The DL Advocate: playing the devil's advocate with hidden systematic uncertainties

Golutvin, Andrei ; Iniukhin, Aleksandr ; Mauri, Andrea ; Owen, Patrick ; Serra, Nicola ; Ustyuzhanin, Andrey

THE EUROPEAN
PHYSICAL JOURNAL C

The DL Advocate: playing the devil's advocate with hidden systematic uncertainties

Andrei Golutvin[1]**, Aleksandr Iniukhin**[2,a]**, Andrea Mauri**[1]**, Patrick Owen**[3]**, Nicola Serra**[3]**,
Andrey Ustyuzhanin**[4,5]

[1] Imperial College of London, London, UK
[2] Yandex School of Data Analysis, Moscow, Russia
[3] Physik-Institut, Universität Zürich, 8057 Zurich, Switzerland
[4] Constructor University, Campus Ring 1, 28759 Bremen, Germany
[5] Institute for Functional Intelligent Materials, National University of Singapore, Singapore 117544, Singapore

Abstract We propose a new method based on machine learning to *play the devil's advocate* and investigate the impact of unknown systematic effects in a quantitative way. This method proceeds by reversing the measurement process and using the physics results to interpret systematic effects under the Standard Model hypothesis. We explore this idea with two alternative approaches: the first one relies on a combination of gradient descent and optimisation techniques, its application and potentiality is illustrated with an example that studies the branching fraction measurement of a heavy-flavour decay. The second method employs reinforcement learning and it is applied to the determination of the P_5' angular observable in $B^0 \to K^{*0}\mu^+\mu^-$ decays. We find that for the former, the size of a hypothetical hidden systematic uncertainty strongly depends on the kinematic overlap between the signal and normalisation channel, while the latter is very robust against possible mismodellings of the efficiency.

1 Introduction

The potential for hidden systematic effects, so-called "unknown-unknowns", in a physics measurement is difficult to address. This problem can be alleviated by an independent confirmation in a different experiment. A good example, in particle physics, is the discovery of the Higgs boson, whereby the simultaneous announcement from both the ATLAS and CMS experiments of an excess led to high confidence in the discovery [1,2]. However, the size of the collaborations and the complexity of the experiments involved can make such independent confirmations prohibitively expensive for future particle physics experiments. The confidence in the Higgs discovery was also aided by the fact that it was not completely unexpected: it was predicted by the Standard Model (SM) of particle physics to emerge with a distinctive pattern of couplings to the known particles.

The discovery of physics beyond the SM is expected to answer many of the open questions of the SM and it is therefore the main focus of experimental particle physics. This discovery will require even more experimental evidence to be confirmed, particularly if it manifests itself in ways that are unexpected. The question in this case is therefore: under which conditions can one claim a physics discovery in an experiment which has unique physics sensitivity and therefore no direct competitors? The answer to this question is normally qualitative, such as seeing a particular control channel pass a set of compatibility tests or observe new physics appear with multiple and complementary experimental signatures. The goal of this paper is to introduce a method that provides quantitative answers to these questions.

The philosophy this paper follows is to apply deep learning techniques to *play the devil's advocate* (DL Advocate), with respect to deviations from the SM. This implies assuming the true value of the parameters of interest is indeed SM-like and determine what experimental effects could cause the observed set of measurements. Concretely, one can consider the set of measurements of the experiment as a system of equations:

Aleksandr Iniukhin and Andrea Mauri contributed equally to this work.

[a] e-mail: aleksandr.iniukhin@cern.ch (corresponding author)

$$\mathcal{F}(\tilde{\eta}, \Omega_i) = M_i \implies \begin{cases} \mathcal{F}(\tilde{\eta}, \Omega_1) = M_1 \\ \mathcal{F}(\tilde{\eta}, \Omega_2) = M_2, \\ \dots \end{cases} \quad (1)$$

where $\mathcal{F}$ represents the measurement process, $\tilde{\eta}$ are parameters in common to all measurements, such as the detector response, and Ω_i are parameters specific to a particular measurement, such as theoretical parameters. When a measurement deviates from its SM prediction, it is tempting to interpret the observed deviation as a sign of physics beyond the SM. However, such an important claim must be supported by an equally strong confidence in the understanding of the experimental apparatus. The idea developed in this paper reverses the classical reasoning and, instead of attributing the observed deviation to physics beyond the SM, it starts from the SM hypothesis by fixing the theory parameters to their predicted values, uses the simulation to model $\mathcal{F}$ and uses a neural network to find possible values of $\tilde{\eta}$ that reproduce the observed measurements M_i. In other words, it tries to find possible detector effects that can cause the observed deviation. For the examples illustrated in this paper, the parameters $\tilde{\eta}$ represent a mismodelling of the detector efficiency, but can be extended to any assumption in the analysis. The system in Eq. 1 could in principle include all measurements by the experiment. However, in practice one can only consider measurements which are correlated either through theory or experiment.

The advantage of this approach is that the resulting values of $\tilde{\eta}$ can then be used to make quantitative predictions of mismodelling that can be falsified by additional cross-checks. There are two categories of information that $\tilde{\eta}$ can represent. One is *high-level* information, such as kinematic information of particles. A mismodelling as a function of a particle momentum would fall into this class of information. Such information can be readily implemented with existing tools, but requires physics intuition and therefore has to be tuned to each specific case. The other category of information is *low-level* quantities, such as the material budget and hit resolution. Exploring low-level information would require tuning simulation in real time and would be a challenge to implement, but would be fully general, applicable for any measurement that relies on simulation.

In recent years increasingly more interest has been devoted to apply machine learning techniques to systematic uncertainties. The main emphasis however has been on the optimisation of statistical inference in the presence of systematic uncertainties (see e.g. [3–5]) or incorporating known systematic effects in simulation (see e.g. [6–8]). In this work we propose a method to bound the size of unknown or underestimated systematic effects. A quantitative and incontrovertible demonstration of the control of systematic uncertainties is in fact essential for any scientific discovery, whose claim must

be supported with a confidence higher than one-in-a-million chance not to be a fluctuation of statistical or systematic origin.

In order to demonstrate the potential of such an approach, we apply this method to a simple toy example of a branching fraction measurement of a particle decay and restrict our attention to a potential mis-modelling of the efficiency. Such measurements are often normalised to a decay mode with a known branching fraction and ideally the same final state as the signal, which cancels systematic uncertainties due to efficiency mismodelling to a high degree. It is therefore an ideal testing ground for our approach. We also consider a more concrete example which is the observable P_5', which arises in the angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ decays and is highly sensitive to physics beyond the SM [9,10]. In both cases we show in a simplified scenario the use of the DL Advocate technique to investigate the impact of possible efficiency mismodelling on the considered measurements. Finally, another interesting system to test this methodology would be the W mass, recently been measured by the CDF Collaboration [11] to be significantly different from previous measurements [12–14] and the Standard Model prediction [15].

This paper is structured as follows: The general idea and the algorithm implementation is described in Sect. 2. The concrete example of the branching fraction measurement of a particle decay is briefly summarised in Sect. 3. An implementation with Deep Reinforcement learning applied to the observable P_5' is discussed in Sect. 4, which is followed by a summary Sect. 5 and a conclusion.

2 Methodology

Measurements in experimental particle physics typically involve the determination of a signal yield, either integrated over a particular decay channel or differentially in regions of phase-space. Due to imperfections of the detection process, the recorded signal yield must be corrected for the finite detector efficiency in order to retrieve the *true* number of signal events originally produced. This efficiency is typically estimated based on simulation or calibration samples.

Taking the example laid out in Eq. 1, we consider a set of measurements M_i, where one is the measurement of interest and the others are control channels which are used to validate the analysis and to check for systematic uncertainties. For each of these measurements there are some observed candidates N_i and an associated efficiency e_i, so that $M_i = \frac{N_i}{e_i}$. The candidates of each channel are characterised by a set of variables (features) such as the kinematics of the produced particles. Differences in these distributions, together with a detection efficiency which can depend on the same kinematic

variables, can result in different total efficiencies between the signal and control channels.

Broadly speaking, mismodelling of the efficiency and unaccounted or mismodelled backgrounds can create a bias in the measurements, which are accounted for by introducing ad-hoc systematic uncertainties. As a consequence, partial or incorrect evaluation of these effects would result in underestimated systematic uncertainties. In this paper we will focus on the role of the efficiency.

We describe possible mismodelling of the efficiency with a weighting function $w(x)$, which depends on the kinematic variables x of the event. Values of $w(x) = 1$ correspond to a perfect modelling of the efficiency, while values below/above unity correspond to efficiency under/over estimated. The key idea is that, while the detector response depends entirely on the kinematics of the single event, the total signal/control channel efficiency can suffer from different biases once integrated over the individual kinematic distribution of each decay channel. We can then define the *true* total efficiency for a given channel i as

$$e_i = \mathbb{E}_{x \sim p(x|i)} \left[w(x) \times \hat{\epsilon}(x) \right] , \tag{2}$$

where $\hat{\epsilon}(x)$ is the per-event *estimated* efficiency in the experiment and the expectation value indicates the weighted average over the kinematic distribution of each decay channel $p(x|i)$. Since our goal is to study the impact of possible mismodelling of the efficiency in a given set of measurements we can safely assume $\hat{\epsilon}(x) = 1$ without loss of generality. This simplifies the expression of the per-channel efficiency to

$$e_i = \mathbb{E}_{x \sim p(x|i)} \left[w(x) \right] \equiv \frac{1}{n_i} \sum_{k=1}^{n_i} w(x_{k,i}), \tag{3}$$

where we approximated the expectation value with a sum over a large number n_i of simulated events $\{x_k\}_i$, where i labels the different decay channels.

