Lorentz invariant has a metric measurement of 0, and the larger the metric the more the symmetry is broken. The metric we use is the absolute sum of the difference between the spatial derivative term coefficients and 1, plus the absolute difference of the terms proportional to the electromagnetic fields and their mean. We normalize by the number of nontime derivative terms in the model. The vertical error bars correspond to ±one standard deviation, and the statistics come from training each model five times with different realizations of the random data sampling. On the horizontal axis are the first ten unique two-fluid plasma equations. We see that all ten models inferred with Lorentz-transformed data abide by Lorentz symmetry better than their laboratory-frame data counterparts beyond a standard deviation.

APPENDIX H: SPARSE REGRESSION AND NEURAL NETWORK PERPENDICULAR PRESSURE CLOSURE DISCOVERY FOR MAGNETIC RECONNECTION

In Sec. II B, we demonstrated the significant improvements when incorporating Galilean data augmentation in the sparse regression and neural network inference of a parallel pressure closure for collisionless pair-plasma magnetic reconnection. However, for a complete closure model we must also discover a pressure closure perpendicular to the local magnetic field. For example, both CGL and Le closures have separate functional forms, descriptions, for the parallel and perpendicular pressure components.

We repeated the sparse regression algorithm with laboratory-frame data, but now with our target variable as $P_{e\perp}$ rather than $P_{e\parallel}$. We allow $P_{e\perp}$ to depend on $\{n_e, \overline{v_{ex}}, \overline{v_{ey}}, \overline{v_{ez}}, B_x, B_y, B_z\}$ and their second-order polynomial combinations; again, we do not include the electric field as a building block catalog term given its correlation with $\overline{\boldsymbol{v_e}} \times \boldsymbol{B}$, which crashes our sparse regression algorithm. As in the parallel pressure closure case, we trained the sparse regression algorithm with 1000 data points sampled every $10w_{pe}^{-1}$ from $t = 0w_{pe}^{-1}$ to $t = t_{\max} = 250w_{pe}^{-1}$ and only selected data points where $P_{e\perp}$ is greater than its 70th percentile value at its sampled time step. SR identified a five-term electron perpendicular pressure model of the form

$$P_{e\perp\text{Lab}} = -0.004 + 0.021n_e + 0.002n_e^2 - 0.152B_x^2 - 0.011B_z n_e. \tag{H1}$$

We immediately notice that the SR inferred closure in Eq. (H1) is Galilean invariant, and thus there are no nonsymmetry abiding terms to be removed like in the parallel pressure closure example. By performing Galilean data augmentation, we double the training dataset size with random Galilean boosts of each sampled point with $|\beta_j| \leqslant 0.1c$ and inferred the same five-term electron perpendicular pressure model:

$$P_{e\perp\text{Boost}} = -0.004 + 0.021n_e + 0.002n_e^2 - 0.152B_x^2 - 0.011B_z n_e. \tag{H2}$$

While incorporating Galilean-augmented data into sparse regression had no impact on the functional form of the perpendicular pressure closure, discovering the same model with Galilean-augmented data gives reassurance that the functional form is indeed Galilean invariant. We note that since the model trained with laboratory-frame data is already symmetry abiding, there is no amount of laboratory-frame data or severity of maximum boost velocity that could alter the original model. We see that the sparse regression inferred closures outperformed commonly used analytical closures (see Fig. 10). Next, we trained neural networks to predict $P_{e\perp}$ with and without Galilean data augmentation to compare with the sparse regression model. We allowed $P_{e\perp}$ to depend on $\{n_e, \overline{v_{ex}}, \overline{v_{ey}}, \overline{v_{ez}}, E_x, E_y, E_z, B_x, B_y, B_z\}$. Again, we sample quantities at 1000 spatial locations every $10w_{pe}^{-1}$ from $t_{\min} = 0$ to $t_{\max} = 250w_{pe}^{-1}$. We only sample from spatial locations where $P_{e\parallel}$ is greater than its 70th percentile value at its sampled time step. When performing Galilean data augmentation, we double the training dataset size by including random Galilean boosts of each sampled point with $|\beta_j| \leqslant 0.1c$. Our network geometry is three hidden layers with 25 nodes per layer. Each network is trained in PyTorch for 2000 epochs with an AdamW optimizer, learning rate of 0.001, and batch size of 10. Additionally, each network is trained and tested five times to develop statistics for the performance of each model. We see that including Galilean-augmented data does not greatly improve model generalization (Fig. 10).

