PAPER • OPEN ACCESS

Machine learning inspired models for Hall effects in non-collinear magnets

To cite this article: Jonathan Kipp *et al* 2024 *Mach. Learn.: Sci. Technol.* **5** 025060

View the article online for updates and enhancements.

MACHINE
LEARNING
Science and Technology

PAPER

Machine learning inspired models for Hall effects in non-collinear magnets

Jonathan Kipp[1,2,*] , Fabian R Lux[3,6] , Thorben Pürling[1,2], Abigail Morrison[4,5], Stefan Blügel[1,2] , Daniele Pinna[1] and Yuriy Mokrousov[1,3]

1 Peter Grünberg Institut and Institute for Advanced Simulation (IAS-1), Forschungszentrum Jülich and JARA, Jülich 52425, Germany
2 Department of Physics, RWTH Aachen University, Aachen 52056, Germany
3 Institute of Physics, Johannes Gutenberg University Mainz, Mainz 55099, Germany
4 Institute for Computational and Systems Neuroscience (INM-6), Forschungszentrum Jülich, 52425 Jülich, Germany
5 Department of Computerscience, RWTH Aachen University, 52056 Aachen, Germany
6 Department of Physics, Yeshiva University, New York, NY 10016, United States of America
* Author to whom any correspondence should be addressed.

E-mail: j.kipp@fz-juelich.de

Abstract

The anomalous Hall effect has been front and center in solid state research and material science for over a century now, and the complex transport phenomena in nontrivial magnetic textures have gained an increasing amount of attention, both in theoretical and experimental studies. However, a clear path forward to capturing the influence of magnetization dynamics on anomalous Hall effect even in smallest frustrated magnets or spatially extended magnetic textures is still intensively sought after. In this work, we present an expansion of the anomalous Hall tensor into symmetrically invariant objects, encoding the magnetic configuration up to arbitrary power of spin. We show that these symmetric invariants can be utilized in conjunction with advanced regularization techniques in order to build models for the electric transport in magnetic textures which are, on one hand, complete with respect to the point group symmetry of the underlying lattice, and on the other hand, depend on a minimal number of order parameters only. Here, using a four-band tight-binding model on a honeycomb lattice, we demonstrate that the developed method can be used to address the importance and properties of higher-order contributions to transverse transport. The efficiency and breadth enabled by this method provides an ideal systematic approach to tackle the inherent complexity of response properties of noncollinear magnets, paving the way to the exploration of electric transport in intrinsically frustrated magnets as well as large-scale magnetic textures.

1. Introduction

The anomalous Hall effect (AHE) has been an essential measurement tool to study magnetic matter since its discovery over 100 years ago, exhibiting a plethora of mechanisms responsible for transverse electric transport [1]. The theoretical description of this ubiquitous phenomenon has an equally long history of new revelations, offering an ever modernizing view of electric transport in magnetic materials, with deeper insight into material characteristics gradually accumulating.

While extrinsic, scattering driven contributions to transverse resistivity can dominate in certain regimes, it is the contribution rooted in the intrinsic electronic structure that provides a large part of the signal in many cases, especially in strongly spin–orbit coupled materials [2, 3]. Investigations of intrinsic AHE are inseparably linked to the study of the geometry and topology of electronic states, most prominently condensing in Berry-phase effects emerging as a result of various flavors of emergent fields. Although AHE is traditionally associated with Berry phase in reciprocal space, the celebrated topological Hall effect (THE) is related to the Berry phase electrons acquire in magnetic textures with nonzero scalar spin chirality

$\xi = \vec{s}_A \cdot (\vec{s}_B \times \vec{s}_C)$ in real space [4]. While the emergent field picture has had great success particularly due to its conceptual elegance, experimental evidence suggesting the necessity to expand this approach, especially in materials with strong spin–orbit interaction, is accumulating rapidly [5–9].

In the past years, the field of AHE in complex magnetic materials has been experiencing a true uprising [10]. The key observation is that a combination of crystal symmetries with time-reversal symmetry breaking by non-collinear magnetic order can result in contributions to the AHE which go beyond those traditionally associated with ferromagnetic magnetization or simplified real-space emergent field pictures. One of suggested guiding principles in keeping a systematic overarching counting of possible contributions to the AHE in a system which undergoes drastic changes in its magnetic configuration lies in a symmetry expansion of the anomalous Hall conductivity in terms of products of spins on different atomic sites of arbitrary power [11]. This approach has been applied in the past to the case of a bipartite lattice, demonstrating that it is possible to conceptually separate the AHE in systems with two spins into chiral and crystal contributions [12, 13], which collect different powers of the spin expansion. Demonstrating the intricate physics that lies behind this decomposition, it was in particular shown that the corresponding chiral Hall effect (CHE) intertwines together the Berry phases of Bloch electrons in real and reciprocal spaces [6]. While being very rigorous and potentially very promising in uncovering the physics of higher-order contributions to the AHE, applying this method to systems containing more spins ideally requires an automated approach.

Generally, the research in the area of AHE can benefit greatly from a deep understanding of the implications that crystal symmetry has on the geometry and topology of Bloch electrons, which can be modified excessively by tuning parameters in the enormous phase space of the magnetic configurations. While this phase space offers intriguing opportunities for exploration [14, 15], the numerical complexity grows exponentially with the number of magnetic moments in the system. Increased computational cost for large system sizes or complex combinatorics are problems many disciplines suffer from as a whole. Consequently, natural sciences, and especially computational solid state physics, are witnessing a paradigm shift to find a way around increasingly costly simulations. Machine learning techniques have been gaining more and more attention over the past 10 years and have found application particularly in the search for pretrained electron potentials for accelerated density functional theory (DFT) calculations [16], computational materials design [17–20], rapid exploration of large parameter spaces [21], or interpretation of extensive experimental databases [22]. However, comparable studies focusing on magnetic degrees of freedom are few and far between [23].

The machine learning techniques have matured so far that feature selection, feature extraction and model regularization can distill relevant features from high-dimensional input both effectively and reliably, and therefore aid algorithms which are in principle unaware of any physical intuition in finding models which we would accept as physical. In essence, using the right tools out of a huge, readily available toolbox, it is possible to fit models depending on just the relevant fraction of a large number of features, thereby reducing a phase space which might be too large to conquer to manageable size. Consequently, the main objective of this work is to train minimal, regularized models predicting the electric transport properties of complex magnets in terms of suitable descriptors with machine learning techniques.

At the heart of this modeling framework an expansion in symmetric invariants is employed as a basis for the descriptor space, closely related to other universal descriptor spaces derived from atomic cluster expansions (ACE) or cluster multipole expansions [24, 25]. This kind of expansion is a powerful ingredient in the sense that it confines the shape of the terms being used, but at the same time provides us with all possible spin interaction terms respecting the crystal symmetry. It is therefore a valuable alternative to spin modeling techniques, where product terms are constructed 'by hand' [23], because it is complete with respect to the point group symmetry. In magnetism, such atomistic spin models traditionally offer great utility for the study of magnetic properties in realistic materials, and recently, efforts have been made to optimize these spin Hamiltonians with the help of machine learning techniques [26]. However, the issues surrounding the role of the magnetic reference state used for the calculations, performed in order to extract the interactions parametrizing the spin Hamiltonian have been subject to debate [27, 28]. Accordingly, expansion techniques which do not refer to a fixed reference state, like the one presented in this work, might offer a new perspective on one of the staple methods in magnetism research.

Furthermore, the expansion method is not bound to a specific limit, neither microscopically canted spins nor mesoscopically extended magnetic textures, and might therefore be employed to derive a description which transitions smoothly between the theories for the two limits. Ultimately, the main goal of this project is to obtain linear models with reasonable generalization statistics on unseen data for calculations of the AHE in complex canted magnets with strong spin–orbit interaction, using a minimal number of descriptors constructed from the symmetric invariant expansion. From these models we aim to extract, first, the already familiar contributions to the AHE, as well as further, previously unknown terms in higher orders of the

expansion. Second, we will test whether these models can be reliably and efficiently employed for extrapolation of the AHE in the magnetic phase space. Third, by varying the Fermi energy, we will examine whether the model construction is consistent enough to enable the exploration of the AHE in the combined phase space of spin moments and that of Bloch states with intricate reciprocal space geometry.

Mapping the fitted model coefficients as functions of the tight-binding model parameters as well as the system size will give us insight into the behavior of the different contributions to the AHE in different regimes. We classify the many higher order terms in low dimensional, microscopically canted magnets with respect to vector chirality $\vec{\xi} = \vec{s}_A \times \vec{s}_B$ in order to derive a recipe for predicting the magnetotransport properties in large scale magnetic textures, such as spin spirals, skyrmions, or multi-q states. While we focus here on the chirality-even contributions to the AHE, which correspond e.g. to the so-called crystal Hall effect in case of antiferromagnets [11, 29], one of our goals lies in providing an ability to identify the differences in transport signatures between different textures and associate these differences with specific $\vec{\xi}$-odd or -even terms in the model. This will contribute to identification and distinction of magnetic textures by measurement of electric transport, which might play a key role in development of novel computing architectures.

In our work we bring together aspects form multiple fields, namely the calculation of the AHE from an electronic model, the expansion of a tensor in symmetric invariants, in this specific case the anomalous Hall conductivity tensor, and the modeling of the tensor based on invariant expansion utilizing machine learning methods. Therefore, we begin by introducing the electronic model and its symmetries in section 2.1 as well as the linear response method used in the calculation of the anomalous Hall effect in section 2.5. We continue with illustrating the representation-based expansion machinery in section 2.2, before explaining the details of the model selection pipeline in section 2.6. Finally, we present and analyze fitting results in section 3.4. We comment on the feature scaling with respect to parameters of the tight-binding model and offer a concluding discussion in section 4.

2. Method

In our work we draw inspiration from three different perspectives, which define the three levels of our approach: on the surface, this is a machine learning approach to fitting sparse, linear models. The second level reveals, that input features to the learning algorithm are supplied by an expansion of the target in orders of spin, firmly tied to the underlying lattice symmetry. The base level of this framework provides the target to the algorithm, by electronic structure calculations on a discrete, magnetic lattice. Defining a working process to successfully combine these three perspectives in a modeling pipeline is at the core of our manuscript. In order to clarify the multifaceted approach presented here, we define the working parts of the pipeline first, before illustrating the methods in obtaining training data later in section 3.

2.1. Electronic model and its symmetries

In this work, we aim at predicting the electric transport properties of electrons on a bipartite honeycomb lattice of magnetic spins. To model the electronic structure, we employ an effective two-dimensional lattice tight-binding (TB) Hamiltonian (in the xy-plane) which reads:

$$H = -t \sum_{\langle ij \rangle \alpha} c_{i\alpha}^{\dagger} c_{j\alpha} + i\,\alpha_{\mathrm{R}} \sum_{\langle ij \rangle \alpha\beta} \hat{\mathbf{e}}_z \cdot \left(\boldsymbol{\sigma} \times \mathbf{d}_{ij} \right)_{\alpha\beta} c_{i\alpha}^{\dagger} c_{j\beta} + \lambda_{\mathrm{ex}} \sum_{i\alpha\beta} \left(\hat{\mathbf{s}}_i \cdot \boldsymbol{\sigma} \right)_{\alpha\beta} c_{i\alpha}^{\dagger} c_{i\beta}, \tag{1}$$

where $c_{i\alpha}^{\dagger}$ ($c_{i\alpha}$) denotes the creation (annihilation) of an electron with spin α at site i, $\langle \ldots \rangle$ restricts the sums to nearest neighbors, the unit vector $\mathbf{d}_{ij}$ points from j to i, and $\boldsymbol{\sigma}$ stands for the vector of Pauli matrices. Besides the hopping with amplitude t, equation (1) contains the Rashba spin–orbit coupling of strength α_{R} originating for example in the surface potential gradient perpendicular to the plane (i.e. along $\hat{z}$). The remaining term in equation (1) is the local exchange term with λ_{ex} characterizing the strength of exchange splitting and $\hat{\mathbf{s}}_i$ stands for the direction of spin on site i. Here, we work with the following parameters of the model: $t = 1.0\,\mathrm{eV}$, $\alpha_{\mathrm{R}} = 0.4\,\mathrm{eV}$, and $\lambda_{\mathrm{ex}} = 1.4\,\mathrm{eV}$.

The analysis presented in this work also heavily relies on the crystallographic point group symmetry of the bipartite, nonmagnetic, honeycomb lattice, which is illustrated in figure 1. We can identify five nontrivial families of symmetry, or conjugacy classes, in the lattice, comprising the point group of C_{6v}, see table 1: on one hand, the lattice shows three classes of rotational symmetry, namely the six-, three-, and twofold rotation around an axis perpendicular to the lattice plane, which is commonly chosen to be the z-axis, with the lattice lying in the xy-plane. On the other hand, the nonmagnetic lattice is invariant under two classes of mirror operations, with the mirror plane being aligned with either the corners of the hexagon or the midpoints of the hexagons edges.
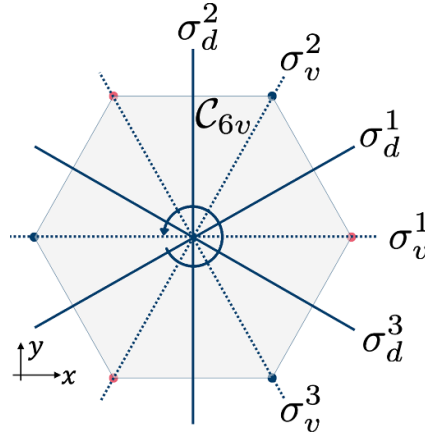
Figure 1. Illustration of symmetry operations of the honeycomb lattice. The symmetry operations leaving the bipartite honeycomb lattice (atom types in red and blue) invariant can be illustrated on the plane containing the lattice, here the *xy*-plane. The point group of the honeycomb lattice, C_{6v}, shows three classes of rotation around the *z*-axis, namely two-, three- and sixfold rotations $C_2(z)$, $C_3(z)$ and $C_6(z)$. Furthermore, there are two classes of mirror operations, σ_v and σ_d, indicated here by the mirror plane going through the corners of the hexagon (σ_v) or through the midpoint of the hexagons edges (σ_d).