Control channels provide important constraints on how well the efficiency is estimated. They are typically selected with topology and kinematics similar to the signal decay mode in order to maximise the phase space overlap between channels. The result of the measurements obtained on such control channels can then be compared to known reference values, e.g. existing precise measurements from other experiments or clean SM predictions. If a good agreement is found, a certain level of confidence can be ascribed to the estimation of the efficiency, at least for what concerns the kinematic regions populated by the control channels. The requirement that measurements performed on the control channels must be compatible with a certain reference can be formulated as

$$M_i \in [M_i^{low}; M_i^{high}] \tag{4}$$

which reduces to

$$e_i \in [V_i^{low}; V_i^{high}], \tag{5}$$

once we restrict the attention to the sole role of the efficiency in the measurement. Here V_i^{low} and V_i^{high} are the values that bound the efficiency for the control measurements to pass scrutiny.

The goal of this paper is to find regions of the kinematic space $\vec{x}$ where a mismodelling of the efficiency can have a significant impact in the signal measurement while the effect on the control channels remains within the constraints of Eq. 5. Given these constraints, the intuition is that this problem can be thought as a classification task between the signal and control channels using the kinematic variables provided in the space $\vec{x}$.

2.1 DL Advocate algorithm

The algorithm is formed of two main parts, as shown schematically in Fig. 1. The first is a fully connected neural network which resembles a multi-classification algorithm. The inputs are a set of features x whereas the output h_j is a classification score for each decay hypothesis. The last layer of the NN is normalised with a softmax activation function which enforce $h_j(x) \geq 0$ and $\sum_j h_j(x) = 1$. Details on the technical implementation of the neural network are given in Appendix A.

The second part of the algorithm consists of a linear combination of the NN output, also referred to as linear programming (LP), which defines the final per-event weight

$$w(x) = \sum_j \alpha_j h_j(x) \quad \text{with} \quad \alpha_j \geq 0. \tag{6}$$

Combining this with Eq. 3 and moving to a vectorial notation we can express the total per-channel efficiency as

$$e_i = \frac{1}{n_i} \sum_k \vec{\alpha} \cdot \vec{h}(x_{k,i}), \tag{7}$$

or, in an even more compact form,

$$\vec{e} = H\vec{\alpha}, \tag{8}$$

where we have introduced the H matrix which is defined as

$$H_{i,j} = \frac{1}{n_i} \sum_k h_j(x_{k,i}). \tag{9}$$

Here, H is a quadratic $M \times M$ matrix where M is the total number of decay channels. From the previous equations it is evident the role of the coefficients α_j which relate the NN classification response to the different decay-channel efficiencies. The meaning of the α_j coefficients can be easily understood in the ideal case of a perfectly discriminating network. In such a scenario H takes the form of the identity matrix and α_j can be individually chosen to satisfy
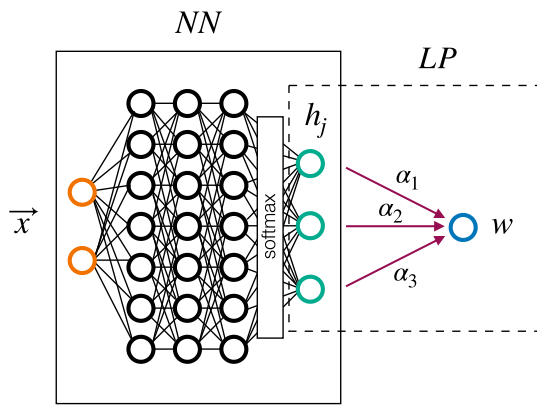
NN

Fig. 1 Schematic view of the DL Advocate algorithm. The algorithm is formed by two parts: the output of a neural network (NN) is passed to a linear programming (LP) solver which returns the final weight for each event

all possible combination of efficiencies, i.e. we can choose $\alpha_{(j|j\in\mathcal{C})} = 1$ for all control channels $\mathcal{C}$ in order to perfectly satisfy their efficiency constraints, and arbitrarily move $\alpha_{(j|j=s)}$ for the signal channel s to get any possible values for the signal efficiency e_s.

In general, however, the H matrix will have non-diagonal terms that correlate the efficiency of the different decay channels. The measurements carried out on the control channels, therefore, provide non trivial constraints on the signal efficiency.

2.2 Optimisation procedure

The goal of the algorithm is to find the solution that maximises possible shifts in the signal efficiency while maintaining the control measurements within their allowed range, as defined in Eqs. 4 and 5. This is achieved with an iterative procedure:

(i) the NN is pretrained as a simple classifier, i.e. minimising the cross-entropy loss between channels;
(ii) for a given set of NN parameters θ the matrix H is determined and the optimal values of the coefficients $\vec{\alpha}$ are calculated;
(iii) with the obtained values of $\vec{\alpha}$, the NN parameters are updated to improve the current solution;

with step (ii) and (iii) repeated for 1000 iterations. In each iteration the value of e_s is recorded and the solution that manifests the largest bias in the signal efficiency is returned at the end of the procedure. More details on the individual steps (ii) and (iii) are given in the next subsections. In general, mismodelling of the efficiency can result in both over and underestimation of the total signal efficiency. In the following, we focus on the minimisation of the signal efficiency;

oppositely, in order to get the maximum allowed positive shift it will be sufficient to target $-e_s$ instead of e_s in the minimisation process below.

2.2.1 Determination of the coefficients α_j

In order to find the values of $\vec{\alpha}$ that allow the largest deviation in e_s while satisfying the constraints from the control channels we have to solve the following system of linear equations

$$\begin{cases} \sum_j \alpha_j H_{i,j} \in [V_i^{low}; V_i^{high}] & \forall i \in \mathcal{C}, \\ e_s = \sum_j \alpha_j H_{i,j} \to \min & \text{for } i = s. \end{cases} \quad (10)$$

This linear programming (LP) problem can be solved numerically and the minimisation of e_s is performed with the SciPy python package with the use of the `scipy.optimize.linprog` function [16].

2.2.2 Update of the neural network

The values of $\vec{\alpha}$ obtained in the previous step correspond to the optimal result for a fixed set of NN parameters, however, these can be modified to further improve the overall solution. The neural network weights θ are then updated based on the following loss function

$$\ell(\theta) = \ell_s(\theta) + \ell_d(\theta), \quad (11)$$

with

$$\ell_s(\theta) = e_s \quad (12)$$

driving the minimisation of e_s with respect to the NN parameters θ and

$$\ell_d(\theta) = -\log|\det(H)| \quad (13)$$

is a regulariser term which is added to avoid $\det(H) = 0$ and it keeps H invertible during the optimisation process. This term is found to help the stability of the process since the determination of $\vec{\alpha}$ obtained in the previous step implicitly requires H matrix inversion ($\vec{\alpha} = H^{-1}\vec{e}$) which can be numerically unstable. The gradient of the loss $\ell(\theta)$ can then be calculated using the chain rule, i.e.

$$\frac{\partial\ell}{\partial\theta} = \sum_j \sum_i \frac{\partial H_{i,j}}{\partial\theta}\left(\frac{\partial\ell_s(\theta)}{\partial H_{i,j}} + \frac{\partial\ell_d(\theta)}{\partial H_{i,j}}\right), \quad (14)$$

with

$$
\frac{\partial \ell_s(\theta)}{\partial H_{i,j}} = \begin{cases} \alpha_j & \text{for } i = s, \\ 0 & \forall i \in \mathcal{C}. \end{cases} \tag{15}
$$

$$
\frac{\partial \ell_d(\theta)}{\partial H_{i,j}} = -\left[H^{-1} \right]_{j,i} \tag{16}
$$

where Eq. 16 follows from Eq. 13 as shown in [17], while $\frac{\partial H_{i,j}}{\partial \theta}$ is computed with standard backpropagation techniques. Finally, the NN parameters θ are updated based on the overall loss function improvement $\theta \to \theta - \eta \frac{\partial \ell}{\partial \theta}$, where η is the learning rate set to 0.0001, and the previous step is repeated. The overall optimisation procedure is schematised in Algorithm 1.

2.3 Smoothening of the output weighting function

The solution obtained following the algorithm description presented in the previous section represents the largest possible variation of e_s that keeps the control-channel measurements within their limits. However, the resulting weighting function $w(x)$ can display extreme trends, such as fast-changing values, which can be mitigated by adding the following regulariser to the loss function

$$
\ell_g(\theta) = \frac{1}{n} \sum_k \left[\left(\frac{\left\| \nabla \vec{h}(x_k; \theta) \right\|}{p} - 1 \right)_+ \right]^2, \tag{17}
$$

where $(\cdot)_+$ denotes a positive cut, i.e. $(y)_+ = \max(y, 0)$, x_k is a dataset sampled randomly from the domain $\vec{x}$ and p is a tunable parameter indicating the turn on of this gradient penalty (GP) term and is set to 0.5 in the rest of the paper.

3 A concrete example: a branching fraction measurement

In this section, we illustrate the use of the DL Advocate method with the example of a branching fraction measurement. Branching fractions are typically measured as a ratios of a given signal decay channel with respect to a normalisation channel

$$
\mathcal{B}_{\text{sig}} = \frac{N_{\text{sig}}}{N_{\text{norm}}} \cdot \frac{e_{\text{norm}}}{e_{\text{sig}}} \cdot \mathcal{B}_{\text{norm}} \tag{18}
$$

where $\mathcal{B}$ is the branching fraction, N is the observed yield and e is the detector efficiency. It is evident from Eq. 18 that a problem in the efficiency estimation would unavoidably lead to an incorrect determination of the branching ratio.