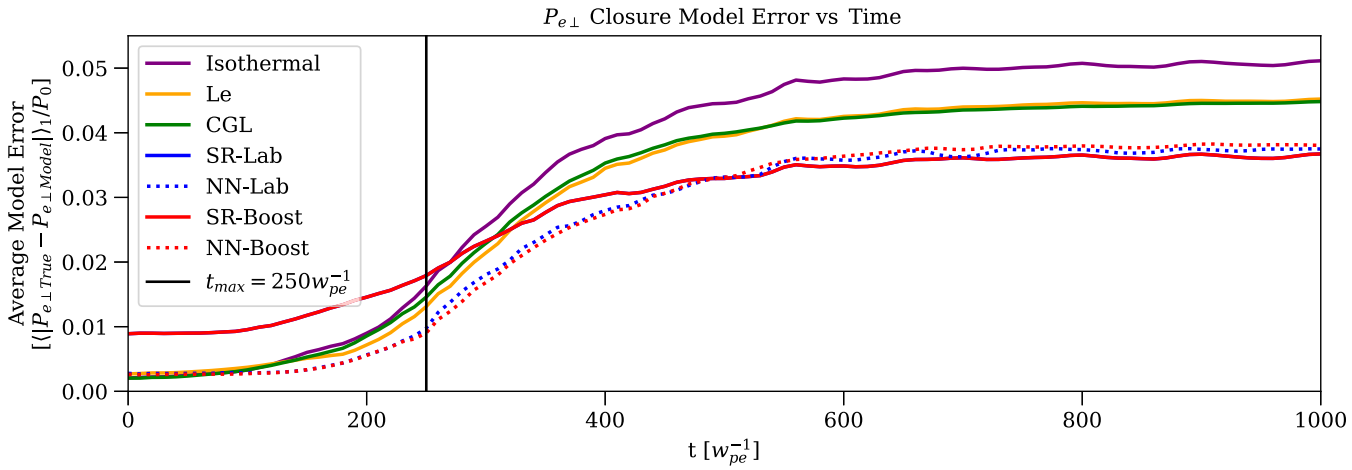


FIG. 10. Sparse regression and neural network $P_{e\perp}$ closures inferred from collisionless magnetic reconnection outperform commonly used analytical closures, however no improvement from Galilean augmentation. The vertical axis corresponds to model error measured as the spatial mean of the absolute L1 difference between the true simulation pressure and the pressure closure model. We normalize the average pressure error to a characteristic pressure value in the simulation $P_0 = 0.1 n_0 m_e c^2$. The horizontal axis is time. The sparse regression and neural network models are trained until $t = t_{\max} = 250 w_{pe}^{-1}$. As time increases and reconnection becomes more nonlinear and complicated, all closures perform worse. At late, generalization times the sparse regression and neural network models with and without boosting have the lowest model error. Since the sparse regression closures with and without Galilean data augmentation are identical, their error is identical. There is also very little difference between the neural network closures trained with and without data augmentation.

Both neural network models outperform commonly used analytical pressure closures. We see that the machine learning algorithms—both sparse regression and neural networks—did not rely on the non-Galilean invariant inputs to predict $P_{e\perp}$. Including Galilean-augmented data did not change model generalization. This is possibly because dynamics are more rapid and kinetic in the $\hat{\parallel}$ direction along magnetic field lines than in the $\hat{\perp}$ direction; thus, the $P_{e\perp}$ evolution is much less interesting and easier to predict without the need for symmetry embedding to further restrict the solution space. We are operating in the medium guide field regime where particles are significantly tied to the field lines, but if we ventured into the low guide field limit the $P_{e\perp}$ dynamics would become more varied and perhaps symmetry embedding would become more important for perpendicular closure identification.

APPENDIX I: ANALYSIS OF SPATIAL ERROR DISTRIBUTION OF ML CLOSURE MODELS FOR $P_{e\parallel}$

We can evaluate the analytical and machine-learned pressure closure models discovered in Sec. II B on the PIC MR data and examine the spatial distributions of their predictions (see Fig. 11). The pressure values are normalized to a characteristic downstream pressure value in the simulation $P_0 = 0.1 n_0 m_e c^2$. We plot the spatial distributions at $t = 600 w_{pe}^{-1}$, significantly later than the ML training cutoff time $t_{\max} = 250 w_{pe}^{-1}$. The true PIC simulation $P_{e\parallel}/P_0$ is plotted in Fig. 11(a). We see that the analytical closure models—isothermal [Fig. 11(b)], CGL [Fig. 11(c)], and Le [Fig. 11(d)]—all greatly under predict the pressure inside of the plasmoids at this time. The ML closure models—SR-Lab [Fig. 11(e)], SR-Boost [Fig. 11(f)], NN-Lab [Fig. 11(g)], and NN-Boost [Fig. 11(h)]—predict larger, more accurate pressure values than the analytical models inside of the plasmoids.

To better understand where in the spatial domain different models succeed or fail, we also visualize the differences between model predictions and the true PIC $P_{e\parallel}$ (see Fig. 12). For reference, the true PIC pressure is again shown in the upper left panel, Fig. 12(a). From this error plot, it becomes easier to see that the analytical models—isothermal [Fig. 12(b)], CGL [Fig. 12(c)], and Le [Fig. 12(d)]—greatly underestimate the pressure inside of the plasmoids, while correctly modeling the upstream regions outside of the plamoids. In contrast, the ML models—SR-Lab [Fig. 12(e)], SR-Boost [Fig. 12(f)], NN-Lab [Fig. 12(g)], and NN-Boost [Fig. 12(h)]—better capture the high pressure plasmoid regions, yet perform worse outside of the plasmoids. Such behavior is likely a result of our training data sampling method, which selects only those measurements where the target variable $P_{e\parallel}$ is greater than its 70th percentile value at its sampled time step. This data bias leads to an imbalance in the dataset, causing the ML models to prioritize fitting larger pressure regions inside of the plasmoids and de-emphasizing smaller pressure regions outside of the plasmoids. Future work should explore the impact of more balanced datasets.

The main improvement of SR-Boost compared with SR-Lab is its improved accuracy in predicting regions outside of the plasmoids (Fig. 12), which leads to the lower mean absolute error (Fig. 6). As for NN-Boost over NN-Lab, the predominant improvement is inside of the plasmoids. Finally, it is interesting to note that the SR-Boost and NN-Boost models have similar mean absolute errors at $t = 600 w_{pe}^{-1}$, yet their spatial error patterns have different structures. This indicates that SR and NN models have identified different functional forms for the closure, yet the NN's highly nonlinear expressive functional form is superfluous as it does not improve mean absolute closure error relative to SR in this phase.

