# Hunting for vampires and other unlikely forms of parity violation at the Large Hadron Collider

**Christopher G. Lester,**[a] **Radha Mastandrea,**[b] **Daniel Noel**[a] **and Rupert Tombs**[a]

[a]*Cavendish Laboratory, University of Cambridge,*
*Cambridge, CB3 0HE, U.K.*

[b]*Department of Physics, University of California,*
*Berkeley, CA 94720, U.S.A.*

*E-mail:* lester@hep.phy.cam.ac.uk, rmastand@berkeley.edu,
dn322@cam.ac.uk, rt500@cam.ac.uk

ABSTRACT: Non-Standard-Model parity violation may be occurring in LHC collisions. Any such violation would go unseen, however, as searches for it are not currently performed. One barrier to searches for parity violation is the lack of model-independent methods sensitive to all of its forms. We remove this barrier by demonstrating an effective and model-independent way to search for parity-violating physics at the LHC. The method is data-driven and makes no reference to any particular parity-violating model. Instead, it inspects data to construct sensitive parity-odd event variables (using machine learning tools), and uses these variables to test for parity asymmetry in independent data. We demonstrate the efficacy of this method by testing it on data simulated from the Standard Model and from a non-standard parity-violating model. This result enables the possibility of investigating a variety of previously unexplored forms of parity violation in particle physics.

Data and software are shared at https://zenodo.org/record/6827724.

## Contents

## 1 Introduction

Sources of parity violation which are due to effects beyond the Standard Model have the potential to be observed in Large Hadron Collider data. Alas, there are many new and unexpected ways in which parity can be violated, and no continuous parity-odd event

variable is sensitive to them all [1]. Furthermore, since there are not many commonly-studied[1] theoretical models which predict forms of parity violation which could be visible at the LHC, there are few guides as to where and how to look for non-standard parity-violating signals. The absence of practical, general, *and* scalable ways of performing such searches has been, until very recently, a significant barrier to performing them.[2]

The primary purpose of this paper is to demonstrate that the data-driven parity-violation search strategies proposed by [4] and [5] are, in fact, capable of removing the above barrier to realistic searches for parity violation in particle physics. Our work is necessary because the methods proposed in [4] are demonstrated only with toy examples from outside of particle physics, while those in [5] use neither real nor simulated particle physics datasets holding signatures of parity-violation. The present work will demonstrate the viability of the new approaches by showing sensitivity to a range of parameters within an example parity-violating model, which is described within the Lorentz-invariance violating framework of the minimal Standard Model Extension [6]. For this model, we generate Monte Carlo simulated events assuming a barrel-shaped LHC detector such as ATLAS, CMS, or ALICE.

Specifically, [4] points out that a clear signal for parity violation can be obtained from a parity-odd event variable if it is seen to have an asymmetric distribution. An optimization using one part of a dataset can be used to create such a variable, while another part of the dataset can be used to test for the suspected asymmetry in the distribution of that variable. For the reasons given in [4], this method presents a very general, flexible, and scalable way of maintaining sensitivity to all potential source of new physics which are able to generate events which are distinguishable (in distribution) from their parity-flipped partners — that is to say any models with 'manifestly' parity violating signatures. Whether any given source is observable would, of course, always depend on the sizes of systematic uncertainties and the amount of data taken. It would also depend on whether the particular choice of parity-odd function(s) that someone implementing [4] had decided to use were sensitive to the form of parity violation which the given source induces in its events. However, those are questions of implementation rather than intrinsic limitations. The generality of the method itself is an unavoidable consequence of the fact that it is, in effect, simply a codification of what it means for a model to have 'manifestly' parity violating consequences. In this sense, electing to use [4] does not place limits on what can be seen.

Known inefficiencies and acceptance effects, which act to filter data from observation in the detector, can also be easily corrected for in the manner described in [5] if they are well understood. Such filtering effects are not, however, included in any detector models in this work.

---

[1] See later comments that BSM parity-violation models, if they are to have observable signals in the context of this work, must abandon at least one of: (i) locality, (ii) Lorentz-invariance, or (iii) a basis in quantum field theory. Vampires are not visible when mirrored [2] and so presumably the laws of physics describing them derive their parity-violation from the loss of at least one of the above three properties!

[2] Only a single search, [3], has been performed on real LHC data so far! Furthermore, [3] described many of its own choices as arbitrary and lacking in any strong theoretical motivation or generality. As a result, its null result is of limited value, even if there is value in the existence of the work as a proof-of-principle that data-data searches for non-standard sources of parity-violation can be performed.

Our approach can be viewed as an adoption of standard practice from the field of machine learning: in training, a general-purpose 'machine' generates hypotheses about the structure of the data distribution. Then in testing, those hypotheses are tested on independent data. By doing this with parity-odd algorithms, success in the testing phase becomes evidence for parity violation.

Various end-stage techniques could be used to assess the strength of that evidence. To follow existing practice for LHC searches, the parity-odd variables can be histogrammed and interpreted with conventional statistical methods. This approach, which we illustrate in appendix A.1, allows one to implement systematic uncertainties at testing time as histogram variations that allow weakly broken parity symmetry

The nature of parity violation and its possible visibility at the LHC is discussed in section 2. Parity-violating physics in an example quantum field theory, and the simulation of its effects, are described in section 3. Machine learning implementations of parity-odd, symmetry-invariant event variables are presented in section 4. The varied successes of these variables in detecting parity violation in simulated data are displayed in section 5.

## 2   Parity violation at the LHC

It is experimentally possible to see parity violation in particle physics data. Whether it *will* be seen depends both on whether nature produces parity violating effects and whether the appropriate data analysis is performed.

Particles and particle physics data are typically described as being sampled from a differential cross-section $\sigma(x)$, where $x$ is some representation of those data. We would like to test whether $\sigma(x)$ violates parity symmetry; that is, we want to answer the question "is $\sigma(x)$ different from $\sigma(\mathrm{P}x)$?" for the parity operator P. However, we have finite data. With finite data, asymmetries may be too subtle to notice, and with continuous distributions asymmetries may hide in intricate patterns which we as analysts would never think to check. We can, however, take opportunities to disprove parity symmetry if nature happens to give clear signals.

We can maintain sensitivity to all forms of parity violation by considering only parity-odd event variables [1].[3] To test parity symmetry, we can therefore take any parity-odd event variable $f(x)$, for which $f(x) = -f(\mathrm{P}x)$, and compare the positive and negative halves of its distribution under $\sigma(x)$. If parity symmetry is respected by $\sigma(x)$, then events and their parity-flipped versions must all be produced at equal rates. Equivalently, there must be equal event rates at all $\pm f$ points. However, if an experiment sees an asymmetric distribution in $f(x)$, for example by histogramming many data, then parity symmetry must be violated in $\sigma(x)$.

Hadron-hadron collisions, such as those currently produced at the LHC, have many symmetries. Two identical, unpolarized beams are collided head-on (to a good approximation), and, since physical space is known to be isotropic (to a very good approximation),

---

[3]This statement is precisely formulated in theorem 2.26 and summarized in section 2.9 of [1]. Essentially, the absolute value of a parity-odd variable is parity-even, and all parity-asymmetric variables can be constructed as functions of even and odd parts.

the scattering processes must be invariant to all rotations which do not change the beams. These rotations include both those about the beam axis (by any angle), and those which swap the two beams by rotating 180° about a perpendicular axis. Furthermore, data often comprise unordered sets of four-momenta for otherwise-identical particles; although these always take some conventional order in practice, nature does not care about any ordering we assign.

When analysing such symmetric data, it is helpful to be blind to transformations under their symmetries; otherwise, the handling of symmetrically equivalent, yet differently represented data introduces unnecessary complexity. This is one benefit of rotation-invariant event variables at the LHC, such as those based on masses or transverse momenta $p_{\mathrm{T}}$, and to collapsing the permutation symmetry of particles by $p_{\mathrm{T}}$-ordering their representations.

Furthermore, our aim is only to test for violation of parity symmetry, and not for violations of other symmetries such as spacetime isotropy. Blinding our analysis to violations of other symmetries, therefore, helps the method to focus more precisely on the task of testing parity violation itself. For these reasons, when constructing parity-odd event variables in this paper, we choose to make them also rotation-, beam-swap- and permutation-invariant. This invariance is achieved by design in both data representations and function architectures.

One example of a parity-odd, rotation-, beam-swap- and permutation-invariant event variable is

$$\alpha = \arcsin\left( \frac{\vec{p}^{j_1} \times \vec{p}^{j_2}}{|\vec{p}^{j_1} \times \vec{p}^{j_2}|} \cdot \frac{\vec{p}^{j_3}}{|\vec{p}^{j_3}|} \right) \tag{2.1}$$

in which $\vec{p}^{j_i}$ is the momentum of the $i$th hardest jet. This $\alpha$ variable was proposed and used in [3] to demonstrate that parity violation can be searched for in events with three or more jets. It has rotation invariance from the basis-independence of dot and cross products, permutation invariance from the $p_{\mathrm{T}}$-ordering of jets, and parity-oddness since the three momentum vectors each pick up a factor of $-1$ under the parity flip. It also has, however, "no claims of generality or optimality" [3]. Indeed, figure 1 illustrates that it does not have good sensitivity to parity violation in a sample of simulated data, described below in section 3, while an observable $f$, created using the methods of *this* paper, does.

Unfortunately, parity violation in *data* does not necessarily imply parity violation in *nature*. Data result from compositions of natural physical processes with the actions of experimental hardware and software, and any of these parts could introduce asymmetries. Fortunately, any detector effect that can be anticipated and that acts as a filter (reducing data rates in certain selections), can be removed (i.e., the possibility of false positive excluded) using the method of [5].