Table 1. Character table of C_{6v}. Shown are the characters of each irreducible representation for each conjugacy class, alongside linear and quadratic basis functions which generate the respective representations. A tensor inducing a particular representation of the point group transforms accordingly to the characters of this representation. This allows us to easily augment the dataset by applying the symmetry operations to the spin configurations and multiply the target value by the according character value of the representation. In the case of the anomalous Hall conductivity (AHC) tensor, which belongs to the representation A_2, the tensor changes sign under the mirror operations with the mirror planes cutting through the edges and the middle of the sides of the hexagon, σ_v and σ_d.

C_{6v}	E	$2C_6(z)$	$2C_3(z)$	$C_2(z)$	$3\sigma_v$	$3\sigma_d$	linear	quadratic
A_1	1	1	1	1	1	1	z	x^2+y^2, z^2
A_2	1	1	1	1	-1	-1	n_z	
B_1	1	-1	1	-1	1	-1		
B_2	1	-1	1	-1	-1	1		
E_1	2	1	-1	-2	0	0	(x,y)	(xz, yz)
E_2	2	-1	-1	2	0	0		(x^2-y^2, xy)

2.2. Expansion in terms of symmetric invariants

Given an observable of interest, which can be of tensorial nature, we have to invoke the representation of corresponding tensor in terms of the crystallographic point group and then expand it in the symmetric invariants of this representation, which correspond to the eigenvalues of the invariant subspaces in the representation. Below, we briefly present the basics of representation theory that we make use of in our study [30].

We call a representation a map $\rho : G \to GL(n, \mathbb{C})$, where $GL(n, \mathbb{C})$ is the general linear group of $n \times n$ matrices with complex elements. n is in this case called the dimension of the representation of ρ. Every representation has to be a group homomorphism, meaning that $\rho(g_1 g_2) = \rho(g_1)\rho(g_2)$ for all $g_1, g_2 \in G$. We speak of two representations ρ and ρ' as equivalent, if there exists a matrix A such that $A\rho(g)A^{-1} = \rho'(g)$, for all $g \in G$. All representations of the same crystallographic point group have an equivalent unitary representation, fulfilling $\rho(g^{-1}) = \rho(g)^\dagger$, where $\dagger$ denotes complex conjugate transposition. The character of a representation ρ is defined as the matrix trace $\chi_\rho(g) = \mathrm{tr}\rho(g)$. The character map is constant under all conjugacy classes, because the trace is cyclic, $\chi(h^{-1}gh) = \chi(g)$, for all $g, h \in G$. For any two representations ρ_1 and ρ_2 we find

$$\chi_{\rho_1 \otimes \rho_2} = \chi_{\rho_1}(g)\,\chi_{\rho_2}(g) \tag{2}$$

$$\chi_{\rho_1 \oplus \rho_2} = \chi_{\rho_1}(g) + \chi_{\rho_2}(g), \tag{3}$$

where $\otimes$ denotes the Kronecker or tensor product and $\oplus$ is the direct sum of matrices. We call a representation reducible if it is equivalent to a direct sum of representations, and irreducible if it is not. The representations ρ of a crystallographic point group can be decomposed into a direct sum of irreducible representations Γ_i as follows:

$$\rho = \bigoplus_{i=1}^{n_c} \lambda_i \Gamma_i, \tag{4}$$

where n_c is the number of conjugacy classes of the group. The multiplicity coefficients can be extracted as $\lambda_i = \langle \Gamma_i | \rho \rangle$, since the Schur orthogonality theorem states that

$$\langle \Gamma_i | \Gamma_j \rangle = \frac{1}{|G|} \sum_{g \in G} \chi_{\Gamma_i}(g) \chi_{\Gamma_j}^*(g) = \delta_{i,j}, \tag{5}$$

with the Kronecker delta $\delta_{i,j}$. One specific representation we are interested in is the magnetic representation, i.e. the product of site and spin representation:

$$\Gamma_{\text{mag}} = \Gamma_{\text{site}} \otimes \Gamma_{\text{spin}}. \tag{6}$$

This representation has the dimension $d = 3 \cdot n_s$, with n_s the number of three-dimensional pseudovectors describing classical spins in the system, acting on vector of dimension d obtained from stacking the spins into one column as $(s_1^x, s_1^y, s_1^z, s_2^x, s_2^y, s_2^z, \cdots, s_{n_s}^x, s_{n_s}^y, s_{n_s}^z)^T$, commonly referred to as the image. The matrix $\Gamma_{\text{mag}}(g)$, with g being any element of the crystallographic point group, describes the combined action of the operation on the lattice sites and the magnetic moments via a reshuffling of the sites and the local rotation of magnetic moments in spin space under g. For example, in the honeycomb lattice with pointgroup C_{6v} and two atoms in the unit cell, the rotation by π around the z-axis, denoted by C_2, rotates both pseudovectors locally by $180°$ and interchanges the atomic sites A and B, resulting in the matrix:

$$\Gamma_{\text{mag}}(C_2) = \begin{array}{c} \\ A \\ \\ \hline \\ B \\ \\ \end{array} \left(\begin{array}{ccc|ccc} \multicolumn{3}{c|}{A} & \multicolumn{3}{c}{B} \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right). \tag{7}$$

The blockdiagonal form that all matrices in this representation have is obvious here.

Turning our attention now to a generic tensor we wish to expand in spin interaction terms, we require a representation of the group in higher dimensional spin space, to describe the action on terms comprising the product of any number of spins. Specifically, we calculate the invariant subspaces and corresponding eigenvalues of a projector P, which is the sum of all symmetry operations in the group calculated in the representation:

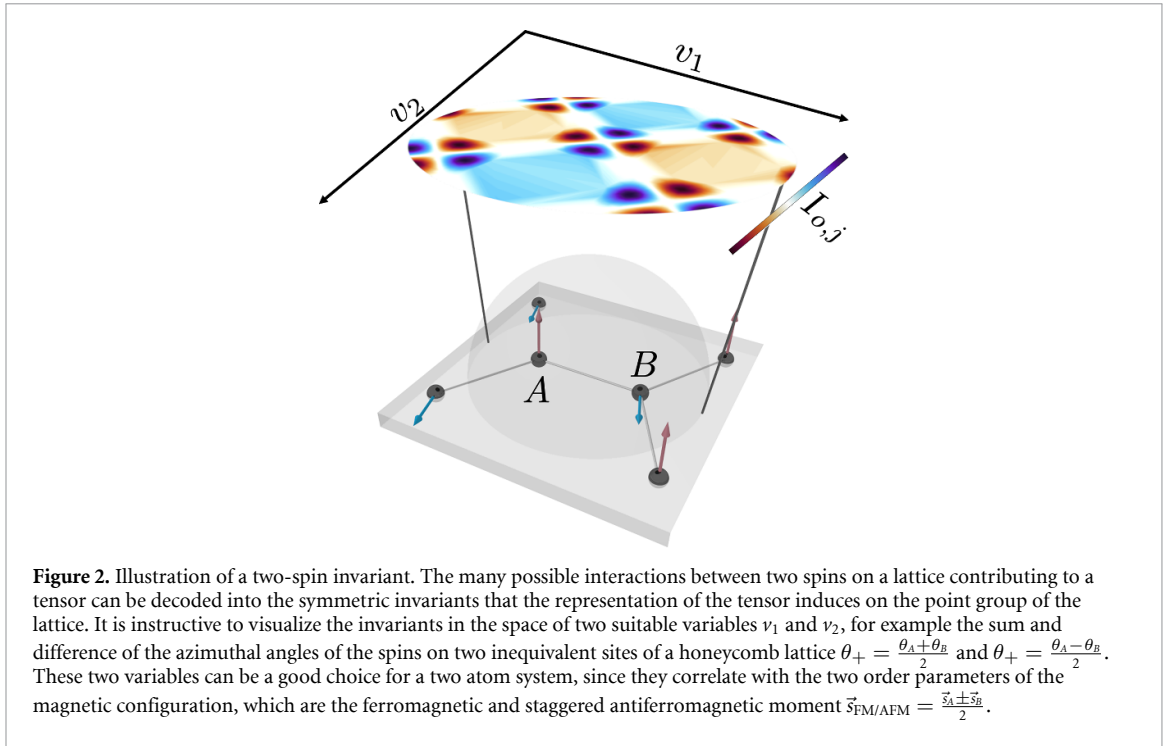$$\rho(P) = \frac{1}{|G|} \sum_{g \in G} \chi_\rho(g)^\dagger \cdot \rho(g), \tag{8}$$

where by ρ we denote the representation corresponding to the order of spin o. We call the eigenvalues equal to one *symmetric invariants*, since they encode the spin product terms we construct in a certain order of spin in scalar values, which are invariant under operations of the crystallographic point group.

For a given tensor expressed as a product of o spins, the magnetic representation induced by this tensor on the point group, $\Gamma_o = (\Gamma_{\text{mag}})^{\otimes o}$, will have the dimension d^o. Accordingly, the exponential growth of the representation's dimensionality renders the calculation of the necessary matrix operations impractical for larger system or high orders of spin.

Therefore, we make use of the fact that the product of scalars is symmetric under permutation of the scalars, and carry out the calculations in the *symmetric* space, which holds only the symmetric components of the product space. This reduces the dimensionality of the resulting representation from d^o to $\frac{(n+d-1)!}{n!(d-1)!}$. For example, this space does not hold all 36 components of the representation $\Gamma_{\text{mag}} \otimes \Gamma_{\text{mag}}$ for the two spin system, but only the 21 components not related by changing the order of vector components in the product. Accordingly, instead of calculating both $s_1^x \cdot s_2^y$ and $s_2^y \cdot s_1^x$, we only calculate one term, and project back from this symmetric space after the calculation.

We thus compute the projector in the symmetric space as

$$\rho_s(P) = \frac{1}{|G|} \sum_{g \in G} \chi_{\rho_s}(g)^\dagger \cdot \rho_s(g), \tag{9}$$

Figure 2. Illustration of a two-spin invariant. The many possible interactions between two spins on a lattice contributing to a tensor can be decoded into the symmetric invariants that the representation of the tensor induces on the point group of the lattice. It is instructive to visualize the invariants in the space of two suitable variables v_1 and v_2, for example the sum and difference of the azimuthal angles of the spins on two inequivalent sites of a honeycomb lattice $\theta_+ = \frac{\theta_A + \theta_B}{2}$ and $\theta_+ = \frac{\theta_A - \theta_B}{2}$. These two variables can be a good choice for a two atom system, since they correlate with the two order parameters of the magnetic configuration, which are the ferromagnetic and staggered antiferromagnetic moment $\vec{s}_{\text{FM/AFM}} = \frac{\vec{s}_A \pm \vec{s}_B}{2}$.

where by ρ_s we denote the representation in the symmetric space corresponding to product space of order o. The symmetric invariants are now given by the eigenvalues of the invariant subspaces of this projector equal to one, hence we diagonalize the projector to find a total of m such spaces, where m is the multiplicity. Consequently, we find a number of m such invariants for each order of spin o, which we denote by the symbol $I_{i,j}$, where $i \in 0, \cdots, o$ and $j \in 0, \cdots, m$.

2.3. Distribution of spin samples for two atom system

We sample the magnetic moments uniformly on a sphere to cover the whole phase space offered by the magnetic configuration. The method used for this uniform sampling on a curved surface is called *sphere-point picking* [31], which avoids bunching of samples near the poles when simply drawing the spherical coordinates φ, θ from uniform distributions. For N moments, let u_i and v_i, $i = 1, \ldots, N$ be random variables on $(0, 1)$. Then

$$\begin{aligned} \varphi_i &= 2\pi\, u_i, \\ \theta_i &= \cos^{-1}\left(2v_i - 1\right), \end{aligned} \tag{10}$$

give the spherical coordinates with the azimuth angle φ in the xy-plane, $0 \leqslant \varphi < 2\pi$, and the polar angle θ from the positive z-axis, $0 \leqslant \theta \leqslant \pi$.

We can further augment the data computed with the TB model by applying the operations of the point group, see section 2.1, on the spins and the target. By taking a look at the character table of the group, the behavior of the target, belonging to a one-dimensional representation, can simply be deduced from the character: for the case of A_2, the target will pick up a sign corresponding to the character of the operation in this representation. We therefore calculate the spin configurations resulting from applying the group operations and apply the sign to the target accordingly. Additionally utilizing the known behavior of the tensor component under time reversal symmetry, we can effectively scale the size of the dataset by a factor of thirteen without the need for further TB calculations, with eleven nontrivial group operations in C_{6v} and the time reversal symmetry.

2.4. Chirality and lattice site exchange

In our following analysis, the notion of chirality—specifically the vector chirality $\vec{\xi}$ defined as a vector product between spins on two inequivalent sites of the honeycomb lattice, $\vec{\xi} = \vec{s}_A \times \vec{s}_B$, see figure 2—and corresponding symmetric and antisymmetric in $\vec{\xi}$ contributions will play a central role.