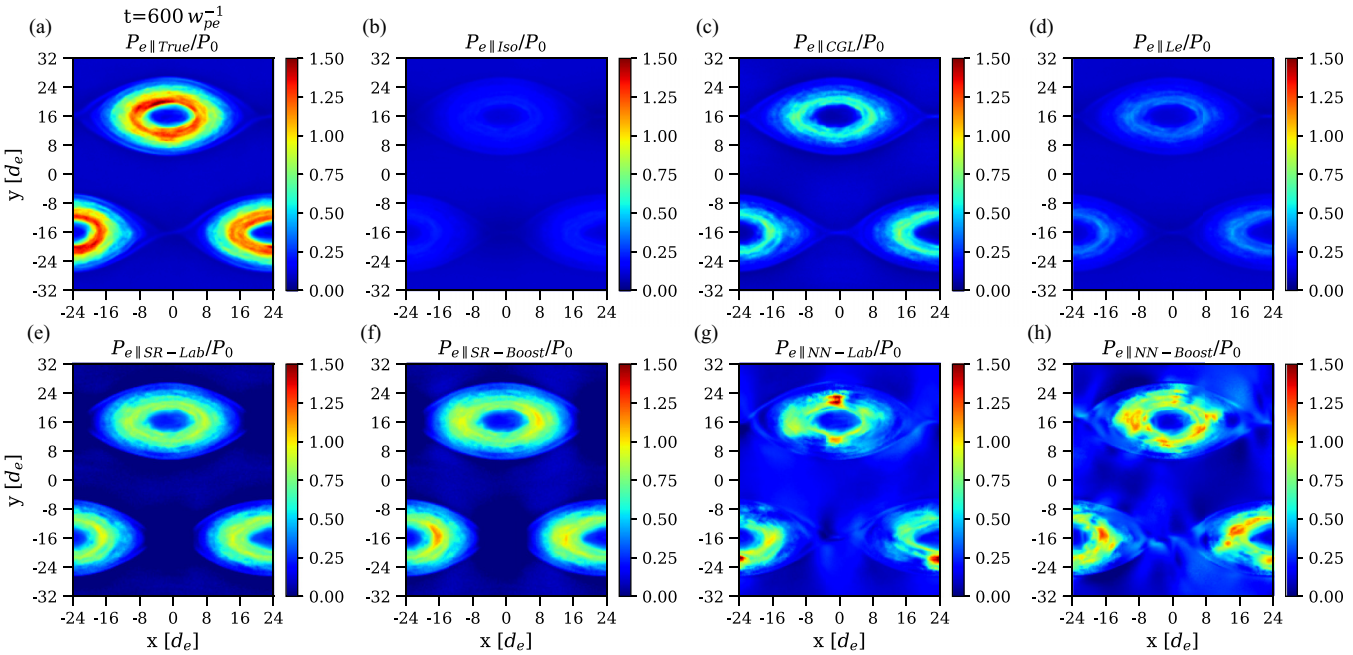


FIG. 11. Comparison of spatial patterns for analytical and ML inferred $P_{e\parallel}$ closure models. In this plot, we examine the spatial distributions of all closure models, evaluated on the PIC MR data in the (x, y) plane at $t = 600w_{pe}^{-1}$. This time is well into the generalization/testing phase. Panel (a) shows the true PIC pressure. Panel (b) depicts the isothermal closure model evaluated on the PIC data, constructed with the PIC density. Likewise, panels (c) and (d) show the CGL and Le closure models, respectively. The lower row is the ML inferred pressure closure models. Panels (e), (f), (g), and (h) plot the pressure as predicted by the SR-Lab, SR-Boost, NN-Lab, and NN-Boost models, respectively. All pressure values are normalized to $P_0 = 0.1n_0m_ec^2$, a characteristic downstream pressure value in the simulation.