For this example, we consider the branching fraction measurement of a hypothetical decay of the type $P \to VC$, where $V \to AB$ is a hypothetical intermediate resonance

decaying into two given particles A and B, and C is a companion final-state particle. Branching ratios are typically normalised relative to decays with similar topology, which we denote here with $P \to XC$ channel, with $X \to AB$, where the branching ratios $P \to XC$ and $X \to AB$ are assumed to be well known. For instance the Particle Data Group (PDG) lists several branching fractions for various systems, some of which have uncertainties as low as 3% [18]. In addition to the normalisation itself, crosschecks involving other control channels with well known branching fractions are often performed during experimental analyses. In the following, we consider the existence of a second control channel denoted with $P \to YC$ with $Y \to AB$.

The advantage of the existence of a second control channel stands in the possibility to build ratio of branching fractions, i.e. the same procedure developed for the signal is employed on the control channel by replacing in Eq. 18 the signal efficiency and observed yield with the corresponding control channel counterparts. Ratios of branching fractions, in fact, are typically characterised by smaller uncertainties compared to single branching fractions, since common sources of uncertainties are cancelled out in the ratio. For instance, the PDG reports ratios of branching fractions which are known with a precision at percent level [19].

In conclusion, in our toy example we assume the absolute branching ratio of the normalisation channel to be known with a precision of 3%, the ratio of branching ratios between the two control channels to be known with a precision of 1% and we train the DL Advocate to place an upper bound on the systematic associated to the measurement of signal branching ratio while obeying these constraints.

3.1 Simulated dataset

In order to explore this method in a concrete setting, we generate simulated events based on a typical measurement performed by the LHCb experiment [20], as many heavy-hadron decays are measured by LHCb and it is therefore a natural testbed for our approach. In order to generate the representitive simulation we employ a fast simulation based on the the RapidSim package [21], which simulates decays of heavy hadrons with parameterised momentum smearing and representitive production kinematics for high energy collisions. Basic acceptance criteria are applied such as requiring each final state particle to be within $2 < \eta < 5$, which is a crude estimate of the LHCb acceptance, and to have a minimum transverse momentum of 300 MeV/c^2. Finally, final state radiation effects are simulated with PHOTOS [22], which play a minor role in the analysis.

We use the decay of a B^+ meson into a kaon and a pair of oppositely charged muons. This decay is chosen as a proxy for a generic three-body decay as several control channels exist such as decays involving the J/ψ and $\psi(2S)$ mesons
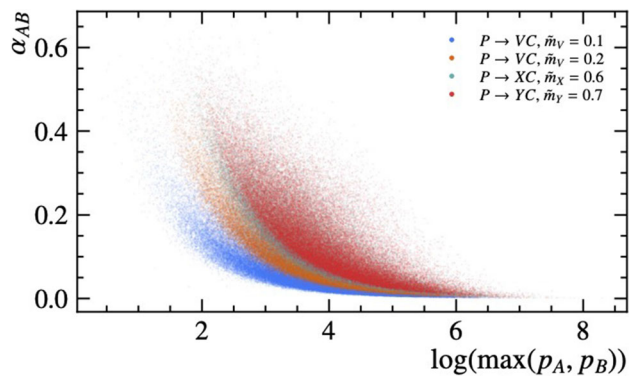
Fig. 2 Distribution of events in the kinematic plane $\max(p_A, p_B)$, α_{AB} for two choices of signal mass $\tilde{m}_V = 0.1$ and 0.2, as well as for the control channels

Fig. 3 Evolution of the signal efficiency e_s for the different values of $\tilde{m}_V$ when trained with the set of features $x = \{p_T^A, p_T^B\}$

[23] which make it a plausible test case for the method. As there is no detailed particle identification response in our analysis, we use this decay as a representation of the general three-body $P \to ABC$ particle decay introduced above, where the two muons, represented here as A and B, originate from an intermediate resonance of variable mass, m_V, and the kaon represents particle C. One hundred thousand events are generated for each considered value of m_V.

While the generated samples do not take into account reconstruction effects specific to the LHCb detector, they provide an excellent template to demonstrate the proposed method on a simplified example.

3.2 Analysis setup

In order to determine how a potential mismodelling of the efficiency could affect the branching fraction measurement of $P \to VC$ decays while keeping unchanged the value of the crosscheck provided by the control channel $P \to YC$, we need to consider all possible differences among the considered decay channels. Due to the different resonance masses m_R, with $R \in \{V, X, Y\}$, signal and control modes are characterised by different kinematic distributions. In the following we use the normalised resonance mass $\tilde{m}_R = m_R/m_P$ as a more generalisable kinematic variable. This resonance mass $\tilde{m}_R$ can be expressed as $\tilde{m}_R^2 = 2p_A p_B (1 - \cos \alpha_{AB})/m_P^2$, i.e. the mass of the resonance is unequivocally determined by the momentum of the two final state particles AB and their opening angle. As an example, we take X and Y particles to have similar normalised masses, i.e. $\tilde{m}_X \simeq 0.6$ and $\tilde{m}_Y \simeq 0.7$, which results in similar kinematic distributions, while $\tilde{m}_V$ is allowed to span all the kinematically allowed range, with the result that the further $\tilde{m}_V$ is from $\tilde{m}_{X,Y}$ the bigger the difference in their kinematic distributions will be (see Fig. 2).

For the example illustrated in this paper, we will therefore focus on the study of the momentum, transverse momentum
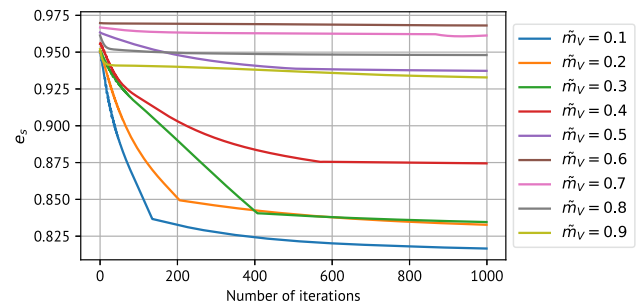
and opening angle, or combination of those. In high-energy-physics experiments, the efficiency to reconstruct and select a given particle is typically dependent on its (transverse) momentum, which makes particularly important – and potentially prone to hidden systematics – to have good control of the detector efficiency as function of those variables.

In the following, we will therefore train the DL Advocate to evaluate what is the maximum impact that a hypothetical mismodelling of the efficiency can have on the determination of the $P \to VC$ branching ratio under different values of $\tilde{m}_V$. It is also important to stress that the goal of the algorithm shall not be to reconstruct $\tilde{m}_V$, which is what occurs if the two momenta and the opening angle of the AB particles are given to the network, but rather to explore possible undetected patterns in the efficiency response of the detector which may lie hidden in a subset of such kinematic variables.

Three different sets of input variables are therefore considered for this study:

- $x = \{p_T^A, p_T^B\}$, the two transverse momenta of particles A and B;
- $x = \{\max p, \alpha_{AB}\}$, the maximum momentum of the two particles A and B and their opening angle;
- $x = \{\max p_T, \alpha_{AB}\}$, the maximum transverse momentum of the two particles A and B and their opening angle;

The DL Advocate algorithm is then trained following the methodology described in Sect. 2 with the constraints imposed by the normalisation and control channels taken into account as discussed in Sect. 2.2.1. In particular, the constraints takes the form

$$\mathcal{B}(P \to XC) \propto e_{P \to XC} \in [-3\%, 3\%], \tag{19}$$

$$\frac{\mathcal{B}(P \to YC)}{\mathcal{B}(P \to XC)} \propto \frac{e_{B \to YC}}{e_{P \to XC}} \in [-1\%, 1\%]. \tag{20}$$

where e_i should be interpreted as relative variations of the efficiency with respect to the hypothesis of perfectly modelled efficiency (as defined in Eq. 3) and the former equation assumes perfect knowledge of the production of the parent particle P. Finally, in order to evaluate the dependence of

the maximum allowed bias on $\tilde{m}_V$, the training is repeated for different values of the normalised resonant mass ranging from 0.1 to 0.9.

3.3 Training and results

Figure 3 shows the evolution of e_s during the training process for the different values of $\tilde{m}_V$ and for the selected input features $x = \{p_T^A, p_T^B\}$. A good learning curve is observed for all cases, with an almost-optimal solution obtained after less than 500 iterations. A very similar pattern is also seen for the other pairs of features used.

Figure 4 shows the resulting weighting functions obtained from the training of the network for the three considered sets of variables and for different values of $\tilde{m}_V$. For cases where $\tilde{m}_V < \tilde{m}_X$, the obtained solution shows a clear split of the 2D plane, with a large fraction of signal events that receive a weight close to zero. This is due to the larger separation in the kinematic space between signal and control channels, which allows to find a solution that strongly affects the former while keeping unchanged the other two. On the other hand, for signal masses between or above the X and Y particles, the strong overlap in the kinematic distributions leaves very little to no room for efficiency mismodelling to be able to modify the signal branching ratio while satisfying the constraints from the control channels.

