



מכון ויצמן למדע

WEIZMANN INSTITUTE OF SCIENCE

THESIS FOR THE MSc DEGREE
עבודת גמר לתואר מוסמך למדעים

SUBMITTED TO THE SCIENTIFIC COUNCIL
OF THE WEIZMANN INSTITUTE OF SCIENCE
МОНОГРАФИЯ МАСТЕРСКОГО КЛASSA
МОНОГРАФІЯ МАСТЕРСКОГО КЛАССА
МОНОГРАФІЯ МАСТЕРСКОГО КЛАССА
МОНОГРАФІЯ МАСТЕРСКОГО КЛАССА

In search for new resonances and lepton flavor violation at the LHC

בוחינוש אחר פיזיקה חדשה במאיצ הדרוני הגדול

Federico De Vito Halevy

Weizmann Institute of Science,
Rehovot, Israel

E-mail: federico.devitohalevy@weizmann.ac.il

פדריקו דה ויטו הלי

מכון ויצמן למדע,
רחובות, ישראל

מייל:

Date:
FEBRUARY 28, 2022

תאריך:
כ"ז באדר א', תשפ"ב

Advisors:
DR. SHIKMA BRESSLER,
PROF. YOSEF NIR

מנחים:
ד"ר שיקמה ברסלר,
פרופ' יוסף ניר

Abstract

Two works are presented on the topic of physics beyond the Standard Model. In the first, we study, within the Standard Model Effective Field Theory, the possibility that the $R_{D^{(*)}}$ anomaly is explained by lepton flavor-violating processes. We find that such processes can account for, at most, a few percent of the deviation from the Standard Model. Yet, no special flavor alignment is required to explain the anomaly. As concerns collider searches, we investigate the reach of LHC experiments in probing related processes in final states involving tau leptons. Compared to existing measurements from the B -factories, we expect that LHC searches could provide complementary, albeit weaker, bounds on certain lepton flavor-violating operators. In the second work, we examine the prospects of adopting data-directed signal hypotheses in searches for new particles. As a broad strategy, we formulate a data-directed paradigm for collider searches which aims to identify promising exclusive selections of data, without primarily relying on Monte Carlo simulations. We implement a neural network to demonstrate this strategy for bump-hunting. Using artificially generated data, we demonstrate that bumps can be efficiently identified in many final state selections, while avoiding background and systematic uncertainty estimation, with only a minor degradation in sensitivity compared to a state-of-the-art analysis method.

Preface

The Standard Model of particle physics, after all its success, is still incomplete. Despite predicting the outcomes of almost all particle experiments to date with remarkable accuracy, it can nonetheless only be viewed as a low-energy effective embodiment of a more general theory. Searches for signs of “new physics” extending beyond the Standard Model follow two main branches: the high-*energy* and high-*precision* frontiers. The former directly probes new physics scales in collisions at high center-of-mass energies by looking for new particle resonances in final states. The latter, involves precise measurements of low-energy processes mediated by virtual particle states, where new particles may take on the role of these mediators, and probes the ratio of new physics couplings to their mass scale. Theory and experiment have worked hand in hand at forging the foundations of physics ever since Newton, and this is still the case in our pursuit for the understanding of physics beyond the Standard Model. It is at the intersection of the above approaches—theory and the high-precision frontier, experiment and the high-energy frontier—where the focus of this research lies.

In this report we give an account of our work on two fronts. Part I is motivated by one peculiar inconsistency between the Standard Model and precise low-energy measurements of certain semi-leptonic decays of the B meson—the $R_{D^{(*)}}$ anomaly. We discuss the reasons why these measurements constitute an anomaly in our comprehension of particle physics, and develop our own study into the phenomenology of one possible solution from the theoretical point of view. Part II shifts to recent developments in our quest to bring new discovery potential to strategies currently being applied in experimental high energy physics. We adopt a data-directed approach to analyzing observations from colliders and discuss the prospects of a particular implementation of ours, based on artificial neural networks, aimed at “bump-hunting”—the search for new particle resonances.

Acknowledgements

Firstly, I would like to thank my supervisors, Dr. Shikma Bressler and Prof. Yossi Nir, for their constant support and patient guidance throughout this project. The insights gained from this research are due to a collaborative effort with them and Sergey Volkovich, who deserves much credit for the work presented in Part II. Last but not least, I thank my fellow group members, Gal Sela and Mattias Birman, for many fruitful discussions and joyful moments—not to mention the comfort of sharing many days debugging code together.

Nomenclature

Symbols

| | |
|---------------|------------------------------------|
| $SU(3)_C$ | Standard Model gauge symmetries |
| $SU(2)_L$ | |
| $U(1)_Y$ | |
| $U(1)_{EM}$ | Gauge symmetry of electromagnetism |
| l | e , μ or τ lepton |
| ℓ | e or μ lepton |
| Γ | Decay rate |
| \mathcal{B} | Branching fraction |

Abbreviations

| | |
|--------|---------------------------------------|
| SM | Standard Model |
| NP | New physics |
| LFU | Lepton flavor universality |
| LFV | Lepton flavor violation |
| CC | Charged current |
| NC | Neutral current |
| FCCC | Flavor changing charged current |
| FCNC | Flavor changing neutral current |
| CKM | Cabibbo–Kobayashi–Maskawa |
| SMEFT | Standard Model Effective Field Theory |
| WC | Wilson coefficient |
| LHC | Large Hadron Collider |
| HL-LHC | High-Luminosity Large Hadron Collider |
| MC | Monte Carlo |
| DDP | Data-directed paradigm |
| NN | Neural network |
| ROC | Receiver operating characteristic |

Contents

| | |
|---|-----------|
| I Violation of lepton flavor symmetries: From B-decays to collider searches | 1 |
| 1 Introduction | 2 |
| 1.1 Lepton flavor universality in the Standard Model | 2 |
| 1.2 The $R_{D^{(*)}}$ anomaly | 4 |
| 1.3 Our goals | 6 |
| 2 Phenomenology of $b \rightarrow c\tau\nu_\ell$ | 7 |
| 2.1 General considerations | 7 |
| 2.2 $R_{D^{(*)}}$ in the SMEFT | 7 |
| 2.3 Bounds from constrained processes | 9 |
| 2.4 Other affected processes | 11 |
| 2.5 Expected bounds from LHC searches | 12 |
| 2.6 Discussion | 14 |
| II A data-directed paradigm for new physics searches: The bump-hunting example | 16 |
| 3 Introduction | 17 |
| 3.1 The blind analysis paradigm | 17 |
| 3.2 A data-directed paradigm | 18 |
| 3.3 Our goals | 19 |
| 4 Bump-hunting | 20 |
| 4.1 Bump-hunting in the DDP | 20 |
| 4.2 Neural network implementation | 21 |
| 4.3 Results | 23 |
| 4.4 Discussion | 26 |
| References | 28 |

PART I

Violation of lepton flavor symmetries: From B -decays to collider searches

SECTION 1

Introduction

Within the *Standard Model* (SM), the three flavors of leptons of a given charge— e , μ and τ —are equal under the electroweak interactions, up to calculable phase space effects due to differences in mass. This property is known as *lepton flavor universality* (LFU). It has provided the testing ground for much research into *new physics* (NP) extending beyond the SM. Before introducing its relevance to searches for NP, however, let us begin by examining how LFU emerges in the SM.

1.1 Lepton flavor universality in the Standard Model

Fundamental interactions of fermions¹ are said to be “flavor universal” when they do not distinguish between copies of fermion fields in the same gauge representation, i.e. flavors (or generations). Renormalizable interactions of fermions are described by Yukawa interactions with scalar bosons and gauge interactions with vector bosons. Among the latter class, those corresponding to unbroken gauge symmetries are flavor universal by definition. The strong and electromagnetic interactions correspond to such symmetries (local $SU(3)_C \times U(1)_{EM}$ invariance) and do not distinguish between flavors, hence are universal in this sense. Physics in the space of flavors enters the picture when considering the former class, Yukawa interactions, and the gauge interactions with massive vector bosons associated with broken gauge symmetries.

Given the SM symmetry breaking pattern and scalar sector, the interactions involving the Higgs, h , and the three weak vector bosons, W^\pm and Z , are candidates for exhibiting non-universal flavor structure. However, the particular fermionic field content of the SM results in flavor non-universality only in the *charged current* (CC) weak interactions of the quark sector ($W^\pm +$ quarks) and in the Yukawa interactions of both the quark and lepton sectors ($h +$ quarks/leptons). We give a brief overview of the reasons underlying this below. In particular, we aim to explain how flavor universality arises in the SM for *neutral current* (NC) weak interactions ($Z +$ quarks/leptons) and leptonic CC weak interactions ($W^\pm +$ leptons).

¹A preprint of the book by [Grossman and Nir \(n.d.\)](#) and the review by [Bifani, Descotes-Genon, Vidal, and Schune \(2018\)](#) serve as primary references for this section, among others.

Yukawa interactions

In general, the Yukawa couplings of a given scalar with n fermion flavors are given by a complex $n \times n$ coupling matrix. Thus, they are not flavor universal. If, for a particular fermion sector, the Yukawa interactions with a single scalar are the only source of mass terms (via spontaneous symmetry breaking), then they are diagonal in the mass basis of that sector. Such a scenario occurs for all fermions of the SM (with h the only scalar field) because their representations are both chiral (prohibiting bare Dirac masses) and non-real (prohibiting bare Majorana masses). The diagonality of Yukawa interactions would break with the addition of either vector-like or real (e.g. gauge-singlet) fermion representations, or with an extended scalar sector.

Neutral current weak interactions

As a direct consequence of the SM Higgs mechanism, the Z couplings are proportional to $T_3 - \sin^2 \theta_W Q$, where $Q = T_3 + Y$ is the generator of $U(1)_{\text{EM}}$, T_3 and Y are the generators of the 3-axis of $SU(2)_L$ and of $U(1)_Y$, respectively, and θ_W is the weak mixing angle. The interaction eigenstates mixing into a given fermion sector share the same electromagnetic charge (and color representation), i.e. the sum $T_3 + Y$ is fixed, since this is part of what defines the sector (together with Lorentz representation). However there are two possibilities for the value of T_3 (and Y) when comparing these states:

1. *The interaction eigenstates share the same T_3 (and Y) value.*

In this case, the Z coupling matrix is proportional to the unit matrix in the interaction basis, and remains so after rotating to the mass basis. In other words, the mass eigenstates of a sector are formed from interaction states in the same $SU(2)_L \times U(1)_Y$ representation, resulting in flavor universal Z interactions.

2. *The interaction eigenstates have different T_3 (and Y) values (but same Q).*

As a result, the Z coupling matrix is diagonal and non-universal in the interaction basis, and does not generally maintain its diagonality under a unitary transformation to the mass basis. Therefore, the mass eigenstates of a sector mix interaction states with different $SU(2)_L \times U(1)_Y$ representations, resulting in flavor non-universal (and in general non-diagonal) Z interactions.

The SM fermions belong to the first class: each sector is formed by a mixture of states from only a single class of $SU(2)_L \times U(1)_Y$ representations. Thus, the Z couples universally in each fermion sector of the SM. As an example, Z interactions could break LFU with the addition of left-handed $SU(2)_L$ -singlet leptons ($T_3 = 0$) with $Y = -1$ to the SM, which would mix with the charged left-handed leptons originating from the $SU(2)_L$ -doublet (with $T_3 = -1/2$).

Charged current weak interactions

The W^\pm couplings are proportional to T^\pm , the ladder operators on T_3 eigenstates, defined as $T^\pm := (T_1 \pm iT_2)/\sqrt{2}$. Thus, the W^\pm couples the components of $SU(2)_L$ -

multiplet fermions with $\Delta T_3 = \pm 1$. Since these components mix into different sectors ($\Delta Q = \pm 1$ is implied), only if they do so equally (as determined by the respective mass matrices) does the W coupling matrix remain universal in the mass basis. In general, this is not the case.

In an interaction basis of the SM, the up-quark and down-quark mass matrices cannot both be diagonalized with a simultaneous transformation of the quark $SU(2)_L$ -doublet components. Thus, the *Cabibbo–Kobayashi–Maskawa* (CKM) matrix of W couplings in the quark sector is non-diagonal and generates quark mixing. On the other hand, the SM neutrino and charged lepton mass matrices can be diagonalized in such a way. This is due to the degeneracy of neutrino masses, which allows any neutrino interaction basis to also be a mass basis. Thus, the CC weak interactions of the SM obey LFU. If extensions of the SM were to break neutrino mass degeneracy (e.g. via dimension-five terms or seesaw mechanisms), the W couplings to leptons would break LFU and generate lepton mixing.