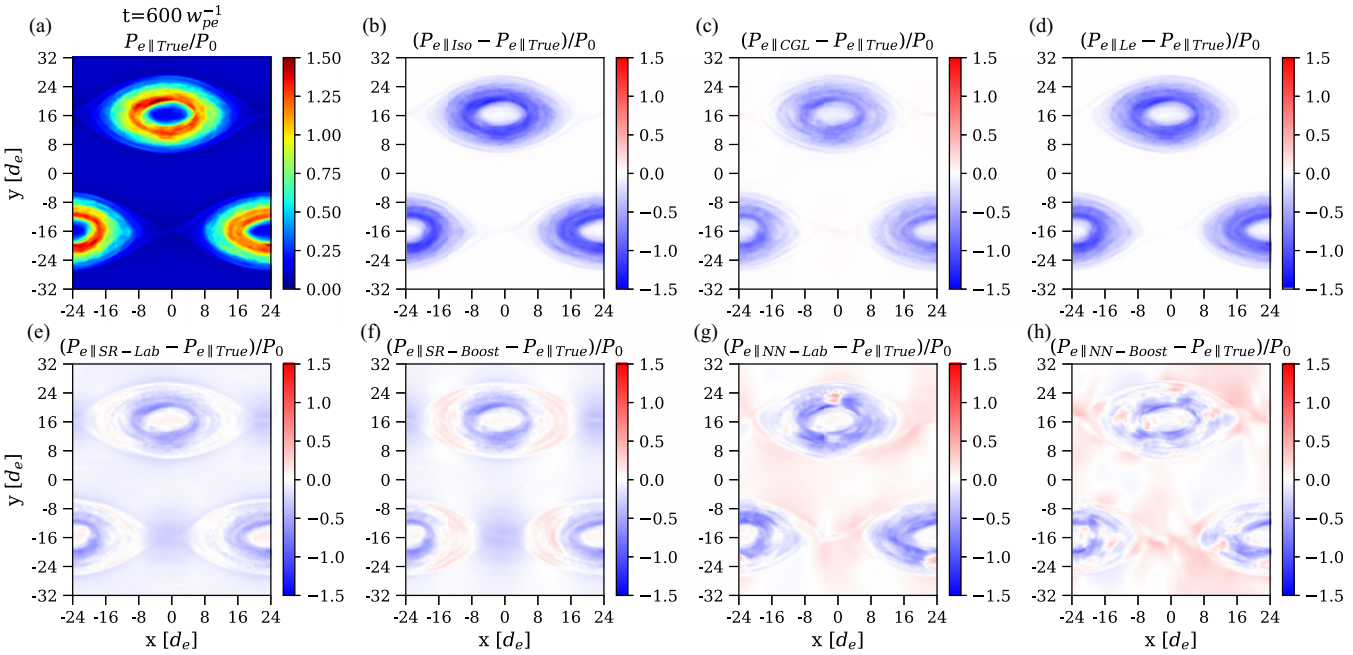


FIG. 12. Comparison of spatial errors for analytical and ML inferred $P_{e\parallel}$ closure models. In this plot, we examine the spatial distribution of error of all closure models, evaluated on the PIC MR data in the (x, y) plane at $t = 600w_{pe}^{-1}$. This time is well into the generalization/testing phase. Panel (a) shows the true PIC pressure, for reference. Panel (b) depicts the difference between the isothermal closure model and the PIC pressure. Likewise, panels (c) and (d) plot the error of CGL and Le closure models, respectively. The lower row is the error from the ML inferred pressure closure models. Panels (e), (f), (g), and (h) plot the errors of the SR-Lab, SR-Boost, NN-Lab, and NN-Boost models, respectively. All pressure values are normalized to $P_0 = 0.1n_0m_ec^2$, a characteristic downstream pressure value in the simulation. Positive (red) values indicate that the model pressure is larger than the true PIC pressure, while negative (blue) values indicate that the model pressure underestimates.