Nonetheless, unknown unknowns could still cause problems. A positive signal for parity violation therefore provides either: (a) evidence for parity violation in *nature*, or (b) evidence for detector effects and/or calibrations that need to be improved. Both possibilities are interesting and valuable, since one tells us something new about nature, while the other tells us how to improve our detector calibrations which will benefit other analyses (even those not interested in symmetry violations) being performed by the same detector.
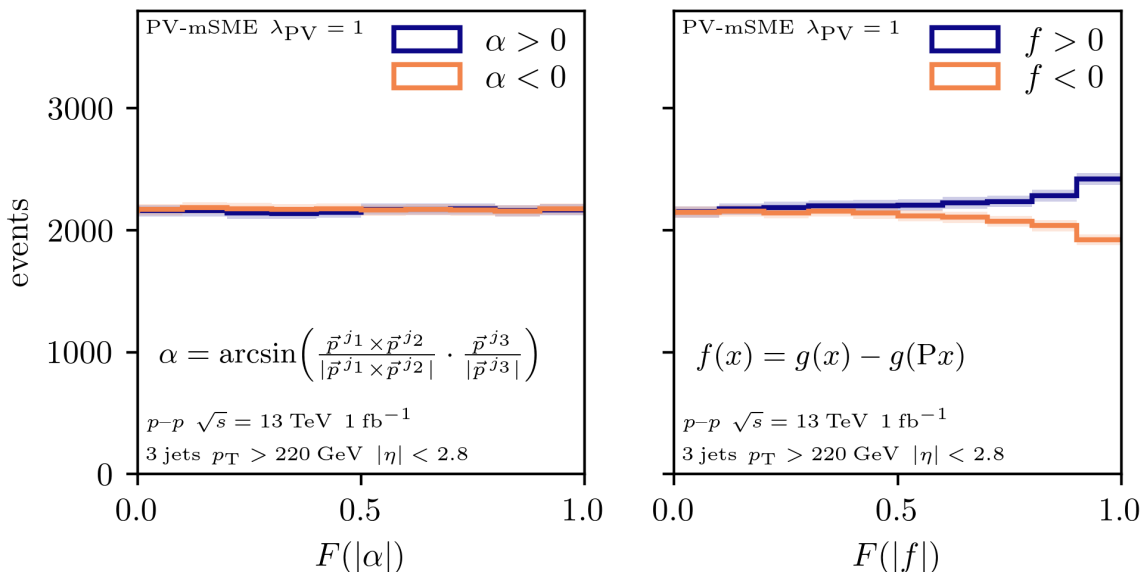
**Figure 1**. Distributions of two different parity-odd variables for the same parity-violating model of physics: (left) $\alpha$ as defined in equation (2.1) and (right) a parity-odd neural network looking at reconstructed jet momenta. Samples are normalized to the Standard Model cross-section at 1 fb$^{-1}$ of LHC $\sqrt{s} = 13$ TeV proton-proton collision data. The histograms contain only data from an independent testing set containing 20% of the total yield. Each function $F(\cdot)$ is a transformation fitted to training data to make each distribution close to uniform. Error bars are of size $\pm\sqrt{n}$ to show the size of statistical uncertainty at the chosen luminosity.

## 3   Parity violating physics

### 3.1   PV-mSME (Parity Violating minimal Standard Model Extension)

Nature may, one day, not be described by a quantum field theory (QFT). At present, however, the majority of existing theories are QFTs, and so we choose to illustrate our method with one. Locally Lorentz invariant QFTs which violate parity[4] must necessarily violate charge-parity (CP) symmetry for the reason explained in footnote 1 of [3]. Although one could therefore choose to illustrate our method by searching for parity violation in a CP-violating model, doing so is unlikely to be competitive with pre-existing methods as tests of CP-violation are highly developed with a mature field; in any case, the amount of CP-violation in the Standard Model is known to be small.

We therefore choose to use a Lorentz-violating extension of the Standard Model, namely the 'minimal Standard Model Extension' (mSME) of [6]. We do not work with the full generality of the mSME, but instead use only a small subset of its parity-violating sector, which we name the 'PV-mSME'. Within the PV-mSME, we control the strength of parity violation with a single real parameter $\lambda_{\mathrm{PV}}$, which is described below.

---

[4]To be precise, one should say: 'Locally Lorentz invariant QFTs for which parity violation is theoretically observable in data derived from collisions of unpolarized particles in the case that the final-state event data recorded contain only four-momenta which are not decorated with additional labels (such as helicities or flavours)'.

The mSME makes many extensions to the Standard Model, among which is a term $\mathcal{L}_{\text{quark}}^{\text{CPT-even}}$ (defined in equation 11 of [6]) which modifies the quark-quark-gluon Feynman Rule. We choose to use only this extension from the mSME, in order to generate parity violation in strongly produced multi-jet events, which are attractive to study since they have large cross-sections at the LHC and are not expected to show parity violation in the Standard Model. Multi-jet events were also previously studied parity violation in [3]. Furthermore, we choose couplings to switch on only its axial vector part with the intention of violating parity in interference with the Standard Model vector interactions.

The Standard Model quark-quark-gluon vertex from takes the form

$$f_\nu^{\text{SM}} = -ig_s \frac{\lambda_{ij}^a}{2}\gamma_\nu, \tag{3.1}$$

to which the PV-mSME adds an axial vector part shaped by a coupling matrix $(c_A)_{\mu\nu}$,

$$f_\nu^{\text{PV-mSME}} = -ig_s \frac{\lambda_{ij}^a}{2}\left(\gamma_\nu + \gamma^\mu (c_A)_{\mu\nu}\gamma^5 \lambda_{\text{PV}}\right), \tag{3.2}$$

in which the strength parameter $\lambda_{\text{PV}}$ is varied to control the magnitude of parity violation in the model. Motivated by reasons given in appendix C, we define $(c_A)_{\mu\nu}$ to be

$$(c_A)_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{3.3}$$

such that the value $\lambda_{\text{PV}} = 1$ results in PV-mSME terms having similar magnitudes to Standard Model terms in matrix elements. Clearly, the Standard Model is recovered for $\lambda_{\text{PV}} = 0$.

Although we are not aware of direct constraints on this exact usage of the mSME, data from astrophysics and deep inelastic scattering have been used to limit the couplings in this mSME quark sector to very small values of at most $10^{-4}$ in magnitude [7, 8]. Large $\lambda_{\text{PV}}$ values would therefore be excluded by current data if tested.

Unlike the common background-plus-signal scenario, PV-mSME cross-sections do not add linearly since they interfere with the Standard Model, nor do they scale linearly with the coefficient $\lambda_{\text{PV}}$. Indeed, squared matrix elements computed from a diagram with $n$ quark-quark-gluon vertices will, in general, contain terms with all powers from $(\lambda_{\text{PV}})^0$ up to and including $(\lambda_{\text{PV}})^{2n}$.

Couplings in the mSME break spatial isotropy; they induce special spatial directions in which particles have differing behaviours. Having constant couplings, as we do from equation (3.3), therefore ignores rotations of the detector through space. Detectors at the LHC do, in fact, rotate — with Earth through the day, and about the sun throughout the year, and so on. Rotations could be controlled for, for example, by selecting data from only times when the detector is aligned within a small solid angle of a special direction, or by relocating the LHC to a non-rotating spaceship. Neither solution is practical, since the first slashes the event rate, and the second is beyond CERN's budget. Alternatively, one could

employ a different, fully rotation-aware method of data generation and analysis, which we discuss further in appendix A.4.

Faced with the decision of whether to account for the Earth's rotation, we elect to make no correction and choose to leave our PV-mSME coupling matrixes constant in the laboratory frame. In effect, this means we are working with a model in which Lorentz violating couplings are 'dragged around' by Earth, or where Earth does not spin. We make this choice, despite it being unphysical, because making those corrections would complicate the analysis without adding anything useful. Moreover, if we were to correct for Earth's rotations we would be building assumptions into our analysis that are not explicitly related to parity, and we do not wish to do so.

Alternatively, our choice of constant couplings could be implemented by selecting data collected at dates and times when the detector aligned with a special direction. In reality, direction dependence might be handled better by encoding the detector's alignment in additional event variables, so that algorithms can learn alignment-dependent parity violation. We demonstrate a simple form of this rotation-encoding in appendix A.4, and find that weaker signals of parity violation can still be observed in the PV-mSME on a rotating planet.

Despite this unphysical use of the mSME, our PV-mSME remains a parity violating quantum field theory, and provides a parity violating model with which we can demonstrate our methods

## 3.2 Simulation

A complete simulation pipeline is performed to produce realistic simulations of proton-proton collisions scattering to jets in the PV-mSME with various settings of $\lambda_{\mathrm{PV}}$ between 0 and 1. This pipeline comprises several standard steps: partons are first simulated in a hard scatter of protons to three or four gluons or light quarks,[5] then dressed with an underlying event, parton shower and pileup overlay, and hadronized into observable particles. A detector response to these energetic particles is simulated and processed into event variables.

We introduce PV-mSME effects through matrix elements in the hard scatter only; any changes that it should make to parton distributions, showering, hadronization, and detector physics are ignored in this work.

Triggering is modelled by selecting only events with at least three central jets with $p_{\mathrm{T}} > 220$ GeV; this selection is designed to approximate the efficiency plateau of the three-jet trigger in a general-purpose LHC detector.

Details of the simulation software and settings are given in appendix B.

Although we primarily target events with three jets, we include four-parton processes at truth level as a sub-leading effect to increase physical accuracy. At reconstruction level after kinematic selections, the Standard Model cross-section of this multi-jet process is 0.22 nb of which 0.7% is attributed to four-parton processes at truth level. This increases with $\lambda_{\mathrm{PV}}$ to 0.31 nb (9%) for $\lambda_{\mathrm{PV}} = 0.3$ and 2.6 nb (27%) for $\lambda_{\mathrm{PV}} = 1$. Evidently, large $\lambda_{\mathrm{PV}}$