1.2 The $R_{D^{(*)}}$ anomaly

Several tests of LFU have been carried out, comparing the rates of processes with interchanged lepton flavor and examining the validity of the SM prediction motivated above. Most have shown exceptional agreement with the SM (better than 1% accuracy), for example in the decays $Z \rightarrow l^+l^-$, $W^- \rightarrow l^-\bar{\nu}_l$, $K^- \rightarrow \ell^-\bar{\nu}_\ell$, $\pi^- \rightarrow \ell^-\bar{\nu}_\ell$, $\tau^- \rightarrow \ell^-\bar{\nu}_\ell\nu_\tau$ and $J/\Psi \rightarrow \ell^+\ell^-$ ([Bifani et al., 2018](#); [Particle Data Group et al., 2020](#); [ATLAS Collaboration, Aad, Abbott, Abbott, Abud, et al., 2021](#)), where $l = e, \mu, \tau$ and $\ell = e, \mu$ (henceforth). Over the past decade, however, significant tensions have been observed in the decays of B mesons ($q\bar{b}$, with $q = u, d, s, c$) to leptons. Known as the “ B flavor anomalies”, these observations constitute a violation of LFU, if taken at face value, at an energy scale $m_b \sim 5$ GeV, well within the SM’s regime of applicability.

The observed anomalies in B decays are categorized into two classes, according to the underlying transition at the quark level: $b \rightarrow c$ decays, involving *flavor-changing charged currents* (FCCCs), which are mediated in the SM at *tree*-level by virtual W bosons; and the much rarer $b \rightarrow s$ decays, involving *flavor-changing neutral currents* (FCNCs), which only occur at *loop*-level in the SM via virtual Z/γ^* penguin or WW box diagrams. For the cases with final state leptons in particular, these consist of $b \rightarrow cl\bar{\nu}$ transitions and $b \rightarrow sl^+\ell^-$ or $b \rightarrow s\bar{\nu}_l\nu_l$ transitions, respectively. In this context, the most sensitive observables are often constructed as ratios of rates, such that any common hadronic matrix elements and phase space integrals cancel, and for theoretical uncertainties (e.g. from the evaluation of form factors) and experimental systematics (e.g. from efficiencies) to be minimized, as the dependency is not on absolute knowledge of their sources but on relative differences. Significant deviations from LFU in such ratios are thus indications that some NP is involved and that it must couple non-universally to the leptons.

For example, $R_{K^{(*)}}$, defined as

$$R_{K^{(*)}} := \frac{\int_{p_{\min}^2}^{p_{\max}^2} \frac{d\Gamma(B \rightarrow K^{(*)}\mu^+\mu^-)}{dp^2} dp^2}{\int_{p_{\min}^2}^{p_{\max}^2} \frac{d\Gamma(B \rightarrow K^{(*)}e^+e^-)}{dp^2} dp^2}, \quad (1.1)$$

where p^2 is the dilepton invariant mass squared, is currently one of the strongest outstanding tests of universality between electrons and muons. The most precise measurement of R_K to date (measured in the range $1.1 < p^2 < 6.0 \text{ GeV}^2$) is $R_K = 0.846_{-0.041}^{+0.044}$, showing a 3.1σ tension with the SM expectation of 1.00 ± 0.01 ([LHCb collaboration et al., 2021](#)). It seems to increasingly hint at NP in $b \rightarrow s\ell^+\ell^-$, most likely causing destructive interference with the SM in the muonic decay of B to K ($q\bar{s}$, with $q = u, d$) mesons, and is under active research along with small anomalies in other observables of this decay mode.

Another significant hint at a manifestation of NP has come to us through the pair of observables known as $R_{D^{(*)}}$. Similarly to the example of $R_{K^{(*)}}$ above, these ratios relate the semi-leptonic decays of B to D ($c\bar{q}$, with $q = u, d, s$) or D^* mesons, i.e. $B \rightarrow D^{(*)}l\bar{\nu}_l$, with different lepton flavors in the final state:

$$R_{D^{(*)}} := \frac{\Gamma(B \rightarrow D^{(*)}\tau\bar{\nu})}{\Gamma(B \rightarrow D^{(*)}\ell\bar{\nu})}, \quad (1.2)$$

comparing the rate to τ with the rate to e or μ (taken as an average of the two). The SM predictions for these ratios are²

$$R_D^{\text{SM}} = 0.299 \pm 0.003, \quad R_{D^*}^{\text{SM}} = 0.258 \pm 0.005, \quad (1.3)$$

where the error is due to theoretical uncertainty; while the experimental world averages are

$$R_D^{\text{exp}} = 0.340 \pm 0.027 \pm 0.013, \quad R_{D^*}^{\text{exp}} = 0.295 \pm 0.011 \pm 0.008, \quad (1.4)$$

where statistical and systematic errors are denoted, respectively, first and second. Thus, R_D^{exp} and $R_{D^*}^{\text{exp}}$ exceed the SM values by 1.4σ and 2.5σ respectively, corresponding to a combined disagreement at the 3.1σ level, given the measurement correlation between the two observables ([Heavy Flavor Averaging Group et al., 2021](#), using the results of [BaBar Collaboration et al., 2012a](#); [Belle Collaboration et al., 2015](#); [LHCb Collaboration et al., 2015](#); [Belle Collaboration, Hirose, et al., 2017](#); [LHCb Collaboration et al., 2018](#); [Belle Collaboration et al., 2020](#); [Bigi & Gambino, 2016](#); [Bernlochner, Ligeti, Papucci, & Robinson, 2017](#); [Bigi, Gambino, & Schacht, 2017](#); [Jaiswal, Nandi, & Patra, 2017](#)). If higher statistics experiments firmly establish these results (currently, the p -value is $2.7 \cdot 10^{-3}$), then we will be faced with the demand to explain

$$\frac{R_{D^{(*)}}}{R_{D^{(*)}}^{\text{SM}}} \approx 1.14 \pm 0.05, \quad (1.5)$$

² $R_{D^{(*)}}$ are far from being 1 since the tau lepton occupies a large portion of the phase space with its mass of 1.78 GeV compared to the electron or muon.

i.e. an unknown contribution of around 14% of the SM value.

This discrepancy has prompted much theoretical study into possible NP contributions to $B \rightarrow D^{(*)}\tau\bar{\nu}$ (being the less constrained factor in the ratio of Eq. (1.2)), via the quark-level transition $b \rightarrow c\tau\bar{\nu}$. Most studies so far have considered additions to $b \rightarrow c\tau\bar{\nu}_\tau$, where the neutrino is perfectly aligned with the τ flavor, in search of the missing contribution to $R_{D^{(*)}}$ and have achieved varying degrees of success (e.g. recently [Shi, Geng, Grinstein, Jäger, & Camalich, 2019](#); [Hu, Li, & Yang, 2019](#); [Greljo, Robinson, Shakya, & Zupan, 2018](#); [Babu, Dutta, & Mohapatra, 2019](#)). However, in all current experiments, and specifically in measurements of $R_{D^{(*)}}$, the flavor of neutrinos is unobservable. It is certainly feasible that NP breaks LFU by enhancing $b \rightarrow c\tau\bar{\nu}_\tau$, but preserves the subset of accidental symmetries known as “lepton flavor conservation” (the $U(1)^3$ subgroup of the $U(3)$ of LFU). Yet, there is no reason to stipulate this to be the case. It is possible that the neutrino is e - or μ -flavored, or some combination of the three flavors. These scenarios incorporate explicit *lepton flavor violation* (LFV)—the violation of lepton flavor conservation—and have seen very little light in the literature³. The phenomenology of new $b \rightarrow c\tau\bar{\nu}_{e,\mu}$ contributions to $R_{D^{(*)}}$ has thus been the focus of our research on the topic.

1.3 Our goals

Our main research questions are the following:

1. Could the $R_{D^{(*)}}$ puzzle be solved by new contributions to $b \rightarrow c\tau\bar{\nu}$ with $\nu = \nu_e, \nu_\mu$?
2. If not, how closely aligned should ν be to ν_τ ?
3. How sensitive is the high- p_T LHC to such LFV signals?

We perform our analysis in the framework of *Standard Model Effective Field Theory* (SMEFT), where all new degrees of freedom are assumed to be much heavier than the electroweak scale. This allows us to be model-agnostic, while obtaining information on new degrees of freedom only via the ratio of their couplings and scale. Any extension will have to be compatible with constraints on all affected observables, beyond $R_{D^{(*)}}$ itself, in addition to the requirement of invariance under the SM gauge group. Our final question will involve understanding the relation between measurements at the *Large Hadron Collider* (LHC) and the bounds these provide on NP considered in the context of $R_{D^{(*)}}$.

³See [Feruglio, Paradisi, and Paltori \(2017\)](#) for an exception, who consider NP coupled to the third generation lepton doublet only, but study the hierarchical LFV contributions to $R_{D^{(*)}}$ (and beyond) induced via the mixing of lepton mass eigenstates.

SECTION 2

Phenomenology of $b \rightarrow c\tau\nu_\ell$

2.1 General considerations

Our analysis of a NP explanation of $R_{D^{(*)}}$ data aims to be independent of any specific model-building choices. However, it is useful to note some key features of the relevant extensions to the SM in setting the context for our analysis. Given that the missing contribution to $b \rightarrow c\tau\bar{\nu}$ is of $\mathcal{O}(10\%)$ of a (SM) tree-level amplitude squared, NP models attempting to account for all (or most) of this discrepancy also typically generate this transition at tree-level. Assuming a minimal extension with only a single new tree-level mediator, the nature of this transition allows only a few viable options.

Given that the new particle couples to quarks and leptons at tree-level, it must be a boson (by dimensional analysis). Additionally, to do so while avoiding direct production constraints implies the particle must be heavy. Since it must mediate a FCCC process, it must also be charged. By Lorentz invariance, the four fermions involved interact in pairs and we require an additional scalar or vector field in the interaction. SU(3)_C representation is determined by whether the interaction involves two quarks/leptons, and thus a colorless new particle, or one quark and one lepton, with a new colored particle.

Each of these possibilities has been proposed and studied in the literature: heavy charged colorless vector bosons (W' , e.g. [Boucenna, Celis, Fuentes-Martín, Vicente, & Virto, 2016](#)), heavy charged colorless scalars (H' , e.g. [Chen & Nomura, 2017](#)), and heavy charged colored scalar or vector bosons (leptoquarks, e.g. [Angelescu, Bećirević, Faroughy, Jaffredo, & Sumensari, 2021](#)). Global fits to data from various flavor physics processes ($R_{D^{(*)}}$, $R_{K^{(*)}}$, and other FCNC observables) have also been performed (see e.g. [Murgui, Peñuelas, Jung, & Pich, 2019](#); [Aebischer et al., 2020](#); [Alok, Kumar, Kumar, & Sharma, 2019](#)), comparing multiple models (including scenarios with two or more mediators) as potential solutions to several flavor anomalies simultaneously.

2.2 $R_{D^{(*)}}$ in the SMEFT

In discussing NP within the SMEFT framework, we consider a Lagrangian

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \mathcal{L}_{\text{NP}}^{\text{eff}}, \quad (2.1)$$

extending the renormalizable SM Lagrangian, \mathcal{L}_{SM} , where the effective NP Lagrangian, $\mathcal{L}_{\text{NP}}^{\text{eff}}$, is composed of Lorentz-invariant operators of dimension-five and above, built from SM fields and respecting the SM gauge group $(\text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y)$. New terms in Eq. (2.1) will be of the form

$$\mathcal{L}_{\text{NP}}^{\text{eff}} = \sum_i \frac{C_i}{\Lambda^{d-4}} \mathcal{O}_i^{(d)}, \quad (2.2)$$

where operators of dimension $d > 4$, $\mathcal{O}_i^{(d)}$, are suppressed by powers of $1/\Lambda$ and accompanied by so-called *Wilson coefficients* (WCs), C_i . The scale Λ determines the energy above which degrees of freedom are integrated out, making our description in Eq. (2.1) *effectively* valid at energies below Λ , and model-agnostic above Λ .