We can now quantitatively define the observed bias in each decay mode as the deviation of its integrated efficiency with respect to the hypothesis of perfect efficiency modelling, i.e. $(1 - e_i)$. The maximum allowed shift on the determination of the signal efficiency, and hence the signal branching ratio, obtained in the different tested configurations is illustrated in Fig. 5. We can draw the following conclusions:

- as expected, the lower the normalised signal mass $\tilde{m}_V$ is, the bigger the bias is allowed to be; this is due to the large separation in the feature space between signal and control channels, which allows a mismodelling of the efficiency that only affects the signal decay channel;
- vice versa, it is nearly impossible for signal with mass similar to the ones of the normalisation/control channels to suffer from uncontrolled systematic effects;
- in all cases, the set of variables that allows the largest bias is given by the combination $\{\max p, \alpha_{AB}\}$, which is natural being the one with the largest correlation with the resonant mass;
- as expected, the gradient penalty term applied to the loss function as described in Sect. 2.3 reduces the overall allowed systematic effect.

We note that, in most of the cases the obtained solution presents region of the parameter space where the mismodelling of the efficiency is maximal (i.e. an efficiency very

close to zero and/or infinity). In addition, despite the gradient penalty term added to the loss function as in Eq. 17, rather steep transitions between well-modelled and badly-modelled regions are still visible in Fig. 4. Nevertheless, the effectiveness of this term is clearly visible from the comparison of the solution obtained with and without it as shown in Fig. 5 (right) and a further tuning of this parameter goes beyond the scope of this paper.

3.4 Implications on differential control-channel measurements

In the toy study presented above, we limited the role of the control channels to the sole contraint provided by their integrated branching fractions. In real analyses, however, the use of simulated datasets undergoes a rigorous validation process, which once again makes use of well-known and largely produced calibration channels. One of the additional crosschecks which can be naturally performed on the control channels consists in looking for possible biases as function of the relevant kinematic variables. A systematic uncertainty can in fact escape the bound imposed by the total branching fraction measurement while appearing in its differential distribution. This is exactly the case studied in the previous section. Figure 6 shows the bias expected for the different obtained solutions as function of the minimum transverse momentum and opening angle of the considered control channel. The strong visible trend demonstrates that, assuming a sufficient statistical precision, the systematic uncertainty postulated by the DL Advocate algorithm in the previous section would have been revealed by analysing the consistency of the control-channel differential branching fraction measurement. For example, in Ref. [23], which measures the branching fraction of $B \rightarrow K^{(*)}\mu^+\mu^-$ decays, the biases shown in Figs. 5 and 6 would have been detected by the experimental crosschecks performed in the analysis thanks to the large datasets collected for the $B \rightarrow J/\psi K^{(*)}$ normalisation channels, which are of the order of hundreds of thousands of events.

One of the limitation of the standard analysis procedure, which consists in manually validating the simulation by looking at different relevant variables, is that it is limited to the statistical power and kinematic distribution of the control channels. Therefore, a systematic effect may still escape this validation procedure if it is limited to a region poorly populated by the control samples or it appears with complex correlations among different or unexpected sets of variables. On the opposite, the DL Advocate algorithm presented in the previous section can be extended to include any arbitrary set of constraints, such as the differential control channel measurements discussed above, and provide a solution that would pass all the standard crosschecks of this kind of analyses. A complete implementation of all the available constraints goes
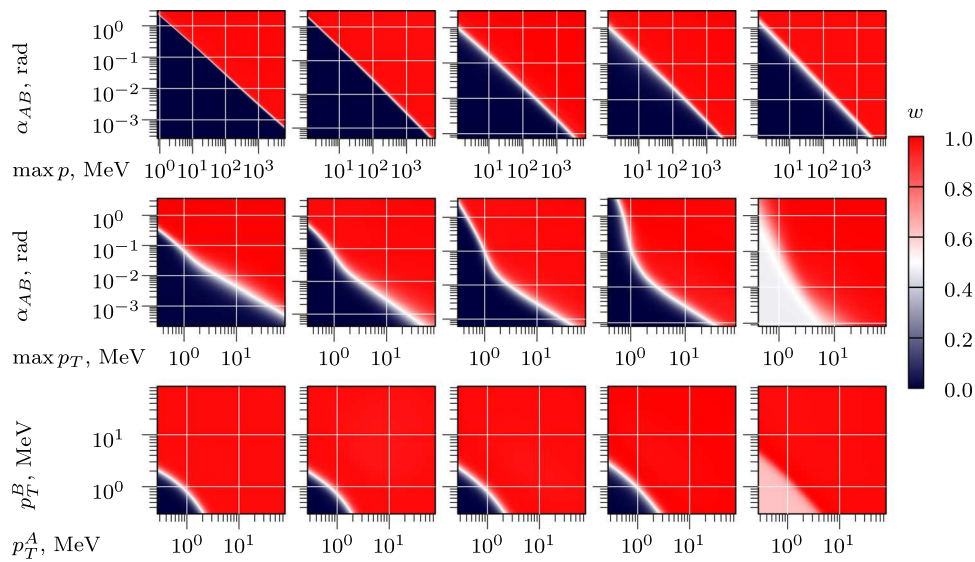
Fig. 4 Resulting weighting maps describing the efficiency mismodelling as function of the input variables for the three sets of considered features {max p, α_{AB}} (top), {max p_T, α_{AB}} (middle) and {p_T^A, p_T^B} (bottom) for $\tilde{m}_V = \{0.1, 0.2, 0.3, 0.4, 0.5\}$ (from left to right). Values of $\tilde{m}_V \geq 0.6$ are not shown since no significant deviations from unity are visible in the entire feature plane
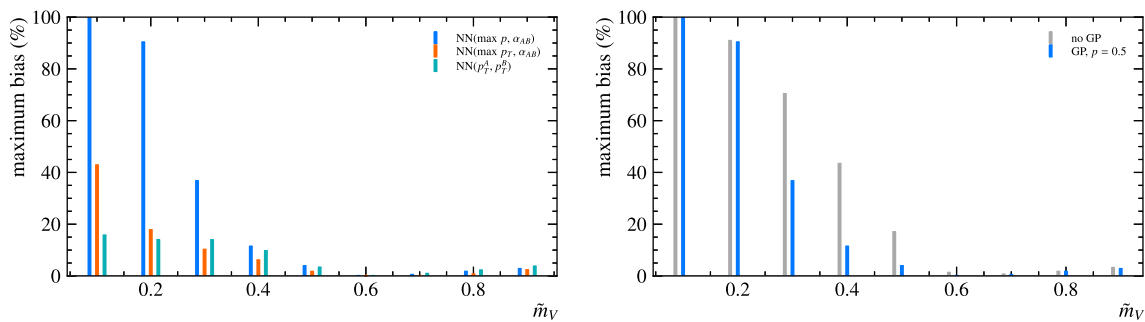

Fig. 5 Left: maximum allowed bias obtained on $P \to VC$ branching ratio by running the DL Advocate algorithm with the three studied sets of features at different masses. All cases have been trained with a gradient penalty term set to 0.5. Right: maximum allowed bias obtained using {max p, α_{AB}} as input features with and without the gradient penalty term introduced in Eq. 17
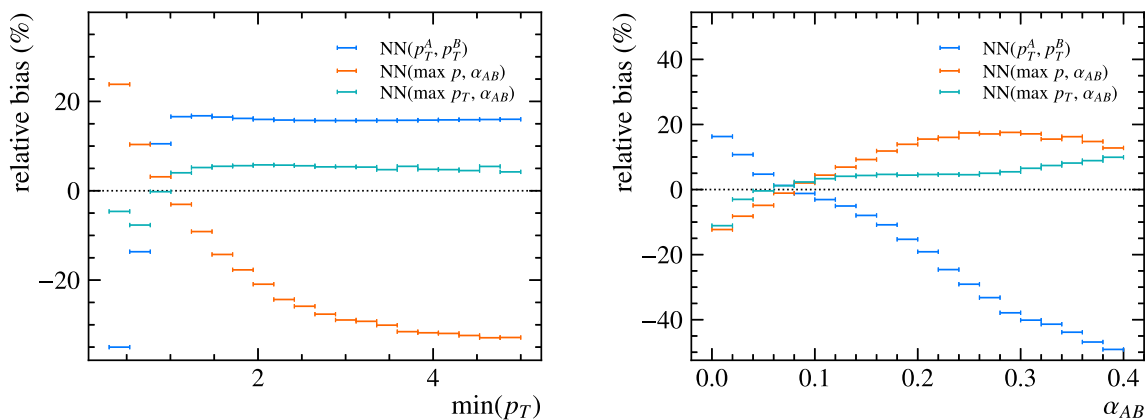

Fig. 6 Expected relative bias on differential branching fraction measurement for the $P \to XC$ control channel as function of the minimum transverse momentum (left) and opening angle (right). The algorithm has been trained with a signal of $\tilde{m}_V = 0.3$, while the result for the three tested configurations is shown with different colours

beyond the scope of this paper, however, we can expect them to play a significant role in reducing the room for possible undetected hidden systematic uncertainties.

Finally, beyond all internal crosschecks that can be developed within a given analysis, one of the strength of the proposed method is that one can systematically explore all the repercussions that a solution found by the DL Advocate may have on other observables and/or decay channels. For example, one can check the impact that a certain solution obtained investigating systematic uncertainties on the signal branching fraction has on the mismodelling of the angular distributions of the same decay or, vice versa, how measuring the angular distribution of the decay can further constrain the impact of hidden systematic uncertainties on the branching fraction determination.