---

[5]Note that a $2 \to 2$ scattering process could never be parity violating, as all participating 4-momenta lie in a plane.
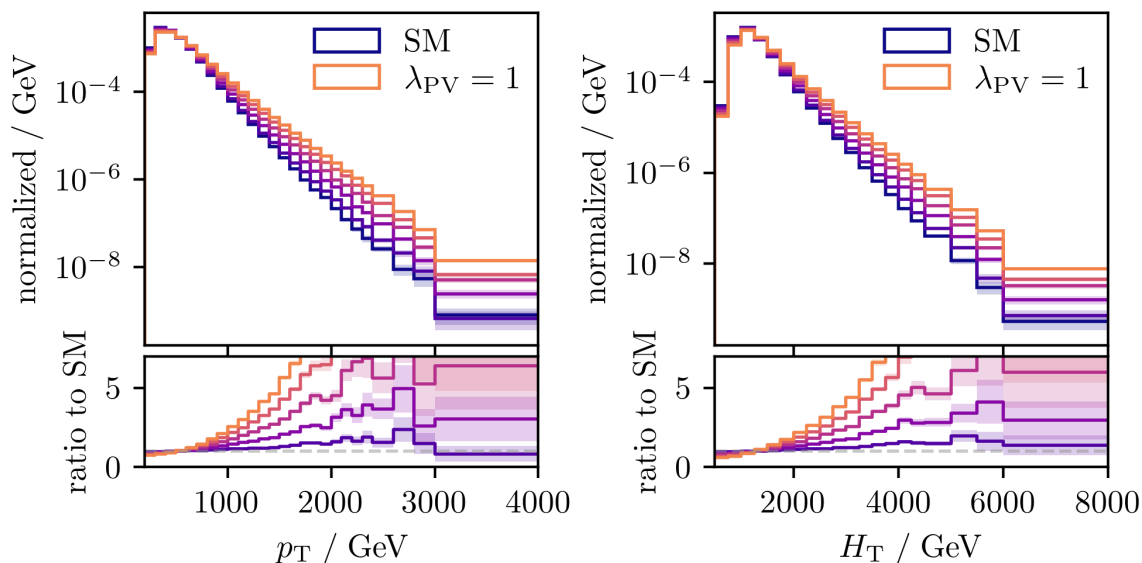
**Figure 2**. Distributions of kinematic variables of reconstructed jets with varied $\lambda_{PV}$ in the PV-mSME. (left) Transverse momentum of the hardest jet and (right) the scalar sum of transverse momenta of the hardest four jets in the event. The five coloured lines linearly interpolate from the Standard Model ($\lambda_{PV} = 0$) to $\lambda_{PV} = 1$.

drastically increases production rates and jet multiplicities. Kinematic shapes also change significantly, as demonstrated in figure 2. Our interest, however, is in the parity violation of these kinematic distributions, independent of other effects. We therefore normalize the PV-mSME distributions to the Standard Model cross-section wherever relevant.

### 3.2.1 Data formats

Simulation results are processed into two kinds of event variables for analysis; the first is standard, using estimated four-momenta of jets, while the second bypasses jet reconstruction to produce low-resolution images of energy deposits in the (cylindrical) calorimeters, projected into the $\eta$–$\phi$ plane, where $\eta$ is the pseudorapidity and $\phi$ is an angle about the beam pipe.

These images consist of $32 \times 32$ pixels that cover the entire $2\pi$ range in $\phi$ and $|\eta| <$ 3.2, which corresponds to the extent of the ATLAS calorimeters, excluding the forward calorimeter [9]. Images are constructed either directly from calorimeter energy deposits or indirectly from jets by histogramming their transverse momenta in two-dimensional $\eta$–$\phi$ histograms, where every jet with $p_T > 30$ GeV and $|\eta| < 2.8$ is included. Images of an example event are shown in figure 3.

To assess both how sensitivity degrades with noise and how well machine learning tools perform with different data formats, we use three representations of the simulated data. These are:

- 'truth-jet': the true momenta of partons from the hard scatter,

- 'reco-jet': reconstructed momenta of the four hardest jets, and

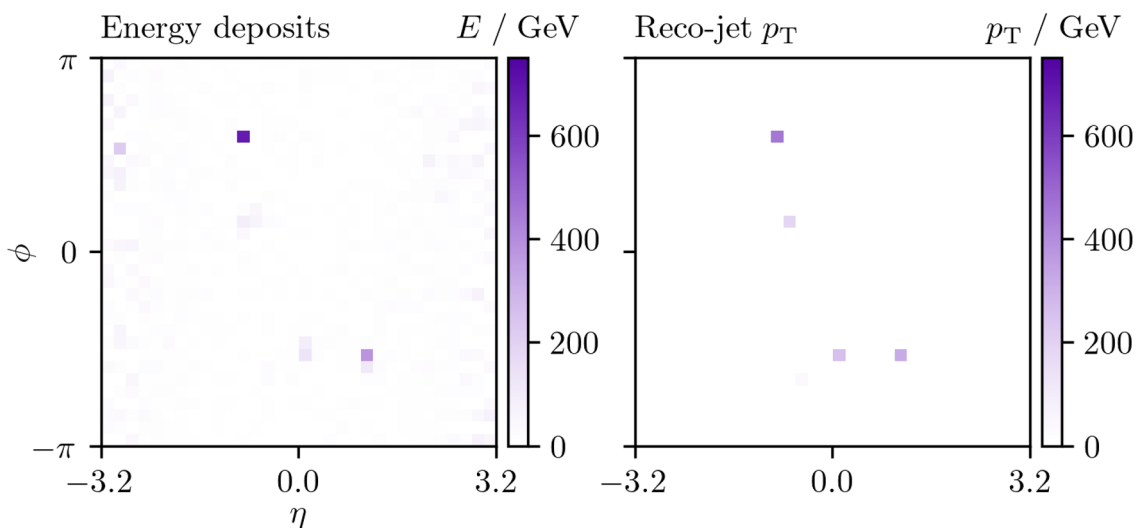- 'calo-image': images of calorimeter energy deposits.

**Figure 3**. Energy deposits in the calorimeter (left) and the $p_T$ of reconstructed jets (right) for an example Standard Model event.

The next section describes how of these data representations is used with each machine learning model.

## 4 Methods

As stated in section 2, we can search for parity-violating physics using a parity-odd event variable $f(x)$. Such a variable can easily be constructed from an arbitrary function $g(x)$ since

$$f(x) = g(x) - g(\mathrm{P}x) \tag{4.1}$$

is parity odd.[6] All of our models use this construction. Machine learning algorithms present practical, general, and scalable ways to assign functional forms to $g(x)$, and our method does not depend on the exact algorithm chosen. To demonstrate that, we use boosted decision trees (BDTs), neural networks (NNs), and convolutional neural networks (CNNs) to build the functions $g(x)$.

The BDT and NN approaches use vectors of jet momentum features derived from truth-jet and reco-jet data, whereas the CNN approach uses either calo-image data or jet features transformed into images as described in section 4.1.2.

### 4.1 Symmetries

We ensure that rotational and permutation symmetries are preserved in our parity-odd event variables, for reasons discussed in section 2. This section describes how those symmetries are enforced for the various models and data formats.

---

[6]Numerically, this works for all numbers except signed zeros, since if $g(x) - g(\mathrm{P}x) = +0.0$, then $g(\mathrm{P}x) - g(x) = +0.0$, not $-0.0$, in standard floating-point arithmetic. Zeros of $f(x)$ must therefore be excluded from parity-odd interpretations.

### 4.1.1 Invariant jets

Jet permutation symmetry is ensured by sorting objects from hardest to softest in $p_{\mathrm{T}}$. Invariance to rotations (along and about the beam axis) is ensured by rotating all momenta to an orthonormal basis $\{\hat{x}, \hat{y}, \hat{z}\}$ based on the hardest jet momentum $\vec{p}^{j_1}$; $\hat{x}$ is taken in the transverse direction of $\vec{p}^{j_1}$, $\hat{z}$ is taken along the beam axis with the sign of its longitudinal momentum $\vec{p}_z^{j_1}$, and the third basis vector is taken as their cross product; $\hat{y} = \hat{z} \times \hat{x}$.

Mutual transformation of the hardest jet with all other momenta under rotations means that the projection of momenta onto this basis is unchanged under rotations (up to numerical precision). Defining each element of momentum in this new rotation-invariant basis as $\vec{q}_a^{j_i} = \vec{p}^{j_i} \cdot \hat{a}$ for $a \in \{x, y, z\}$, it is clear that $\vec{q}_y^{j_1} = \vec{p}^{j_1} \cdot \hat{y} = 0$. That is, the $y$ component of the hardest jet is zero by definition and not useful, so we discard it.

Parity flipping momenta in laboratory coordinates leads only to changing the signs of $\vec{q}_y^{j_i}$ components, since under parity

$$\hat{x} \to -\hat{x}, \tag{4.2}$$

$$\hat{z} \to -\hat{z}, \text{ but} \tag{4.3}$$

$$\hat{y} \to -\hat{z} \times -\hat{x} = +\hat{y}; \tag{4.4}$$

the $\vec{q}_x^{j_i}$ and $\vec{q}_z^{j_i}$ components are parity-even, and $\vec{q}_y^{j_i}$ components are parity-odd. Both parity-odd and parity-even event variables can be useful in the search for parity violation; although parity-flipped events differ only in their parity-odd parts, the even parts are context which can be required to separate regions in which the parity flipped versions occur at different rates [10].

### 4.1.2 Invariant images

Invariance to rotations (about and along the beam axis) is built into the CNN structure. For invariance to $\phi$-rotations, note that a rotational symmetry in $\phi$ corresponds to a wrap-around translational symmetry in $\phi$ when considering the $\eta$–$\phi$ plane.

Each image is padded cyclically around the $\phi$ axis before applying each convolutional layer; this maintains equivariance. Invariance to $\phi$ rotations is achieved by performing a max-pool operation over the entire $\phi$-axis after the convolutions. This pooling selects the largest pixel value for each slice in $\phi$, which is clearly invariant to $\phi$ rotations of those pixels. This design gives perfect invariance to discrete rotations of $2\pi/32$, but the CNN is not perfectly blind to other rotations prior to pixelization; this could incur some costs for the reasons discussed in section 2 which encourage invariance.