In order to disentangle chiral and nonchiral contributions to the AHE, we (anti)symmetrize the data obtained from tight-binding calculations with respect to the rotational sense, or vector chirality $\vec{\xi}$, of the

canted magnetic configuration. This symmetrization is achieved by exchange of the atomic sites *A* and *B*, as can be easily seen from the definition of the vector chirality $\vec{\xi}$, which is antisymmetric under this operation:

$$\vec{\xi} = \vec{s}_A \times \vec{s}_B = -\left(\vec{s}_B \times \vec{s}_A\right). \tag{11}$$

We therefore perform the following decomposition of the anomalous Hall conductivity tensor $\sigma_{\alpha\beta}$—or *any* given (scalar or tensor) quantity of interest for that matter—into even (non-chiral, σ^{nc}) and odd (chiral, σ^c) parts with respect to the vector chirality:

$$\sigma_{\alpha\beta}^{c(nc)} = \frac{\sigma_{\alpha\beta}\left(\vec{\xi}\right) \mp \sigma_{\alpha\beta}\left(-\vec{\xi}\right)}{2}. \tag{12}$$

While in this work, we focus predominantly on the non-chiral part of the AHC, by separating chiral and non-chiral parts, we aim at distinguishing chiral and non-chiral terms in the symmetrically invariant expansion as well, since this construction allows us to fit models which use $\vec{\xi}$-even or $\vec{\xi}$-odd features respectively. Consequently, we can write the equation for the fitted model as follows:

$$\sigma^{c(nc)} = \vec{c}_{c(nc)} \cdot \vec{I}_{c(nc)}, \tag{13}$$

where we construct the $\vec{\xi}$-even or $\vec{\xi}$-odd invariants in the exact same way as above:

$$\vec{I}_{c(nc)} = \frac{\vec{I}\left(\vec{\xi}\right) \mp \vec{I}\left(-\vec{\xi}\right)}{2}. \tag{14}$$

Accordingly, we make use of the symmetric and antisymmetric angles θ_+ and θ_-, defined as $\theta_+ = \frac{\theta_A + \theta_B}{2}$ and $\theta_- = \frac{\theta_A - \theta_B}{2}$, to visualize the symmetric invariants and anomalous Hall effect and their dependence on the magnetic configuration. These two variables can be an especially good choice of latent variables for the two atom system, since they correlate with the two order parameters of the magnetic configuration, which are the ferromagnetic and staggered antiferromagnetic moment $\vec{s}_{\text{FM/AFM}} = \frac{\vec{s}_A \pm \vec{s}_B}{2}$.

2.5. Calculation of the anomalous Hall conductivity

Given specific electronic structure, we calculate the transverse anomalous Hall conductivity at zero temperature using the Kubo formalism, which allows us to take into account the effect of disorder in the system on the conductivity tensor. In order to do so, we replace the retarded and advanced Green functions G_0 of the perfect crystal by the full Green function $G = \frac{1}{G_0^{-1} - \Sigma}$, where $\Sigma(E, \vec{k})$ is the self-energy representing the effect of disorder. Here, we use a constant broadening model such that $\Sigma(E, \vec{k}) = -i\Gamma$. With the constant broadening Γ we obtain a Green function diagonal in the eigenspace of the Hamiltonian:

$$G^{R/A}\left(E, \vec{k}\right)_{mn} = \frac{\delta_{mn}}{E - \epsilon_{n\vec{k}} \pm i\Gamma}, \tag{15}$$

where $\epsilon_{n\vec{k}}$ are the single-electron eigenenergies. The antisymmetric part of the conductivity tensor, which can be expressed in terms of $G^{R/A}$ [3], splits into two contributions:

$$\sigma_{[\alpha\beta]}^I = -\frac{1}{2\pi} \int \frac{d^3k}{(2\pi)^3} \sum_{\substack{mn \\ m \neq n}} \Im\left\{v_{mn}^\alpha\left(\vec{k}\right) v_{nm}^\beta\left(\vec{k}\right)\right\} \frac{\left(\epsilon_{m\vec{k}} - \epsilon_{n\vec{k}}\right)\Gamma}{\left(\left(E_f - \epsilon_{m\vec{k}}\right)^2 + \Gamma^2\right)\left(\left(E_f - \epsilon_{n\vec{k}}\right)^2 + \Gamma^2\right)} \tag{16}$$

and

$$\begin{aligned}
\sigma_{[\alpha\beta]}^{II} = -\frac{1}{\pi} \int \frac{d^3k}{(2\pi)^3} \sum_{\substack{mn \\ m \neq n}} \Im\left\{v_{mn}^\alpha\left(\vec{k}\right) v_{nm}^\beta\left(\vec{k}\right)\right\} &\frac{\Gamma}{\left(\epsilon_{m\vec{k}} - \epsilon_{n\vec{k}}\right)\left(\left(E_f - \epsilon_{m\vec{k}}\right)^2 + \Gamma^2\right)} \\
&- \frac{1}{\left(\epsilon_{m\vec{k}} - \epsilon_{n\vec{k}}\right)^2} \Im\left\{\ln\left(\frac{E_f - \epsilon_{m\vec{k}} + i\Gamma}{E_f - \epsilon_{m\vec{k}} + i\Gamma}\right)\right\},
\end{aligned} \tag{17}$$

where α and β are the Cartesian indices and E_f is the Fermi energy. We refer to $\sigma_{\alpha\beta}^I$ as the Fermi-surface term in the anomalous Hall conductivity (AHC), since it only picks up contributions from the Fermi surface. The term $\sigma_{\alpha\beta}^{II}$ collects terms from all occupied states up to the Fermi level and is therefore referred to as the Fermi-sea term in the AHC. In evaluating the Kubo expressions for the conductivity of the two-atom system we have used roughly 65×10^4 *k*-points to perform the Brillouin zone integrals and a value of 25 meV for the broadening parameter Γ.

2.6. Algorithm layout

In this work, we aim at fitting a linear model of the symmetric invariants $I_{o,j}$ to the target y,

$$y = \sum_{o}^{o_{\max}} \sum_{j=0}^{n_o} c_{o,i} \cdot I_j^o, \tag{18}$$

where $I_{o,j}$ indicates the jth invariant out of a total of n_o invariants in the order o, up to a maximal order $o_{\max}$, and $c_{o,j}$ is the corresponding fitted model coefficient. Stacking all invariants and coefficients into vectors in increasing order, $\vec{x} = [x_0^0, \ldots, x_{n_0}^0, x_0^1, \ldots, x_{n_1}^1, \ldots, x_0^{o_{max}}, \ldots, x_{n_{o_{\max}}}^{o_{\max}}]$, we can write:

$$y = \vec{c} \cdot \vec{I}. \tag{19}$$

The symmetric invariants are possibly correlated, as are the features in many model selection tasks. Further, most of the common algorithms are optimized towards working with normal distributions with unit variance and zero mean as inputs. Therefore, we need to perform a number of tasks in order to obtain a robust feature space as a solid basis for selecting the model.

First and foremost, the dataset is split into training and test set to avoid training a model which just repeats the labels (overfitting). The training pipeline, see black dashed box in figure 3, consists of a number of preprocessing steps (light blue dashed box) and the model selection step. The preprocessing involves scaling features to unit variance and zero mean (standardization), see section 2.6.1, and decorrelation and feature selection by principal component analysis (PCA), see section 2.6.2, as well as a variance threshold removing all features with variance below a given threshold and standardization after PCA. We then perform the model selection step by training regularized models on the features selected by the PCA method, see section 2.6.5. We take care to avoid data leakage along the pipeline: if we perform e.g. the PCA on the whole dataset before splitting into train and test set, the prediction of the model will be overly optimistic. Therefore, the preprocessing transformations are trained on the training set only.

2.6.1. Standardization

The model selected by cross validation maps the input features $\vec{x}_m$ onto the single input label y_m via the model coefficients $\vec{c}_m$, where the M true features are obtained from N observations:

$$y_m = \vec{c}_m^T \cdot \vec{x}_m. \tag{20}$$

Input features and labels have been obtained from the true features and labels by standardization, subtracting the mean μ and dividing by the standard deviation, or scale, s, before training the model:

$$y_t^j = s_y \cdot y_m^j + \mu_y, \tag{21}$$

$$\vec{x}_t^j = \mathcal{D}_{[\vec{s}_x]} \cdot \vec{x}_m^j + \vec{\mu}_x, \tag{22}$$

$$\vec{\mu}_x = \frac{1}{N} \sum_j \vec{x}_t^j = \frac{1}{N} \sum_j \left(\mathcal{D}_{[\vec{s}_x]} \cdot \vec{x}_m^j + \vec{\mu}_x \right), \tag{23}$$

$$\mu_y = \frac{1}{N} \sum_j y_t^j = \frac{1}{N} \sum_j \left(\vec{c}_t^T \vec{x}_t^j \right) = \vec{c}_t^T \vec{\mu}_x, \tag{24}$$

where $\mathcal{D}_{[\vec{s}_x]}$ is the matrix containing the standard deviation of each true feature on the diagonal, $\mathcal{D}_{[\vec{s}_x]}^{ii} = s_x^i$, with all other entries equal to zero. The vector $\vec{\mu}_x$, contains the mean values of each true feature, and the summation index j runs over the samples.

2.6.2. Principal component analysis

Further, the standardized features are decorrelated by using *principal component analysis* (PCA) [32]. This method finds the singular value decomposition of the feature matrix, which is in essence a generalization of finding the diagonal of the feature matrix. The method of singular value decomposition, on which PCA is based, relies on a following geometric consideration [33]. Considering a unit sphere in n dimensions, the image of this unit sphere under any $m \times n$ matrix is a hyperellipse. A hyperellipse is the m dimensional equivalent of the two-dimensional ellipse, obtained by stretching the two-dimensional unit circle by some factors $\sigma_1, \ldots, \sigma_m$ along some orthogonal directions $u_1, \ldots, u_2 \in \mathbb{R}$, which are, for convenience, normalized to unit length, $||u_i||_2 = 1$. We refer to the vectors $\sigma_i u_i$ as the principal semiaxes of the hyperellipse, with length σ_i, and for a matrix A of rank r, exactly r of these lengths turn out to be nonzero. In particular, if $m \geqslant n$, at most n of the will be nonzero.

With this geometric picture in mind, let us define the singular value decomposition (SVD) of $A \in \mathbb{C}^{m \times n}$: let m and n be arbitrary natural numbers $\geqslant 0$, then the SVD of A is defined as:

$$A = U \Sigma V^*, \tag{25}$$

where

$$
\begin{aligned}
& U \in \mathbb{C}^{m \times m} \text{is unitary}, \\
& V \in \mathbb{C}^{n \times n} \text{is unitary}, \\
& \Sigma \in \mathbb{R}^{m \times n} \text{is diagonal}. \tag{26}
\end{aligned}
$$

Additionally, we assume the diagonal entries σ_j of Σ to be non-negative and ordered in nonincreasing order, $\sigma_1, \geqslant \ldots, \sigma_p \geqslant 0$, where $p = \min(m,n)$. Looking at this definition, above statement is made quite clear: the unit sphere $S \in \mathbb{R}^n$ is transformed to the hyperellipse $SA \in \mathbb{R}^m$, since the unitary map V^* preserves the sphere, the diagonal matrix Σ stretches it and the unitary matrix U rotates and reflects the ellipse. For a proof, that every matrix A has such an SVD, see also reference [33].

The usefulness of the SVD for machine learning problems can now be illustrated by investigating the diagonal matrix Σ. Considering the SVD of a feature matrix of shape $m \times n$, constructed from stacking the vectors of n features for all the m samples, the matrix Σ is potentially very large, especially for problems with many samples, many features, or both. However, two characteristics of this matrix make the SVD worth calculating. On one hand, the fact that Σ is diagonal implies that the corresponding singular vectors are orthogonal, or in statistical terms, uncorrelated. On the other hand, the number of non-trivial singular values is not given by the dimensions m or n, but by the rank r of the matrix, which can be significantly smaller than m and n for features that are correlated. Accordingly, for matrices with rank $r < n, m$, the resulting feature space is potentially much smaller than the original feature space.

In the context of machine learning, methods involving singular value decomposition are referred to as principal component analysis (PCA) methods, where the principal components are equivalent to the coordinates of the principal axes, and the measure of explained variance corresponds to the singular value, or length, of the principal axes. This correspondence lends itself to the analysis of variance in the feature space, as the singular values, as the norm of the singular axis they correspond to, determine how much overall feature variance a change along a certain axis induces.

2.6.3. Variance threshold

Features with low variance will very likely not have a large influence on the model target. To compare feature variances in a meaningful way, we first scale all the features to the range $[0,1]$ individually by using a MinMaxScaler. We then discard features with a variance below a certain threshold. This allows us to reduce the dimension of the feature space with very small effort. The threshold for the data shown in this work is $v_t = 10^{-5}$.

2.6.4. Feature selection with statistical methods

Since the modeling technique introduced in this work relies on an expansion agnostic of any physical intuition regarding the presence of certain interactions like exchange or spin–orbit coupling in the model Hamiltonian equation (1), condensation of the input feature space to the relevant components is a key ingredient for meaningful modeling of the electric transport.

To this end, several techniques can be utilized in conjunction with the regularized regression algorithm, which prefers a sparse description of the target by penalizing models with many nonzero coefficients. While the PCA-transformation can be used to extract portions of the latent space explaining overall variance of the feature space, we suggest to use a statistical measure which incorporates also the correlation with the target variable.

Specifically, we focus here on a technique estimating the linear correlation between single input features and the target variable, the *f-regression test*, since the working hypothesis is that the electric transport can be described by a linear model of the input features.

This quantity measures the cross-correlation between two variables x and y over N observations, given by the Pearson correlation coefficient:

$$
\begin{aligned}
R &= \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} \\
&= \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2} \sqrt{\sum_i^N (y_i - \bar{y})^2}}. \tag{27}
\end{aligned}
$$

Internally, the Pearson correlation coefficient is converted to an F-score (a number between 0 and 1). Computing the f-regression score for each component and ranking the components according to this metric, we can select a top scoring percentile to fit the model and disregard features scoring below a threshold. This approach nicely complements the PCA transformation step, which ranks the constructed components by explained variance ratio in the feature space. The f-regression score presents a systematic way to assess the importance of input features for the model based on their correlation with the target, which the PCA transformation is not able to do.