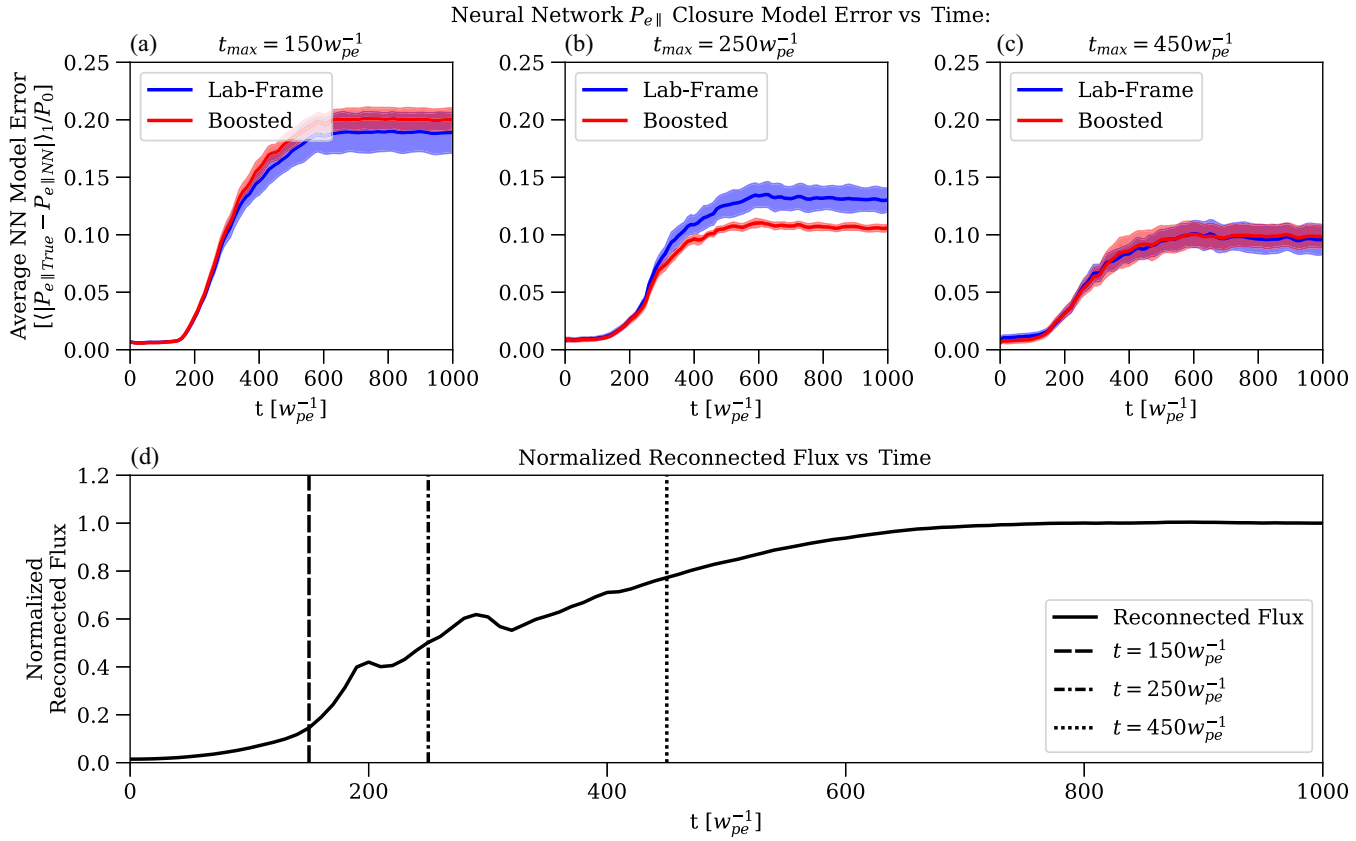


FIG. 13. Galilean data augmentation improves neural network closure model for $P_{e\parallel}$. We train neural networks with and without Galilean-augmented data from $t = 0w_{pe}^{-1}$ to differing $t_{\max}$. $t_{\max}$ is varied between $150w_{pe}^{-1}$, $250w_{pe}^{-1}$, and $450w_{pe}^{-1}$ across the top three panels (a), (b), and (c), respectively. The vertical axes correspond to model error measured as the spatial mean of the absolute L1 difference between the true simulation pressure and the pressure closure model. The statistics come from training each model five times with different realizations of the random data sampling. We normalize the average pressure error to a characteristic pressure value in the simulation $P_0 = 0.1n_0m_ec^2$. The horizontal axis is time. In panel (d), we display percentage of reconnected magnetic flux with the corresponding values of $t_{\max}$ overlaid. We observe that for the models trained up to $t_{\max} = 250w_{pe}^{-1}$ in panel (b), Galilean data augmentation improves generalization. For the other two models in panels (a) and (c), the performance differences between laboratory frame and boosted neural networks are within the variance.

APPENDIX J: IMPACT OF MAXIMUM TRAINING TIME ON THE RELATIVE PERFORMANCE OF LABORATORY-FRAME AND BOOSTED NEURAL NETWORKS

In Sec. II B, we discovered $P_{e\parallel}$ closures with data $t < t_{\max} = 250w_{pe}^{-1}$ (corresponding to a time when ∼50% of the magnetic flux has reconnected) and then examined pressure closure model generalization to later times. $t_{\max}$ is a free parameter that we can vary, and here we examine neural network laboratory-frame and boosted models trained with different values of $t_{\max}$. We vary $t_{\max} = \{150, 250, 450\}w_{pe}^{-1}$ to understand when Galilean data augmentation improves model generalization and when it does not (Fig. 13). We additionally train each model five separate times with different realizations of the random data sampling to ensure our results are statistically significant given the spread in model performance for different samplings. We find that for the models trained with data up to $t_{\max} = 150w_{pe}^{-1}$, the network is not able to improve with the addition of Galilean-boosted data [Fig. 13(a)]. We believe that this is due to the lack of sufficiently rich (nonlinear) dynamics in the training data, which prevents the ML model from effectively capturing the underlying physics. As a result, Galilean-augmented data do not lead to improved model performance. For the models trained with data up to $t_{\max} = 250w_{pe}^{-1}$, Galilean-augmented data give a significant generalization improvement [Fig. 13(b)]. Finally, for the models trained with data up to $t_{\max} = 450w_{pe}^{-1}$, we again observe no improvement with the inclusion of Galilean-augmented data [Fig. 13(c)]. This is presumably because the dynamics at times greater than $t_{\max} = 450w_{pe}^{-1}$ do not significantly differ than those during the training data; hence, the model trained with laboratory-frame data is not greatly penalized by not abiding by the symmetry and grasping onto non-symmetry-preserving terms. There appears to be a Goldilocks zone ($t = 250w_{pe}^{-1}$) in which the models have been exposed to enough, but too much, dynamics allowing Galilean-augmented data to meaningfully improve generalization. The onset and duration of the Goldilocks interval depend on the characteristics of the plasma system. However, despite this Goldilocks interval existing for model performance, the use of symmetry augmented data plays an important role in helping eliminate spurious, frame-dependent correlations that violate frame invariance, regardless of dataset size.

[1] J. Nuckolls, L. Wood, A. Thiessen, and G. Zimmerman, Laser compression of matter to super-high densities: Thermonuclear (CTR) applications, Nature (London) **239**, 139 (1972).

[2] J. Ongena, R. Koch, R. Wolf, and H. Zohm, Magnetic-confinement fusion, Nat. Phys. **12**, 398 (2016).

[3] L. Sironi and A. Spitkovsky, Relativistic reconnection: An efficient source of non-thermal particles, Astrophys. J. **783**, L21 (2014).

[4] F. Guo, Y.-H. Liu, W. Daughton, and H. Li, Particle acceleration and plasma dynamics during magnetic reconnection in the magnetically-dominated regime, Astrophys. J. **806**, 167 (2015).

[5] G. W. Hammett and F. W. Perkins, Fluid moment models for Landau damping with application to the ion-temperature-gradient instability, Phys. Rev. Lett. **64**, 3019 (1990).

[6] O. Ohia, J. Egedal, V. S. Lukin, W. Daughton, and A. Le, Demonstration of anisotropic fluid closure capturing the kinetic structure of magnetic reconnection, Phys. Rev. Lett. **109**, 115004 (2012).