For invariance to discrete beam-flip rotations, $R_\pi$ (rotation by 180° about an axis perpendicular to the beams), we construct $R_\pi$-invariant $g(x)$ functions (of equation (4.1)) as $g(x) = h(x) + h(R_\pi x)$, similarly to how parity-oddness is achieved. Then,

$$f(x) = [h(x) + h(R_\pi x)] - [h(\mathrm{P}x) + h(R_\pi \mathrm{P}x)], \tag{4.5}$$

and we learn a form for $h(x)$ with a CNN. This maintains the parity-odd property $f(x) = -f(\mathrm{P}x)$ and introduces the beam-flip symmetry $f(x) = f(R_\pi x)$, since $R_\pi R_\pi x = x$.
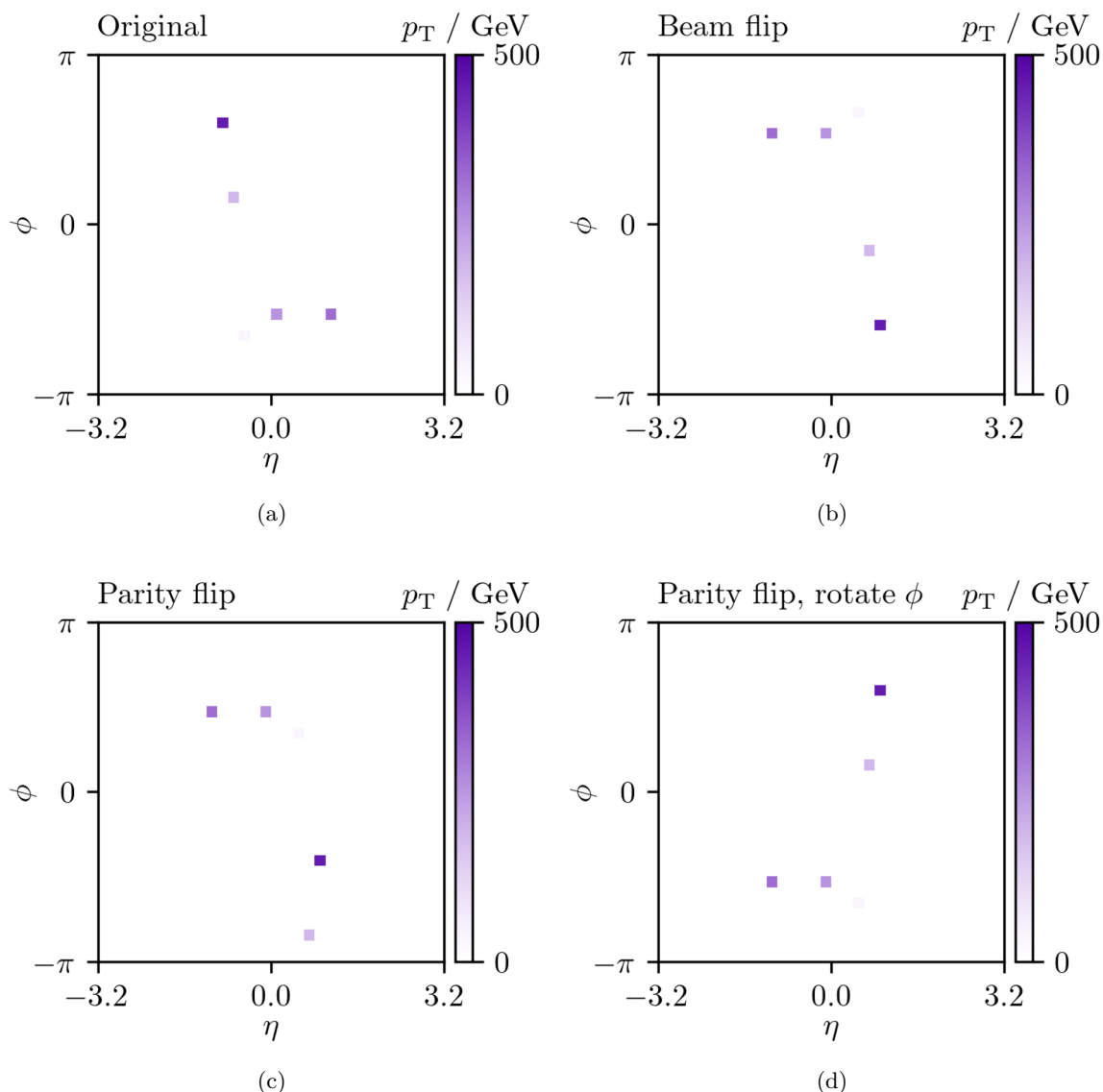
**Figure 4**. Histogrammed $p_T$ of reconstructed jets for a Standard Model event projected in the $\eta$–$\phi$ plane. (a) The original event image (b) rotated 180° about the $x$-axis (c) parity flipped (d) parity flipped and rotated by $\pi/2$ in $\phi$. The net output $f(x)$ is $-0.132$ for (a) and (b), and $0.132$ for (c) and (d); the symmetries we build into the network can be identified here.

The parity flip operator transforms $\eta \to -\eta$ and $\phi \to \phi + \pi \mod 2\pi$. With the rotationally invariant CNN, $\phi$ changes are inconsequential, so the parity flip operator is just a negation of $\eta$, or equivalently a mirror image through the line $\eta = 0$.

An example event image is shown in figure 4, along with its transformed copies. From the network outputs reported in the caption, we see that the network is indeed parity-odd whilst being symmetric to beam-flips. In addition, by comparing the outputs with a translation in $\phi$, we can see that the network is invariant to translations in $\phi$. This illustrates how a parity flip on an image in the $\eta$–$\phi$ plane is equivalent to a flip in $\eta$.

### 4.2 Training

Models are trained from gradients of the 'which is real?' objective function introduced in [5], which is a special case of the standard 'cross-entropy' (negative binomial log-likelihood) loss function [11] for binary classification. Its class labels are the 'real-fake' or 'fake-real' orderings of an $\{x, \mathrm{P}x\}$ or $\{\mathrm{P}x, x\}$ pair, where each pair contains both a real observed event $x$ and its parity-flipped 'fake' representation $\mathrm{P}x$. Since the function being trained is parity odd, taking the form of equation (4.1), this loss is the same whether looking at orderings $\{x, \mathrm{P}x\}$ with label real-fake or $\{\mathrm{P}x, x\}$ with label fake-real. We can therefore make every label real-fake without consequence, and use the loss function

$$\mathcal{L}_{\text{which-is-real?}} = -\frac{1}{N}\sum_{i=1}^{N} \log p(\text{real-fake} \mid \{x_i, \mathrm{P}x_i\}) \tag{4.6}$$

$$= -\frac{1}{N}\sum_{i=1}^{N} -\log(1 + e^{-f(x)}) \tag{4.7}$$

for a batch of $N$ events, which follows from our choice to convert $f$ to a probability through the logistic function. Functional forms for $f(x)$ are assigned by machine learning models through their stochastic optimizations of this loss function.

Just as classifiers end up learning divergences between data distributions, this process learns divergences between the data distribution and its parity flipped version. Since probabilities assigned by classifiers are invertible to likelihood ratios, classifiers approximate optimal event variables for separating their target classes, and indeed they have become standard practice for this task in high energy physics. By the same reasoning, the parity-odd functions constructed here approximate optimal event variables for separating events from their parity-mirror images.

As an evaluation metric, we compare each learned model to the parity-symmetric hypothesis, which assigns probability $p_{\text{sym}} = 1/2$ to each ordering. The mean log-likelihood ratio between these models is

$$Q = \frac{1}{N}\sum_{i=1}^{N} \log p(\text{real-fake} \mid \{x_i, \mathrm{P}x_i\}) - \log \frac{1}{2}, \tag{4.8}$$

which is positive only if the parity-odd model has made better predictions than the parity-even $p_{\text{sym}}$.

For perfect classification, where each event is unambiguously distinguished from its parity-flipped counterpart, $p(\text{real-fake} \mid \{x_i, \mathrm{P}x_i\}) = 1$ for all events, and $Q$ has a maximum of $\log 2 \approx 0.693$. If both asymmetric and symmetric models predict equally well, then $Q$ takes a value of 0, and $Q$ is not bounded from below.

A positive $Q$ with sufficiently large $N$ is evidence for parity violation, since they imply a large likelihood ratio. We also present error bars for estimates of the limiting value of $Q$ as $N \to \infty$, which are constructed in the usual way from the mean and standard deviation of its summands.

## 4.3 Machine learning models

This section details the BDT, NN and CNN machine learning algorithms we use for the results in this paper. The dataset is split into subsets training : validation : testing in the ratio 60 : 20 : 20. Initial tuning was performed on the training and validation sets, primarily aiming for sensitivity to the $\lambda_{\mathrm{PV}} = 1$. Early stopping methods are used to increase robustness to data from different models. The testing set was not looked at until all models and data were finalized in preparation of the results presented in this paper.

For input features, both the BDT and NN models use only jet momenta from the truth-jet or reco-jet datasets in the rotation-invariant coordinates $\vec{q}_a^{ji}$ of section 4.1.1. Where no fourth jet exists, we set its momenta to zero. No derived features are included. Separately, the CNN model processes only image data.

### 4.3.1 Boosted decision tree

The BDT uses XGBoost [12] in its scikit-learn interface [13, 14] with the loss function of equation (4.6) implemented as described in [5]. Tuning found that parameters that slow the learning process were effective, plausibly due to the subtlety of parity violation in the PV-mSME; all BDTs are trained with a learning rate of 0.1, with `n_estimators=1000`, `min_child_weight=10000`, and `tree_method="hist"`.

To implement early stopping, we evaluate every 50th iteration of the BDT on validation data, and choose the iteration with the best $Q$ on the validation set to use in testing.

### 4.3.2 Neural network

The NN is a multilayer perceptron with ReLU activation functions, three hidden layers of widths $(100, 100, 10)$, and $50\%$ dropout between the second and third hidden layers. It is implemented using the Haiku [15] neural network library in JAX [16] and optimized with Adam [17] (default settings and learning rate 0.001) implemented in Optax [18]. Network inputs are pre-scaled by a mean and standard deviation in each coordinate.

Training is performed in steps with $10\,000$ examples per batch, and is evaluated on a validation set after each 1000th step. Training is terminated when the validation score does not increase for 10 consecutive rounds of evaluation, up to a maximum of 100 rounds in total.

### 4.3.3 Convolutional neural network

The CNN is implemented using PyTorch [19]. It is trained with Adam [17] using a learning rate of 0.001 and L2 regularization penalty of 0.1. The CNN design consists of two convolutional layers with $5 \times 5$ kernels, each outputting 6 channels. These are followed by two fully connected layers of widths 96 and 10. The Leaky ReLU activation function is used, with a negative slope of 0.01, to generate non-linearity between layers. Invariance to rotations are built into the CNN design as described in section 4.1.2.