2.6.5. Model selection by regularized regression
We regularize this model by the *elastic net* penalty [34]. The regularized loss-function $F(w)$ to minimize is therefore the sum of loss-function $L(\vec{c}^T \cdot \vec{x}, y)$ and regularization term $R(\vec{c})$:

$$
\begin{aligned}
F(c) &= L\left(\vec{c}^T \cdot \vec{x}, y\right) + \alpha \cdot R(\vec{c}) \\
&= \frac{1}{2N} \cdot ||y - \vec{I} \cdot \vec{c}||_2^2 + \alpha \cdot l_1 ||\vec{c}||_1 + \frac{1}{2}\alpha \cdot (1 - l_1) \, ||\vec{c}||_2^2,
\end{aligned}
\tag{28}
$$

where N is the number of samples, and α is the parameter controlling the strength of regularization, with $\alpha = 0$ corresponding to an ordinary least squares (OLS) fit. The OLS fit can be adapted to an *squared-epsilon-insensitive* loss, by allowing the model to ignore errors below a certain threshold of ϵ, and applying the standard squared loss above the threshold. This adapted error function allows for slightly improved description of outliers in some cases. We use a value of $\epsilon = 10^{-2}$ here.

The parameter l_1 is the ratio interpolating between two kinds of regularization related to a norm on the weights or coefficients of the model, LASSO and RIDGE, by L_1-norm ($l_1 = 1$, LASSO) or by L_2-norm ($l_1 = 0$, RIDGE). While the RIDGE regression penalizes the absolute size of coefficients, LASSO penalizes the number of nonzero coefficients. In general, RIDGE regression gravitates towards models where the size of all coefficients is optimal. In contrast, LASSO prefers sparse models, where the number of nonzero coefficients is optimal.

Regularization as a mix of the two methods helps us to find a sparse model, with the optimal number of nonzero coefficients, while keeping the size of coefficients in check, and is also commonly referred to as *compressive sensing*. The models are ranked by a scoring method of choice, in this case the coefficient of determination, R^2, calculated from the residual sum and total sum of squares:

$$
u = \sum_i^N \left(y_{\text{true}}^i - y_{\text{predict}}^i\right)^2,
$$

$$
v = \sum_i^N \left(y_{\text{true}}^i - \bar{y}_{\text{true}}\right)^2.
$$

$$
R^2 = \left(1 - \frac{u}{v}\right).
\tag{29}
$$

Here, for a sample numbered by the index i, we denote the true value of the target by y_{true}^i and the model prediction for the sample by y_{predict}^i. The mean of all true target values is calculated as $\bar{y}_{\text{true}} = 1/N \sum_j^N y_{\text{true}}^j$, where N is the total number of samples.

The best possible score is 1.0, and the score can be negative, since the model can be arbitrarily bad. The score on either train or test set is not our only measure of the models fidelity: we can additionally probe the functional dependence along specific paths in the feature space. This allows us to quantify not only statistical quantities like the MSE, but also to probe how faithfully the model reproduces the functional dependence of the target along certain paths or in certain regions, for example varying only the polar angle θ on an arc starting from the north pole, $\theta \in \{0, \pi\}$, keeping the rest the same. Especially when the model extrapolates from one region, overfitting in this regard must be closely monitored.

2.6.6. Stochastic gradient descent regression
The regressor is the tool that minimizes the loss function defined by the regularization, elastic net defined in equation (28) in this case, on the given training data. We use the Stochastic Gradient Descent (SGD) regressor here, which is an especially effective optimization tool for large quantities of training data, usually $>10^4$ samples. SGD estimates the true gradient of the loss function on one training sample, as opposed to other methods like batch gradient descent, and updates the weights along the gradient. One pass over the

whole training set is referred to as one iteration here. After an estimate of the gradient $\partial F(\vec{c})/\partial \vec{c}$ has been computed on one sample, the model coefficients, or weights, are updated as follows:

$$\vec{c} \leftarrow \vec{c} - \mu \frac{\partial F(\vec{c})}{\partial \vec{c}}$$

$$\vec{c} \leftarrow \vec{c} - \mu \left[\alpha \frac{\partial R(\vec{c})}{\partial \vec{c}} + \frac{\partial L\left(\vec{c}^T \cdot \vec{x}, y\right)}{\partial \vec{c}} \right], \tag{30}$$

where by $R(\vec{c})$ we refer to the regularization term of the elastic net and by $L(\vec{c}^T \cdot \vec{x}, y)$ we refer to the loss-function, as defined in equation (28), and μ is the learning rate.

2.6.7. Hyperparameter optimization via cross validation

SGD is influenced by a number of hyperparameters, which we choose based on a cross-validation search. This method is used to optimize the values of hyperparameters for a particular algorithm by repetitively running the algorithm for different parameter values and thereby determining the highest scoring point in the parameter space. Note that the score here does not necessarily need to be the test score of the resulting model found by the algorithm, but can be any metric supplied by the user. Some obvious examples would be to optimize for minimal runtime, optimal model score, or even a trade off between the two. We make use of this tool to finetune the regularization parameter α_{reg}, and to find the optimal l_1-ratio between Lasso and Ridge in the elastic net, as well as the optimal learning rate μ. In general, different loss functions can be probed in order to better handle outliers or zero inflation, like the *squared epsilon insensitive* loss function, which ignores errors below a given epsilon.

In order to achieve the highest efficiency for the cross validation method, we employ the *Halving Grid Search* algorithm. A given set, or grid, of parameters, serves as candidates. All candidates are evaluated in the first pass with a limited amount of resources, and a top fraction of the candidates with respect to a scoring method of the users choice is kept while the rest is discarded. The surviving candidates are then evaluated again with an increased amount of resources and the process is repeated. While the number of candidates carried over to the following iterations is divided by factor *s*, the amount of resources is multiplied by the same factor *s*.

A hyperparameter search using this method is especially efficient, since not all candidates are evaluated on equally large amounts of resources, in this case training samples. In order to avoid overfitting, the algorithm employs cross validation as well by using the *k*-fold strategy, where the available resources are split into *k* consecutive folds, each of the used as a validation set once while all other folds form the training set.

2.6.8. Backtransformation of model coefficients

The model coefficients are not the true physical coefficients connecting true physical features and labels. However, to interpret the trained model in physical terms, we require the coefficients in the physical feature space:

$$y_t = \vec{c}_t^T \cdot \vec{x}_t. \tag{31}$$

A mapping between model coefficients and true coefficients is found by applying the inverse transformation along the pipeline to the coefficients. This means inserting equations (20) and (21) from section 2.6.1 into equation (31), and multiplying by $\vec{x}_m^T \cdot \mathcal{D}_{[\vec{s_x}]}^{-1}$, to arrive at the coefficients for the PCA-transformed features $\vec{x}_p$:

$$\vec{c}_p^T \cdot \vec{x}_p = s_y \cdot \left(\vec{c}_m^T \cdot \vec{x}_m\right) + \mu_y$$

$$\Rightarrow \vec{c}_p^T \cdot \mathcal{D}_{[\vec{s_x}]} \cdot \vec{x}_m = s_y \cdot \left(\vec{c}_m^T \cdot \vec{x}_m\right) + \mu_y - \vec{c}_t^T \vec{\mu}_x$$

$$\Rightarrow \vec{c}_p = s_y \cdot \left(\vec{c}_m^T \cdot \mathcal{D}_{[\vec{s_x}]}^{-1}\right). \tag{32}$$

Then, applying the inverse transformation of the PCA, we arrive at the true coefficients of the original symmetric invariants:

$$\vec{x}_t = P^{-1}\left(\vec{x}_p\right). \tag{33}$$

We can therefore extract the physical parameters from the trained model by multiplying with the standard deviation of the target and dividing by the standard deviation of the feature to reverse a standardization step or by applying the inverse transform of the exactly invertible transformations along the pipeline, such as PCA.
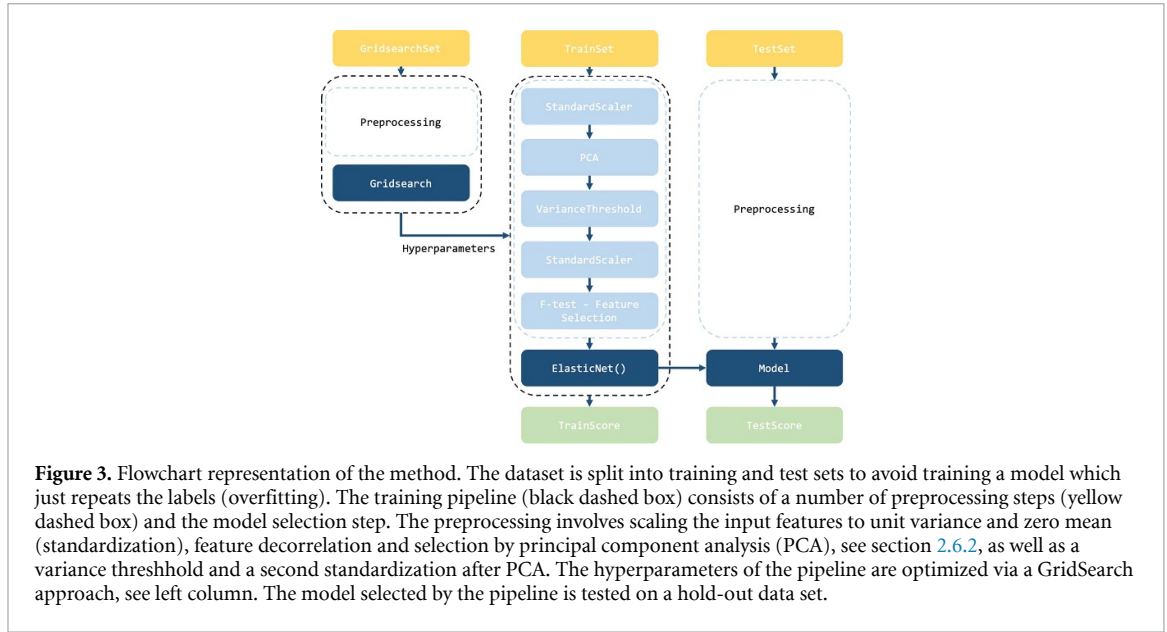
Figure 3. Flowchart representation of the method. The dataset is split into training and test sets to avoid training a model which just repeats the labels (overfitting). The training pipeline (black dashed box) consists of a number of preprocessing steps (yellow dashed box) and the model selection step. The preprocessing involves scaling the input features to unit variance and zero mean (standardization), feature decorrelation and selection by principal component analysis (PCA), see section 2.6.2, as well as a variance threshhold and a second standardization after PCA. The hyperparameters of the pipeline are optimized via a GridSearch approach, see left column. The model selected by the pipeline is tested on a hold-out data set.
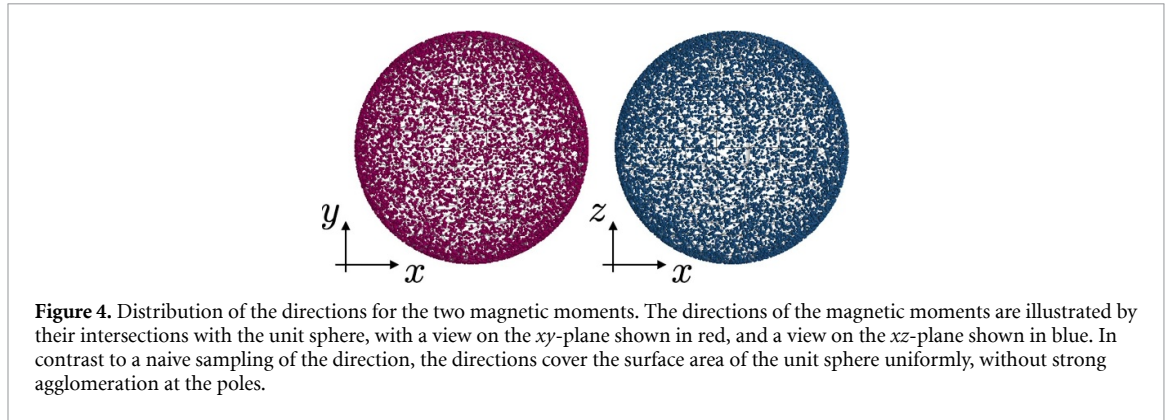
Figure 4. Distribution of the directions for the two magnetic moments. The directions of the magnetic moments are illustrated by their intersections with the unit sphere, with a view on the *xy*-plane shown in red, and a view on the *xz*-plane shown in blue. In contrast to a naive sampling of the direction, the directions cover the surface area of the unit sphere uniformly, without strong agglomeration at the poles.

3. Results

3.1. Input data

By making use of the method presented in section 2.3, we generate a total number of 10^4 samples, each describing the two magnetic moments on sites *A* and *B* of the bipartite honeycomb lattice. The moments point in randomly chosen directions, such that the overall distribution of directions is uniform over the unit-sphere, see figure 4. The anomalous Hall conductivity, which represents the target quantity, is numerically calculated by using the Kubo approach detailed in section 2.5. The calculation provides the target value for different values of the Fermi energy, which allows the analysis of the fitted models with respect to changes when probing the transport at different points in the electronic structure, characterized by different regimes of reciprocal geometry.

An explicit implementation of the symmetry expansion, outlined in section 2.2, provides input features to the modeling pipeline. Since the anomalous Hall conductivity tensor contains an odd number of spins, the even expansion terms for the anomalous Hall conductivity tensor vanish. In principle, objects of arbitrary order in spin can appear in the model. For the sake of practicability, we calculate the first six non-vanishing orders of the expansion, namely orders $\{1, 3, 5, 7, 9, 11\}$ with $\{1, 7, 26, 76, 185, 392\}$ terms respectively, representing a complete description of possible interaction terms up to order eleven in spin, containing 687 different terms. We clarify specifically, that higher orders in spin do not incorporate a higher number of distinct magnetic moments. The order in spin corresponds plainly to the power of the image, which is the pseudovector of dimension *d* introduced in section 2.2, representing the magnetic state of the system.