4 A reinforcement learning approach

As already mentioned, the method proposed in this paper can only reach its full potential when applied to *low-level* detector information. The use of detector-related quantities, such as hits positions, energy clusters, magnetic fields, etc., introduces several additional complications to the problem, since it requires an accurate description of the detector and, most importantly, an iterative refinement of the simulation, i.e. the interaction of the particles with the detector has to be re-evaluated for every considered modification. On the other hand, the inclusion of low-level quantities would provide a general tool that is applicable to any measurement that relies on simulation, enabling a systematic evaluation of all possible mismodellings of the detector response. While a complete solution to this problem is out of the scope of this paper, we illustrate with a simplified example the use of reinforcement learning (RL) to potentially tackle this difficult task.

Reinforcement learning [24,25] recently achieved impressive results in many domains of applied research, such as robotics [26], self-driving cars [27], gaming [28]. Coming to high energy physics, it has been mainly suggested for jets reconstruction [29,30] and on-line control system for accelerator machines [31]. In this section we discuss the possibility to train a RL *agent* to play the role of the devil's advocate (RL Advocate). The goal of the algorithm is unchanged, i.e. trying to find possible mismodelling of the efficiencies that can affect a certain set of observed measurements; however, we need to introduce some new concepts in order to formalise the problem within a RL approach.

Reinforcement learning algorithms are designed to train an *agent* via continuous interaction with an external environment. At each time step t, the agent makes an observation of the environment's *state*, S_t, and, based on such observation, it undertakes a certain *action* A_t. In turn, as a consequence of this action, the agent will find itself in a new state S_{t+1},
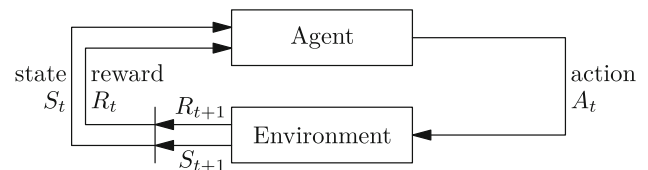

Fig. 7 Schematic representation of agent–environment interaction [24]

while receiving from the environment a numerical reward R_{t+1}. Figure 7 illustrates the described process. Finally, the decision-making of the agent, also called *policy*, π, is trained to maximise the expected total reward over the long run.

4.1 Environment's setup

Identically to the classifier-based approach discussed in Sect. 2, also in this case we intend to describe possible mismodelling of the efficiency by introducing a per-event weighting function $w(x)$ which can depend on a certain set of input features x. However, instead of training a neural network to determine the mismodelling map, we define a parametric expression of the weighting function itself, $w(\tilde{\eta}, x)$, where $\tilde{\eta}$ is the set of parameters that describes the detector response which needs to be determined. This approach has some advantages, e.g. we avoid the risk of having sharp (unphysical) changes of efficiencies, as well as disadvantages, e.g. the requirement of a parametric expression in the formulation of the problem, first, looses generality compared to the NN output (given the enormous number of parameters that characterises a NN allowing it to be able to approximate any function in its domain), and second, it definitely requires some physics intuition.

The goal of the RL agent can therefore be formulated as finding the values of the parameters $\tilde{\eta}$ that best satisfies the agreement with the experimental measurements. We can then define:

- the *state*, $s \equiv \tilde{\eta}$ (renamed for convenience), the list of the parameters used to describe the weighting function;
- the possible *actions*, which consist of increasing or decreasing each of the $\tilde{\eta}$ parameter by a discrete quantity;
- the *reward* system,

$$r = 0.01 \times \begin{cases} -\chi^2/N_{\text{meas}}, & \text{if } \chi^2/N_{\text{meas}} > 3, \\ -\chi^2/N_{\text{meas}} + 10\,(3 - \chi^2/N_{\text{meas}}), \\ \quad \text{if } \chi^2/N_{\text{meas}} \in [0.1,\ 3], \\ 10^3, & \text{if } \chi^2/N_{\text{meas}} < 0.1, \end{cases}$$

(21)

with N_{meas} the number of considered measurements and $\chi^2 \equiv \sum_{i=1}^{N_{\text{meas}}} \left(\frac{M_i - \mu_i}{\sigma_i} \right)^2$ where, recalling the notation

introduced in Eq. 1, $M_i = \mathcal{F}(\tilde{\eta})$ are the values of the different measurements evaluated at each time step based on the evolving values of the $\tilde{\eta}$ parameters, while μ_i and σ_i are the corresponding central values and uncertainties measured by the experiment.

The scheme of Eq. 21 is designed to return a negative reward (proportional to χ^2) when the agent is far from finding a good solution that accommodates all the considered measurements (large χ^2 values). On the other hand, the agent is *encouraged* with an increasingly positive reward once it reaches the condition $\chi^2/N_{\text{meas}} < 3$. Finally, in the event the agent finds a good compatibility between the obtained solution and the experimental measurements, which is defined in Eq. 21 as $\chi^2/N_{\text{meas}} \leq 0.1$, a large positive reward is assigned and the episode is terminated. Ultimately, since χ^2 can take on very large values, a scaling factor of 0.01 is added to improve the convergence of the algorithm.

Each episode is run for a maximum of 200 time steps, while the size of each parameter's update is adjusted to decrease from 0.1 to 0.01, according to the returned χ^2/N_{meas} value, i.e. the closer the algorithm is to the best solution the smaller the parameter's update will be.

We implemented the described environment within the RLlib python library [32], which offers a great variety of Reinforcement Learning algorithms ready for use. In the following studies, we will employ deep Q-networks (DQN) [33–35], which are designed to learn the action-value function, or Q-value, defined as

$$Q(s, a) = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \middle| S_t = s, A_t = a \right] \tag{22}$$

where γ is a discount factor that weights future rewards compared to present ones, whose default value is set to 0.99. In general, the Q-value quantifies *how good* is to take action a while being in state s; learning the Q-value function is the target of the DQN training procedure. Once a good approximation of the Q-values is achieved, it will be sufficient to always select the action with the highest Q-value, namely following a *greedy* policy, in order to find the optimal solution. However, better training performances are typically achieved by following the so-called ε-*greedy* policy, where a random action is selected a fraction ε of the times. This behaviour allows more exploration for new – and potentially better – actions and it is found to help the convergency of the training process. In the following, the value of ε is chosen to decrease exponentially from 1.0 to 0.1 with a decay length of 10^4 steps.

4.2 RL Advocate applied to P_5'

This section illustrates the use of Reinforcement Learning to play the devil's advocate with a second concrete example

inspired by a set of measurements that shows some tension with respect to the Standard Model prediction. In principle, the same study could be conducted with the linear programming method presented in Sect. 2 (after an appropriate modification of the loss definition) as well as RL be applied to the branching fraction example of Sect. 3. In both cases similar results and conclusions have to be expected given the simple nature of the examples under study and agreement between the simple optimisation and RL results shown later. In the following, we present the efficacy of the RL Advocate applied on a physics example of high interest for the community, i.e. the P_5' anomaly [36].

4.2.1 The P_5' measurement

Rare decays proceeding via $b \to s\ell\ell$ transitions are a sensitive probe of physics Beyond the Standard Model. In particular, potentially yet-undiscovered particles may contribute to the decay process and modify, among other properties, the angular distribution of the final state particles [36–39]. Precision measurements of these angular observables are therefore a fundamental test of the Standard Model.

In recent years, different experiments [40–43] measured a deviation in one of the angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ decays, namely P_5', in a region of q^2 between 4 and 8 GeV^2, where q^2 is defined as the di-muon invariant mass squared. In the following, we will use the example of the LHCb measurement of P_5' in the [4.0, 6.0] GeV^2 q^2 bin [40]

$$P_5' = -0.439 \pm 0.117 \tag{23}$$

which is found to be higher than the Standard Model value by a factor of 30–40%, depending on the considered theoretical predictions [44–47].

In the following, we will run the RL Advocate to test whether an uncontrolled modification of the efficiency can cause a shift in the P_5' observable capable of explaining the observed discrepancy.

4.2.2 Determination of P_5' and choice of the weighing function

Three decay angles are necessary to describe the angular distributions of $B^0 \to K^{*0}\mu^+\mu^-$ decays, namely $\vec{\Omega} = \{\theta_K, \theta_\ell, \phi\}$, defined as in Ref. [48]. Starting from this angular definition, one can define a basis of angular functions, whose coefficients – the angular observables – depend on the underlying physics model [49]. The P_5' observable is defined as a combination of the angular coefficient S_5 and the fraction of longitudinal polarisation of the K^{*0} meson F_L, i.e. $P_5' = \frac{S_5}{\sqrt{F_L(1-F_L)}}$ [9,10].

In the case of a pure signal sample, F_L, S_5 and consequently P_5' can be easily extracted using the *method of moments* [48], i.e. averaging the value of the relevant angular functions f_j over the collected sample, which gives $F_L = 2 - \frac{5}{2}M_{1s}$ and $S_5 = \frac{5}{2}M_5$, where the relevant angular moments are defined as

$$M_{1s(5)} = \frac{1}{N}\sum f_{1s(5)}(\vec{\Omega}_i), \tag{24}$$

with $f_{1s} = \sin 2\theta_K$ and $f_5 = \sin 2\theta_K \sin\theta_\ell \cos\phi$.

The loss of efficiency due to imperfect experimental detection can be taken into account by introducing a series of weights and modifying the determination of the moment M_j to

$$M_{1s(5)} = \frac{1}{\sum w_i}\sum w_i f_{1s(5)}(\vec{\Omega}_i), \tag{25}$$

where w_i represents a per-event weighting function. In general the weight w_i encodes the distorsion occurred to the signal angular distributions due to the finite detector efficiency, which is studied with high accuracy by the different experiments [40–43]. In the following, however, we will uncouple all known detector effects, i.e. as in the previous sections we assume perfect estimated efficiency, and we entirely identify the weight w_i with the mismodelling of the efficiency under investigation. Finally, for this study we employ a fast simulation of one million events of $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays, with $K^{*0} \rightarrow K^+\pi^-$, obtained with the RapidSim package [21] run under the same conditions described in Sect. 3.1.