Dimension-six operators

Since we aim to generate $b \rightarrow c\tau\bar{\nu}$ transitions involving four fermions, the minimum operator dimension to consider is six. Therefore, we begin by expanding dimension-six SMEFT operators coupling two quarks to two leptons¹. Specifically, we consider two combinations of left-handed fermion bilinears with a Lorentz vector-like structure, which together generate both $b \rightarrow c$ and $b \rightarrow s$ transitions at tree-level. They are:

$$\mathcal{O}_1^{ijkm} := (\overline{L}_i \gamma_\sigma L_j) (\overline{Q}_k \gamma^\sigma Q_m), \quad (2.3)$$

which generates $b \rightarrow s$ transitions, and

$$\mathcal{O}_3^{ijkm} := (\overline{L}_i \gamma_\sigma \tau^a L_j) (\overline{Q}_k \gamma^\sigma \tau^a Q_m), \quad (2.4)$$

which generates both $b \rightarrow c$ and $b \rightarrow s$ transitions, and is currently among the preferred candidates in the literature for NP explanations of $R_{D^{(*)}}$ (see e.g. Shi et al., 2019). We use the notation L for $\text{SU}(2)_L$ -doublet lepton fields, Q for $\text{SU}(2)_L$ -doublet quark fields, τ for Pauli matrices, and i, j, k, m for flavor indices. Lorentz, $\text{SU}(3)_C$ and $\text{SU}(2)_L$ indices are contracted within brackets and not shown.

After expanding the $\text{SU}(2)_L$ structure of Eqs. (2.3) and (2.4), and rotating to the usual d -type quark mass basis (\mathcal{O}_1 and \mathcal{O}_3 do not introduce any new mass terms), we have the redefinitions

$$\mathcal{O}_1^{ijkm} := V_{k'k} V_{m'm}^* (\overline{\nu}_{Li} \gamma_\sigma \nu_{Lj} + \overline{e}_{Li} \gamma_\sigma e_{Lj}) (\overline{u}_{Lk'} \gamma^\sigma u_{Lm'} + \overline{d}_{Lk'} \gamma^\sigma d_{Lm'}) \quad (2.5)$$

and

$$\begin{aligned} \mathcal{O}_3^{ijkm} := & 2V_{m'm}^* (\overline{\nu}_{Li} \gamma_\sigma e_{Lj}) (\overline{d}_{Lk} \gamma^\sigma u_{Lm'}) + 2V_{k'k} (\overline{e}_{Li} \gamma_\sigma \nu_{Lj}) (\overline{u}_{Lk'} \gamma^\sigma d_{Lm}) \\ & + V_{k'k} V_{m'm}^* (\overline{\nu}_{Li} \gamma_\sigma \nu_{Lj} - \overline{e}_{Li} \gamma_\sigma e_{Lj}) (\overline{u}_{Lk'} \gamma^\sigma u_{Lm'} - \overline{d}_{Lk'} \gamma^\sigma d_{Lm'}), \end{aligned} \quad (2.6)$$

where V is the CKM quark mixing matrix. The leading contribution to $b \rightarrow c\tau\bar{\nu}_l$ comes from \mathcal{O}_3^{lslb} . With this in mind, and in order to avoid strong FCNC constraints,

¹A full list of dimension-six operators in the SMEFT can be found in Grzadkowski, Iskrzyński, Misiak, and Rosiek (2010).

we consider non-zero WCs only for $C_{1,3}^{\tau ls b}$, denoted by the shorthand $C_{1,3}^\ell$. We absorb any implicit factors (e.g. from renormalization group flow) into the definitions of these two coefficients, such that we have an effective NP Lagrangian at the scale $m_b \sim 5 \text{ GeV}$ given by

$$\begin{aligned}\Lambda^2 \mathcal{L}_{\text{NP}}^{\text{eff}} = & (C_1^\ell + C_3^\ell) V_{is} V_{jb}^* (\bar{u}_{Li} \gamma^\mu u_{Lj}) (\bar{\nu}_\tau \gamma_\mu \nu_l) \\ & + (C_1^\ell - C_3^\ell) V_{is} V_{jb}^* (\bar{u}_{Li} \gamma^\mu u_{Lj}) (\bar{\tau}_L \gamma_\mu l_L) \\ & + (C_1^\ell - C_3^\ell) (\bar{s}_L \gamma^\mu b_L) (\bar{\nu}_\tau \gamma_\mu \nu_l) \\ & + (C_1^\ell + C_3^\ell) (\bar{s}_L \gamma^\mu b_L) (\bar{\tau}_L \gamma_\mu l_L) \\ & + 2C_3^\ell V_{is} (\bar{u}_{Li} \gamma^\mu b_L) (\bar{\tau}_L \gamma_\mu \nu_l) \\ & + 2C_3^\ell V_{jb} (\bar{u}_{Lj} \gamma^\mu s_L) (\bar{\tau}_L \gamma_\mu \nu_l) + \text{h.c..}\end{aligned}\quad (2.7)$$

The NP contribution to $R_{D^{(*)}}$

Given Eq. (2.7), and taking into account the W -mediated diagram in the SM, the Lagrangian terms contributing to $b \rightarrow c\tau\bar{\nu}_l$ are

$$\mathcal{L} \supset \left(\frac{4G_F V_{cb} \delta_{l\tau}}{\sqrt{2}} + \frac{2C_3^\ell V_{cs}}{\Lambda^2} \right) (\bar{c}_L \gamma^\mu b_L) (\bar{\tau}_L \gamma_\mu \nu_l). \quad (2.8)$$

Since we can write $R_{D^{(*)}}$ relative to the SM prediction as

$$\frac{R_{D^{(*)}}}{R_{D^{(*)}}^{\text{SM}}} = \frac{\Gamma^{\text{SM+NP}}(B \rightarrow D^{(*)}\tau\bar{\nu}_\tau)}{\Gamma^{\text{SM}}(B \rightarrow D^{(*)}\tau\bar{\nu}_\tau)} + \frac{\sum_\ell \Gamma^{\text{NP}}(B \rightarrow D^{(*)}\tau\bar{\nu}_\ell)}{\Gamma^{\text{SM}}(B \rightarrow D^{(*)}\tau\bar{\nu}_\tau)}, \quad (2.9)$$

we see that it only depends on ratios of the coefficients in Eq. (2.8),

$$= 1 + \frac{\sqrt{2}}{G_F} \text{Re} \left(\frac{V_{cs}}{V_{cb}} \frac{C_3^\tau}{\Lambda^2} \right) + \frac{|C_3^e|^2 + |C_3^\mu|^2}{2G_F^2 \Lambda^4} \left| \frac{V_{cs}}{V_{cb}} \right|^2, \quad (2.10)$$

where G_F is the Fermi constant, and we assume the contribution from $|C_3^\tau|^2$ is negligible compared to the interference term linear in C_3^τ .

Therefore, to account for the 14% shift observed in $R_{D^{(*)}}$ data with NP in $b \rightarrow c\tau\nu_\ell$ only, we require

$$\frac{\sqrt{|C_3^e|^2 + |C_3^\mu|^2}}{\Lambda^2} = (0.24 \pm 0.04) \text{ TeV}^{-2} = \frac{1}{((2.0 \pm 0.2) \text{ TeV})^2}, \quad (2.11)$$

and to account for the shift with NP purely in $b \rightarrow c\tau\nu_\tau$, we need

$$\frac{C_3^\tau}{\Lambda^2} = (0.046 \pm 0.016) \text{ TeV}^{-2} = \frac{1}{((4.7 \pm 1.1) \text{ TeV})^2}. \quad (2.12)$$

2.3 Bounds from constrained processes

We consider now the feasibility of the scenario where $R_{D^{(*)}}$ is explained by $|C_3^\ell|/\Lambda^2 \sim 1/(2 \text{ TeV})^2$. From Eq. (2.7), we see that C_3^ℓ contributes to various FCNC and LFV decays of the B meson which are forbidden in the SM, most notably through the $\bar{s}b\tau\ell$ and $\bar{s}b\bar{\nu}_\tau\nu_\ell$ terms. In the following, we obtain the constraints on C_3^ℓ imposed by the experimental upper bounds on these processes, measured at the “ B -factories”: LHCb, Belle, and BaBar.

$$B_s \rightarrow \tau^\pm \mu^\mp$$

The $B_s \rightarrow \tau^+ \mu^-$ decay rate is given by

$$\Gamma(B_s \rightarrow \tau^+ \mu^-) = \frac{|C_1^\mu + C_3^\mu|^2}{\Lambda^4} \frac{f_{B_s}^2 m_\tau^2 m_{B_s}}{64\pi} \left(1 - \frac{m_\tau^2}{m_{B_s}^2}\right)^2. \quad (2.13)$$

The experimental upper bound (LHCb Collaboration et al., 2019),

$$\mathcal{B}(B_s \rightarrow \tau^\pm \mu^\mp) < 4.2 \cdot 10^{-5}, \quad (2.14)$$

implies

$$\frac{|C_1^\mu + C_3^\mu|}{\Lambda^2} < 0.073 \text{ TeV}^{-2}. \quad (2.15)$$

$$B^+ \rightarrow K^+ \tau^+ \mu^-$$

The $B^+ \rightarrow K^+ \tau^+ \mu^-$ branching ratio is given by (Bordone, Cornella, Fuentes-Martín, & Isidori, 2018; Crivellin et al., 2015)

$$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \mu^-) = 8.2 \cdot 10^{-3} \text{ TeV}^4 \cdot \frac{|C_1^\mu + C_3^\mu|^2}{\Lambda^4}. \quad (2.16)$$

The experimental upper bound (BaBar Collaboration et al., 2012b),

$$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \mu^-) < 2.8 \cdot 10^{-5}, \quad (2.17)$$

implies

$$\frac{|C_1^\mu + C_3^\mu|}{\Lambda^2} < 0.058 \text{ TeV}^{-2}. \quad (2.18)$$

$$B^+ \rightarrow K^+ \tau^+ e^-$$

The $B^+ \rightarrow K^+ \tau^+ e^-$ branching ratio is given by

$$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ e^-) = 8.2 \cdot 10^{-3} \text{ TeV}^4 \cdot \frac{|C_1^e + C_3^e|^2}{\Lambda^4}. \quad (2.19)$$

The experimental upper bound (BaBar Collaboration et al., 2012b),

$$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ e^-) < 1.5 \times 10^{-5}, \quad (2.20)$$

implies

$$\frac{|C_1^e + C_3^e|}{\Lambda^2} < 0.044 \text{ TeV}^{-2}. \quad (2.21)$$

$$B^+ \rightarrow K^+ \bar{\nu}_\tau \nu_\ell$$

The $B^+ \rightarrow K^+ \bar{\nu}\nu$ branching ratio, relative to the SM rate, is given by (Buras, Girrbach-Noe, Niehoff, & Straub, 2015)

$$R_{K\bar{\nu}\nu} := \frac{\mathcal{B}(B^+ \rightarrow K^+ \bar{\nu}\nu)}{\mathcal{B}^{\text{SM}}(B^+ \rightarrow K^+ \bar{\nu}\nu)} = 1 + 3.5 \cdot 10^3 \text{ TeV}^4 \cdot \frac{|C_1^\mu - C_3^\mu|^2 + |C_1^e - C_3^e|^2}{\Lambda^4}. \quad (2.22)$$

The experimental upper bound (BaBar Collaboration et al., 2013; Belle Collaboration, Grygier, et al., 2017),

$$\mathcal{B}(B^+ \rightarrow K^+ \bar{\nu}\nu) < 1.6 \cdot 10^{-5}, \quad (2.23)$$

which corresponds to $R_{K\bar{\nu}\nu} \lesssim 4$, implies

$$\frac{|C_1^\ell - C_3^\ell|}{\Lambda^2} < 0.031 \text{ TeV}^{-2}. \quad (2.24)$$

Interpretation

Given the allowed ranges for $|C_1^\ell \pm C_3^\ell|$, we can extract upper bounds on C_3^ℓ alone. Assuming the WCs are real, we obtain:

$$\frac{|C_3^\mu|}{\Lambda^2} < 0.044 \text{ TeV}^{-2}, \quad \frac{|C_3^e|}{\Lambda^2} < 0.037 \text{ TeV}^{-2}. \quad (2.25)$$

Comparing to the requirement of Eq. (2.11), we find that contributions from $b \rightarrow c\tau\bar{\nu}_{e,\mu}$ can account for up to $\sim 4\%$ of the shift in $R_{D^{(*)}}$ from the SM expectation.

2.4 Other affected processes

The terms in Eq. (2.7) also affect the rates of other constrained processes, not included in the above. With our choice of flavor structure for $C_{1,3}^{ijkm}$, the contributions to these are CKM suppressed and hence weak. However, they might provide relevant limits given NP in the form of Eqs. (2.3) and (2.4) with a different flavor structure.

Forbidden t decays

Decays such as $t \rightarrow q\tau\ell$ with $q = c, u$ are currently unbounded by experiment (Particle Data Group et al., 2020).