In figures 5 and 6 we present an overview of the symmetric invariants calculated for the two-atom system. They show the original features symmetrized and antisymmetrized with respect to vector chirality $\vec{\xi}$ respectively, the panels show one invariant each as a function of azimuth angles θ_+ and θ_-, sorted by
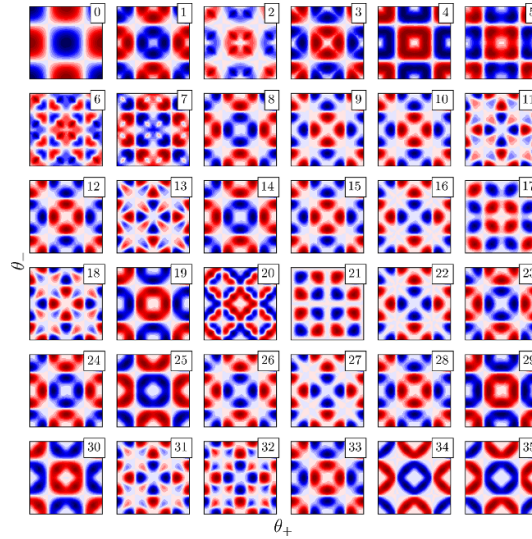
Figure 5. Invariants symmetrized with respect to exchange of atomic sites. Invariants are shown one per panel, with order increasing from top left to bottom right, as functions of θ_+ and θ_-, with both variables spanning the range $[0, \pi]$. The center of the plot is located at the coordinates $(0, 0)$. In this parameter space the invariants show distinct features, which allows to categorize them into a handful of classes.
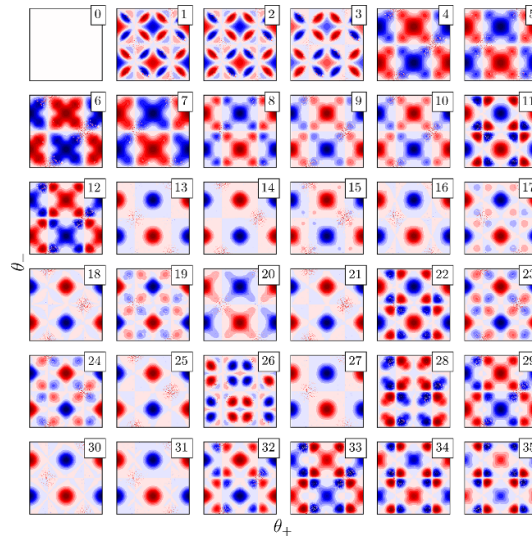

Figure 6. Invariants antisymmetrized with respect to exchange of atomic sites. Invariants are shown one per panel, with order increasing from top left to bottom right, as functions of θ_+ and θ_-, with both variables spanning the range $[0, \pi]$. The center of the plot is located at the coordinates $(0, 0)$. For the discussion of morphologies, see text.

increasing expansion order from top left to bottom right. Inspecting the data shown in figure 5, the invariants reveal their structure as functions of the spherical coordinates. All $\vec{\xi}$-even invariants share the more or less sharp nodal lines in both horizontal and vertical direction, at coordinates $\theta_+ \in n \cdot \pi + \pi/2, n \in \mathbb{N}$ or $\theta_- \in n \cdot \pi + \pi/2, n \in \mathbb{N}$. Between these nodal lines, blocks with alternating sign are present, which grow in complexity for higher order, displaying multiple sign changes (panel 6), polar features (panel 16) and fine modulations (panel 11). Investigating the data presented in the first panel, displaying the most simple structure, the correlation with the overall magnetization, or sum of the two magnetic moments of the system, is evident. This in turn promotes the idea, that this first invariant should be assigned large weight in the description of the anomalous Hall effect, since this phenomenon is conventionally associated with the overall magnetization. Invariants beyond the 36 shown here display quite similar morphology, with the same building blocks repeating with slight variations.

The interpretation of these patterns becomes more intuitive when associating the values of $\theta_\pm$ with specific magnetic configurations. In the center of the panels, at coordinates $(0, 0)$, we find the standard

ferromagnetic alignment of moments along the *z*-axis in positive direction. Traversing the panel in the horizontal direction corresponds to a collective rotation of both spins around an axis, with the nodal lines indicating the moments passing through the *xy*-plane and the AHE consequently vanishing. A similar explanation holds for the vertical direction, and traversing along this axis corresponds to a rotation by equal amounts in different directions per site, which we refer to as spin canting. However, the nodal lines along the θ_--axis are not associated with the moments passing through the *xy*-plane, but rather with the moments falling into an antiferromagnetic alignment. When either rotation or canting angle has reached the value of π, the magnetic configuration has been reversed to the ferromagnet aligned with the *z*-axis in negative direction.

If we imagine now, that we only rotate one of the spins around an axis in the *xy*-plane, then both θ_+ and θ_- will increase or decrease with the same rate. In this way, depending on the rotation direction, we reach the four nodes placed at the corners of the tiles. These correspond to the antiferromagnetic alignment along the *z*-axis, with one spin aligned in positive and negative *z*-direction each. As can be easily imagined, traversing from here in the θ_+ direction mimics the collective rotation of the antiferromagnet around an axis in the *xy*-plane, with the nodal lines again representing the position where the moments pass through this plane.

Focusing now on the $\vec{\xi}$-odd invariants shown in figure 6, the most obvious feature of the data presented here is that the first invariant is exactly zero everywhere. Considering the fact that this invariant is the only one of first order in spin, it is conceivable that this invariant is related to the overall magnetization. As the magnetization, or ferromagnetic moment of the configuration, is insensitive to exchange of the lattice sites, the antisymmetrization leads to a cancellation. As is the case of $\vec{\xi}$-symmetric invariants, there are dominant nodal lines, however at different coordinates given by $\theta_+ \in n \cdot \pi + \pi/2, n \in \mathbb{N}$ or $\theta_- \in n \cdot \pi, n \in \mathbb{N}$. These nodal lines transform into quite different morphologies in higher order, namely the kind similar to the data shown in panel 11, with quadrupoles surrounding the critical points of $(\theta_+, \theta_-) \in (n \cdot \pi + \frac{\pi}{2}, m \cdot \pi))$, $n \in \mathbb{N}, m \in \mathbb{N}$ and isotropic features at coordinates $(\theta_+, \theta_-) \in (n \cdot \pi, m \cdot \pi + \frac{\pi}{2})), n \in \mathbb{N}, m \in \mathbb{N}$. Additionally, only the isotropic characteristics are present, e.g. in panels $11, 13, 14, 15, 16$ at points shifted by $\pi/2$ from the nodal points mentioned before.

Referring to explicit magnetic arrangements, we can relate the characteristics of the invariants to specific properties of the magnetic configuration. The horizontal nodal line, visible in all of the antisymmetric invariants, marks the ferromagnetic configurations with $\theta_- = 0$, connected by collective rotation of spins. The two vertical nodal lines then correspond to the arrangements with ferromagnetic moment rotated into the *xy*-plane. In contrast to the symmetrized data, the antisymmetrized components are all exactly vanishing for perfect ferromagnetic alignment of spins, but not for the antiferromagnetic arrangement, with the exception of coordinates $(\pm \pi/2, \pm \pi/2)$. These coordinates mark an antiferromagnetic arrangement in the *xy*-plane. The isotropic features, appearing at coordinates $(\theta_+, \theta_-) \in (n \cdot \pi, m \cdot \pi + \frac{\pi}{2})), n \in \mathbb{N}, m \in \mathbb{N}$, correspond to the perfect antiferromagnetic configurations along the *z*-axis, perpendicular to the lattice plane.

Finally, the dataset is augmented by applying the lattice symmetries, as detailed in section 2.1. From each of the spin configurations, 12 additional configurations can be generated, one for each nontrivial group operation and one for the time reversal operation. Multiplication with the operations character in the representation yields the corresponding values of the anomalous Hall conductivity tensor and the symmetric invariants from the original values.

In conclusion, the modeling pipeline works on a set of 13×10^4 samples—obtained from augmenting the 10^4 initial TB calculations by exploiting the crystal symmetries and time reversal symmetry—where each sample is characterized by a magnetic configuration of $\vec{s}_A$ and $\vec{s}_B$, a corresponding value of the anomalous Hall conductivity, and a collection of 687 symmetric invariants. For training purposes, $3/4$ of these samples are used, while $1/4$ is held out as a set for gridsearch analysis. The training data are split into $3/4$ for training ($73\,125$ samples) and $1/4$ ($24\,325$) for testing. Before continuing with any further steps, the feature data are standardized by subtracting the mean and scaling by the variance, as illustrated in section 2.6.

3.2. PCA transformation

After standardizing the features according to section 2.6, a PCA transformation constructs a decorrelated and greatly condensed feature space, as clarified in section 2.6.2. This transformation is especially useful for the kind of data considered here, as the features show strong correlation, see off-diagonal terms in figure 7(a). Turning our attention to the data shown in figure 7(b), we realize that the correlation matrix assumes a block diagonal form after performing the transformation found by PCA. The matrix displays a square block in the bottom left corner, indicating that the PCA transformation finds a number of $\sim$110 uncorrelated components in the feature space. While the amount of off-diagonal weight in this matrix is visibly less than before, the off-diagonal terms are quite large, especially in higher orders, for feature indexes above 110. Closer inspection of the components in this region reveals that the explained variance of the components is
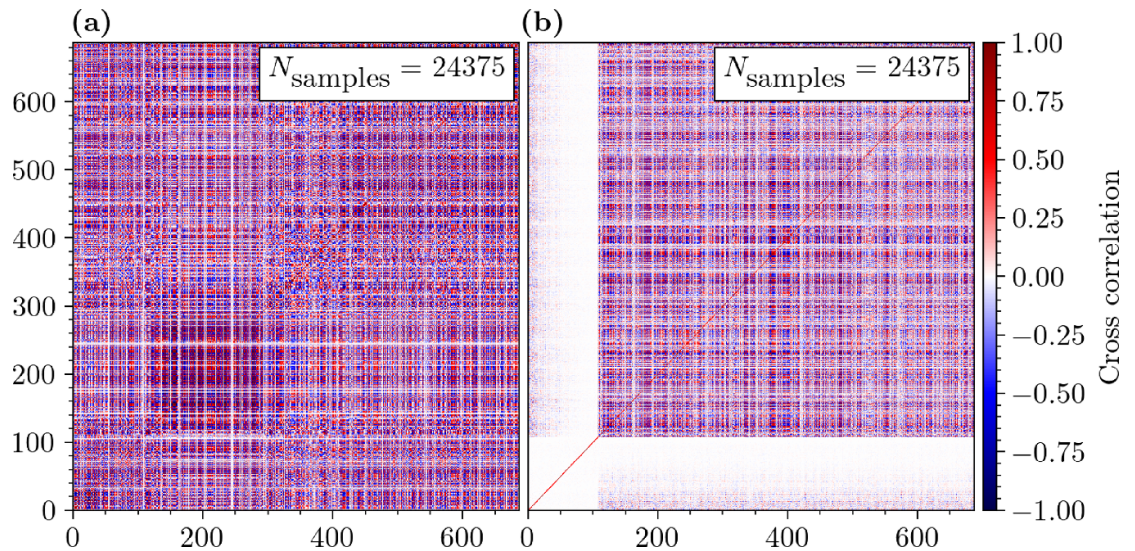
Figure 7. Cross-correlation matrix of the input features. (a) The correlation matrix of symmetric invariants symmetrized with respect to vector chirality $\vec{\xi}$. The cross-correlation is calculated over the test set, which contains 24 375 samples. Obviously, the symmetric invariants are, as already apparent from the similar structure in the space of (θ_+, θ_-), heavily correlated amongst each other. Perfectly uncorrelated features would result in a matrix with ones along the diagonal and zeros everywhere else. (b) The PCA transformation is trained on 73 125 samples and 24 375 samples are held out as a test set, shown in the top right corner. The correlation matrix assumes a block diagonal form after performing the transformation found by PCA.
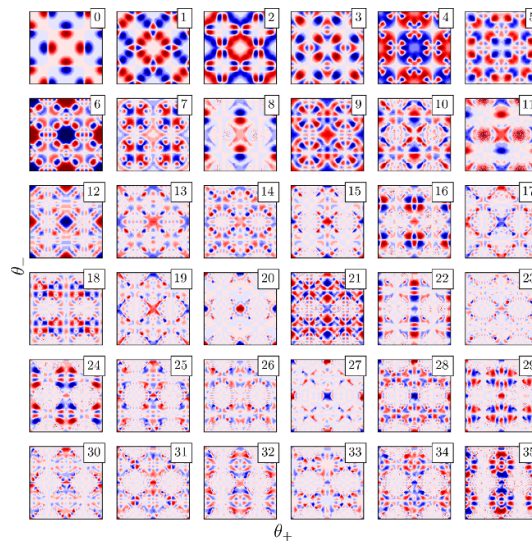
Figure 8. PCA-transformed invariants symmetrized with respect to exchange of atomic sites. The components are normalized to one in each panel, otherwise higher order features would not be visible due to their small scale. The transformation effectively merges by linear combination the multiple, quite similar symmetric invariants shown in figure 5, into a few relevant morphologies as the dominant directions in the resulting latent space with respect to explained variance. This allows to reduce the input feature space to a handful of relevant features.

orders of magnitude smaller than for the first components, and therefore the cross-correlation is inflated by the inverse dependence on the variance scale. Accordingly, we discard features beyond ~ 110 by applying a variance threshold after the PCA transformation.