We again use early stopping to train on batches of 512 images until the validation score saturates, The validation score $Q$ is evaluated every 1000th step, and termination occurs when it has not increased for 10 consecutive evaluation rounds.
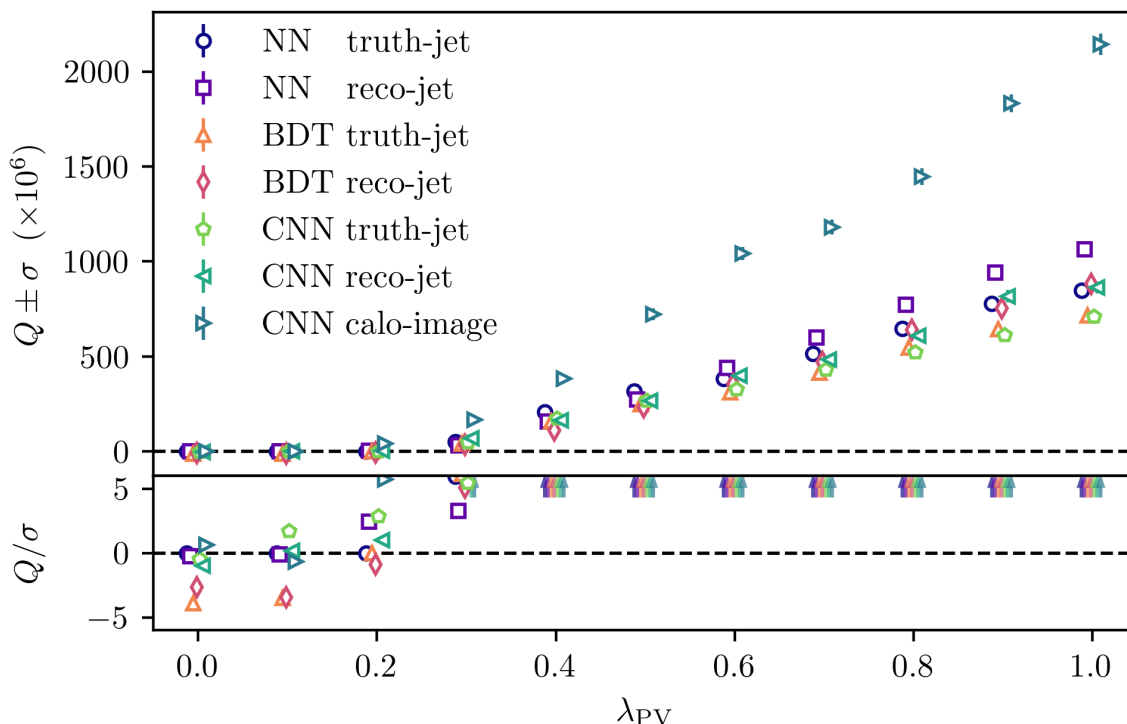
**Figure 5**. Test set evaluation score $Q$ as a function of the coupling $\lambda_{\mathrm{PV}}$, displayed for BDT, NN and CNN models applied to truth-jet, reco-jet and calorimeter-image datasets. A new model is trained and tested for each value of $\lambda_{\mathrm{PV}}$. All points use approximately two million events in their testing sets. Markers at each $\lambda_{\mathrm{PV}}$ point are slightly offset to aid visibility.

## 5 Results

Test results displayed in figure 5 show positive $Q$ values, which demonstrate sensitivity of our method to parity violation in simulated PV-mSME data for $\lambda_{\mathrm{PV}} \approx 0.3$ and above. All models and data representations are effective, but some are more effective than others. As might be hoped, $Q$ are seen to increase with $\lambda_{\mathrm{PV}}$.

A representative output distribution from the NN trained on reco-jet outputs with $\lambda_{\mathrm{PV}} = 1$ is shown in figure 6. On the right of this plot, the variable is re-scaled to target a uniform distribution, as in figure 1. The clear asymmetry between the two halves of the histogram demonstrates that parity violation in this dataset is visible with the trained event variable. Here, the evaluation score (from equation (4.8)) is $Q = (1.06 \pm 0.03) \times 10^{-3}$ with 2.3 million testing data. Since the training and validation sets consume four times as many data again, this corresponds to a luminosity about of 53 fb$^{-1}$ at the Standard Model cross-section of 0.22 nb.

It can be helpful to understand some of what a classifier is doing. For this NN model and dataset, visualizations of what has been learnt in kinematic space are given in appendix A.2.

Despite the considerable noise from the various phases of physics and detector simulation, there is a trend for models to perform *better* on reco-jets than truth-jets. Not only is the method robust to simulation and reconstruction noise, but it also appears to be able to scrape additional information from that noise!
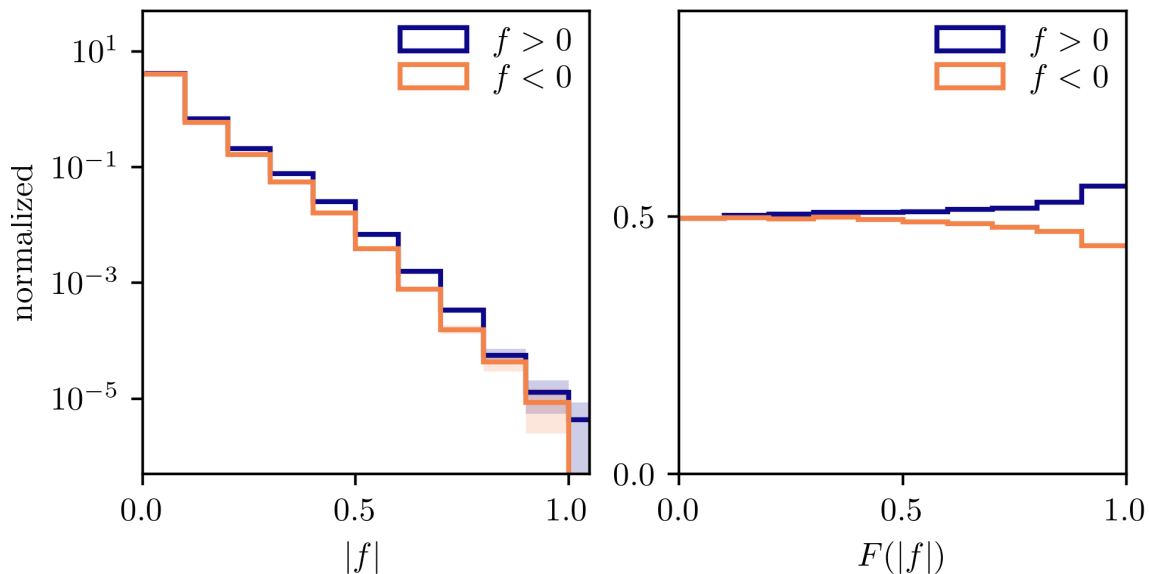
**Figure 6**. Distribution of the parity-odd NN variable $f$: (left) unscaled, and (right) transformed to target a uniform distribution. The positive and negative halves of the distributions are overlaid. Testing data are plotted with statistical error bars.

Flavour information may play a significant part; flavour is hidden from the truth-jet event variables, but does affect showering and hadronization into reco-jets. Different characteristics between gluons and the quark flavours may leave traces in the distributions of reconstructed jets, which the algorithms see. Further study of flavour information in truth data, presented in appendix A.3, finds that flavour information greatly increases sensitivity to parity violation.

Similarly, effects due to colour connections between partons may also play a role. The strikingly stronger performance for the CNN with calo-images is consistent with these ideas, as it preserves some more details than clustered jets. The CNN working with calo-images additionally accesses softer and more numerous jets than the other representations, which only have jets selected above $p_T$ thresholds.

However, this discussion of information content may be too hasty; since all results depend entirely on the successes of learning algorithms, all differences can also be explained by how well those algorithms have performed. That performance is influenced both by algorithm design and tuning (on the training and validation sets), which was conducted manually by us as users.

## 6 Discussion and conclusions

We have demonstrated a method to perform model-independent searches for parity-violating non-Standard-Model physics in LHC data. This method works by training parity-odd event variables on one dataset, and evaluating them on another. Various technologies may be used to implement these parity-odd functions — BDTs, NNs, and CNNs are all demonstrated to be effective.

Previous work [4, 5] has been extended with closer approximations to real particle physics data by simulating collider events in a parity violating quantum field theory, with reconstruction from a detector simulation and approximated triggering. We see from figure 1 that this method achieves greater sensitivity than the previous search for parity violating physics at the LHC [3].

Parity violation may manifest itself in unexpected and unforeseen ways in nature, and in ways that are detectable at the LHC. Indeed, parity-violating signals may already have been produced, but they have not yet been sought; to find parity violation, it is first necessary to search. The methods developed in this paper shrink a boundless space of previously unexplored parity violating models to a manageable size.

## Acknowledgments

## A  Additional content

### A.1  Interpretations

The two lines in figure 6 represent histogram yields in different bins, which are separated by the parity flip operation. An example statistical interpretation of histograms like these is illustrated in figure 7, which uses the common method of fitting two likelihood models; in this case, one model is one parity-even and the other is not. The parity even model must assign equal background expectations to both bins, up to modifications by known biases or systematic uncertainties. The parity-odd model can assign different expectations, so fits both bins perfectly.

### A.2  Output visualization

Visualizations of what has been learned by an algorithm can be interesting and useful checks. In this appendix, we visualize the NN parity variable trained on reco-jet data with scatter plots in which every point $x$ is coloured by its NN output $f(x)$. These plots are shown in figure 8.

Most events have $|f(x)| \ll 1$, as shown in figure 6, so parity-violating phase space is highlighted by the opacity of scattered data points with $|f(x)|$.

In the rotation-invariant coordinates $\vec{q}_a^{j_i}$, described in section 4.1.1, only the non-zero $\hat{y}$ components are parity odd; all other event variables are parity even. To show parity violation, we therefore include at least one $y$ component in each plot. Any scatter plot in only parity-even variables cannot see parity violation and appears as a monochrome blob.
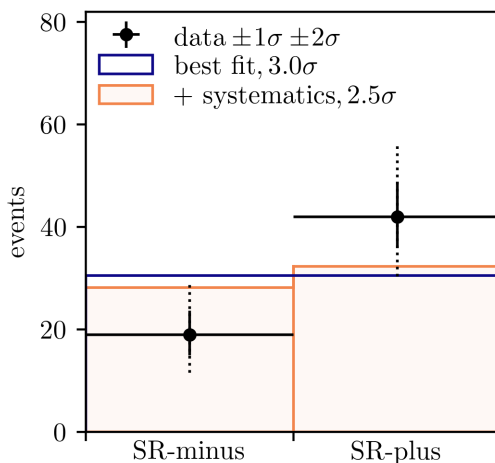
**Figure 7**. Illustrated analysis of parity-opposite histograms in a minimal two-bin example. A parity-asymmetric model assigns separate expectations in each bin, so is able to fit both perfectly. The purple line shows the best fit from a parity-symmetric model, for which both expectations must be equal. The orange histogram modifies the parity-symmetric model with constrained 'systematic' variations, which can allow small amounts of parity violation to account for known biases or uncertainties; these variations give a slightly better fit, reducing the significance. We assign significances $\sigma$ from maximum likelihood ratios by $\sigma^2 = -2 \log[\max L_1(\theta)/\max L_2(\theta)]$, where $L_1(\theta)$ and $L_2(\theta)$ are the likelihood functions of two alternative models.