Before entering the core of the problem, it is interesting to test whether the solution obtained in the previous section can have any sort of impact on the angular observables P_5'. We note that none of the solutions found in Sect. 3.3 can cause a shift in P_5', this is because P_5' (and the angular observables in general) does not depend on the total signal efficiency, as it was for the branching fraction, but it can only be affected by variations of the efficiency in the angular space. Therefore, in order to create a shift in P_5', we need a mismodelling of the efficiency that goes beyond the simple dependency on the total kinematic of the event. One possibility would be to introduce a detector response that is different for the two muons. In particular, one can consider a muon efficiency that depends on the relative charge between the muon and the kaon. Such scenario can in principle occur if there are hidden correlations between the hadronic system and the muons in the reconstruction of the decay. In the ideal case, one may want the *agent* to reach this conclusion by itself and autonomously take the initiative to split the efficiency weighting function by muon charge. While this possibility may appear exciting, it introduces an enormous amount of complications in the definition of the *state–action* space, e.g. a variable number of available actions and/or variable dimensionality of the state, and it is left to future work.

In the following, we will therefore program the *agent* to have a mismodelling function which is different for the two muons and we define the total per-event weight

$$w_i(x) = w^+ \cdot w^-, \tag{26}$$

where w^+ is the mismodelling function associated to the muon with the same charge of the kaon, e.g. $\mu^{+(-)}$ for a $K^{+(-)}$ in the final state, while w^- the one associated to the oppositely charged muon, e.g. $\mu^{-(+)}$ for a $K^{+(-)}$ in the final state.

Concerning the choice of the functional form for $w^+(w^-)$, we limit the study reported in this example to a simple linear dependence on the muon transverse momentum, i.e.

$$w^+(p_T) = (1 - k_+) + 2k_+ \cdot p_T^{(\mu^{SS})}, \tag{27}$$

$$w^-(p_T) = (1 - k_-) + 2k_- \cdot p_T^{(\mu^{OS})}, \tag{28}$$

where k_+ and k_- are the two parameters to be determined by the RL Advocate and $\mu^{SS}(\mu^{OS})$ indicates the same-sign (opposite-sign) muon with respect to the kaon. In addition, in order to avoid unphysical negative values of the efficiency, a lower bound of 0.1 is applied to the weights w^+ and w^-, i.e. $w^\pm = \max(0.1, w^\pm)$. In principle the mis-modelling function can be of any form. However, we assume this simple parameterisation as a proof-of-principle of the proposed method for ease of comparison with simple optimisation for validation purposes.

4.2.3 Training and results

We train the RL Advocate under the configuration described above with two different targets

(I) First, we only focus on signal $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays and try to see if we can find a mismodelling of the efficiency that could describe a shift in P_5' of the order of what it is seen in the experiment;

(II) second, we introduce the control channel $B^0 \rightarrow J/\psi K^{*0}$, with $J/\psi \rightarrow \mu^+\mu^-$, and check if a systematic bias in the determination of P_5' is still possible once we impose the constraints derived from the precise measurement of the angular observables in this control channel. Similarly to the previous example, in fact, the resonant mode provides an important validation of the efficiency estimation.

In concrete terms, the difference between the two training configurations stands in the composition of the *reward* system defined in Eq. 21. In the first case, only one measurement is considered and the χ^2 entering in the reward is simply

$$\chi^2_{\text{case-I}} = \left(\frac{P_5'(k^+, k^-) - P_5'^{meas}}{\sigma_{P_5'}^{meas}}\right)^2, \tag{29}$$

where $P_5^{'meas}$ and $\sigma_{P_5'}^{meas}$ are the central value and uncertainty measured by the experiment, while $P_5'(k^+, k^-)$ is the modified value of P_5' based on the variation of the mismodelling parameters k^+, k^-. On the other hand, in order to include the constraint derived from the control channel, it is sufficient to add to the χ^2 of Eq. 29 the term corresponding to the control channel measurement

$$\chi^2_{\text{case-II}} = \left(\frac{P_5'(k^+, k^-) - P_5^{'meas}}{\sigma_{P_5'}^{meas}} \right)^2$$
$$+ \left(\frac{P_5^{'J/\psi}(k^+, k^-) - P_5^{'SM, J/\psi}}{\sigma_{P_5'^{J/\psi}}} \right)^2, \qquad (30)$$

where $P_5^{'SM, J/\psi} = 0$ is Standard Model value of P_5' for $B^0 \rightarrow J/\psi K^{*0}$ decays, $\sigma_{P_5'^{J/\psi}}$ is assumed to be 0.0015, which is the typical uncertainty associated to the control channel measurement [50,51] and $P_5^{'J/\psi}(k^+, k^-)$ is the alteration of the control channel value of P_5' obtained by running the RL Advocate.

Table 1 shows the results obtained by running the RL Advocate for 3000 episodes. We note that, in this simplified example, finding the best solution for $\{k_+, k_-\}$ is equivalent to find the minimum of the χ^2. Therefore, we can validate the result obtained by the trained RL agent with the one derived by a simple χ^2 minimisation, also shown in Table 1 for completeness. We can draw the following conclusions

- in case-I, the solution $\{k^+ = 0.04, k^- = -1.09\}$ found by the RL Advocate can result in a positive shift in the value of P_5' of about 0.24, which corresponds to approximately 34 % of its Standard Model prediction. This shift is of the order of the deviation observed by the LHCb experiment.
- the solutions found by the RL Advocate and the χ^2 minimisation provide very similar results. By definition, the RL solution cannot achieve values of χ^2 lower than 0.1, since episodes are considered to be successfully terminated when the χ^2 reaches the 0.1 threshold. The threshold of 0.1 is chosen as it is significantly smaller than one standard deviation and indeed decreasing this threshold to smaller values has a neglgible effect on the results. In general, it is sufficient to have a χ^2/d.o.f. of the order of 1 to claim a good statistical description of a given series of measurements, however, due to the extremely simple example considered here, with only one input measurement, a threshold of 0.1 is chosen in order to get a solution that is as close as possible to the true minimum.
- the solutions obtained for case-I (trained only with the signal) would result in a shift on the P_5' measurement for the control mode that is almost as large as the one

Table 1 Solutions obtained by running the RL Advocate on the P_5' example

	SciPy	RL I	RL II
k^+	0.14	0.04	2.38
k^-	−1.58	−1.09	0.38
$\Delta P_5^{'sig}$	0.28	0.24	0.01
$\Delta P_5^{'J/\psi}$	0.14	0.15	0.0002
χ^2/N_{meas}	0.0	0.1	5.3

The first two columns refer to the case where only the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ measurement in $q^2 \in [4.0, 6.0]$ GeV^2 is considered, with the first column (SciPy) showing the values obtained by running an analytic χ^2 minimisation with the BFGS method from the SciPy python package [16], while the second column (RL I) report the solution found by the RL Advocate algorithm. The third column (RL II) shows the best solution obtained in presence of the constraint imposed by the control channel measurement

on the signal, up to $\Delta P_5^{'J/\psi} \simeq 0.15$, which is more than one order of magnitude larger than the uncertainty on the control channel measurement.

- when explicitly including the constraints from the control channel (case-II) we find that it is impossible for a mismodelling of the efficiency to cause a shift in the value of P_5' for $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays while keeping unchanged the measurement in the control channel, with the largest allowed variation of the order of 0.01.

5 Discussion

The two methods presented above demonstrate the ability to account for potentially hidden systematic uncertainties in measurements that may show a discrepancy between the values observed in the experiments and those predicted by the theory. While in this paper we employ simplified assumptions, which are necessary due to the impossibility to access realistic detector simulation and all the crosschecks performed during physics analyses that are not made public by the experiments, we set the ground for a novel methodology to access potentially hidden systematic uncertainties in high-energy physics measurements. A final and more accurate implementation of the proposed method is therefore left to the individual experiments.

The first method relies on a combination of gradient descent and optimisation techniques and its application is illustrated with the example of a branching fraction measurement. This method can be extended to incorporate additional assumptions on the mismodelling of the efficiency which can be derived from physics considerations. As an example, we demonstrated how to impose a smoothening of the mismodelling function by adding an appropriate penalty term to the

loss function. These coarse assumptions can be avoided if low-level detector information is used rather than the high-level features used for this paper.

In principle, the first method can be extended to account for latent theoretical or detector parameters. Those parameters can become available within the simulation with the help of automatic differentiation [52]. For example, one can explore cases where the measurement uncertainty depends on the detector's alignment or some coupling coefficient. However, most simulation packages are written in a way that doesn't support such techniques. This is precisely the motivation for studying RL-based approach, which can deal with non-differentiable function optimisations. The downside of this approach is that it is computationally more expensive compared to a classifier-based method. However, the flexibility afforded might be crucial in certain cases, particularly when dealing with effects difficult to parametrise. We illustrated the use of Reinforcement Learning with a simple example applied to the measurement of the angular observable P_5' in $B^0 \to K^{*0}\mu^+\mu^-$ decays. Despite the fact that the limited complexity of the considered example makes the examined case solvable by more standard approaches, we find that formulating the problem in terms of Reinforcement Learning can open a new avenue of Machine Learning applications in experimental particle physics.