When we focus now on the data shown in figures 8 and 9, we observe that the PCA transformation successfully reduces the dimensionality of the feature space, while keeping the characteristics of the different types of invariants intact and separated. Figures 8 and 9 display the features obtained by PCA-transformation from the $\vec{\xi}$-symmetric and $\vec{\xi}$-antisymmetric features as functions of the spherical coordinates $\theta_\pm$. The panels show one invariant each, sorted by decreasing explained variance from top left to bottom right. The principal components of the feature space scoring high in explained variance represent the mixtures of types that were identified in the original feature space. Most strikingly however, the polar structures, more associated with higher order invariants, score highest in explained variance. From the data presented in this section, it is
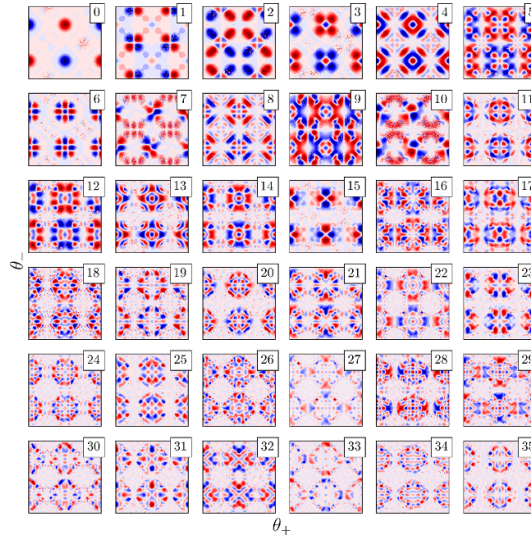
Figure 9. PCA-transformed invariants antisymmetrized with respect to exchange of atomic sites. Overall, the effect of the PCA is similarly drastic as in the case for the $\vec{\xi}$-symmetric features, leaving us with a handful of relevant features as input.
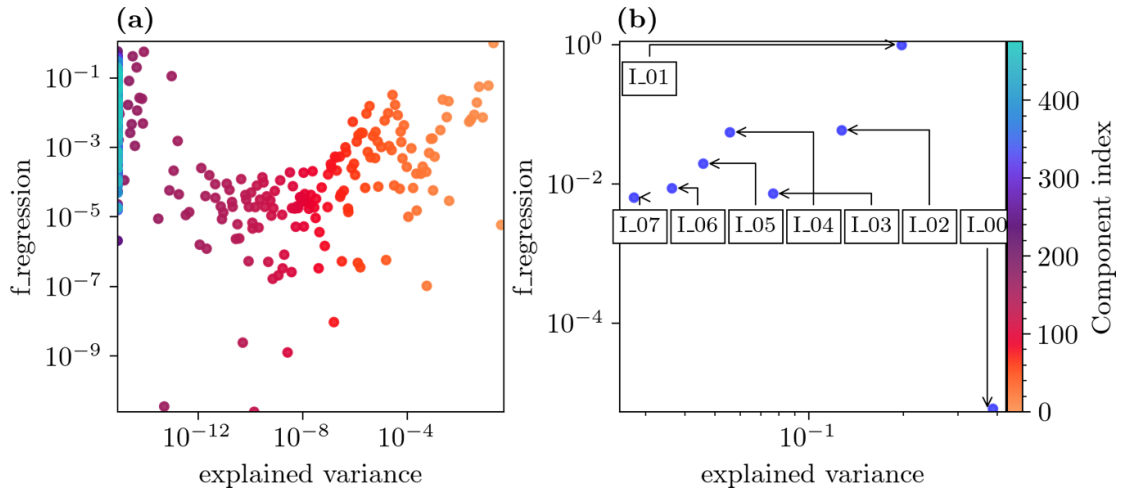

Figure 10. Explained variance ratio and f-regression score for the $\vec{\xi}$-symmetrized PCA components. (a) Explained variance ratio of the PCA components on a logarithmic scale on the *x*-axis and f-regression score on a logarithmic scale on the *y*-axis. The PCA transformation finds a number of nontrivial components with nonzero explained variance ratio, corresponding to the true rank of the feature matrix. The sharp step in explained variance around index 110 indicates the true rank, as the values for the explained variance ratio drop to zero for components beyond this index. Correspondingly, the data fall into two clusters, one with relevant influence on the feature space in the right cluster, and one with negligible influence in the left cluster, correspondingly expressed in the values of explained variance. Notably, the relationship between f-regression score and explained variance shows a linear trend in the right cluster, however the data points show sizable spread. The overall negligible variance and magnitude of components beyond the true rank of the feature matrix, i.e. data points in the left cluster, lead to inflated f-regression scores. We therefor disregard these components scores as uninformative. (b) Zoom in on the top right corner, showing the 8 components with highest explained variance and f-regression, which are the components I_0-I_7.

apparent that efficient dimensionality reduction can be achieved by the PCA-transformation, which additionally provides us with invertability (so we can easily transform back to the original space) and conceptual simplicity (since it is a linear transformation).

3.3. Feature selection with statistical methods

Investigating the f-regression scores presented in figure 10, indicating the correlation between single feature and target, we can form a clear picture of which input features are relevant for describing the electric transport. The entries are scattered on the *y*-axis against the values of explained variance on the *x*-axis, in order to relate the component's importance regarding feature space variance and target correlation respectively. On both axes, the data are shown on a logarithmic scale. Overall, the data are split into two clusters with respect to the explained variance: on the left hand side, redundant components accumulate at the limit of vanishing explained variance. As we are forcing the PCA transformation to construct as many

components as there are features, components beyond the true rank of the feature matrix are redundant, having zero explained variance and being constant, as we already established in section 2.6.2. Due to the vanishing variance of these components, the f-regression scores in this cluster are highly inflated.

On the right hand side, a cluster of possibly relevant components forms an almost linear trend on the double logarithmic scale. Components below the true rank of the feature matrix are assigned non-vanishing values of explained variance, which show some correlation with the size of the f-regression score. This correlation is a confirmation of the intuitive reasoning, that components with larger portions of explained feature space variance should be more relevant when determining the value of the target on the basis of the features. In contrast to this intuition, the data shows a sizable spread around the trend-line. Accordingly, components associated with higher-order characteristics of the feature space with lower overall variance, still show significant correlation with the target. Measuring the linear correlation of single features and the target is therefore yielding valuable information when performing modeling, which cannot be retrieved from feature analysis alone.

3.4. Model selection: anomalous Hall effect

Below, we present the results obtained from fitting linear models of the symmetric invariant expansion to the intrinsic part of the anomalous Hall effect (AHE) at a fixed value of Fermi energy $E_F = -1.5$ eV. To this end, for compactness, we focus only on the $\vec{\xi}$-symmetric non-chiral part of the AHE, noting that while the conclusions formulated below generally hold true also for the $\vec{\xi}$-odd part of the AHC, the much smaller average magnitude of the chiral AHC in our system makes the numerical analysis in the chiral case generally more cumbersome. In figure 11 we present the data fitted by two $\vec{\xi}$-even models with different number of PCA-transformed components, selected on the basis of f-regression score, in order to illustrate the importance of using all components up to the true rank of the feature matrix. While the larger model has access to the top 90% of components with respect to the f-regression test (blue crosses), the smaller model only uses the top 30% of components (red dots). This amounts to 99 and 33 components out of 110, respectively.

The first question to be addressed is whether the method used here is able to reproduce the training data with sufficient confidence. In order to rate the model performance we employ the well-known coefficient of determination R^2, defined in equation (29). The best score possible is 1, and a score of 0 would correspond to a model simply predicting the mean of the training data for every data point. From the correlation plots presented in figures 11(a) and (b) we can immediately see that both models perform reasonably well, with the large model scoring 0.99 (blue crosses) and the smaller model 0.97 (red dots) respectively. Obvious deviations from perfect linear correlation (black line) occur in the pockets above and below the point $(0,0)$ in the center of the plot for both models. Such deviations, ranging up to 0.6 for the model with less coefficients, with the target value being normalized to 1.0, would be a reason for concern, if they occurred for relevant portions of the test data. However, inspecting the distribution of residuals, shown in figure 11(c), proves that only a small number of samples deviate significantly from the center. Inspecting the variance of the residuals for the larger model reveals, that the deviations from the model approximately follow a normal distribution (blue line). From the indicated variance ranges (black dotted lines) it is apparent, that deviations greater than $3 \cdot \sigma \approx 0.18$ are quite rare. By comparing the blue and red distribution, the effect of restricting the model to less components is illustrated by the larger spread of the red distribution. More weight is allocated to the tails of the curve, corresponding to samples with larger difference between target and prediction value.

Turning our attention now to the model coefficients, presented in figure 11 (d), it becomes immediately obvious that the models are sparse. Utilizing the combination of PCA, statistical feature selection and regularization, the method is able to condense the input feature space, containing a total number of 687 features, to 33 and 99 features used in predicting the target value respectively. Allowing the model to use a larger fraction of the input features, as we have seen in discussing the distribution of residuals above, leads to a significant reduction in deviations of predictions from the target. However, the values of the coefficients used in both models do not change drastically, indicating stability of the models. Similarly, when removing small valued coefficients from the fitted models by hand (not shown here), which we call *feature annealing*, has a modest effect on the prediction. Depending on the threshold below which coefficients are annealed, the predictions are relatively stable. This indicates that the annealed coefficients are not relevant to the model and could be suppressed by adjusting the regularization parameter when fitting the model, see cross-validation approach in section 2.6.

A special remark can be made for the coefficients of components 108 and 109, appearing in figure 11(d) in 5th and 7th place. Keeping in mind that the coefficients are ordered from left to right by f-regression score, these two coefficients are extreme cases of a small explained variance ratio, but comparatively large correlation with the target. Investigating the influence of only these two coefficients on the model by hand (not shown here) reveals their small influence on the model, in spite of their high f-regression score ranking.
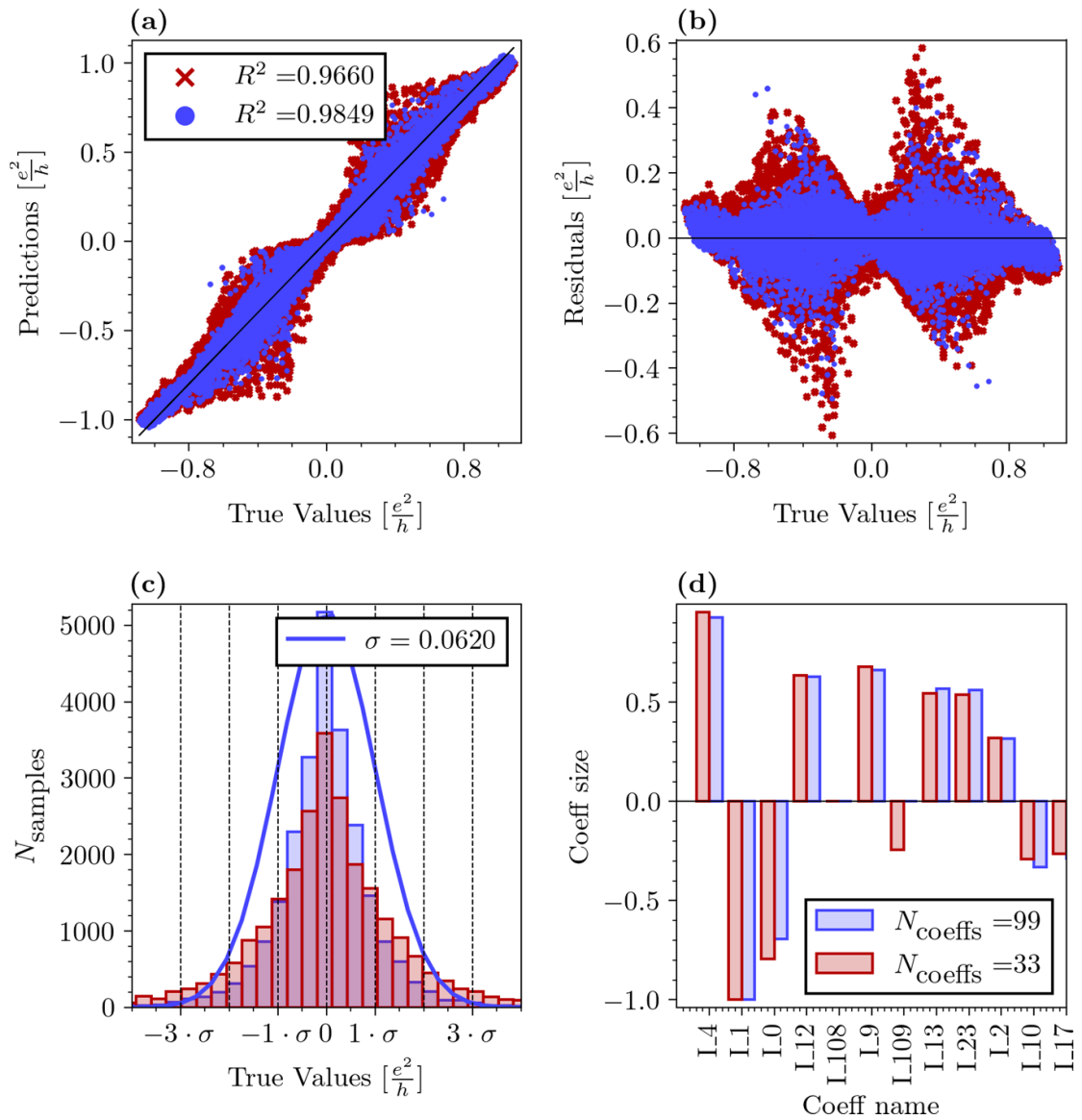
Figure 11. Fitting results for the anomalous Hall effect. The $\vec{\xi}$-even part of the anomalous Hall conductivity can be well described by a linear model of a few variables. The figure shows correlation plots between true value y_{true} and fitted target y_{fit} (a) as well as residual $y_{\text{res}} = y_{\text{true}} - y_{\text{fit}}$ (b), where the red line indicates perfect linear correlation. The distribution of residuals is given in panel (c). Panel (d) presents the coefficients of the model sorted by feature selection score, indicating the linear correlation of one feature and the target estimated by the f-regression test. The model score, shown in the top left corner of panel (a), indicates a strong performance of the model, as the top score that can be achieved is 1. Investigating the coefficients shown in panel (d), it is evident that effective condensation of the input feature space can be achieved by combining PCA, statistical feature selection and model regularization.

Annealing these two coefficients from the model leaves the model predictions essentially unchanged, as can already partly be seen in figure 11, where coefficient 109 is fitted to be zero in the smaller model (red bars). In contrast, removing a single coefficient of higher explained variance score, such as the coefficient I_9, has drastic effects on the model fidelity. Annealing this coefficient from the model increases the maximal residuals by a factor of 1.5, and—most importantly—the fraction of samples with residuals larger than the 3σ-range is more than doubled. We suspect, that the cutoff in the feature matrix referring to the principal values was chosen slightly too large, including the redundant invariants 108 and 109 in the model. In conclusion, f-regression score in conjunction with principal component analysis—taking special care in choosing the right number of components—is a reliable and effective metric for assessing the importance of single components for the model, controlling the overall prediction quality and number of outliers.