[7] L. Comisso and L. Sironi, Particle acceleration in relativistic plasma turbulence, Phys. Rev. Lett. **121**, 255101 (2018).

[8] E. P. Alves, J. Zrake, and F. Fiuza, Efficient nonthermal particle acceleration by the kink instability in relativistic jets, Phys. Rev. Lett. **121**, 245101 (2018).

[9] B. Laperre, J. Amaya, S. Jamal, and G. Lapenta, Identification of high order closure terms from fully kinetic simulations using machine learning, Phys. Plasmas **29**, 032706 (2022).

[10] Y. Qin, J. Ma, M. Jiang, C. Dong, H. Fu, L. Wang, W. Cheng, and Y. Jin, Data-driven modeling of Landau damping by physics-informed neural networks, Phys. Rev. Res. **5**, 033079 (2023).

[11] Z. Huang, C. Dong, and L. Wang, Machine-learning heat flux closure for multi-moment fluid modeling of nonlinear Landau damping, Proc. Natl. Acad. Sci. USA **122**, e2419073122 (2025).

[12] A. S. Joglekar and A. G. R. Thomas, Machine learning of hidden variables in multiscale fluid simulation, Mach. Learn.: Sci. Technol. **4**, 035049 (2023).

[13] N. Barbour, W. Dorland, I. G. Abel, and M. Landreman, Machine-learning closure for Vlasov-Poisson dynamics in Fourier-Hermite space, J. Plasma Phys. **91**, E140 (2025).

[14] D. D. Carvalho, D. R. Ferreira, and L. O. Silva, Learning the dynamics of a one-dimensional plasma model with graph neural networks, Mach. Learn.: Sci. Technol. **5**, 025048 (2024).

[15] J. Bongard and H. Lipson, Automated reverse engineering of nonlinear dynamical systems, Proc. Natl. Acad. Sci. USA **104**, 9943 (2007).

[16] M. Schmidt and H. Lipson, Distilling free-form natural laws from experimental data, Science **324**, 81 (2009).

[17] S.-M. Udrescu and M. Tegmark, AI Feynman: A physics-inspired method for symbolic regression, Sci. Adv. **6**, eaay2631 (2020).

[18] M. Cranmer, A. Sanchez Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho, Discovering symbolic models from deep learning with inductive biases, in *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Curran Associates, Inc., Red Hook, NY, USA, 2020), Vol. 33, pp. 17429–17442.

[19] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, Predicting catastrophes in nonlinear dynamical systems by compressive sensing, Phys. Rev. Lett. **106**, 154101 (2011).

[20] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. USA **113**, 3932 (2016).

[21] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Data-driven discovery of partial differential equations, Sci. Adv. **3**, e1602614 (2017).

[22] H. Schaeffer, Learning partial differential equations via data discovery and sparse optimization, Proc. R. Soc. A **473**, 20160446 (2017).

[23] E. P. Alves and F. Fiuza, Data-driven discovery of reduced plasma physics models from fully kinetic simulations, Phys. Rev. Res. **4**, 033192 (2022).

[24] J. Donaghy and K. Germaschewski, In search of a data-driven symbolic multi-fluid ten-moment model closure, J. Plasma Phys. **89**, 895890105 (2023).

[25] E. R. Ingelsten, M. C. McGrae-Menge, E. P. Alves, and I. Pusztai, Data-driven discovery of a heat flux closure for electrostatic plasma phenomena, J. Plasma Phys. **91**, E64 (2025).

[26] I. Abramovic, E. Alves, and M. Greenwald, Data-driven model discovery for plasma turbulence modelling, J. Plasma Phys. **88**, 895880604 (2022).

[27] M. Dam, M. Brøns, J. Juul Rasmussen, V. Naulin, and J. S. Hesthaven, Sparse identification of a predator-prey system from simulation data of a convection model, Phys. Plasmas **24**, 022310 (2017).

[28] A. A. Kaptanoglu, K. D. Morgan, C. J. Hansen, and S. L. Brunton, Physics-constrained, low-dimensional models for magnetohydrodynamics: First-principles and data-driven approaches, Phys. Rev. E **104**, 015206 (2021).

[29] M. Golden, K. Satapathy, and D. Psaltis, Scalable discovery of fundamental physical laws: Learning magnetohydrodynamics from 3d turbulence data, arXiv:2501.04094.

[30] A. A. Kaptanoglu, T. Qian, F. Wechsung, and M. Landreman, Permanent-magnet optimization for stellarators as sparse regression, Phys. Rev. Appl. **18**, 044006 (2022).

[31] A. A. Kaptanoglu, C. Hansen, J. D. Lore, M. Landreman, and S. L. Brunton, Sparse regression for plasma physics, Phys. Plasmas **30**, 033906 (2023).

[32] J. D. Lore, S. De Pascuale, P. Laiu, B. Russo, J. S. Park, J. M. Park, S. L. Brunton, J. N. Kutz, and A. A. Kaptanoglu, Time-dependent SOLPS-ITER simulations of the tokamak plasma boundary for model predictive control using SINDy, Nucl. Fusion **63**, 046015 (2023).

[33] B. Bayón-Buján and M. Merino, Data-driven sparse modeling of oscillations in plasma space propulsion, Mach. Learn.: Sci. Technol. **5**, 035057 (2024).