Forbidden τ decays

There exist experimental upper bounds for two of these (Belle Collaboration et al., 2007; BaBar Collaboration et al., 2007):

$$\begin{aligned} \mathcal{B}(\tau \rightarrow \mu\pi^0) &< 1.1 \cdot 10^{-7}, \\ \mathcal{B}(\tau \rightarrow e\pi^0) &< 8.0 \cdot 10^{-8}. \end{aligned} \quad (2.26)$$

Comparing Eq. (2.7) to the allowed $\tau \rightarrow \nu_\tau \bar{u}q$ ($q = d, s$) process, whose branching ratio is ~ 0.65 , we notice a CKM suppression of $|V_{us}V_{ub}|^2 \sim 7 \cdot 10^{-7}$ in the contributions to the above. Therefore, limits of $\mathcal{O}(10^{-7})$ on these branching ratios do not provide significant constraints.

Allowed τ decays

The ratio $R_{K/\pi}$, defined and measured to be (Particle Data Group et al., 2020)

$$R_{K/\pi} := \frac{\mathcal{B}(\tau \rightarrow K^-\nu)}{\mathcal{B}(\tau \rightarrow \pi^-\nu)} = (6.44 \pm 0.09) \cdot 10^{-2}, \quad (2.27)$$

is affected by a new contribution to the numerator. The SM prediction (BaBar Collaboration et al., 2010)

$$R_{K/\pi} = \frac{f_K^2 |V_{us}|^2}{f_\pi^2 |V_{ud}|^2} \frac{(1 - m_K^2/m_\tau^2)^2}{(1 - m_\pi^2/m_\tau^2)^2} (1 + \delta_{\text{LD}}), \quad (2.28)$$

has a theoretical uncertainty dominated by the long distance correction, $\delta_{\text{LD}} = (0.03 \pm 0.44)\%$. Compared to the $\tau \rightarrow K^-\nu$ rate in the SM, the new contribution has a CKM suppression of $|V_{ub}/V_{us}|^2 \sim 3 \times 10^{-4}$. Therefore, it is much smaller than the uncertainty from δ_{LD} and cannot be constrained.

Forbidden J/ψ decays

There are new contributions to the decays in (BES Collaboration et al., 2004; BESIII Collaboration et al., 2021):

$$\begin{aligned} \mathcal{B}(J/\psi \rightarrow \mu^\pm \tau^\mp) &< 2.0 \cdot 10^{-6}, \\ \mathcal{B}(J/\psi \rightarrow e^\pm \tau^\mp) &< 7.5 \cdot 10^{-8}. \end{aligned} \quad (2.29)$$

Compared to the allowed $J/\psi \rightarrow \ell^+ \ell^-$ process, with branching ratio ~ 0.06 , we are CKM suppressed by $|V_{cb} V_{cs}|^2 \sim 1.6 \times 10^{-3}$. In addition, we have that (Abada, Bećirević, Lucente, & Sumensari, 2015; Hazard & Petrov, 2016)

$$\frac{\mathcal{B}(J/\psi \rightarrow \ell^\pm \tau^\mp)}{\mathcal{B}(J/\psi \rightarrow \ell^+ \ell^-)} \propto \left(\frac{m_{J/\psi}}{\Lambda} \right)^4. \quad (2.30)$$

Therefore, the constraints these bounds imply are of $\mathcal{O}(6m_{J/\psi} \sim 20 \text{ GeV})$ on $\Lambda/\sqrt{|C_3^e|}$ and even weaker on $\Lambda/\sqrt{|C_3^\mu|}$, hence are not relevant.

2.5 Expected bounds from LHC searches

The existence of NP in the form of Eq. (2.7) would be a source of LFV in pp collisions. Here, we estimate the upper bound on $|C_1^\mu + C_3^\mu|/\Lambda^2$, denoted $1/\Lambda_{\tau\mu}^2$, that we could expect to obtain at the LHC and *High-Luminosity LHC* (HL-LHC), with a dedicated search for $pp \rightarrow \tau^\pm \mu^\mp X_h$, where X_h stands for any final state hadrons. Our estimate is based on the ATLAS search for NP in $pp \rightarrow \mu^+ \mu^-$ (with up to one b -jet) at $\sqrt{s} = 13 \text{ TeV}$, with 139 fb^{-1} of data (ATLAS Collaboration, Aad, Abbott, Abbott, Abed Abud, et al., 2021b). Building upon this search, we deduce the reach of ATLAS for $\Lambda_{\tau\mu}$ ².

²For related work, see Choudhury, Kumar, and Kundu (2019) and Kumar (2020).

The phenomenological framework for the ATLAS analysis (Afik, Cohen, Gozani, Kajomovitz, & Rozen, 2018) considers new $b\bar{s}\ell\ell$ operators of the form in Eqs. (2.3) and (2.4). ATLAS reports a bound $\Lambda_{\mu\mu} > 2.4$ TeV on the scale that suppresses dimension-six $b\bar{s}\mu\mu$ contact interaction. This limit is obtained in the analysis with a b -veto ($pp \rightarrow \mu^+\mu^- + 0b$)³, by searching for events with high dimuon mass, $m_{\mu\mu} > 1800$ GeV. The analogous bounds $\Lambda_{ee} > 2.0(1.8)$ TeV are also attained, from the analysis of $pp \rightarrow e^+e^- + 0b(1b)$ events with high dielectron mass, $m_{ee} > 1900(1500)$ GeV, showing a slightly lower sensitivity to dimension-six $b\bar{s}ee$, and thus $b\bar{s}\tau e$ contact interaction.

The bound on $\Lambda_{\mu\mu}$ is inferred from the upper bound on $\sigma_{\mu\mu}$, the $\mu\mu$ signal cross section. In translating the search to $pp \rightarrow \tau_h^\pm \mu^\mp + 0b$, one muon is replaced by a hadronically decaying tau lepton. Given that $\mathcal{B}(\tau \rightarrow \text{hadrons}) \approx 2/3$, then at similar energy and for $\Lambda_{\tau\mu} = \Lambda_{\mu\mu}$, we expect

$$\frac{\sigma_{\tau_h\mu}}{\sigma_{\mu\mu}} \approx \frac{2}{3}, \quad (2.31)$$

where $\sigma_{\tau_h\mu}$ is the $\tau_h\mu$ signal cross section.

The leading SM background at high $m_{\mu\mu}$ is the Drell-Yan process to two muons, which is suppressed in a $\tau\mu$ final state selection. In this latter case, the background contribution is from the Drell-Yan process to two tau leptons, and is suppressed by requiring that one decays hadronically (with branching ratio $\approx 2/3$) and the other muonically (with branching ratio $\approx 1/6$). Hence, we have

$$\frac{\sigma_{\tau_h\mu}^{Z/\gamma^*}}{\sigma_{\mu\mu}^{Z/\gamma^*}} \approx 2 \cdot \frac{2}{3} \cdot \frac{1}{6} = \frac{2}{9}. \quad (2.32)$$

The next most significant SM background is the top-quark contribution, composed of $t\bar{t}$, Wt and $Wt\bar{t}$ processes. In each of these, two final state leptons arise from independent decay chains, *e.g.* $t \rightarrow l^+$ and $\bar{t} \rightarrow l'^-$ in $t\bar{t}$ events. Thus,

$$\frac{\sigma_{\tau_h\mu}^t}{\sigma_{\mu\mu}^t} \approx 2 \cdot \frac{2}{3} = \frac{4}{3}. \quad (2.33)$$

Defining $r_{Z/t} := \sigma_{\mu\mu}^{Z/\gamma^*}/\sigma_{\mu\mu}^t$, the ratio of total SM backgrounds from the contributions above can be written

$$\frac{\sigma_{\tau_h\mu}^{\text{bg}}}{\sigma_{\mu\mu}^{\text{bg}}} \approx \frac{(2/9)r_{Z/t} + 4/3}{r_{Z/t} + 1}. \quad (2.34)$$

Therefore, the relative sensitivity can be approximated as

$$\frac{(s/\sqrt{b})_{\tau_h\mu}}{(s/\sqrt{b})_{\mu\mu}} \approx \frac{2}{3} \sqrt{\frac{r_{Z/t} + 1}{(2/9)r_{Z/t} + 4/3}}, \quad (2.35)$$

³A slightly weaker bound, $\Lambda_{\mu\mu} > 2.0$ TeV, is obtained in the b -tag category ($pp \rightarrow \mu^+\mu^- + 1b$) with $m_{\mu\mu} > 1600$ GeV.

where s and b are, respectively, the signal and background yields above a given m_{ll} threshold. For $r_{Z/t} > 1$, as is the case at high m_{ll} , we thus have

$$\frac{(s/\sqrt{b})_{\tau_h \mu}}{(s/\sqrt{b})_{\mu \mu}} \gtrsim \frac{3}{4}, \quad (2.36)$$

which gives a decrease in sensitivity by a factor $3/4$ as our conservative estimate. Note that if $r_{Z/t} \gtrsim 4$, the sensitivity to $\tau \mu$ is in fact stronger than to $\mu \mu$.

Finally, given that the signal sensitivity is fixed by experiment, and using our conservative estimate of the relative sensitivity to $\tau \mu$, we expect the observational significance in $\tau \mu$ to be the same as in $\mu \mu$ for $s_{\tau \mu} \sim (4/3)s_{\mu \mu}$. Since $s \propto 1/\Lambda^4$, we therefore expect that current data ($L = 139 \text{ fb}^{-1}$) can provide a lower bound of

$$\Lambda_{\tau \mu} > 2.2 \text{ TeV}. \quad (2.37)$$

If the HL-LHC achieves its target integrated luminosity, $L = 4000 \text{ fb}^{-1}$, and no signal is observed, this bound would be strengthened to

$$\Lambda_{\tau \mu} > 3.3 \text{ TeV}, \quad (2.38)$$

where we estimate a scaling of $s/\sqrt{b} \propto \sqrt{L}$ since $s/\sqrt{b} \sim L\sigma_s/\sqrt{L\sigma_b}$.

2.6 Discussion

Determining $R_{D^{(*)}}$ is an important test of LFU and current observations hint at NP violating this property at tree level. However, the puzzle also concerns LFV, as $R_{D^{(*)}}$ is affected by processes which do not conserve lepton flavor. In this study, our goal was to investigate whether the missing piece of the puzzle could come from purely lepton flavor violating NP.

We conclude that accounting for the observed discrepancy in $R_{D^{(*)}}$ with $b \rightarrow c\tau\bar{\nu}_{e,\mu}$ transitions requires $|C_3^\ell|/\Lambda^2 \simeq 0.24 \text{ TeV}^{-2}$. In contrast, bounds from other LFV and FCNC decays of the B meson impose the limit $|C_3^\ell|/\Lambda^2 \lesssim 0.04 \text{ TeV}^{-2}$. Therefore, these transitions can contribute at most $\sim 4\%$ of the required shift.

On the other hand, accounting for the discrepancy with new $b \rightarrow c\tau\bar{\nu}_\tau$ transitions requires $|C_3^\tau|/\Lambda^2 \simeq 0.046 \text{ TeV}^{-2}$. Since all phenomenological constraints can be satisfied with $|C_3^\mu|/\Lambda^2 < 0.044 \text{ TeV}^{-2}$ and $|C_3^e|/\Lambda^2 \simeq 0.037 \text{ TeV}^{-2}$, we reason that no special flavor structure is required of NP with the effective form of Eq. (2.7) to explain $R_{D^{(*)}}$.

If NP generates C_3^τ , C_3^μ and C_3^e all of the same order of magnitude, with $|C_3^\ell|/\Lambda^2 \sim 0.04 \text{ TeV}^{-2}$, $R_{D^{(*)}}$ can be satisfied within less than 1σ while complying with all relevant constraints. In such a scenario, the contribution to the shift in $R_{D^{(*)}}$ from C_3^τ would be larger by a factor ~ 30 relative to C_3^ℓ .

Searches for LFV at the LHC can provide complementary information on NP operators considered in the context of $R_{D^{(*)}}$. Our estimated bounds on $\Lambda_{\tau \mu}$ from ATLAS, Eqs. (2.37) and (2.38), can be compared to the limit measured by BaBar in Eq. (2.18), corresponding to $\Lambda_{\tau \mu} \gtrsim 4.2 \text{ TeV}$. From this comparison, we infer that

future searches at the (HL-)LHC will have to achieve an improvement in signal sensitivity by a factor $\sim (2.5)13$ to compete with existing constraints on $|C_1^\ell + C_3^\ell|/\Lambda^2$ from the B -factories. Future work on this topic could involve an analysis of the constraints on effective $bb\tau\ell$ operators.