### A.3 Truth information

Although parton flavour and helicity are features of the simulations, they are not included in the truth-jet dataset. Flavour and helicity are, however, parity-even event variables which could be approximately 'tagged' for with data in appropriate detectors. This appendix investigates the effects of including flavour and helicity as additional features in variable training. Flavour is encoded with the PDG Monte Carlo Particle Numbering Scheme [20], and helicity is encoded by $\pm 1$, which are both pre-scaled by a mean and standard deviation for the NN.

Flavour information drastically boosts sensitivity, as demonstrated in figure 9 for the same $\lambda_{\mathrm{PV}} = 0 \ldots 1$ scan as used figure 5.

To investigate both the switch-on in sensitivity and the joint effect with helicity, a narrower scan is performed in figure 10 for $\lambda_{\mathrm{PV}} = 0 \ldots 0.2$. Helicity appears to not help, and in fact hurts performance, as might be expected from the inclusion of irrelevant information.

### A.4 Rotated PV-mSME

The use of a constant coupling matrix $(c_A)_{\mu\nu}$ in section 3 means that the main PV-mSME results of the paper assume either: (i) that the detector does not rotate in space; or (ii) that the detector rotates with the earth, but the data being analysed are only those which were recorded when the detector had a particular orientation in space; or (iii) that the frame in which the Lorentz-violating effects are constant is 'earth centric' or 'dragged around' by the earth. The first is not physical, the second would decimate the cross-section available for analysis, while the third is unlikely on physical grounds. Despite those drawbacks, we
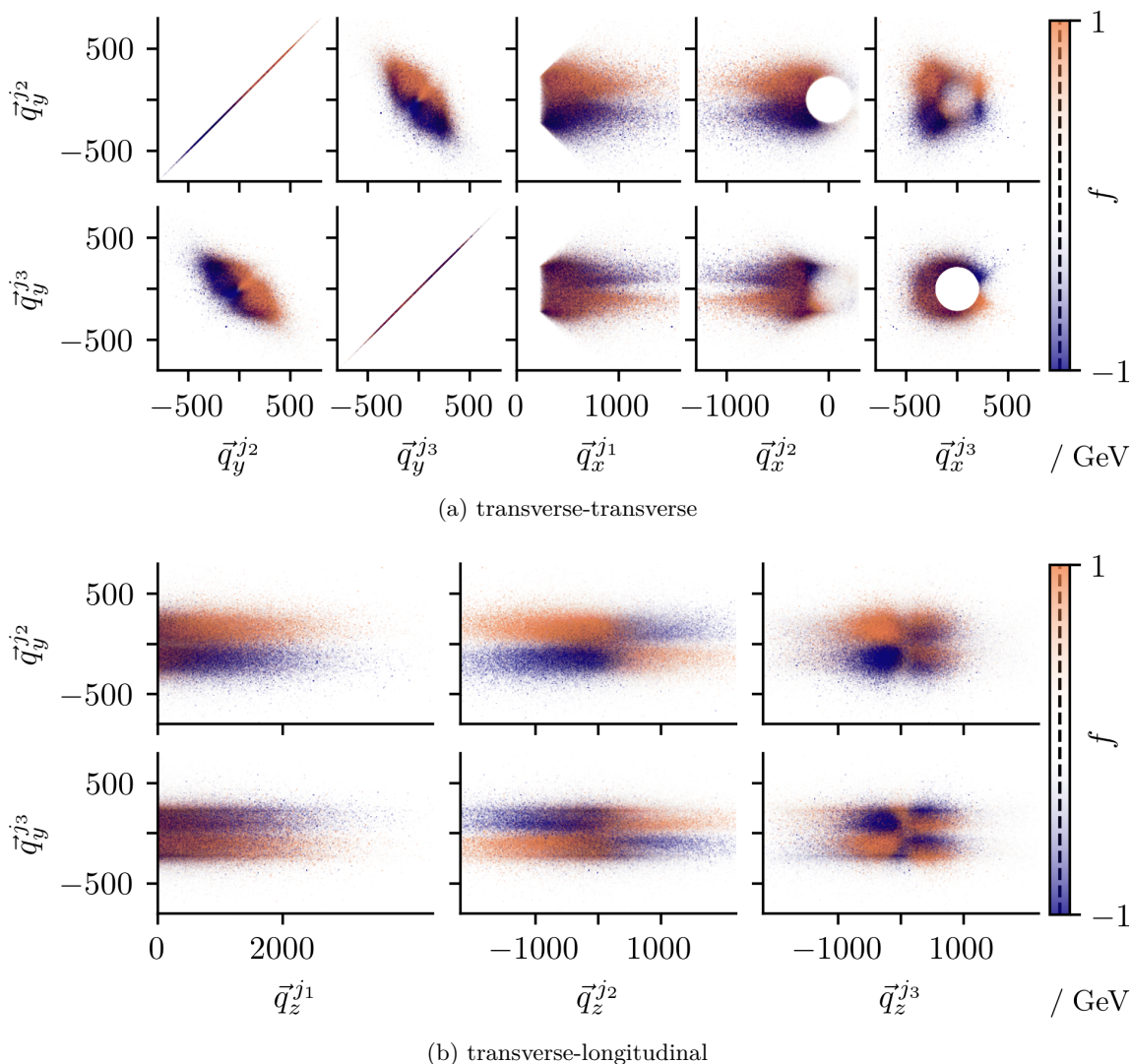
(a) transverse-transverse



(b) transverse-longitudinal

**Figure 8**. Kinematic distributions of the parity-odd variable $f$ from the reco-jet NN with $\lambda_{\mathrm{PV}} = 1$ in the rotation-invariant $\vec{q}_a^{j_i}$ coordinates defined in section 4.1.1. The plots contain one million data from the testing dataset, scattered with colour and transparency set by the value of $f$. Although each point has relatively low transparency, overlaid points accumulate to the observed distributions. As seen from figure 6, the vast majority of data have small $|f|$; only the most parity-violating phase space is visible here. If the NN is accurate, then colours in blue ($f < 0$) should have lower density than those in orange ($f > 0$); as seen, the difference is subtle.

justify the decision to use a constant coupling matrix $(c_A)_{\mu\nu}$ in section 3 on the grounds that it allows a proof-of-principle to be demonstrated in the simplest manner with the fewest distractions.

However: since the earth does rotate, it is interesting to ask what magnitude of degradation would be seen in our results if the earth's rotation were allowed to induce a corresponding time-dependence into the couplings $(c_A)_{\mu\nu}$ in the detector frame. Practically, this means adding the orientation of the detector to each event in the data so that the
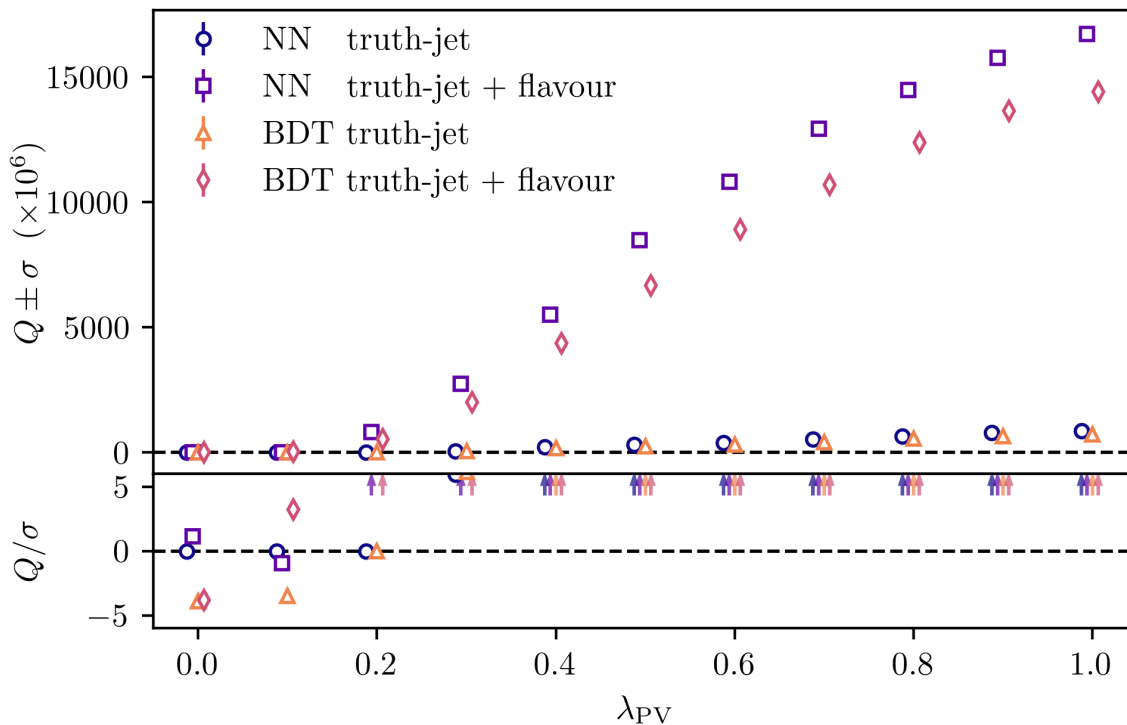
**Figure 9**. Test set evaluation score $Q$ as for small values of $\lambda_{PV}$, displayed for BDT and NN models applied to truth-jet data with and without the inclusion of parton flavour information. Small values of $\lambda_{PV}$ are further investigated in figure 10. All points use approximately two million events in their testing sets. Markers at each $\lambda_{PV}$ point are slightly offset to aid visibility.

learning algorithms can learn about alignment-dependent parity violation. If, for example, the data were only parity violating in alignments close to a special axis, the algorithms could learn that and automatically implement an effective selection by mapping all other data to 0 in their parity-odd output. And they could learn about other alignment-dependent parity-violating effects if they existed.