[34] B. Sanderse, P. Stinis, R. Maulik, and S. E. Ahmed, Scientific machine learning for closure models in multiscale problems: A review, Found. Data Sci. **7**, 298 (2025).

[35] A. A. Vlasov, On vibration properties of electron gas, J. Exp. Theor. Phys. **8**, 291 (1938).

[36] A. A. Vlasov, The vibrational properties of an electron gas, Usp. Fiz. Nauk **93**, 444 (1967).

[37] R. Hazeltine and F. Waelbroeck, *The Framework of Plasma Physics* (CRC Press, Boca Raton, FL, 1998).

[38] D. R. Gurevich, M. R. Golden, P. A. Reinbold, and R. O. Grigoriev, Learning fluid physics from highly turbulent data using sparse physics-informed discovery of empirical relations (SPIDER), J. Fluid Mech. **996**, A25 (2024).

[39] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, arXiv:2104.13478.

[40] T. Cohen and M. Welling, Group equivariant convolutional networks, in *Proceedings of The 33rd International Conference on Machine Learning*, edited by M. F. Balcan and K. Q. Weinberger (PMLR, New York, 2016), Vol. 48, pp. 2990–2999.

[41] J. Spinner, V. Bresó, P. de Haan, T. Plehn, J. Thaler, and J. Brehmer, Lorentz-equivariant geometric algebra transformers for high-energy physics, Adv. Neural Info. Proc. Sys. **37**, 22178 (2024).

[42] M. Zhdanov, D. Ruhe, M. Weiler, A. Lucic, J. Brandstetter, and P. Forré, Clifford-steerable convolutional neural networks, in *Proceedings of the 41st International Conference on Machine Learning*, edited by R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, and F. Berkenkamp (Proceedings of Machine Learning Research, 2024), Vol. 235. pp. 61203–61228.

[43] M. Raissi, P. Perdikaris, and G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. **378**, 686 (2019).

[44] W. Hao, Symmetry-regularized neural ordinary differential equations, arXiv:2311.16628.

[45] S. Chen, E. Dobriban, and J. H. Lee, A group-theoretic framework for data augmentation, J. Machine Learning Res. **21**, 1 (2020).

[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems*, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Curran Associates, Inc., Red Hook, NY, USA, 2012), Vol. 25.

[47] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, Deep, big, simple neural nets for handwritten digit recognition, Neural Comput. **22**, 3207 (2010).

[48] S. Monaco, S. Barresi, and D. Apiletti, Lorentz-invariant augmentation for high-energy physics deep learning models, in *Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, edited by R. Meo and F. Silvestri (Springer Nature, Cham, 2025), pp. 368–376.

[49] E. N. Parker, Sweet's mechanism for merging magnetic fields in conducting fluids, J. Geophys. Res. **62**, 509 (1957).

[50] J. Dungey, LXXVI. Conditions for the occurrence of electrical discharges in astrophysical systems, London Edinburgh Dublin Philos. Mag. J. Sci. **44**, 725 (1953).

[51] P. A. Sweet, The neutral point theory of solar flares, in *Electromagnetic Phenomena in Cosmical Physics*, *IAU Symposium*, edited by B. Lehnert (Cambridge University Press, London, UK, 1958), Vol. 6, p. 123.

[52] H. E. Petschek, Magnetic field annihilation, in *NASA Special Publication*, edited by W. N. Hess (NASA Special Publication, Washington, DC, 1964), Vol. 50, p. 425.

[53] J. Birn, J. F. Drake, M. A. Shay, B. N. Rogers, R. E. Denton, M. Hesse, M. Kuznetsova, Z. W. Ma, A. Bhattacharjee, A. Otto, and P. L. Pritchett, Geospace environmental modeling (GEM) magnetic reconnection challenge, J. Geophys. Res.: Space Phys. **106**, 3715 (2001).

[54] E. G. Zweibel and M. Yamada, Magnetic reconnection in astrophysical and laboratory plasmas, Annu. Rev. Astron. Astrophys. **47**, 291 (2009).

[55] Y.-H. Liu, P. Cassak, X. Li, M. Hesse, S.-C. Lin, and K. Genestreti, First-principles theory of the rate of magnetic reconnection in magnetospheric and solar plasmas, Commun. Phys. **5**, 97 (2022).

[56] P. A. Cassak, Y.-H. Liu, and M. Shay, A review of the 0.1 reconnection rate problem, J. Plasma Phys. **83**, 715830501 (2017).

[57] S. Selvi, O. Porth, B. Ripperda, F. Bacchini, L. Sironi, and R. Keppens, Effective resistivity in relativistic collisionless reconnection, Astrophys. J. **950**, 169 (2023).

[58] R. A. Fonseca, L. O. Silva, F. S. Tsung, V. K. Decyk, W. Lu, C. Ren, W. B. Mori, S. Deng, S. Lee, T. Katsouleas, and J. C. Adam, OSIRIS: A three-dimensional, fully relativistic particle in cell code for modeling plasma based accelerators, in *Computational Science—ICCS 2002*, edited by P. M. A. Sloot, A. G. Hoekstra, C. J. K. Tan, and J. J. Dongarra (Springer, Berlin, 2002), pp. 342–351.

[59] R. A. Fonseca, S. F. Martins, L. O. Silva, J. W. Tonge, F. S. Tsung, and W. B. Mori, One-to-one direct modeling of experiments and astrophysical scenarios: Pushing the envelope on kinetic plasma simulations, Plasma Phys. Controlled Fusion **50**, 124034 (2008).