In addition, throughout this paper we have only considered systematic effects from mismodelling of the efficiency as this is the simplest case to incorporate into the DL Advocate methodology. However, this can be extended to include several other types of systematic effects, such as signal resolution and backgrounds mismodellings or even theoretical uncertainties. Low-level information would again be very useful in order to generalise this approach to these issues.

Finally, with the growth of sensitivity of measuring devices and an inherent increase in the sensitivity to measured events, the degree of the uncertainty introduced during the measurement process becomes more and more eminent. Thus techniques helping estimating the boundaries of the observables that would confidently distinguish alternative physics hypotheses become an imminent research tool. Another reason pushing towards the development of such methods is that it becomes innately expensive to support a twin experiment to confirm expected discoveries, like CMS and ATLAS did for the discovery of the Higgs boson. The possible development of the Future Circular Collider [53] would be rather difficult to match by any other research facility. Nevertheless, all the results and measurements should be scrutinised from the perspective of possibly unknown systematic uncertainties, which the devil's advocate approach is capable of. An interesting venue for exploring and extending this method would be the measurement of W mass, which has recently been measured by the CDF Collaboration [11] to be significantly different from previous measurements [12–14,54]

and the Standard Model prediction [15]. This would require a large amount of numerical detail to be publicly released by experiments to have meaningful results such as auxiliary inputs assumed and the exact variations that are applied in the measurements.

6 Summary

In summary, we have introduced a method to place quantitative bounds for hidden systematic effects using machine learning. The philosophy behind the approach is to play the devil's advocate by reversing the measurement process and assuming the Standard Model hypothesis, such that systematic nuisance parameters are determined by the measurements themselves.

To demonstrate the method, we have applied the proposed approach to a hypothetical branching fraction measurement to see the maximum allowed bias due to possible mismodelling of the detector efficiency. We find that the bias depends significantly on the kinematic overlap between the signal and the control channels that are used as crosschecks in the analysis.

In addition to the nominal approach, a reinforcement learning technique is also employed. We explore the potentiality of this alternative route by applying it to the measurement of the P_5' angular observable in $B^0 \to K^{*0}\mu^+\mu^-$ decays. We find that in the tested minimal scenario, when including all the crosschecks of the analysis there is no room for hidden systematic uncertainties associated to the modelling of the detector efficiency to explain the deviation observed in data. This result is based on the use of fast simulation, a robust and final statement can only be achieved by considering all possible detector effects which can only be accessed with a full simulation of the detector. The development of a reinforcement learning solution is an essential complement to the results of the paper, but such an approach can have the highest potential when generalised to low-level quantities. The extension of the presented method to all possible aspects directly related to the detector, in fact, will be fundamental to demonstrate the robustness of any future discovery claim.

Appendix A: DL Advocate

The formal definition of the optimization algorithm is presented in Algorithm 1.

Algorithm 1 DL Advocate: linear programming optimisation.

Require: dataset $D = \{(x_i, y_i)\}_{i=1}^N$, efficiency bounds $B = \{(e_j^l, e_j^u)\}_{j=1}^k$, NN $f(x; \theta)$
Require: Learning rate η, gradient penalty p, numeric differentiation step ξ
1: $h(x; \theta) \equiv \text{softmax}(f(x; \theta))$
2: $\theta \leftarrow \arg\min_\theta \sum_{(x,y)\in D} \text{crossentropyloss}(y, h(x; \theta))$
3: **while** not converged **do**
4:    $\partial\ell \leftarrow 0$
5:    **for all** $i, y \in 1 \ldots k$ **do**
6:        $H_{i,y} \leftarrow \frac{1}{N_y} \sum_{j:y_j=y} h_i(x_j; \theta)$
7:    **end for**
8:    $H^+ \leftarrow \text{inverse}(H)$
9:    $\alpha \leftarrow \text{solveLP}(H, B)$
10:    $E \leftarrow \text{estimator}(H, B, \alpha)$
11:    **for** $(x, y) \in D$ **do**
12:        $\partial\ell_{sd} \leftarrow \frac{1}{N_y} \sum_{i=1}^k (E_{y,i} - H_{y,i}^+) \frac{\partial h_i(x;\theta)}{\partial\theta}$
13:        $\partial\ell \leftarrow \partial\ell + \partial\ell_{sd}$
14:        **if** GP is enabled **then**
15:            **for all** $i \in 1 \ldots k$ **do**
16:                $\partial x_i \leftarrow \frac{\partial}{\partial x} h_i(x; \theta)$
17:                $L_i \leftarrow \frac{h_i(x+\xi\partial x_i;\theta) - h_i(x-\xi\partial x_i;\theta)}{2\xi\|\partial x_i\|}$
18:            **end for**
19:            $\partial\ell_g \leftarrow \frac{1}{N} \frac{\partial}{\partial\theta} \left(\frac{1}{k} \sum_{i=1}^k \left(\frac{L_i}{p} - 1 \right)^2 \right)$
20:            $\partial\ell \leftarrow \partial\ell + \partial\ell_g$
21:        **end if**
22:    **end for**
23:    $\theta \leftarrow \theta - \eta\partial\ell$
24: **end while**
25: **for all** $i, y \in 1 \ldots k$ **do**
26:    $H_{i,y} \leftarrow \frac{1}{N_y} \sum_{j:y_j=y} h_i(x_j; \theta)$
27: **end for**
28: $\alpha \leftarrow \text{solveLP}(H, B)$
29: **return** α, θ

Training procedure is implemented using PyTorch framework [55]. The neural network $f(x; \theta)$ is composed of three internal linear layers with 20 outputs each and softplus activations. The final linear layer has three output nodes, which correspond to the number of considered channels. This network is rather simple but it is sufficient to represent nontrivial interpretable weightings. Softplus activations are used to keep the resulting function smooth in the domain. Input data is normalized before training to have same scale between features, while the gradient penalty term is computed after feature normalization.

References

1. G. Aad et al., Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B **716**(1), 1–29 (2012). https://doi.org/10.1016/j.physletb.2012.08.020. arXiv:1207.7214

2. S. Chatrchyan et al., Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B **716**(1), 30–61 (2012). https://doi.org/10.1016/j.physletb.2012.08.021. arXiv:1207.7235

3. S. Wunsch, S. Jörger, R. Wolf, G. Quast, Optimal Statistical inference in the presence of systematic uncertainties using neural network optimization based on binned Poisson likelihoods with nuisance parameters. Comput. Softw. Big Sci. **5**(1), 4 (2021). https://doi.org/10.1007/s41781-020-00049-5. arXiv:2003.07186 [physics.data-an]

4. P. De Castro, T. Dorigo, INFERNO: inference-aware neural optimisation. Comput. Phys. Commun. **244**, 170–179 (2019). https://doi.org/10.1016/j.cpc.2019.06.007. arXiv:1806.04743 [stat.ML]

5. A. Ghosh, B. Nachman, D. Whiteson, Uncertainty-aware machine learning for high energy physics. Phys. Rev. D **104**(5), 056026 (2021). https://doi.org/10.1103/PhysRevD.104.056026. arXiv:2105.08742 [physics.data-an]

6. B. Viren, J. Huang, Y. Huang, M. Lin, Y. Ren, K. Terao, D. Torbunov, H. Yu, Solving simulation systematics in and with AI/ML. In: 2022 Snowmass Summer Study (2022). arXiv:2203.06112

7. C. Englert, P. Galler, P. Harris, M. Spannowsky, Machine learning uncertainties with adversarial neural networks. Eur. Phys. J. C **79**(1), 4 (2019). https://doi.org/10.1140/epjc/s10052-018-6511-8. arXiv:1807.08763 [hep-ph]

8. G. Louppe, M. Kagan, K. Cranmer, Learning to pivot with adversarial networks. (2016). arXiv:1611.01046 [stat.ML]

9. S. Descotes-Genon, J. Matias, M. Ramon, J. Virto, Implications from clean observables for the binned analysis of $B \to K^*\mu^+\mu^-$ at large recoil. JHEP **01**, 048 (2013). https://doi.org/10.1007/JHEP01(2013)048. arXiv:1207.2753 [hep-ph]

10. J. Matias, F. Mescia, M. Ramon, J. Virto, Complete Anatomy of $\bar{B}_d \to \bar{K}^{*0}(\to K\pi)l^+l^-$ and its angular distribution. JHEP **04**, 104 (2012). https://doi.org/10.1007/JHEP04(2012)104. arXiv:1202.4266 [hep-ph]

11. T. Aaltonen, S. Amerio, D. Amidei, A. Anastassov, A. Annovi, J. Antos, G. Apollinari, J.A. Appel, T. Arisawa, A. Artikov, J. Asaadi, W. Ashmanskas, High-precision measurement of the W boson mass with the CDF II detector. Science **376**(6589), 170–176 (2022). https://doi.org/10.1126/science.abk1781

12. T.A. Aaltonen et al., Combination of CDF and D0 W-boson mass measurements. Phys. Rev. D **88**(5), 052018 (2013). https://doi.org/10.1103/PhysRevD.88.052018. arXiv:1307.7627 [hep-ex]

13. M. Aaboud et al., Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Eur. Phys. J. C **78**(2), 110 (2018). https://doi.org/10.1140/epjc/s10052-017-5475-4. arXiv:1701.07240 [hep-ex]. [Erratum: Eur. Phys. J. C 78, 898 (2018)]

14. R. Aaij et al., Measurement of the W boson mass. JHEP **01**, 036 (2022). https://doi.org/10.1007/JHEP01(2022)036. arXiv:2109.01113 [hep-ex]