Considering that the input features do not contain any information about the electronic properties of the system explicitly, the discrepancy in prediction of the true value might be most probably explained by referring to the band structure, consisting of four bands only [11]. Tuning the direction of magnetic moments can result in drastic and abrupt changes in the position of the bands with respect to the Fermi
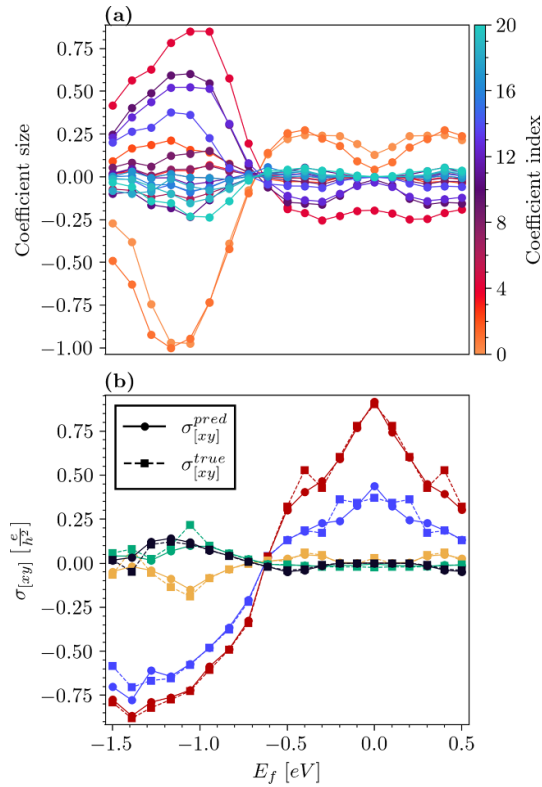
Figure 12. Fermi energy dependence. (a) Model coefficients as a function of the Fermi energy. One model is fitted at each of the sampled Fermi energies E_f. The model coefficients vary depending on the energy E_f at which the model is fitted. Notably, at $E_f \approx 0.7$ eV, all the coefficients change sign, as does the target. (b) Predictions of the $\vec{\xi}$-even part of the anomalous Hall conductivity (AHC) as a function of the Fermi energy. Here, the prediction quality of the model can be inspected for four randomly chosen spin samples (blue, orange, petrol, red), where the true values are shown in dashed lines with square symbols, while the predictions are shown with solid lines and circles. The data presented here are examples, for which the residuals are below 0.1.

energy. By probing the electric transport at constant Fermi energy, the analysis is sensitive to sudden changes in occupation, for example when single bands are pushed above or below the Fermi energy by altering the magnetic configuration. Correspondingly, analysis of the prediction for different energetic positions in the band structure shows, that the amount of deviation in the model prediction depends on the Fermi energy. In order to avoid this nonuniform behavior, the target can be calculated at constant filling. This approach, which we discuss in detail below, requires much more numerical effort, since the calculations of the AHC have to be complemented by precise calculations of the density of states. In contrast, the expansion could be adapted to incorporate certain band structure effects, sacrificing conceptual simplicity for a more accurate description of the transport. On the other hand, considering the fact that in larger systems the electronic structure contains many more bands, the changes in occupation due to the rearrangement of bands might be less drastic, and the general behavior of the coefficients and fitting much smoother. Correspondingly, deviations in model prediction could be dramatically reduced.

3.4.1. Model coefficients as function of the Fermi energy

The AHE is a quantity driven by the occupied states of a system. It is therefore instructive to investigate the influence of the Fermi energy E_F on the selected model − we quantify this relationship by fitting models at several values of the Fermi energy and tracking the values of the model coefficients. Only states below E_F are occupied and therefore contribute to the AHC, which allows us to associate model coefficients and their size with regimes in the electronic structure contributing to the electric transport. As the data presented in figure 12(a) clearly shows, the model coefficients in the $\vec{\xi}$-even case are smooth functions of the energy, which provides a basic sanity check for the model: the selected model transfers from one regime in the electronic structure to another, without selecting a different set of features. We can further identify a few dominating features in the model, with the corresponding model coefficients being as much as five times larger than the other coefficients in size. Perhaps the most significant characteristic of the presented data is the sign change that all coefficients undergo at a Fermi energy of about −0.65 eV, which corresponds to a sign change in the
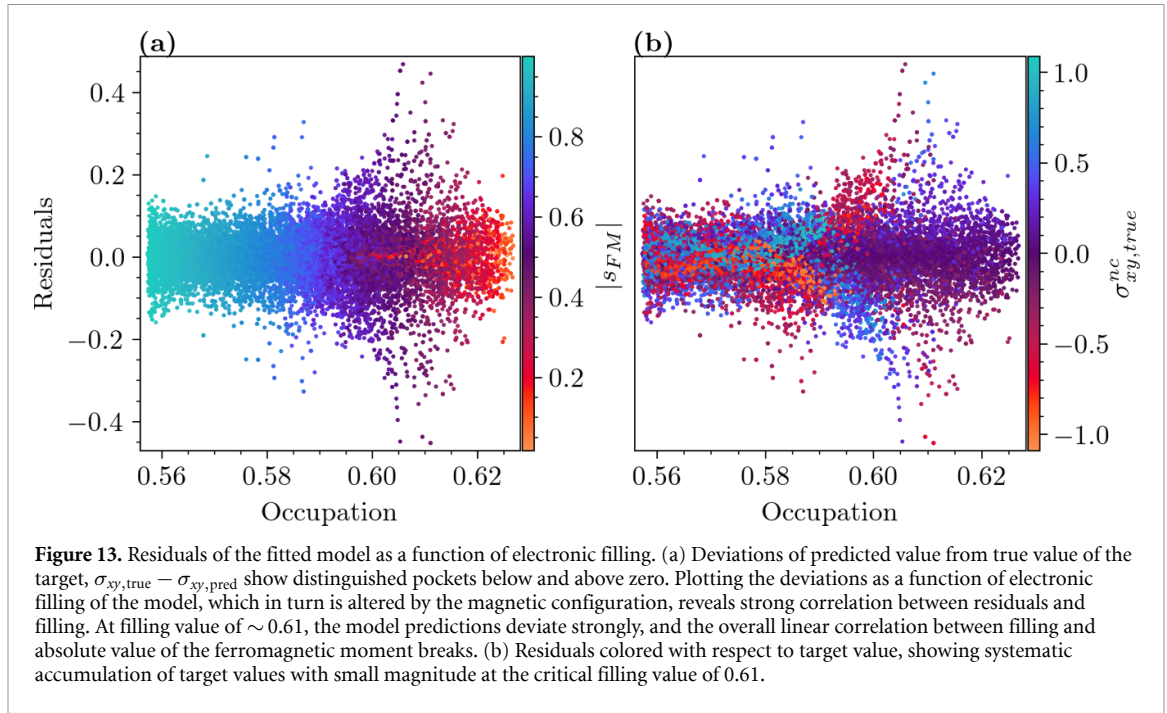
Figure 13. Residuals of the fitted model as a function of electronic filling. (a) Deviations of predicted value from true value of the target, $\sigma_{xy,\text{true}} - \sigma_{xy,\text{pred}}$ show distinguished pockets below and above zero. Plotting the deviations as a function of electronic filling of the model, which in turn is altered by the magnetic configuration, reveals strong correlation between residuals and filling. At filling value of ~ 0.61, the model predictions deviate strongly, and the overall linear correlation between filling and absolute value of the ferromagnetic moment breaks. (b) Residuals colored with respect to target value, showing systematic accumulation of target values with small magnitude at the critical filling value of 0.61.
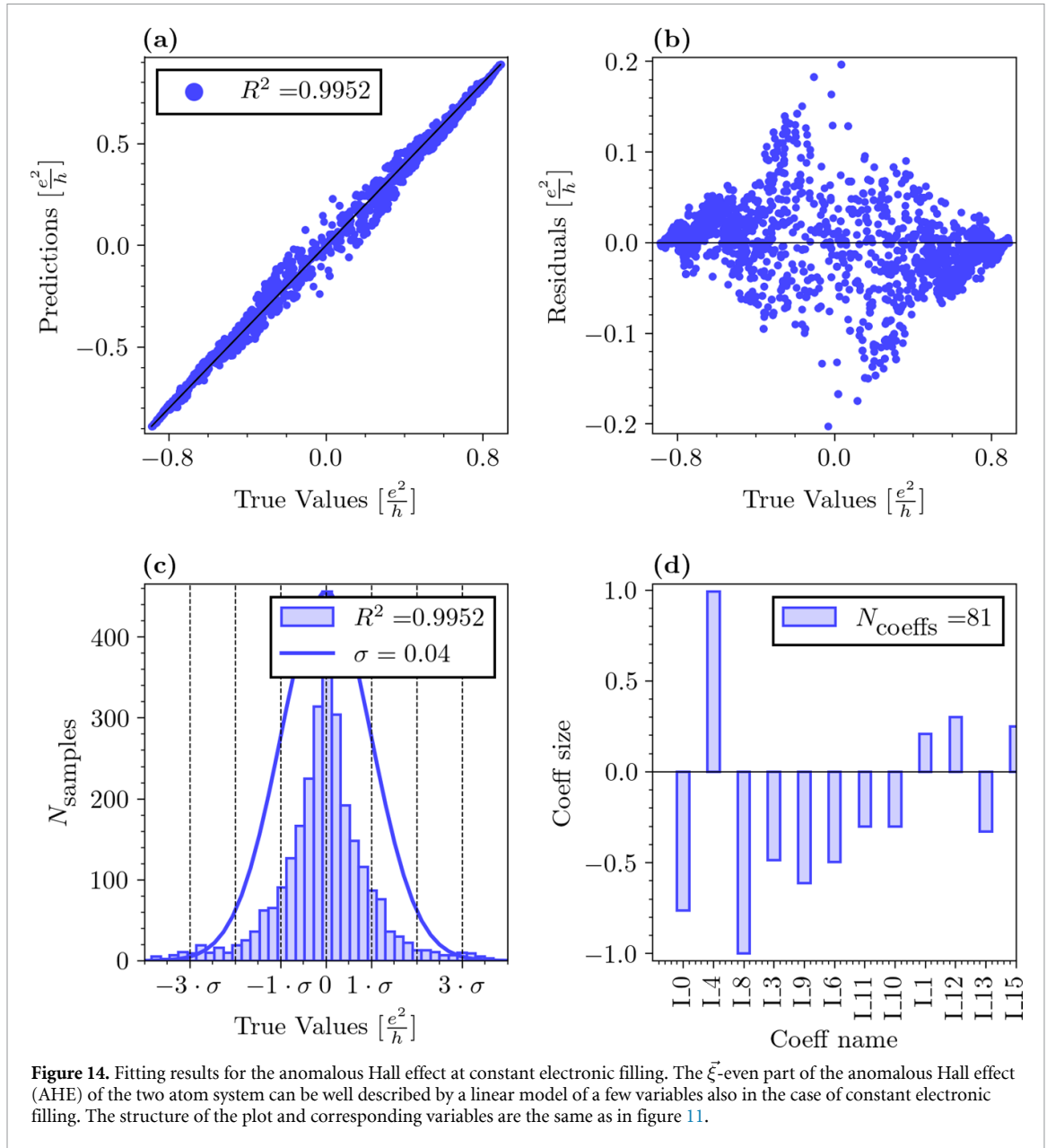
target, or AHE, at that specific energy. Overall, we also observe how the interplay between dominant, lower, and higher-order components of the chiral AHE can sensitively depend on the electronic structure.

Inspecting the behavior of the predictions and target values for four representative spin configurations (blue, orange, cyan, red) as a function of the Fermi energy, we can immediately recognize that in the even case, figure 12(b), the model prediction (circular markers, solid line) resembles the target values (square markers, dashed line) closely across different values of the energy. At the Fermi energy of about 0.7 eV, the target and prediction for all samples change sign. The samples we are referring to here are randomly selected from the samples with residuals below 10% at all values of the Fermi energy, which constitutes about 82% of the training set. The examples presented in figure 12(b) are a strong indicator that the model closely represents the true target value on a significant amount of previously unseen test data.

3.4.2. Transport at constant electronic filling
Coming back to the systematic deviations of model predictions from the true target, shown in figure 11(b) as pockets below and above zero at the center of the plot, we suggested to identify varying electronic filling as one of the underlying difficulties for a linear model. Indeed, tuning the magnetic configuration results, as a direct consequence of the presence of spin–orbit interaction in the electronic model equation (1), in changes to the energetic position and shape of electronic bands in reciprocal space. As a direct consequence, the electronic occupation at a given value of the Fermi energy is not constant under changes in the magnetic configuration. In order to judge whether differences in electronic filling, driven by these changes, introduce systematic deviations in the fitted model, we present the residuals of the fitted data as a function of electronic filling, with a color code indicating the absolute value of magnetization, in figure 13(a), or the true value of the target, figure 13(b).

As basic physical intuition concerning magnetic systems with strong exchange splitting and weak spin–orbit interaction would suggest, the electronic filling of the system, shown on the *x*-axis in the figure, correlates strongly with overall magnetization in the system. For the region of $[0.56, 0.6]$ electrons per atom, the residuals are constrained to the range $[-0.1, 0.1]$. Most interestingly, the deviations of model prediction from true target value increase suddenly around a value of about 0.61 for the electronic filling, with a magnitude of up to 0.4. Furthermore, the strict correlation between occupation and magnetization is lifted at this value, with strongly compensated configuration showing small residuals, and samples with moderate magnetization inhibiting large deviations. Additionally, when inspecting figure 13(b), showing the same data colored according to the true value of the target, smaller oscillations of the model predictions below 0.1 are visible in the region $[0.56, 0.6]$ electrons per atom. Interestingly, target values close to upper and lower limit of the target range accumulate in this region. Samples with small target value are mostly localized around electron filling of 0.61 electrons per atom, where the model predictions deviate strongly. We can thus conclude that the residuals of the fitted model show systematic dependence on electronic filling, which is modulated by the magnetic configuration, in a strongly nonlinear fashion. The data presented in figure 13 is

Figure 14. Fitting results for the anomalous Hall effect at constant electronic filling. The $\vec{\xi}$-even part of the anomalous Hall effect (AHE) of the two atom system can be well described by a linear model of a few variables also in the case of constant electronic filling. The structure of the plot and corresponding variables are the same as in figure 11.

strong evidence, supporting the idea that systematic deviations in the linear model are caused by nonlinear characteristics of the electronic structure, induced by changes to the magnetic configuration and mediated by the coupling of spin to electronic degrees of freedom through spin–orbit interaction.