PART II

A data-directed paradigm for new physics searches: The bump-hunting example

SECTION 3

Introduction

3.1 The blind analysis paradigm

Searches for NP in data collected by the LHC general-purpose experiments, ATLAS and CMS, have not yet discovered any significant deviation from SM predictions. Hundreds of NP scenarios have been hypothesized and tested for, with searches typically focusing on those which carry the highest theoretical incentive. Most have targeted specific signal characteristics, established in advance by theoretical considerations, and have used traditional data analysis techniques (e.g. recently [ATLAS Collaboration, Aad, Abbott, Abbott, Abed Abud, et al., 2021a](#); [ATLAS Collaboration et al., 2022](#); [CMS Collaboration, Sirunyan, Tumasyan, Adam, Andrejkovic, et al., 2021](#)). In particular, with only a few exceptions (e.g. [ATLAS Collaboration et al., 2020a](#); [CMS Collaboration, Sirunyan, Tumasyan, Adam, Ambrogi, et al., 2021](#); [ATLAS Collaboration, Aaboud, et al., 2019](#)), almost all searches have been conducted following the blind analysis paradigm, in which the data of interest is only accessed in the final stage of the analysis.

This is a powerful strategy when searching for signals that are well-defined in a theoretical model. It greatly reduces the risk of biasing results based on any preconceived ideas, by withholding knowledge of the outcome until the analysis is complete. However, when searching for NP more broadly, this method has a substantial disadvantage: the enormous amount of resources required to analyze a single region. Since the observed data is made unavailable, it becomes crucial to precisely estimate and validate the background and systematic errors expected in the signal region—a process which usually requires several person years. Heavy reliance on *Monte Carlo* (MC) event generators also introduces mismodeling effects which are difficult to account for.

Given the limited time, manpower and computing resources available, the cost of such analyses has allowed only a small part of the space spanned by all measured observables (the “observable-space”) to be explored to-date. Indeed, searches generally focus on inclusive final states (e.g. dilepton), ignoring all but a few observables, and avoid exclusive final states (e.g. dilepton + single b -jet + missing transverse energy). Signals which can only be discovered at a given sensitivity when associated with other objects are thus easily missed. Evidently, it is possible that NP events are hidden within existing LHC data and yet to be found.

3.2 A data-directed paradigm

The challenges above have motivated us to consider alternative avenues of discovery in experimental high energy physics. The general search strategies proposed by [ATLAS Collaboration, Aaboud, et al. \(2019\)](#) and [CMS Collaboration, Sirunyan, Tumasyan, Adam, Ambrogi, et al. \(2021\)](#) approach the problem by scanning data in search for discrepancies between observations and MC simulations of the SM in many final states, using various search algorithms. In a similar spirit, we consider a *data-directed paradigm* (DDP) ([Volkovich, De Vito Halevy, & Bressler, 2021](#)), where the primary goal is to fully exploit collider data, directing dedicated analyses towards regions of the data which are the most likely to contain deviations from the SM—and, hence, lead to NP signals. This can be accomplished by complementing the high sensitivity of traditional search methods with the speed and efficiency of automated scans performed on many regions of the observable-space. Crucially, in the DDP we do not focus on comparisons of data to MC simulation, in contrast to the general search strategies mentioned.

Two key ingredients are required in order to implement such a strategy:

1. A well-established property of the SM, with respect to which deviations can be searched for;
2. An efficient algorithm, to scan the observable-space in search for such deviations.

Once the breadth of the data has been scanned, exclusive selections can be prioritized based on information in the data itself, and can then be studied in depth. The selections and characteristics exhibiting the highest probability of disagreement with fundamental SM predictions (independent of MC simulation), according to well-defined statistical measures, thus provide data-directed signal hypotheses for subsequent high-sensitivity analyses.

Data-driven methods are not new to particle physics and are being revisited with the recent rise in innovation in the field of artificial intelligence. The breakthrough analyses which discovered the Higgs boson employed tools from machine learning to increase sensitivity by an amount equivalent to collecting 50% more data ([Radovic et al., 2018](#)). In another relevant study ([ATLAS Collaboration et al., 2020b](#)), these techniques helped reduce dependence on MC simulation in search for NP. Much interest has developed into various other applications to particle physics searches (e.g. [Guest, Cranmer, & Whiteson, 2018](#); [Metodiev, Nachman, & Thaler, 2017](#); [Nachman & Shih, 2020](#)). Here, we investigate the potential of looking at the data and applying similar tools in the first stage of an analysis (when deciding where to search), while also avoiding MC simulation.

We focus our implementation on one of the core objectives of experimental high energy physics in particular: the search for new resonances (“bump-hunting”). This is an area which could benefit greatly from a DDP approach, as there is a vast spectrum of bump-like signal hypotheses that are theoretically plausible. While strong bounds on NP can be placed even when the targeted regions of data are revealed, *a posteriori*, to lack any indication of a signal, it would be infeasible to devote substantial resources to each (as in, e.g., [ATLAS Collaboration, Aad, et](#)

al., 2019; ATLAS Collaboration et al., 2020c). Therefore, we aim to examine the prospects of adopting data-directed signal hypotheses as candidates for discovering new particles.

3.3 Our goals

Based on the work presented in Volkovich et al. (2021), we address the following questions:

1. *Can we correctly identify bumps in invariant mass distributions quickly?*
2. *Specifically, to what extent can we avoid extensive background and systematic error estimation, and reliance on MC physics engines?*
3. *How much signal-to-background separation power would such a method provide?*

We begin by framing a bump-hunt within the DDP concept, and describing our implementation based on an artificial *neural network* (NN). We then use artificially generated datasets to evaluate its capacity to identify bumps through various tests, and perform comparisons to the profile likelihood ratio (commonly used in hypothesis testing). Finally, we discuss possible uses and future developments, aimed at searches at the LHC.

SECTION 4

Bump-hunting

4.1 Bump-hunting in the DDP

In order to realize a bump-hunt within the design of the DDP, as outlined above, we formulate it as follows:

1. Within the SM, in absence of known resonances, invariant mass distributions are generally smoothly falling. Hence, we search for statistically significant deviations from large-scale smoothness and consider these as departures from SM prediction.
2. A neural network can be trained to map an invariant mass distribution into a distribution of statistical significance for excesses of events known as a “ z ” distribution ([Cowan, Cranmer, Gross, & Vitells, 2011](#)). A high z value defines the notion of a bump, and such an algorithm can be used to quickly scan large collections of data.

These two elements can be implemented in a simple fashion to showcase the concept of bump-hunting in the DDP. In this regard, we make the following choices: (*i*) We perform a random sampling of a variety of monotonic functions, such that the qualities of smoothly falling curves can be approximated from the features common to many different samples in a large dataset; (*ii*) We train a single NN with a simple architecture, inspired by models intended for feature-recognition. Furthermore, we do not attempt to enhance bumps by optimizing selections event-by-event (studied in, e.g., [Farina, Nakai, & Shih, 2020](#); [J. H. Collins, Howe, & Nachman, 2019](#); [J. Collins, Howe, & Nachman, 2018](#)).

With the trained NN, we can execute statistical inference from many selections of observed invariant mass data, identifying large bumps. Most will be due to statistical fluctuations and will fade with more data. Those arising from detector effects (trigger thresholds, kinematic edges, etc.) should be visible in MC simulations and excluded. Finally, the most significant bumps remaining could be considered for signal hypotheses. Some may inevitably turn out to be caused by mismodeled systematic effects. Although, such a method may, in fact, help identify such modeling errors in regions less commonly inspected.

4.2 Neural network implementation

Our implementation is based on the predictions of a supervised NN, which we train and test using simulated data. Each data sample consists of an input and a target, both are one-dimensional arrays of 100 numbers. The NN is trained by example to map inputs to targets. Each input array represents a histogram of observed events over 100 invariant mass bins, simulating the full statistics of a search projected and binned onto this particular observable (as opposed to individual events, as in [Farina et al., 2020](#); [J. H. Collins et al., 2019](#); [J. Collins et al., 2018](#)). Each target is a distribution of the test statistic, z , over the same domain as its corresponding input array. The statistic is defined as ([Cowan et al., 2011](#))

$$z := \sqrt{q_0}, \quad (4.1)$$

where q_0 is calculated via a profile likelihood ratio test for a positive signal:

$$q_0 := \begin{cases} -2 \ln \left(\frac{L(0)}{L(\hat{\mu})} \right), & \hat{\mu} > 0, \\ 0, & \hat{\mu} \leq 0, \end{cases} \quad (4.2)$$

where L and $\hat{\mu}$ are defined as follows. $L(\mu)$ is the likelihood of a signal strength μ given an observed distribution of events (the input array), n , a hypothesized background distribution, b , and a hypothesized signal distribution, s , assuming Poisson probabilities in each bin (indexed by the subscript i):

$$L(\mu) = \prod_{i=1}^{100} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}. \quad (4.3)$$

The value $\mu = \hat{\mu}$ maximizes the likelihood function L and is calculated numerically.

For each sample, the distribution of z is calculated using the background and signal shapes used to generate n , by evaluating $z_i(n)$ as the hypothesized signal is shifted such that s is centered, in turn, on each given bin i . The trained NN thus approximates the mapping $n \rightarrow z$ (input to target) for general distributions of observed events. Most importantly, it does so without explicitly estimating the signal and background, and hence takes a negligible amount of time to make a prediction, z_{pred} , of the statistical significance distribution, given any 100-bin invariant mass histogram. Therefore, by training it with data containing bumps of various significance levels, we can use its predictions as quick estimations of bump positions and significances.

Data generation

Our process for generating samples is depicted in Fig. 4.1. We begin by randomly selecting a functional form for the background distribution from the following ten options:

$$\begin{aligned} & a(x - x_2)^2 + y_2, \quad ax + b, \quad \frac{1}{ax} + b, \quad \frac{1}{ax^2} + b, \quad \frac{1}{ax^3} + b, \quad \frac{1}{ax^4} + b, \\ & be^{-ax}, \quad -a \ln x + b, \quad (y_1 - y_2) \cos(ax + b) + y_2, \quad \cosh(a(x - x_2)) + b. \end{aligned} \quad (4.4)$$

The parameters a and b are calculated such that the curve decays from (x_1, y_1) to (x_2, y_2) , where $x_1 < x_2$ are, respectively, the centers of bins 1 and 100 (the edge bins), and $y_1 > y_2$ are event counts chosen randomly from the interval [100, 10000]. This results in a smoothly falling background distribution, which is then taken as a distribution of means in a Poisson sampling, to produce a background with fluctuations. With the chosen range of event counts, these inputs are representative of data with large statistics and dynamic range.

Signals are generated as standard Gaussians with a mean chosen randomly between bins 25 and 76 (to avoid edge effects) and a width set to 3 bins. Each signal is scaled such that, at the chosen mean position, z matches a desired value (within a tolerance of $1 \cdot 10^{-4}$). For the training dataset, this value is picked uniformly from the interval [1, 20]. For the testing dataset, it is fixed to either 0 (no signal) or 3 (3σ signal). These significance values are chosen in order to improve the NN’s bump-detection capabilities, and to test it with signals similar to what would typically be considered “hints” of NP. Furthermore, when injecting signals to the input arrays, the desired z value is calculated relative to the fluctuated background distribution, so that the presence of a signal is ensured. On the other hand, when preparing the target arrays, the z distribution is calculated relative to the smooth background distribution, simulating an ideal analysis.

Finally, signal and background are added to form an input, and the corresponding target is calculated following Eqs. (4.1)–(4.3). All inputs and, separately, all targets are globally scaled to (inverted from) the interval [0, 1] before (after) being processed by the NN, as its internal parameters are initially of $\mathcal{O}(1)$.

Architecture

Our NN architecture is simple: the first hidden layer is fully connected, and the following six layers are one-dimensional convolutions with decreasing kernel sizes. The latter are commonly used in feature-detection models (see e.g. [Dhillon & Verma, 2020](#)) and are the main processors in the NN. The unusual choice of fully connecting the first layer was found to suppress a systematic bias depending on signal location in the 100-bin window. Each layer is activated with a rectified linear unit function. The resulting model contains 73,941 parameters.

To train the NN, i.e. optimize its internal parameters, we define a loss function as the mean squared error between a given prediction (z_{pred}) and known target (z_{true}) of the z distribution, integrated over all bins. This is accomplished with an “Adam” optimizer, using a learning rate of $3 \cdot 10^{-4}$. For our benchmark model, we generate a total of 600,000 training samples—20% of which we use for validation during training—150,000 testing samples with no signal, and 150,000 testing samples with an injected 3σ signal. The optimizer processes the training dataset in batches of 100 samples, and iterates this procedure over the entire dataset 200 times (in “epochs”). The NN does not see the testing data until all parameters are frozen and training is complete.