To demonstrate this approach, we imagine a simple case of an East-West aligned detector (as ATLAS is, approximately), and consider daily rotations of a planet which is otherwise floating stationarily in space (that is, not otherwise orbiting a star about a different axis). Since our analysis is invariant to rotations in $\phi$ (about the beam axis), this daily rotation has an effect equivalent to rotating the detector about a single transverse axis (by an angle $\theta$), independent of latitude.[7]

To study this example, we simulate truth-jet datasets from the PV-mSME with its coupling matrix rotated for 24 different earth-rotation angles $\theta$ spaced evenly in $[0, 2\pi)$,[8] such that an angle of 0 recovers the PV-mSME. In addition to the invariant jet momentum

---

[7]The latitude of Geneva is therefore not an input to the analysis, and no generality is lost as a consequence. Put another way: any approximately east-west-aligned and $\phi$-symmetric detector may be assumed to be near the North Pole, without loss of generality.

[8]Continuous rotations would be more realistic, but we opt for this discrete approximation to avoid substantial changes to the simulation software.
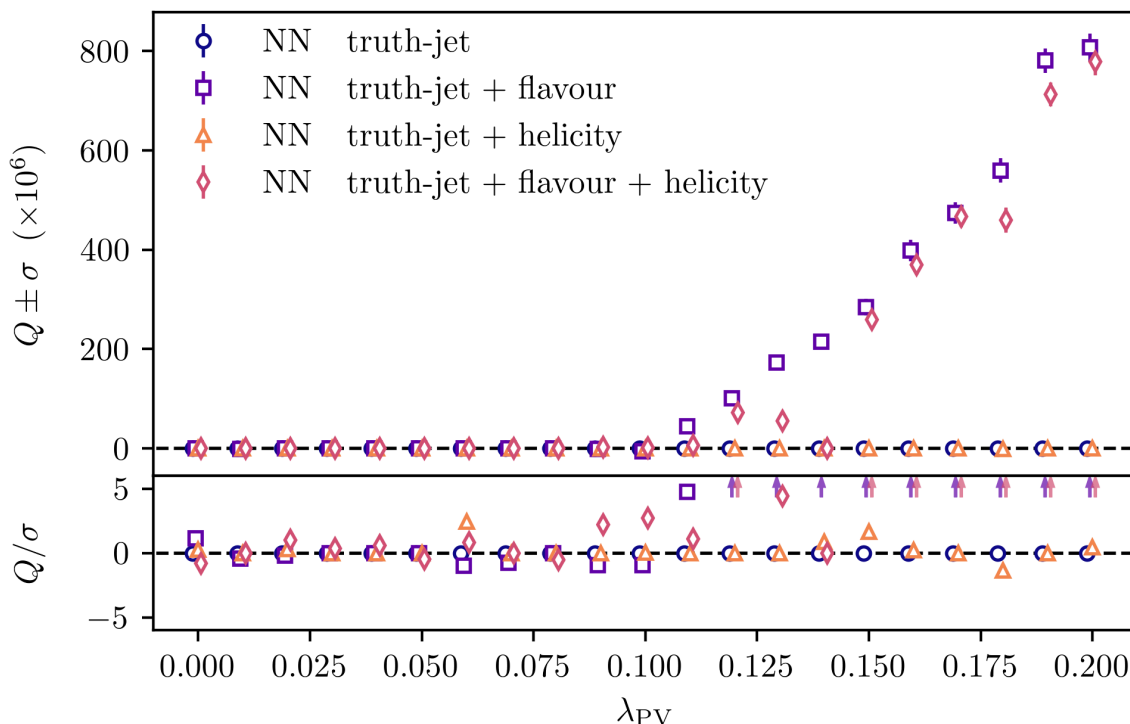
**Figure 10**. Test set evaluation score $Q$ for small values of the coupling $\lambda_{PV}$, displayed for the NN model applied to truth-jet data with and without the inclusion of parton helicity and flavour information. This figure explores small values of $\lambda_{PV}$ to observe the switch-on of sensitivity in the leftmost parts of figure 9. All points use approximately two million events in their testing sets. The algorithm has performed poorly for some points; it is not perfect! Markers at each $\lambda_{PV}$ point are slightly offset to aid visibility.

features described in section 4.1.1 we also encode the $\theta$ rotation of each event with the two features $\sin\theta$ and $\cos\theta$.[9] Training, validation, and testing sets are prepared by subsampling and shuffling portions of these data; we take samples from each rotation with rates proportional to their production cross-sections, which we calculate from MadGraph and the efficiency of our kinematic selections.

Cross-sections in the rotated PV-mSME vary hugely with $\theta$. (Although this anisotropy is a clear sign of new physics, it is not a direct sign of parity violation.) These variations are approximately sinusoidal, with minima at $0$ and $\pi$, maxima at $\pi/2$ and $3\pi/2$, and a maximum-over-minimum ratio of 13.5. Since parity violation is strongest at $0$ and $\pi$ (as designed in appendix C), this drastically dilutes the observable parity violation. When training our standard learning models (described in section 4.3) on these rotated data, they report $Q \leq 0$ in validation, so not find parity violation.

After tuning new model designs towards these different data, however, we do find some evidence for parity violation in these rotated samples. With the diluted parity violation in

---

[9]This $(\sin\theta, \cos\theta)$ pair-encoding avoids the discontinuities of $\theta$ (at $0$ or $2\pi$), and relates linearly to dot products with directions. Empirically, it appeared to improve results in preliminary tests on the validation set.

the rotated sample, the standard NN tended to collapse towards all-0 outputs; this error was avoided by simplifying its design to three hidden layers of widths $(20, 20, 10)$, and no dropout layer. For statistically significant results, we also triple the data size from 10 million (as stated in section B) to 30 million, with the same training : validation : testing split of $60 : 20 : 20$.

After finalizing this tuned NN, we find $Q = (7.6 \pm 2.6) \times 10^{-6}$ in the testing set, which is a positive indication of parity violation. This result is weaker than our other examples of evidence for parity violation, such as the results displayed in figure 5, but still corresponds to a substantial log likelihood ratio of 45.6 in favour of the parity violating NN model.

# B   Simulation details

We simulate parton scattering at leading order using MadGraph5_aMC@NLO v3.3.0 [21]. By default, MadGraph evaluates its matrix elements in the centre of mass frame; since the mSME is not Lorentz invariant, we modify code generated by MadGraph to evaluate its matrix elements in the lab frame instead. We use a custom implementation of the PV-mSME quark-quark-gluon vertices in a model imported into MadGraph. The simulated hard scatter is of LHC-like proton-proton collisions at $\sqrt{s} = 13$ TeV scattering to three or four gluons or light quarks (up, down, strange, or charm) each with $p_\mathrm{T} > 200$ GeV and $|\eta| < 3.2$, and selected for CKKW-L merging with $k_\mathrm{T}^\mathrm{Durham} = 200$ GeV [22]. Other MadGraph parameters of are left at their default values — in particular, partonic jets must be separated by $\Delta R_{jj} > 0.4$, and we use the NNPDF2.3LO1 PDF set [23].

The parton shower and hadronization are simulated with PYTHIA 8.235 [24]. Detector reconstruction is performed with the Delphes 3.5.0 [25] approximation to the ATLAS detector. The output from Delphes includes anti-$k_\mathrm{T}$ jets [26] at $\Delta R = 0.4$, and energy deposits in the calorimeters. Delphes performs jet clustering with the FastJet package [27, 28].

The effect of pileup is simulated in Delphes by overlaying a mean of 50 minimum-bias events to each event. Delphes sub-samples from a single batch of pileup events for each batch of events processed; this raises a risk of bias if pileup events are repeated enough for our learning algorithms to recognize them individually. We mitigate this risk by using large numbers of pileup events and not sharing any between training and testing datasets. We generate events in batches of 200 000, and for each batch simulate 200 000 uniquely-seeded pileup events.

To approximate the efficiency plateau of a three-jet trigger, events are accepted only if they are reconstructed with at least three jets with $p_\mathrm{T} > 220$ GeV and $|\eta| < 2.8$. Additional reconstructed jets are included only if they satisfy $p_\mathrm{T} > 30$ GeV and $|\eta| < 2.8$.

We generate 500 batches of 200 000 truth events for each model specification. These events acquire weights in their processing to reconstruction level, which we unweight by downsampling with respect to the maximum of all weights in the simulation. Depending on $\lambda_\mathrm{PV}$, around 20–30% of events survive triggering and kinematic selections, and of those accepted around 60–40% survive unweighting. After these reductions, the reco sets include 9.2 million events at $\lambda_\mathrm{PV} = 0$, and 11.5 million events at $\lambda_\mathrm{PV} = 1$. Truth sets have more events available since they do not suffer from unweighting, so to approximately match reco we use exactly 10 million truth events for each model.

## C  PV-mSME details

### C.1  Notation

Our notation is related to equation 11 of [6], which defines coupling matrices $(c_X)_{\mu\nu AB}$ for $X \in \{Q, U, D\}$ with generation indices $A$ and $B$. The PV-mSME defines these coupling matrices to be diagonal in the first two generations: $(c_X)_{\mu\nu AB} = (c_X)_{\mu\nu} \operatorname{diag}(1, 1, 0)_{AB}$. Couplings split into an axial part $(c_A)_{\mu\nu} = (c_U)_{\mu\nu} - (c_Q)_{\mu\nu}$, and a vector part which we define to vanish $(c_V)_{\mu\nu} = (c_D)_{\mu\nu} + (c_Q)_{\mu\nu} = 0$ The PV-mSME therefore couples quarks within the first two generations only, has the same couplings within each of those generations, and does not mix between them.

After some rearrangement, the quark-quark-gluon Feynman Rule of equation (3.2) can be read from the Lagrangian in these terms.

As stated in [6], the full coupling matrices must be Hermitian in their generation indices and traceless in their Lorentz indices. Since our couplings are on the generation diagonal, they must be real to satisfy the Hermitivity constraint.

### C.2  Couplings

Within those real and traceless constraints, the specific form of the PV-mSME coupling matrix $(c_A)_{\mu\nu}$ of equation (3.3) is chosen primarily from empirical results of numerical experiments. However, it has some plausible justifications, which we develop here.

We seek couplings which not only generate parity violation, but which make it visible when blind to rotations by $\phi$ about the beam axis and by $180°$ about a perpendicular axis (to swap the beams). Variation of a differential cross-section under these rotations is of no benefit, since the rotation-invariant observer sees only an average. Variations may plausibly, however, cause parity violation to be 'washed out' from a rotation-invariant perspective. We therefore attempt to choose couplings which minimize the dependence of cross-sections on these rotations.