[60] E. G. Harris, On a plasma sheath separating regions of oppositely directed magnetic field, Nuovo Cimento **23**, 115 (1962).

[61] M. C. McGrae-Menge, Embedding physical symmetries into machine-learned reduced plasma physics models via data augmentation: Data, Zenodo (2025), http://doi.org/10.5281/zenodo.15498576.

[62] J. P. Crutchfield and B. S. McNamara, Equations of motion from a data series, Complex Syst. **1**, 417 (1987).

[63] H. Schaeffer and S. G. McCalla, Sparse model selection via integral terms, Phys. Rev. E **96**, 023302 (2017).

[64] P. A. K. Reinbold, D. R. Gurevich, and R. O. Grigoriev, Using noisy or incomplete data to discover models of spatiotemporal dynamics, Phys. Rev. E **101**, 010203(R) (2020).

[65] D. A. Messenger and D. M. Bortz, Weak SINDy for partial differential equations, J. Comput. Phys. **443**, 110525 (2021).

[66] A. Adadi, A survey on data-efficient algorithms in big data era, J. Big Data **8**, 24 (2021).

[67] A. Le, J. Egedal, W. Daughton, W. Fox, and N. Katz, Equations of state for collisionless guide-field reconnection, Phys. Rev. Lett. **102**, 085001 (2009).

[68] J. Ng, A. Hakim, A. Bhattacharjee, A. Stanier, and W. Daughton, Simulations of anti-parallel reconnection using a nonlocal heat flux closure, Phys. Plasmas **24**, 082112 (2017).

[69] J. Ng, A. Hakim, L. Wang, and A. Bhattacharjee, An improved ten-moment closure for reconnection and instabilities, Phys. Plasmas **27**, 082106 (2020).

[70] G. F. Chew, M. L. Goldberger, F. E. Low, and S. Chandrasekhar, The Boltzmann equation and the one-fluid hydromagnetic equations in the absence of particle collisions, Proc. R. Soc. London A **236**, 112 (1956).

[71] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359 (1989).

[72] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, PyTorch: an imperative style, high-performance deep learning library, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, Article, 2019), Vol. 721, pp. 8026–8037.

[73] E. D. Held, J. D. Callen, C. C. Hegna, C. R. Sovinec, T. A. Gianakon, and S. E. Kruger, Nonlocal closures for plasma fluid simulations, Phys. Plasmas **11**, 2419 (2004).

[74] R. Maulik, N. A. Garland, J. W. Burby, X.-Z. Tang, and P. Balaprakash, Neural network representability of fully ionized plasma fluid model closures, Phys. Plasmas **27**, 072106 (2020).

[75] C. Ma, B. Zhu, X.-Q. Xu, and W. Wang, Machine learning surrogate models for Landau fluid closure, Phys. Plasmas **27**, 042502 (2020).

[76] L. Bois, E. Franck, L. Navoret, and V. Vigon, A neural network closure for the Euler-Poisson system based on kinetic simulations, Kinet. Relat. Models **15**, 49 (2022).

[77] M. C. McGrae-Menge, D. D. Carvalho, A. Velberg, M. Almanza, N. Barbour, V. K. Decyk, F. Fiuza, W. D. Dorland, N. F. Loureiro, and E. P. Alves, Data-driven discovery of fluid closures for magnetic reconnection from fully kinetic particle-in-cell simulations, in *66th Annual Meeting of the APS Division of Plasma Physics* (American Physical Society, Atlanta, GA, 2024).

[78] J. Huang, Y. Cheng, A. J. Christlieb, and L. F. Roberts, Machine learning moment closure models for the radiative transfer equation I: Directly learning a gradient based closure, J. Comput. Phys. **453**, 110941 (2022).

[79] R. Greif, F. Jenko, and N. Thuerey, Physics-preserving AI-accelerated simulations of plasma turbulence, arXiv:2309.16400.

[80] C. Chang, V. Deringer, K. Katti, V. Speybroeck, and C. Wolverton, Simulations in the era of exascale computing, Nat. Rev. Mater. **8**, 309 (2023).

[81] M. Bugli, E. F. Lopresti, E. Figueiredo, A. Mignone, B. Cerutti, G. Mattia, L. Del Zanna, G. Bodo, and V. Berta, Relativistic reconnection with effective resistivity: I. Dynamics and reconnection rate, Astron. Astrophys. **693**, A233 (2025).

[82] A. Moran, L. Sironi, A. Levis, B. Ripperda, E. R. Most, and S. Selvi, Effective resistivity in relativistic reconnection: A prescription based on fully kinetic simulations, Astrophys. J. Lett. **978**, L45 (2025).

[83] K. Kaheman, J. N. Kutz, and S. L. Brunton, SINDy-PI: A robust algorithm for parallel implicit sparse identification of nonlinear dynamics, Proc. R. Soc. A **476**, 20200279 (2020).

[84] M. Golden, Scalable sparse regression for model discovery: The fast lane to insight, arXiv:2405.09579.

[85] J.-L. Vay, D. L. Bruhwiler, C. G. R. Geddes, W. M. Fawley, S. F. Martins, J. R. Cary, E. Cormier-Michel, B. Cowan, R. A. Fonseca, M. A. Furman, W. Lu, W. B. Mori, and L. O. Silva, Simulating relativistic beam and plasma systems using an optimal boosted frame, J. Phys.: Conf. Ser. **180**, 012006 (2009).