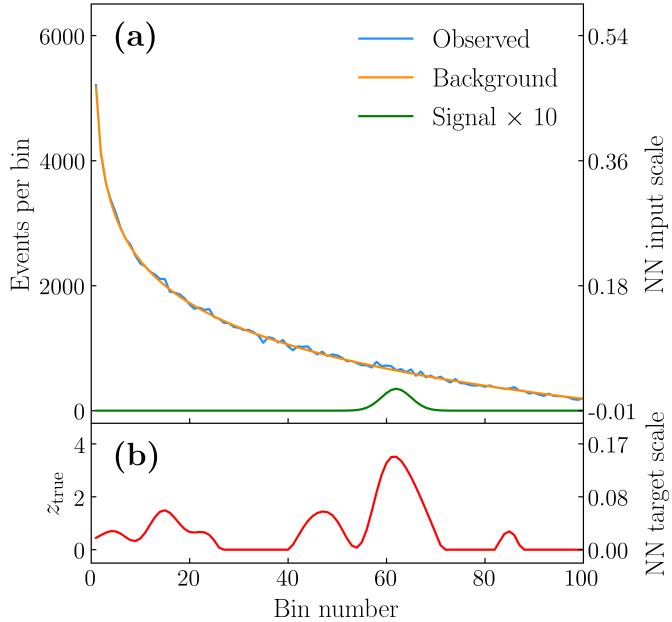


Figure 4.1. Illustration of the procedure for generating data samples. **(a)** A smoothly falling background curve (*orange*) is generated over 100 bins. Each bin is fluctuated with Poisson probabilities. A Gaussian signal (*green*) of a chosen significance (relative to the fluctuated background) is injected, producing a histogram of observed events (*blue*). **(b)** Given these three distributions, the corresponding z_{true} distribution (*red*) is calculated (as described in the text). The right axis of panels (a) and (b) shows corresponding values after a linear transformation to the NN input and target scales, respectively.

4.3 Results

The accuracy of the trained NN’s predictions is shown in Fig. 4.2. For each testing sample, we plot the difference between the predicted maximum significance, z_{pred}^{\max} , and the “true” (target) maximum significance, z_{true}^{\max} . We find that the difference has a mean (μ) of -0.02 and a standard deviation (σ) of 0.46 over our test set. The positions of z_{pred}^{\max} and z_{true}^{\max} were found to be within 1 bin of each other in over 87% of cases. Note that the asymmetry in the figure, appearing as a sharp edge in the third quadrant, is explained by the small number of cases with $z_{\text{pred}}^{\max} < 1$.

The results above show that the NN can accurately identify bumps in distributions where a 3σ signal has been injected. However, this is not a sufficient test of its performance, as we are only interested in finding samples containing a signal, while rejecting samples with large background fluctuations—but no signal. In Fig. 4.3, we show the distributions of z_{pred}^{\max} (dashed line) and z_{true}^{\max} (solid line), for samples with both no signal added (blue) and a 3σ signal added (orange). We observe that the predicted distributions match the target distributions closely. Also, we see a clear separation of signal and background in the predictions. Note that the large widths of the z_{true}^{\max} distributions are caused by the difference between our methods for injecting versus measuring bumps of a given significance level, as mentioned before.

The separation power of the NN is illustrated in Fig. 4.4 in terms of *receiver operating characteristic* (ROC) curves, which are obtained from the histograms of Fig. 4.3. The ROC curves plot the fraction of (3σ) signal-injected samples with

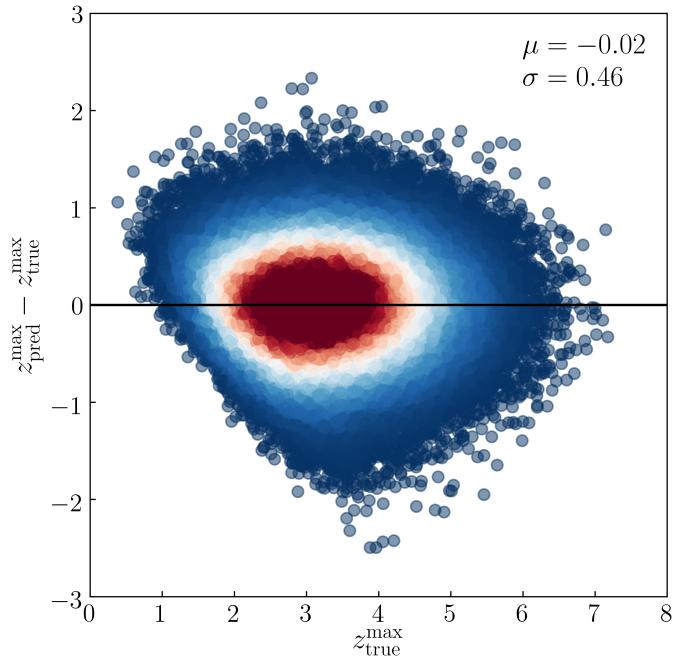


Figure 4.2. Plot of $z_{\text{pred}}^{\max} - z_{\text{true}}^{\max}$ against z_{true}^{\max} . The samples plotted contain 3σ signal injections; measured z_{true}^{\max} values are distributed around this significance. The NN predictions, z_{pred}^{\max} , are spread around the true values with a mean $\mu = -0.02$ and standard deviation $\sigma = 0.46$. Dense regions are shown in red (roughly corresponding to the 1σ region), while sparse regions are shown in blue.

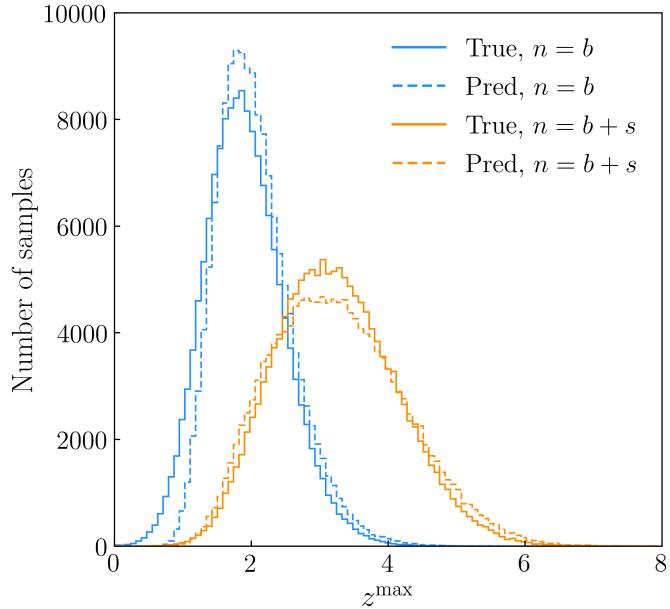


Figure 4.3. The distributions of z_{true}^{\max} (solid line) and z_{pred}^{\max} (dashed line) for samples where the generated data (n) contains background (b) only (blue), and for samples where a 3σ signal (s) is added (orange).

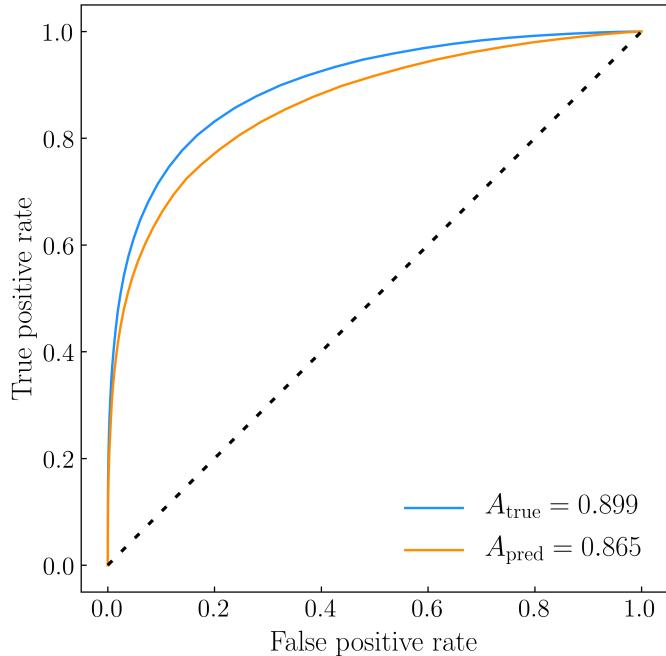


Figure 4.4. True (blue) and predicted (orange) ROC curves for 3σ signals, and their associated areas under curve, A_{true} and A_{pred} , respectively.

z^{\max} above a given threshold value (the true positive rate) against the fraction of background-only samples with z^{\max} above the same value (the false positive rate), as this threshold is decreased. This is shown for both z_{true}^{\max} (in blue) and z_{pred}^{\max} (in orange). According to the Neyman-Pearson lemma (see e.g. [Lehmann & Romano, 2005](#)), given complete knowledge of the shapes of the underlying background and signal, z_{true} provides the highest possible level of discrimination between the two. Yet, despite being given no such information explicitly, the NN achieves almost the same separation power, on average, for 3σ signals. This is quantified by the areas under the true and predicted ROC curves, $A_{\text{true}} = 0.899$ and $A_{\text{pred}} = 0.865$, respectively—a degradation in average performance of less than 4%, based on their ratio ($A_{\text{pred}}/A_{\text{true}} > 0.96$). Thus, the probability that similar distributions will be marked as interesting based on z_{pred} approaches that of an ideal analysis method employed in traditional searches.

Generalization

Additional experiments are performed to test whether the NN is able to generalize its capacity to identify bumps to different scenarios. Smaller datasets and fewer training epochs are used in order to estimate this. Nonetheless, a performance similar to the above, albeit slightly poorer ($|\mu| \lesssim 0.19$, $\sigma \lesssim 0.95$, $A_{\text{pred}}/A_{\text{true}} \gtrsim 0.87$), is obtained given any of the following modifications: (i) Training on nine of the background shapes in Eq. (4.4), excluding either be^{-ax} or $1/ax^4 + b$, and testing with this (unseen) tenth shape; (ii) Testing for bumps over linear combinations of up to four of the background forms; (iii) Testing over background distributions with lower or higher statistics, using either 100–500 events or 5,000–10,000 events

in the edge bins; (*iv*) Extending the bin range available to injected signals from 25–76 to 5–96; (*v*) Training and testing with wider signals, using Gaussians with a width of either 4 or 5 bins. These results show that the NN’s ability to detect bumps efficiently is not entirely restricted to specific background or signal forms—a capacity that goes beyond the potential of traditional techniques.

Validation

We validate both our training configuration and the consistency of predictions with further short tests. The convergence of the training loss value is confirmed by studying the effects of increasing the amount of training. We notice very small changes when extending its duration past 200 epochs—using 100,000 training samples, we observe $\delta\mu \sim 0.09$, $\delta\sigma \sim 0.02$, and $\delta A_{\text{pred}} \sim 0.005$, as well as some overfitting—or when increasing its sample size above 500,000—with 100 epochs, we observe $\delta\mu, \delta\sigma, \delta A_{\text{pred}} < 0.01$.

To ensure consistency, we compare the predictions of four copies of the NN in two ways. The copies are trained for 100 epochs with statistically independent datasets, each consisting of 100,000 samples. Firstly, the NNs are tested on the same dataset (containing 25,000 samples) and performance is compared between NNs. We measure small variations in performance (equal to the effect due to training duration), indicating little sensitivity to any particular training stage. Secondly, a given NN is tested on four different datasets (each is statistically independent, but equivalent, and contains 25,000 samples) and performance is compared between datasets. For each NN copy, we observe that differences depending on the particular testing data are negligible.

4.4 Discussion

In the search for NP we must leave no stone unturned. Given the expected ramp-up of the LHC, and the challenges faced by the blind analysis paradigm, it is important to consider search methods such as the DDP, which provide efficient (and complementary) routes to discovering signals. Our strategy is guided by the principles of exploiting the full potential of collected data and avoiding resource-consuming tasks. With this in mind, our aim here was to demonstrate an implementation of the DDP for bump-hunting, and examine its ability to detect resonant signals over smooth backgrounds.

We conclude that a supervised NN can correctly identify bumps over many smoothly falling curves in seconds, without any explicit information on the underlying background and signal shapes. Compared to a profile likelihood ratio test that relies on exact knowledge of both these shapes, the NN shows competitive levels of accuracy and precision on a simulated dataset containing 3σ signals. We quantify this by the mean and standard deviation of $z_{\text{pred}}^{\max} - z_{\text{true}}^{\max}$, which we measure to be -0.02 and 0.46 , respectively. In terms of separation power between such signals and background fluctuations, the NN’s average performance loss (with respect to z_{true}^{\max}) is found to be less than 4%, based on the ratio of areas under ROC curves, $A_{\text{pred}}/A_{\text{true}}$.