The PV-mSME introduces terms into matrix elements that do not appear in the Standard Model alone. Among these non-standard terms are expressions of the form $(c_A)_{\mu\nu} p^\mu q^\nu$, which we abbreviate here as $\xi = p^T C q$. Parameterizing axial and beam-swap rotations with the matrix

$$S(\phi, \pm) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & \mp\sin\phi & 0 \\ 0 & \sin\phi & \pm\cos\phi & 0 \\ 0 & 0 & 0 & \pm 1 \end{pmatrix}, \tag{C.1}$$

$\xi$ may be seen to transform to $\xi(\phi, \pm) = p^T S(\phi, \pm)^T C S(\phi, \pm) q$. If one were to require that $\xi(\phi, \pm)$ be independent of $\phi$ for all $p$ and $q$, then one would need to take $C$ to be of the form

$$C = \begin{pmatrix} c_{00} & 0 & 0 & c_{03} \\ 0 & c_{22} & -c_{21} & 0 \\ 0 & c_{21} & c_{22} & 0 \\ c_{30} & 0 & 0 & c_{33} \end{pmatrix}. \tag{C.2}$$
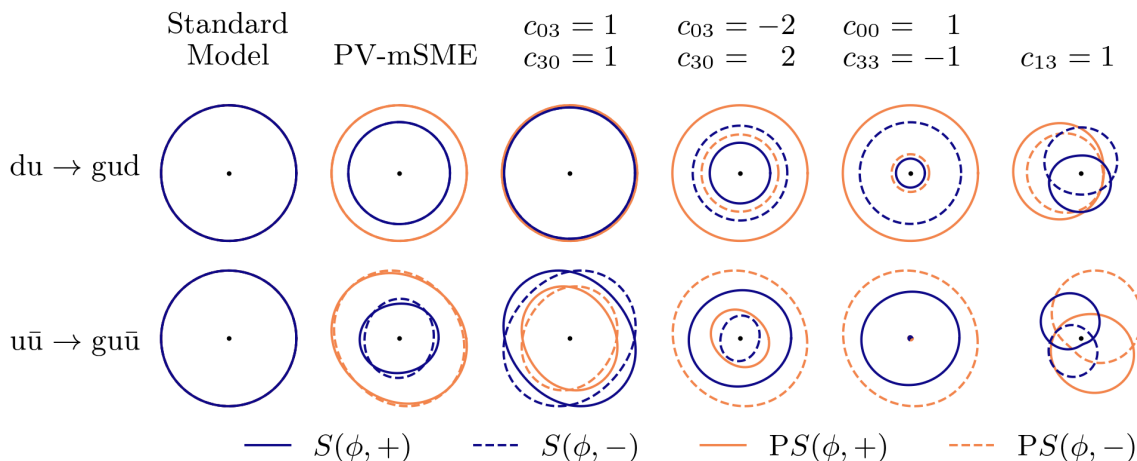
Standard Model  PV-mSME  $c_{03} = 1$ $c_{30} = 1$  $c_{03} = -2$ $c_{30} = 2$  $c_{00} = 1$ $c_{33} = -1$  $c_{13} = 1$

du → gud

uū → guū

—— $S(\phi, +)$ ---- $S(\phi, -)$ —— $\mathrm{P}S(\phi, +)$ ---- $\mathrm{P}S(\phi, -)$

**Figure 11**. Helicity-summed, magnitude-squared matrix elements $|\mathcal{M}|^2$ displayed as radial distances, such that differences between the orange and blue lines indicate parity violation. Momenta are rotated by $S(\phi, \pm)$, where $\phi$ is the angle from the vertical axis. Parton distribution functions are invariant under these rotations, so $|\mathcal{M}|^2$ is proportional to the differential cross-section. The Standard Model is invariant under rotation and parity operations. The PV-mSME has $(c_A)_{\mu\nu}$ couplings from equation (3.3); latter columns modify the $(c_A)_{\mu\nu}$ coupling matrix by their stated assignments. Top: $\vec{p}_\mathrm{g} = (200, 400, -600)$, $\vec{p}_\mathrm{u} = (-200, -150, 200)$, and $\vec{p}_\mathrm{d} = (0, -250, 1200)$ GeV. Bottom: $\vec{p}_\mathrm{g} = (100, 250, 250)$, $\vec{p}_\mathrm{u} = (200, 250, 400)$, and $\vec{p}_{\bar{\mathrm{u}}} = (-300, -500, -300)$ GeV.

If one were to subsequently require that $\xi(\phi, \pm)$ be proportional to $\pm$ for all $p$ and $q$, then the diagonal elements of $C$ would be set to zero, resulting in:

$$\xi(\phi, \pm) = \pm(p_0 q_3 c_{03} + p_3 q_0 c_{30} + p_2 q_1 c_{21} - p_1 q_2 c_{21}) . \tag{C.3}$$

This choice (fixed magnitude but not fixed sign) does not lead matrix elements to be invariant under beam swaps, but does give opportunities for them to have reduced dependence on this operation, perhaps arising from even powers or cancellations against sign-flipped terms.

Pleasingly, the form shown in (C.3) contains both parity even ($p_1 q_2$, $p_2 q_1$) and parity odd ($p_0 q_3$, $p_3 q_0$) terms. Violation of parity requires interference between parity-odd and parity-even terms; parity odd terms alone are not sufficient, since cross-sections are proportional to the modulus of the matrix element squared.

Requiring $\xi(\phi, \pm)$ to be parity asymmetric (for some $p$ and $q$) therefore forces $c_{21}$ and at least one of $c_{30}$ and $c_{03}$ to be non-zero. Without loss of generality we can therefore choose $c_{21} = 1$, absorbing any overall scale into $\lambda_\mathrm{PV}$, and are left with the freedom to choose values for $c_{30}$ and $c_{03}$.

In summary, requiring $\xi(\phi, \pm)$ to be independent of $\phi$, invariant in magnitude to $\pm$ flips, and parity asymmetric implies

$$C = \begin{pmatrix} 0 & 0 & 0 & c_{03} \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ c_{30} & 0 & 0 & 0 \end{pmatrix}, \tag{C.4}$$

where at least one of $c_{30}$ and $c_{03}$ is non-zero.

The PV-mSME choice is $c_{30} = 1$ and $c_{03} = -1$. This empirically generates visible parity violation, and has differential cross-sections which are close to rotation-invariant when explored from individual points in phase space. Numerical matrix elements calculated with MadGraph are illustrated at example phase points in figure 11, along with some effects from differently assigned couplings.

In the top row of figure 11, Standard Model and PV-mSME rings appear to be rotation-invariant, but they are not exactly. On closer inspection, the PV-mSME shows sinusoidal variations of a few parts per million, whereas the Standard Model varies only by numerical rounding errors in parts per $10^{15}$.

# References

[1] C.G. Lester, *Chiral Measurements*, `arXiv:2111.00623` [INSPIRE].

[2] Bram Stoker, *Dracula*, Archibald Constable and Company, Edinburgh U.K. (1899).

[3] C.G. Lester and M. Schott, *Testing non-standard sources of parity violation in jets at the LHC, trialled with CMS Open Data*, *JHEP* **12** (2019) 120 [`arXiv:1904.11195`] [INSPIRE].

[4] C.G. Lester and R. Tombs, *Stressed GANs snag desserts, a.k.a Spotting Symmetry Violation with Symmetric Functions*, `arXiv:2111.00616` [INSPIRE].

[5] R. Tombs and C.G. Lester, *A method to challenge symmetries in data with self-supervised learning*, `arXiv:2111.05442` [INSPIRE].

[6] D. Colladay and V.A. Kostelecky, *Lorentz violating extension of the standard model*, *Phys. Rev. D* **58** (1998) 116002 [`hep-ph/9809521`] [INSPIRE].

[7] V.A. Kostelecky and N. Russell, *Data Tables for Lorentz and CPT Violation*, *Rev. Mod. Phys.* **83** (2011) 11 [INSPIRE].

[8] V.A. Kostelecký, E. Lunghi and A.R. Vieira, *Lorentz violation and deep inelastic scattering*, *Phys. Lett. B* **769** (2017) 272 [`arXiv:1610.08755`] [INSPIRE].

[9] ATLAS collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, 2008 *JINST* **3** S08003.

[10] C.G. Lester, W. Haddadin and B. Gripaios, *Lorentz and permutation invariants of particles III: Constraining nonstandard sources of parity violation*, *Int. J. Mod. Phys. A* **37** (2022) 2250093 [`arXiv:2008.05206`] [INSPIRE].

[11] Kevin P. Murphy, *Machine Learning a probabilistic perspective*, Adaptive Computation and Machine Learning series, MIT Press, Cambridge, U.S.A. (2012) [ISBN: 9780262018029].

[12] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, `arXiv:1603.02754` [INSPIRE].

[13] Fabian Pedregosa et al., *Scikit-Learn: Machine Learning in Python*, J. Mach. Learn. Res. **12** (2011) 2825 [`arXiv:1201.0490`].

[14] Lars Buitinck et al., *API design for Machine Learning software: experiences from the scikit-learn project*, in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, Prague Czech Republic, September 23 – September 27 2013, pp. 108 [arXiv:1309.0238].

[15] Tom Hennigan, Trevor Cai, Tamara Norman and Igor Babuschkin, *Haiku: Sonnet for JAX*, (2020) [http://github.com/deepmind/dm-haiku].

[16] James Bradbury et al., *JAX: composable transformations of Python+NumPy programs*, (2018) [http://github.com/google/jax].

[17] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, (2014) [arXiv:1412.6980] [INSPIRE].

[18] Matteo Hessel et al., *Optax: composable gradient transformation and optimisation, in JAX!*, (2020) [http://github.com/deepmind/optax].

[19] Adam Paszke et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, Advances in Neural Information Processing Systems, Curran Associates, Inc. (2019) [arXiv:1912.01703].

[20] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *PTEP* 2020 **08** 083C01.

[21] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[22] L. Lönnblad, *Correcting the color dipole cascade model with fixed order matrix elements*, *JHEP* **05** (2002) 046 [hep-ph/0112284] [INSPIRE].

[23] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244 [arXiv:1207.1303] [INSPIRE].

[24] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, arXiv:2203.11601 [INSPIRE].

[25] DELPHES 3 collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [arXiv:1307.6346] [INSPIRE].

[26] M. Cacciari, G.P. Salam and G. Soyez, *The anti-$k_t$ jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[27] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[28] M. Cacciari and G.P. Salam, *Dispelling the $N^3$ myth for the $k_t$ jet-finder*, *Phys. Lett. B* **641** (2006) 57 [hep-ph/0512210] [INSPIRE].