15. J. de Blas, M. Ciuchini, E. Franco, A. Goncalves, S. Mishima, M. Pierini, L. Reina, L. Silvestrini, Global analysis of electroweak data in the standard model. Phys. Rev. D **106**, 033003 (2022). https://

doi.org/10.1103/PhysRevD.106.033003. arXiv:2112.07274 [hep-ph]

16. P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 contributors: SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods **17**, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2

17. K.B. Petersen, M.S. Pedersen, The matrix cookbook. Technical University of Denmark. Version 20121115 (2012). http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html. Accessed 23 Aug 2023

18. R.L. Workman et al., Review of particle physics. PTEP **2022**, 083–01 (2022). https://doi.org/10.1093/ptep/ptac097

19. R. Aaij et al., Measurement of the ratio of branching fractions and difference in CP asymmetries of the decays $B^+ \to J/\psi\pi^+$ and $B^+ \to J/\psi K^+$. JHEP **03**, 036 (2017). https://doi.org/10.1007/JHEP03(2017)036. arXiv:1612.06116 [hep-ex]

20. A.A. Alves Jr. et al., The LHCb detector at the LHC. JINST **3**, 08005 (2008). https://doi.org/10.1088/1748-0221/3/08/S08005

21. G.A. Cowan, D.C. Craik, M.D. Needham, RapidSim: an application for the fast simulation of heavy-quark hadron decays. Comput. Phys. Commun. **214**, 239–246 (2017). https://doi.org/10.1016/j.cpc.2017.01.029. arXiv:1612.07489 [hep-ex]

22. N. Davidson, T. Przedzinski, Z. Was, PHOTOS interface in C++: technical and physics documentation. Comput. Phys. Commun. **199**, 86–101 (2016). https://doi.org/10.1016/j.cpc.2015.09.013. arXiv:1011.0937 [hep-ph]

23. R. Aaij et al., Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays. JHEP **06**, 133 (2014). https://doi.org/10.1007/JHEP06(2014)133. arXiv:1403.8044 [hep-ex]

24. R.S. Sutton, A.G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, 2018)

25. Y. Li, Deep reinforcement learning. (2018). arXiv:1810.06339 [cs.LG]

26. D. Kalashnikov, A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke, S. Levine, Qt-opt: salable deep reinforcement learning for vision-based robotic manipulation. (2018). arXiv:1806.10293 [cs.LG]

27. B.R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A.A.A. Sallab, S. Yogamani, P. Pérez, Deep reinforcement learning for autonomous driving: a survey. (2021). arXiv:2002.00444 [cs.LG]

28. D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassabis, Mastering the game of go without human knowledge. Nature **550**, 354–359 (2017). https://doi.org/10.1038/nature24270

29. S. Carrazza, F.A. Dreyer, Jet grooming through reinforcement learning. Phys. Rev. D **100**(1), 014014 (2019). https://doi.org/10.1103/PhysRevD.100.014014. arXiv:1903.09644 [hep-ph]

30. K. Cranmer, M. Drnevich, S. Macaluso, D. Pappadopulo, Reframing jet physics with new computational methods. EPJ Web Conf. **251**, 03059 (2021). https://doi.org/10.1051/epjconf/202125103059. arXiv:2105.10512 [hep-ph]

31. J. St. John et al., Real-time artificial intelligence for accelerator control: a study at the Fermilab Booster. Phys. Rev. Accel. Beams **24**(10), 104601 (2021). https://doi.org/10.1103/PhysRevAccelBeams.24.104601. arXiv:2011.07371 [physics.acc-ph]

32. E. Liang, R. Liaw, P. Moritz, R. Nishihara, R. Fox, K. Goldberg, J.E. Gonzalez, M.I. Jordan, I. Stoica, Rllib: Abstractions for distributed reinforcement learning. (2018). arXiv:1712.09381 [cs.AI]

33. V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing atari with deep reinforcement learning. (2013). arXiv:1312.5602

34. H. van Hasselt, A. Guez, D. Silver, Deep reinforcement learning with double q-learning. (2015). arXiv:1509.06461

35. Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, N. de Freitas, Dueling network architectures for deep reinforcement learning. (2015). arXiv:1511.06581

36. W. Altmannshofer, D.M. Straub, New Physics in $B \to K^*\mu\mu$? Eur. Phys. J. C **73**, 2646 (2013). https://doi.org/10.1140/epjc/s10052-013-2646-9. arXiv:1308.1501 [hep-ph]

37. R. Gauld, F. Goertz, U. Haisch, An explicit Z'-boson explanation of the $B \to K^*\mu^+\mu^-$ anomaly. JHEP **01**, 069 (2014). https://doi.org/10.1007/JHEP01(2014)069. arXiv:1310.1082 [hep-ph]

38. A.J. Buras, J. Girrbach, Left-handed Z' and Z FCNC quark couplings facing new $b \to s\mu^+\mu^-$ data. JHEP **12**, 009 (2013). https://doi.org/10.1007/JHEP12(2013)009. arXiv:1309.2466 [hep-ph]

39. A. Crivellin, L. Hofer, J. Matias, U. Nierste, S. Pokorski, J. Rosiek, Lepton-flavour violating B decays in generic Z' models. Phys. Rev. D **92**(5), 054013 (2015). https://doi.org/10.1103/PhysRevD.92.054013. arXiv:1504.07928 [hep-ph]

40. R. Aaij et al., Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay. Phys. Rev. Lett. **125**(1), 011802 (2020). https://doi.org/10.1103/PhysRevLett.125.011802. arXiv:2003.04831 [hep-ex]

41. S. Wehle et al., Lepton-flavor-dependent angular analysis of $B \to K^*\ell^+\ell^-$. Phys. Rev. Lett. **118**(11), 111801 (2017). https://doi.org/10.1103/PhysRevLett.118.111801. arXiv:1612.05014 [hep-ex]

42. M. Aaboud et al., Angular analysis of $B^0_d \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. JHEP **10**, 047 (2018). https://doi.org/10.1007/JHEP10(2018)047. arXiv:1805.04000 [hep-ex]

43. A.M. Sirunyan et al., Measurement of angular parameters from the decay $B^0 \to K^{*0}\mu^+\mu^-$ in proton–proton collisions at $\sqrt{s} = 8$ TeV. Phys. Lett. B **781**, 517–541 (2018). https://doi.org/10.1016/j.physletb.2018.04.030. arXiv:1710.02846 [hep-ex]

44. A. Bharucha, D.M. Straub, R. Zwicky, $B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules. JHEP **08**, 098 (2016). https://doi.org/10.1007/JHEP08(2016)098. arXiv:1503.05534 [hep-ph]

45. W. Altmannshofer, D.M. Straub, New physics in $b \to s$ transitions after LHC run 1. Eur. Phys. J. C **75**(8), 382 (2015). https://doi.org/10.1140/epjc/s10052-015-3602-7. arXiv:1411.3161 [hep-ph]

46. A. Khodjamirian, T. Mannel, A.A. Pivovarov, Y.-M. Wang, Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^*\gamma$. JHEP **09**, 089 (2010). https://doi.org/10.1007/JHEP09(2010)089. arXiv:1006.4945 [hep-ph]

47. S. Descotes-Genon, L. Hofer, J. Matias, J. Virto, On the impact of power corrections in the prediction of $B \to K^*\mu^+\mu^-$ observables. JHEP **12**, 125 (2014). https://doi.org/10.1007/JHEP12(2014)125. arXiv:1407.8526 [hep-ph]

48. R. Aaij et al., Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity. JHEP **02**, 104 (2016). https://doi.org/10.1007/JHEP02(2016)104. arXiv:1512.04442 [hep-ex]

49. W. Altmannshofer, P. Ball, A. Bharucha, A.J. Buras, D.M. Straub, M. Wick, Symmetries and asymmetries of $B \to K^*\mu^+\mu^-$ decays in the standard model and beyond. JHEP **01**, 019 (2009). https://doi.org/10.1088/1126-6708/2009/01/019. arXiv:0811.1214 [hep-ph]

50. R. Aaij et al., Measurement of polarization amplitudes and CP asymmetries in $B^0 \to \phi K^*(892)^0$. JHEP **05**, 069 (2014). https://doi.org/10.1007/JHEP05(2014)069. arXiv:1403.2888 [hep-ex]

51. K. Chilikin et al., Observation of a new charged charmoniumlike state in $\bar{B}^0 \to J/\psi K^-\pi^+$ decays. Phys. Rev. D **90**(11), 112009 (2014). https://doi.org/10.1103/PhysRevD.90.112009. arXiv:1408.6457 [hep-ex]

52. A.G. Baydin, B.A. Pearlmutter, A.A. Radul, J.M. Siskind, Automatic differentiation in machine learning: a survey. J. Mach. Learn. Res. **18**, 5595–5637 (2018). arXiv:1502.05767

53. A. Abada et al., FCC physics opportunities. Eur. Phys. J. C (2019). https://doi.org/10.1140/epjc/s10052-019-6904-3

54. Improved W boson mass measurement using 7 TeV proton–Proton collisions with the ATLAS detector. Technical report, CERN, Geneva (2023). https://cds.cern.ch/record/2853290. Accessed 23 Aug 2023

55. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: an imperative style, high-performance deep learning library, in Advances in Neural Information Processing Systems 32, ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, R. Garnett (Curran Associates, Inc., Red Hook, NY, 2019), pp. 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf. Accessed 23 Aug 2023