Similar algorithms could be used to scan the overwhelmingly large observable-space that is being measured at the LHC, in search of interesting regions that show potential for discovery. Example usages are bump-hunting in exclusive selections with dilepton, dijet, or diphoton final states, and criteria on a multitude of other observables, such as missing transverse energy, angular observables, etc.. Since no MC simulation would be used, the search would not be sensitive to MC mismodelling or limited by low MC statistics. Instead, modeling would be based on the fundamental SM properties featured in the data.

A first milestone for the DDP could be to establish an efficient sensitivity to bumps in regions previously analyzed by the ATLAS and CMS collaborations. To achieve this, further developments of the algorithm could include adapting to different dynamic ranges (bin ranges set by detector resolutions, ranges of event counts set by physical processes), increasing the flexibility of data generation (e.g. via spline interpolation techniques), using more sophisticated architectures, or optimizing for accurate background estimation. However, the strategy is not limited to bump-hunting. Other forms of NP could be searched for as well, by developing algorithms which exploit, for example, flavour symmetry, CP symmetry or forward-backward symmetry.

References

- Abada, A., Bećirević, D., Lucente, M., & Sumensari, O. (2015, Jun). Lepton flavor violating decays of vector quarkonia and of the Z boson. *Phys. Rev. D*, *91*(11), 113013. doi:[10.1103/PhysRevD.91.113013](https://doi.org/10.1103/PhysRevD.91.113013)
- Aebischer, J., Altmannshofer, W., Guadagnoli, D., Reboud, M., Stangl, P., & Straub, D. M. (2020, Mar). B -decay discrepancies after Moriond 2019. *Eur. Phys. J. C*, *80*(3), 252–27. doi:[10.1140/epjc/s10052-020-7817-x](https://doi.org/10.1140/epjc/s10052-020-7817-x)
- Afik, Y., Cohen, J., Gozani, E., Kajomovitz, E., & Rozen, Y. (2018, Aug). Establishing a search for $b \rightarrow s\ell^+\ell^-$ anomalies at the LHC. *J. High Energy Phys.*, *2018*(8), 56–14. doi:[10.1007/JHEP08\(2018\)056](https://doi.org/10.1007/JHEP08(2018)056)
- Alok, A. K., Kumar, J., Kumar, D., & Sharma, R. (2019, Aug). Lepton flavor non-universality in the B -sector: a global analyses of various new physics models. *Eur. Phys. J. C*, *79*(8), 707–10. doi:[10.1140/epjc/s10052-019-7219-0](https://doi.org/10.1140/epjc/s10052-019-7219-0)
- Angelescu, A., Bećirević, D., Faroughy, D. A., Jaffredo, F., & Sumensari, O. (2021, Sep). Single leptoquark solutions to the B -physics anomalies. *Phys. Rev. D*, *104*(5), 055017. doi:[10.1103/PhysRevD.104.055017](https://doi.org/10.1103/PhysRevD.104.055017)
- ATLAS Collaboration, Aaboud, M., Aad, G., Abbott, B., Abdinov, O., Abeloos, B., ... others (2019, Feb). A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment. *Eur. Phys. J. C*, *79*(2), 120–45. doi:[10.1140/epjc/s10052-019-6540-y](https://doi.org/10.1140/epjc/s10052-019-6540-y)
- ATLAS Collaboration, Aad, G., Abbott, B., Abbott, D. C., Abdinov, O., Abed Abud, A., ... others (2019, Sep). Search for high-mass dilepton resonances using 139 fb^{-1} of pp collision data collected at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector. *Phys. Lett. B*, *796*, 68–87. doi:[10.1016/j.physletb.2019.07.016](https://doi.org/10.1016/j.physletb.2019.07.016)
- ATLAS Collaboration, Aad, G., Abbott, B., Abbott, D. C., Abed Abud, A., Abeling, K., ... others (2020a, Sep). Dijet Resonance Search with Weak Supervision Using $\sqrt{s} = 13 \text{ TeV}$ pp Collisions in the ATLAS Detector. *Phys. Rev. Lett.*, *125*(13), 131801. doi:[10.1103/PhysRevLett.125.131801](https://doi.org/10.1103/PhysRevLett.125.131801)
- ATLAS Collaboration, Aad, G., Abbott, B., Abbott, D. C., Abed Abud, A., Abeling, K., ... others (2020b, Jul). Search for Heavy Higgs Bosons Decaying into Two Tau Leptons with the ATLAS Detector Using pp Collisions at $\sqrt{s} = 13 \text{ TeV}$. *Phys. Rev. Lett.*, *125*(5), 051801. doi:[10.1103/PhysRevLett.125.051801](https://doi.org/10.1103/PhysRevLett.125.051801)
- ATLAS Collaboration, Aad, G., Abbott, B., Abbott, D. C., Abed Abud, A., Abeling, K., ... others (2020c, Mar). Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} =$

- 13 TeV with the ATLAS detector. *J. High Energy Phys.*, *2020*(3), 145–41. doi:[10.1007/JHEP03\(2020\)145](https://doi.org/10.1007/JHEP03(2020)145)
- ATLAS Collaboration, Aad, G., Abbott, B., Abbott, D. C., Abed Abud, A., Abeling, K., ... others (2021a, Jun). Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, *103*(11), 112006. doi:[10.1103/PhysRevD.103.112006](https://doi.org/10.1103/PhysRevD.103.112006)
- ATLAS Collaboration, Aad, G., Abbott, B., Abbott, D. C., Abed Abud, A., Abeling, K., ... others (2021b, Oct). Search for New Phenomena in Final States with Two Leptons and One or No b -Tagged Jets at $\sqrt{s} = 13$ TeV Using the ATLAS Detector. *Phys. Rev. Lett.*, *127*(14), 141801. doi:[10.1103/PhysRevLett.127.141801](https://doi.org/10.1103/PhysRevLett.127.141801)
- ATLAS Collaboration, Aad, G., Abbott, B., Abbott, D. C., Abed Abud, A., Abeling, K., ... others (2022, Jan). Search for Higgs boson decays into a pair of pseudoscalar particles in the $bb\mu\mu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, *105*(1), 012006. doi:[10.1103/PhysRevD.105.012006](https://doi.org/10.1103/PhysRevD.105.012006)
- ATLAS Collaboration, Aad, G., Abbott, B., Abbott, D. C., Abud, A. A., Abeling, K., ... others (2021, Jul). Test of the universality of τ and μ lepton couplings in W -boson decays with the ATLAS detector - Nature Physics. *Nat. Phys.*, *17*(7), 813–818. doi:[10.1038/s41567-021-01236-w](https://doi.org/10.1038/s41567-021-01236-w)
- BaBar Collaboration, Aubert, B., Bona, M., Boutigny, D., Couderc, F., Karyotakis, Y., ... others (2007, Feb). Search for Lepton Flavor Violating Decays $\tau^\pm \rightarrow \ell^\pm \pi^0, \ell^\pm \eta, \ell^\pm \eta'$. *Phys. Rev. Lett.*, *98*(6), 061803. doi:[10.1103/PhysRevLett.98.061803](https://doi.org/10.1103/PhysRevLett.98.061803)
- BaBar Collaboration, Aubert, B., Karyotakis, Y., Lees, J. P., Poireau, V., Prencipe, E., ... others (2010, Jul). Measurements of Charged Current Lepton Universality and $|V_{us}|$ Using Tau Lepton Decays to $e^- \bar{\nu}_e \nu_\tau, \mu^- \bar{\nu}_\mu \nu_\tau, \pi^- \nu_\tau$, and $K^- \nu_\tau$. *Phys. Rev. Lett.*, *105*(5), 051602. doi:[10.1103/PhysRevLett.105.051602](https://doi.org/10.1103/PhysRevLett.105.051602)
- BaBar Collaboration, Lees, J. P., Poireau, V., Tisserand, V., Garra Tico, J., Grauges, E., ... others (2012a, Sep). Evidence for an Excess of $\bar{B} \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ Decays. *Phys. Rev. Lett.*, *109*(10), 101802. doi:[10.1103/PhysRevLett.109.101802](https://doi.org/10.1103/PhysRevLett.109.101802)
- BaBar Collaboration, Lees, J. P., Poireau, V., Tisserand, V., Garra Tico, J., Grauges, E., ... others (2012b, Jul). Search for the decay modes $B^\pm \rightarrow h^\pm \tau \ell$. *Phys. Rev. D*, *86*(1), 012004. doi:[10.1103/PhysRevD.86.012004](https://doi.org/10.1103/PhysRevD.86.012004)
- BaBar Collaboration, Lees, J. P., Poireau, V., Tisserand, V., Grauges, E., Palano, A., ... others (2013, Jun). Search for $B \rightarrow K^{(*)} \nu \bar{\nu}$ and invisible quarkonium decays. *Phys. Rev. D*, *87*(11), 112005. doi:[10.1103/PhysRevD.87.112005](https://doi.org/10.1103/PhysRevD.87.112005)
- Babu, K. S., Dutta, B., & Mohapatra, R. N. (2019, Jan). A theory of $R(D^*, D)$ anomaly with right-handed currents. *J. High Energy Phys.*, *2019*(1), 168–44. doi:[10.1007/JHEP01\(2019\)168](https://doi.org/10.1007/JHEP01(2019)168)
- Belle Collaboration, Caria, G., Urquijo, P., Adachi, I., Aihara, H., Al Said, S., ... others (2020, Apr). Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a Semileptonic Tagging Method. *Phys. Rev. Lett.*, *124*(16), 161803. doi:[10.1103/PhysRevLett.124.161803](https://doi.org/10.1103/PhysRevLett.124.161803)

- Belle Collaboration, Grygier, J., Goldenzweig, P., Heck, M., Adachi, I., Aihara, H., ... others (2017, Nov). Search for $B \rightarrow h\nu\bar{\nu}$ decays with semileptonic tagging at Belle. *Phys. Rev. D*, **96**(9), 091101. doi:[10.1103/PhysRevD.96.091101](https://doi.org/10.1103/PhysRevD.96.091101)
- Belle Collaboration, Hirose, S., Iijima, T., Adachi, I., Adamczyk, K., Aihara, H., ... others (2017, May). Measurement of the τ Lepton Polarization and $R(D^*)$ in the Decay $\bar{B} \rightarrow D^*\tau^-\bar{\nu}_\tau$. *Phys. Rev. Lett.*, **118**(21), 211801. doi:[10.1103/PhysRevLett.118.211801](https://doi.org/10.1103/PhysRevLett.118.211801)
- Belle Collaboration, Huschle, M., Kuhr, T., Heck, M., Goldenzweig, P., Abdesselam, A., ... others (2015, Oct). Measurement of the branching ratio of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle. *Phys. Rev. D*, **92**(7), 072014. doi:[10.1103/PhysRevD.92.072014](https://doi.org/10.1103/PhysRevD.92.072014)
- Belle Collaboration, Miyazaki, Y., Adachi, I., Aihara, H., Anipko, D., Arinstein, K., ... others (2007, May). Search for lepton flavor violating τ^- decays into $\ell^-\eta$, $\ell^-\eta'$ and $\ell^-\pi^0$. *Phys. Lett. B*, **648**(5), 341–350. doi:[10.1016/j.physletb.2007.03.027](https://doi.org/10.1016/j.physletb.2007.03.027)
- Bernlochner, F. U., Ligeti, Z., Papucci, M., & Robinson, D. J. (2017, Jun). Combined analysis of semileptonic B decays to D and D^* : $R(D^{(*)})$, $|V_{cb}|$, and new physics. *Phys. Rev. D*, **95**(11), 115008. doi:[10.1103/PhysRevD.95.115008](https://doi.org/10.1103/PhysRevD.95.115008)
- BES Collaboration, Ablikim, M., Bai, J. Z., Ban, Y., Bian, J. G., Cai, X., ... others (2004, Sep). Search for the lepton flavor violation processes $J/\Psi \rightarrow \mu\tau$ and $e\tau$. *Phys. Lett. B*, **598**(3), 172–177. doi:[10.1016/j.physletb.2004.08.005](https://doi.org/10.1016/j.physletb.2004.08.005)
- BESIII Collaboration, Ablikim, M., Achasov, M. N., Adlarson, P., Ahmed, S., Albrecht, M., ... others (2021, Jun). Search for the charged lepton flavor violating decay $J/\psi \rightarrow e\tau$. *Phys. Rev. D*, **103**(11), 112007. doi:[10.1103/PhysRevD.103.112007](https://doi.org/10.1103/PhysRevD.103.112007)
- Bifani, S., Descotes-Genon, S., Vidal, A. R., & Schune, M.-H. (2018, Dec). Review of lepton universality tests in B decays. *J. Phys. G: Nucl. Part. Phys.*, **46**(2), 023001. doi:[10.1088/1361-6471/aaf5de](https://doi.org/10.1088/1361-6471/aaf5de)
- Bigi, D., & Gambino, P. (2016, Nov). Revisiting $B \rightarrow D\ell\nu$. *Phys. Rev. D*, **94**(9), 094008. doi:[10.1103/PhysRevD.94.094008](https://doi.org/10.1103/PhysRevD.94.094008)
- Bigi, D., Gambino, P., & Schacht, S. (2017, Nov). $R(D^*)$, $|V_{cb}|$, and the Heavy Quark Symmetry relations between form factors. *J. High Energy Phys.*, **2017**(11), 61–23. doi:[10.1007/JHEP11\(2017\)061](https://doi.org/10.1007/JHEP11(2017)061)
- Bordone, M., Cornella, C., Fuentes-Martín, J., & Isidori, G. (2018, Oct). Low-energy signatures of the PS³ model: from B -physics anomalies to LFV. *J. High Energy Phys.*, **2018**(10), 148–50. doi:[10.1007/JHEP10\(2018\)148](https://doi.org/10.1007/JHEP10(2018)148)
- Boucenna, S. M., Celis, A., Fuentes-Martín, J., Vicente, A., & Virto, J. (2016, Dec). Phenomenology of an $SU(2) \times SU(2) \times U(1)$ model with lepton-flavour non-universality. *J. High Energy Phys.*, **2016**(12), 59–43. doi:[10.1007/JHEP12\(2016\)059](https://doi.org/10.1007/JHEP12(2016)059)
- Buras, A. J., Girrbach-Noe, J., Niehoff, C., & Straub, D. M. (2015, Feb). $B \rightarrow K^{(*)}\nu\bar{\nu}$ decays in the Standard Model and beyond. *J. High Energy Phys.*, **2015**(2), 184–39. doi:[10.1007/JHEP02\(2015\)184](https://doi.org/10.1007/JHEP02(2015)184)
- Chen, C.-H., & Nomura, T. (2017, Sep). Charged-Higgs on $R_{D^{(*)}}$, τ polarization, and FBA. *Eur. Phys. J. C*, **77**(9), 631–11. doi:[10.1140/epjc/s10052-017-5198-6](https://doi.org/10.1140/epjc/s10052-017-5198-6)
- Choudhury, D., Kumar, N., & Kundu, A. (2019, Oct). Search for an opposite sign