In order to rectify the approach, we present a model fitted to training data calculated at constant electronic filling in figure 14. As is the case for figure 11, the model is fitted to $\vec{\xi}$-symmetrized data, with a score larger than 0.99, and visibly smaller deviations in prediction when comparing the data presented in panels (a), showing model predictions on the y-axis and true target value on the x-axis, in figures 14 and 11. Under closer inspection of the data shown in panels (b), which present residuals $y_{\text{true}} - y_{\text{pred}}$ on the y-axis and true values on the x-axis, we conclude that the overall scale of deviations relative to the absolute value of the target is reduced by factor of about four when considering the effect of constant filling. The comparison of the model coefficients, presented in panel (d) in each figure, reveals, that very similar components are active in the model, while the precise ordering and assigned weight is slightly different. We can thus conclude that, overall, while accounting for the effect of constant electron filling may somewhat improve the accuracy of the modeling, the qualitative predictions and understanding of the model can be already achieved at the level of a much more computationally efficient approach which relies on fixing the value of the Fermi energy. The two approaches are expected to converge in the limit of many atoms in the unit cell, as it is the case for example for complex spin textures or fluctuating magnets.

4. Discussion

The main goal of this study was to assess whether the electric transport, generated by the intricate changes in electronic structure introduced by complex non-collinear magnetism, can be predicted by using only descriptors of the magnetic pattern realized on a two-dimensional lattice. First, in order to obtain a complete, but non-redundant, description of the general conductivity tensor, an expansion in symmetric invariants of the underlying lattice point group was utilized to find a suitable set of descriptors. Second, a machine-learning pipeline was designed, tested and trained for the purpose of finding a sparse, linear model of the conductivity as a function of the symmetric invariants, which displays reasonable generalization statistics on unseen data. Relevant features of the input data were extracted by utilizing two complementary methods: on the one hand, feature selection was performed by ranking the features with a statistical metric measuring the correlation between feature and target, the f-regression test. On the other hand, regularization of the model search with the elastic net penalty was used to assign lower scores to models using a larger number of features, therefore pushing the search towards models which minimize the number of nonzero coefficients. Combining these two steps has led to significant condensation of the input feature space in conjunction with good overall generalization ability, verified by reasonable scoring of the final model on the test set.

The results of this study illustrate, that an excellent description of electric transport in magnetic materials can be achieved on the basis of the magnetic structure only, when the electronic filling of the system is considered to be unchanged by tuning the magnetic configuration, while the analysis performed for a fixed Fermi energy can already give a very good insight into the qualitative constituents of the model and the relative importance of different terms in the conductivity expansion. Given suitable modeling techniques, explicit calculations of the electronic structure, especially challenging in materials exhibiting strong spin–orbit interaction and hosting complex magnetic textures, can be augmented and enhanced by the symmetric invariants method, encoding the magnetic order parameters of the system.

Commenting especially on the statistical feature analysis, the results suggest the importance of magnetic texture characteristics beyond the established order parameters of ferromagnetic and staggered antiferromagnetic moment. Overall, while the feature space can be condensed considerably, the resulting linear models still show remarkable entanglement of magnetization properties beyond the established order parameters. While the resulting model is linear, the PCA components used as building blocks for the model, are sensitive beyond linear dependence to intricate changes in the orientation of magnetic moments. Our findings thus underline the importance of higher-order in spin contributions for a consistent description of the AHE already in the simplest case of bipartite canted magnets described by elementary electronic models with spin–orbit interactions. This is the main finding of our work concerning the physical properties of quantum spin systems. We can speculate with a high degree of certainly that the situation is even more complex for example in frustrated spin systems of kagome type, where various contributions to the AHE are still under debate (see e.g. [35]). Our approach is thus unique in identifying the magnitude and symmetry of leading AHE contributions without resorting to an excruciating process of guessing complemented by scarce calculations.

At this point, a remark on the amount of training data required for this model is in order. While modern, commercial applications of machine learning or artificial intelligence work on massive amounts of data consisting of millions of records with many dimensions, the project at hand requires the calculation of only 10^4 magnetic configurations. On one hand, the amount of data can be multiplied by applying the symmetry-based augmentation technique. On the other hand, the tools incorporated into the pipeline like PCA are already stable for sample sizes $\sim 10^4$. The results of the fitting technique, especially the decorrelated feature space as well as the dominant coefficients, are therefore robust for smaller sample sizes while the pipeline readily scales to larger amounts of data.

A special word should be said about the potential of suggested methodology in exploring the influence of various further parameters and corresponding more complex phase-spaces on the transport properties of dynamically evolving or fluctuating magnets. Indeed, we expect that the effect of various characteristics of the quantum system as reflected in its Hamiltonian—such as the magnitude and symmetry of the hoppings, strength of correlations, exchange splitting or spin–orbit interaction strength, which can be tuned dynamically e.g. by temperature, laser pulse, or other external means—can be consistently included into our analysis. Going beyond, we dare to suggest a possibility of including the time evolution of the spin system according to some dynamical equation into consideration explicitly, which may result in finding a clear path towards consistent modeling of the memory kernel of the system [36].

Overall, we can conclude that, in spite of our work focusing on a simple four-band electronic structure component, the principal possibility of modeling that we have demonstrated here marks an important step in integrating machine learning methods into the field of spintronics research and in particular

magneto-transport phenomena. Symmetry-enhanced compressive sensing in conjunction with principal component analysis turns out to be a promising candidate to conquer the complexity of magnetic phase-spaces, and the study of spin and electronic transport phenomena offers plenty of avenues for extending the machine learning technique. The symmetric invariants, although already showing to be a feature space with impressive potential for generalizability, can be compared and possibly augmented with features directly extracted from the magnetic texture in real space, or the electronic structure in reciprocal space. Specifically, direct feature extraction, e.g. through variational auto encoders (VAE), lends itself to the case of large textures with many atomic sites, when the construction of local chiral and non-chiral contributions is not trivial, as opposed to the two-atom system studied here. Acquiring a firm ability to predict the transport properties of large spin textures is pertinent for our ability to model and understand the response characteristics of matter in an automated manner, as such presenting a challenge for research at the junction of solid states physics and machine learning.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

Conflict of interests

The authors declare no competing interests.

ORCID iDs

Jonathan Kipp ⦿ https://orcid.org/0000-0002-8400-867X
Fabian R Lux ⦿ https://orcid.org/0000-0003-4658-1878
Stefan Blügel ⦿ https://orcid.org/0000-0001-9987-4733
Yuriy Mokrousov ⦿ https://orcid.org/0000-0003-1072-2421

References

[1] Garello K, Miron I M, Avci C O, Freimuth F, Mokrousov Y, Blügel S, Auffret S, Boulle O, Gaudin G and Gambardella P 2013 Symmetry and magnitude of spin–orbit torques in ferromagnetic heterostructures *Nat. Nanotechnol.* **8** 587
[2] Yao Y, Kleinman L, MacDonald A H, Sinova J, Jungwirth T, Wang D-S, Wang E and Niu Q 2004 First principles calculation of anomalous Hall conductivity in ferromagnetic BCC Fe *Phys. Rev. Lett.* **92** 037204
[3] Czaja P, Freimuth F, Weischenberg J, Blügel S and Mokrousov Y 2014 Anomalous Hall effect in ferromagnets with Gaussian disorder *Phys. Rev.* B **89** 014411
[4] Bruno P, Dugaev V K and Taillefumier M 2004 Topological Hall effect and Berry phase in magnetic nanostructures *Phys. Rev. Lett.* **93** 096806
[5] Lux F R, Freimuth F, Blügel S and Mokrousov Y 2018 Engineering chiral and topological orbital magnetism of domain walls and skyrmions *Commun. Phys.* **1** 60
[6] Lux F R, Freimuth F, Blügel S and Mokrousov Y 2020 Chiral Hall effect in noncollinear magnets from a cyclic cohomology approach *Phys. Rev. Lett.* **124** 096602
[7] Yamaguchi T and Yamakage A 2021 Theory of magnetic-texture-induced anomalous Hall effect on the surface of topological insulators *J. Phys. Soc. Japan* **90** 063703
[8] Hirschberger M *et al* 2021 Geometrical Hall effect and momentum-space Berry curvature from spin-reversed band pairs *Phys. Rev.* B **103** L041111
[9] Bouaziz J, Ishida H, Lounis S and Blügel S 2021 Transverse transport in two-dimensional relativistic systems with nontrivial spin textures *Phys. Rev. Lett.* **126** 147203
[10] Feng Z *et al* 2020 An anomalous Hall effect in altermagnetic ruthenium dioxide *Nat. Electron.* **5** 735
[11] Kipp J *et al* 2021a The chiral Hall effect in canted ferromagnets and antiferromagnets *Commun. Phys.* **4** 99
[12] Kipp J, Lux F R and Mokrousov Y 2021 Chiral response of spin-spiral states as the origin of chiral transport fingerprints of spin textures *Phys. Rev. Res.* **3** 043155

[13] Bac S-K *et al* 2022 Topological response of the anomalous Hall effect in MNBi$_2$Te$_4$ due to magnetic canting *npj Quantum Mater.* **7** 46

[14] von Malottki S, Dupé B, Bessarab P F, Delin A, Delin A and Heinze S 2017 Enhanced skyrmion stability due to exchange frustration *Sci. Rep.* **7** 12299

[15] Bessarab P F, Uzdin V M and Jónsson H 2015 Method for finding mechanism and activation energy of magnetic transitions, applied to skyrmion and antivortex annihilation *Comput. Phys. Commun.* **196** 335

[16] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons *Phys. Rev. Lett.* **104** 136403

[17] Gigli L, Veit M, Kotiuga M, Pizzi G, Marzari N and Ceriotti M 2022 Thermodynamics and dielectric response of BaTiO3 by data-driven modeling *npj Comput. Mater.* **8** 209

[18] Rao Z *et al* 2022 Machine learning-enabled high-entropy alloy discovery *Science* **378** 78–85

[19] Hilgers R, Wortmann D and Blügel S 2023 Application of batch learning for boosting high-throughput ab initio success rates and reducing computational effort required using data-driven processes (arXiv:2311.15430)

[20] Hilgers R, Wortmann D and Blügel S 2023 Machine learning-based estimation and explainable artificial intelligence-supported interpretation of the critical temperature from magnetic ab initio Heusler alloys data (arXiv:2311.15423)

[21] Peano V, Sapper F and Marquardt F 2021 Rapid exploration of topological band structures using deep learning *Phys. Rev. X* **11** 021052

[22] Venderley J *et al* 2022 Harnessing interpretable and unsupervised machine learning to address big data from modern x-ray diffraction *Proc. Natl Acad. Sci.* **119** e2109665119

[23] Li X-Y, Lou F, Gong X-G and Xiang H 2020 Constructing realistic effective spin Hamiltonians with machine learning approaches *New J. Phys.* **22** 053036

[24] Yanagi Y, Kusunose H, Nomoto T, Arita R and Suzuki M-T 2023 Generation of modulated magnetic structures based on cluster multipole expansion: application to α-mn and CoM$_3$s$_6$ *Phys. Rev. B* **107** 014407

[25] Lysogorskiy Y *et al* 2021 Performant implementation of the atomic cluster expansion (pace) and application to copper and silicon *npj Comput. Mater.* **7** 1

[26] Li X, Xu C, Liu B, Li X, Bellaiche L and Xiang H 2023 Realistic spin model for multiferroic NiI$_2$ *Phys. Rev. Lett.* **131** 036701

[27] dos Santos Dias M, Brinker S, Lászlóffy A, Nyári B, Blügel S, Szunyogh L and Lounis S 2022 Reply to comment on proper and improper chiral magnetic interactions *Phys. Rev. B* **105** 026402

[28] Szilva A, Kvashnin Y, Stepanov E A, Nordström L, Eriksson O, Lichtenstein A I and Katsnelson M I 2023 Quantitative theory of magnetic interactions in solids *Rev. Mod. Phys.* **95** 035004

[29] Šmejkal L, González-Hernández R, Jungwirth T and Sinova J 2019 Crystal Hall effect in collinear antiferromagnets (arXiv:1901.00445)

[30] Go D, Sallermann M, Lux F R, Blügel S, Gomonay O and Mokrousov Y 2022 Noncollinear spin current for switching of chiral magnetic textures *Phys. Rev. Lett.* **129** 097204

[31] Weisstein E W 2024 Sphere point picking From MathWorld–A Wolfram Web Resource (available at: https://mathworld.wolfram.com/SpherePointPicking.html)

[32] Szlam A, Kluger Y and Tygert M 2014 An implementation of a randomized algorithm for principal component analysis

[33] Trefethen L and Bau D 1997 Numerical Linear Algebra *Other Titles in Applied Mathematics* (*Society for Industrial and Applied Mathematics*) (SIAM) p 19104

[34] Friedman J H, Hastie T and Tibshirani R 2010 Regularization paths for generalized linear models via coordinate descent *J. Stat. Softw.* **33** 1–22

[35] Rajan A *et al* 2023 Revealing the higher-order spin nature of the Hall effect in non-collinear antiferromagnet Mn$_3$Ni$_{0.35}$Cu$_{0.65}$N (arXiv:2304.10747)

[36] Gouasmi A, Parish E J and Duraisamy K 2017 A priori estimation of memory effects in reduced-order models of nonlinear systems using the Mori-Zwanzig formalism *Proc.: Math. Phys. Eng. Sci.* **473** 1