- muon-tau pair and a b -jet at the LHC in the context of flavor anomalies. *Phys. Rev. D*, **100**(7), 075001. doi:[10.1103/PhysRevD.100.075001](https://doi.org/10.1103/PhysRevD.100.075001)
- CMS Collaboration, Sirunyan, A. M., Tumasyan, A., Adam, W., Ambrogi, F., Bergauer, T., ... others (2021, Jul). MUSiC: a model-unspecific search for new physics in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, **81**(7), 629–49. doi:[10.1140/epjc/s10052-021-09236-z](https://doi.org/10.1140/epjc/s10052-021-09236-z)
- CMS Collaboration, Sirunyan, A. M., Tumasyan, A., Adam, W., Andrejkovic, J. W., Bergauer, T., ... others (2021, Jul). Search for resonant and nonresonant new phenomena in high-mass dilepton final states at $\sqrt{s} = 13$ TeV. *J. High Energy Phys.*, **2021**(7), 208–62. doi:[10.1007/JHEP07\(2021\)208](https://doi.org/10.1007/JHEP07(2021)208)
- Collins, J., Howe, K., & Nachman, B. (2018, Dec). Anomaly Detection for Resonant New Physics with Machine Learning. *Phys. Rev. Lett.*, **121**(24), 241803. doi:[10.1103/PhysRevLett.121.241803](https://doi.org/10.1103/PhysRevLett.121.241803)
- Collins, J. H., Howe, K., & Nachman, B. (2019, Jan). Extending the search for new resonances with machine learning. *Phys. Rev. D*, **99**(1), 014038. doi:[10.1103/PhysRevD.99.014038](https://doi.org/10.1103/PhysRevD.99.014038)
- Cowan, G., Cranmer, K., Gross, E., & Vitells, O. (2011, Feb). Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, **71**(2), 1554–19. doi:[10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0)
- Crivellin, A., Hofer, L., Matias, J., Nierste, U., Pokorski, S., & Rosiek, J. (2015, Sep). Lepton-flavor violating B decays in generic Z' models. *Phys. Rev. D*, **92**(5), 054013. doi:[10.1103/PhysRevD.92.054013](https://doi.org/10.1103/PhysRevD.92.054013)
- Dhillon, A., & Verma, G. K. (2020, Jun). Convolutional neural network: a review of models, methodologies and applications to object detection. *Prog. Artif. Intell.*, **9**(2), 85–112. doi:[10.1007/s13748-019-00203-0](https://doi.org/10.1007/s13748-019-00203-0)
- Farina, M., Nakai, Y., & Shih, D. (2020, Apr). Searching for new physics with deep autoencoders. *Phys. Rev. D*, **101**(7), 075021. doi:[10.1103/PhysRevD.101.075021](https://doi.org/10.1103/PhysRevD.101.075021)
- Feruglio, F., Paradisi, P., & Pattori, A. (2017, Jan). Revisiting Lepton Flavor Universality in B Decays. *Phys. Rev. Lett.*, **118**(1), 011801. doi:[10.1103/PhysRevLett.118.011801](https://doi.org/10.1103/PhysRevLett.118.011801)
- Greljo, A., Robinson, D. J., Shakya, B., & Zupan, J. (2018, Sep). $R(D^{(*)})$ from W' and right-handed neutrinos. *J. High Energy Phys.*, **2018**(9), 169–30. doi:[10.1007/JHEP09\(2018\)169](https://doi.org/10.1007/JHEP09(2018)169)
- Grossman, Y., & Nir, Y. (n.d.). *The Standard Model: a Uniquely Beautiful Theory*. To be published by Princeton University Press.
- Grzadkowski, B., Iskrzyński, M., Misiak, M., & Rosiek, J. (2010, Oct). Dimension-six terms in the Standard Model Lagrangian. *J. High Energy Phys.*, **2010**(10), 85–18. doi:[10.1007/JHEP10\(2010\)085](https://doi.org/10.1007/JHEP10(2010)085)
- Guest, D., Cranmer, K., & Whiteson, D. (2018, Oct). Deep Learning and Its Application to LHC Physics. *Annu. Rev. Nucl. Part. Sci.*, **68**(1), 161–181. doi:[10.1146/annurev-nucl-101917-021019](https://doi.org/10.1146/annurev-nucl-101917-021019)
- Hazard, D. E., & Petrov, A. A. (2016, Oct). Lepton flavor violating quarkonium decays. *Phys. Rev. D*, **94**(7), 074023. doi:[10.1103/PhysRevD.94.074023](https://doi.org/10.1103/PhysRevD.94.074023)
- Heavy Flavor Averaging Group, Amhis, Y., Banerjee, Sw., Ben-Haim, E., Bernlochner, F. U., Bona, M., ... others (2021, Mar). Averages of b -hadron, c -

- hadron, and τ -lepton properties as of 2018. *Eur. Phys. J. C*, *81*(3), 226–326. doi:[10.1140/epjc/s10052-020-8156-7](https://doi.org/10.1140/epjc/s10052-020-8156-7)
- Hu, Q., Li, X., & Yang, Y. (2019, Mar). $b \rightarrow c\tau\nu$ transitions in the standard model effective field theory. *Eur. Phys. J. C*, *79*(3), 264–20. doi:[10.1140/epjc/s10052-019-6766-8](https://doi.org/10.1140/epjc/s10052-019-6766-8)
- Jaiswal, S., Nandi, S., & Patra, S. K. (2017, Dec). Extraction of $|V_{cb}|$ from $B \rightarrow D^{(*)}\ell\nu_\ell$ and the Standard Model predictions of $R(D^{(*)})$. *J. High Energy Phys.*, *2017*(12), 60–24. doi:[10.1007/JHEP12\(2017\)060](https://doi.org/10.1007/JHEP12(2017)060)
- Kumar, N. (2020, Nov). Flavor violation at LHC in events with two opposite sign leptons and a b -jet. *arXiv*. Retrieved from <https://arxiv.org/abs/2011.12810v1>
- Lehmann, E. L., & Romano, J. P. (2005). *Testing Statistical Hypotheses*. New York, NY, USA: Springer, New York, NY. doi:[10.1007/0-387-27605-X](https://doi.org/10.1007/0-387-27605-X)
- LHCb Collaboration, Aaij, R., Abellán Beteta, C., Adeva, B., Adinolfi, M., Aidala, C. A., ... others (2019, Nov). Search for the Lepton-Flavor-Violating Decays $B_s^0 \rightarrow \tau^\pm \mu^\mp$ and $B^0 \rightarrow \tau^\pm \mu^\mp$. *Phys. Rev. Lett.*, *123*(21), 211801. doi:[10.1103/PhysRevLett.123.211801](https://doi.org/10.1103/PhysRevLett.123.211801)
- LHCb Collaboration, Aaij, R., Adeva, B., Adinolfi, M., Affolder, A., Ajaltouni, Z., ... others (2015, Sep). Measurement of the Ratio of Branching Fractions $\mathcal{B}(\overline{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$. *Phys. Rev. Lett.*, *115*(11), 111803. doi:[10.1103/PhysRevLett.115.111803](https://doi.org/10.1103/PhysRevLett.115.111803)
- LHCb Collaboration, Aaij, R., Adeva, B., Adinolfi, M., Ajaltouni, Z., Akar, S., ... others (2018, Apr). Measurement of the Ratio of the $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ and $B^0 \rightarrow D^{*-}\mu^+\nu_\mu$ Branching Fractions Using Three-Prong τ -Lepton Decays. *Phys. Rev. Lett.*, *120*(17), 171802. doi:[10.1103/PhysRevLett.120.171802](https://doi.org/10.1103/PhysRevLett.120.171802)
- LHCb collaboration, Aaij, R., Beteta, C. A., Ackernley, T., Adeva, B., Adinolfi, M., ... others (2021, Mar). Test of lepton universality in beauty-quark decays. *arXiv*. Retrieved from <https://arxiv.org/abs/2103.11769v1>
- Metodiev, E. M., Nachman, B., & Thaler, J. (2017, Oct). Classification without labels: learning from mixed samples in high energy physics. *J. High Energy Phys.*, *2017*(10), 174–18. doi:[10.1007/JHEP10\(2017\)174](https://doi.org/10.1007/JHEP10(2017)174)
- Murgui, C., Peñuelas, A., Jung, M., & Pich, A. (2019, Sep). Global fit to $b \rightarrow c\tau\nu$ transitions. *J. High Energy Phys.*, *2019*(9), 103–45. doi:[10.1007/JHEP09\(2019\)103](https://doi.org/10.1007/JHEP09(2019)103)
- Nachman, B., & Shih, D. (2020, Apr). Anomaly detection with density estimation. *Phys. Rev. D*, *101*(7), 075042. doi:[10.1103/PhysRevD.101.075042](https://doi.org/10.1103/PhysRevD.101.075042)
- Particle Data Group, Zyla, P. A., Barnett, R. M., Beringer, J., Dahl, O., Dwyer, D. A., ... others (2020, Aug). Review of Particle Physics. *Prog. Theor. Exp. Phys.*, *2020*(8). doi:[10.1093/ptep/ptaa104](https://doi.org/10.1093/ptep/ptaa104)
- Radovic, A., Williams, M., Rousseau, D., Kagan, M., Bonacorsi, D., Himmel, A., ... Wongjirad, T. (2018, Aug). Machine learning at the energy and intensity frontiers of particle physics. *Nature*, *560*(7716), 41–48. doi:[10.1038/s41586-018-0361-2](https://doi.org/10.1038/s41586-018-0361-2)
- Shi, R., Geng, L.-S., Grinstein, B., Jäger, S., & Camalich, J. M. (2019, Dec). Revisiting the new-physics interpretation of the $b \rightarrow c\tau\nu$ data. *J. High Energy Phys.*, *2019*(12), 65–29. doi:[10.1007/JHEP12\(2019\)065](https://doi.org/10.1007/JHEP12(2019)065)

Volkovich, S., De Vito Halevy, F., & Bressler, S. (2021, Jul). A Data-Directed Paradigm for BSM searches: the bump-hunting example. *arXiv*. Retrieved from <https://arxiv.org/abs/2107.11573v3>